Optimising the Sampling Methodology for CSO Household Surveys

Theoretical Considerations and Simulation Study

Trutz Haase
Jonathan Pratschke

May 2012

Table of Contents

Figures and Tables

Executive Summary

This report assesses alternative sampling designs for CSO household surveys, such as the Quarterly National Household Survey (QNHS), the annual Survey on Income and Living Conditions (SILC) and Household Budget Survey (HBS), as well as considering the design of a potential future General Household Survey. The aim of the report is to provide recommendations in relation to appropriate sampling designs, including considerations in relation to sample size, the degree of clustering, stratification and selection procedures for areas and households.

Background

The sampling designs currently used by the CSO were inspired by a report written by Prof. David Steel (1997), whilst the report *Variance Estimation for Complex Surveys* (2003) assesses the precision of estimates from two of these surveys, the QNHS and HBS. The development and refinement of Small Areas (SA) - the new census geography - has already led to change in these designs, whilst the publication of Small Area Population Statistics (SAPS) from the 2011 Census of Population is likely to reinforce this trend in coming months. SAs are of a similar size to the blocks that have been used as Primary Sampling Units for CSO surveys for many years, with the advantage that a full range of aggregate-level indicators will be available for these units. The 2011 SA definitions incorporate further improvements when compared to the 2006 boundaries, yielding a more uniform population distribution. This opens up a range of new possibilities for designing and managing household surveys in Ireland.

Methodology

This report is divided into two parts, the first of which discusses various aspects of sampling design for household surveys. Whilst discussing each issue individually, the study highlights the ways in which these are interlinked, including the need to develop a comprehensive strategy that optimises simultaneously across various factors. Part Two builds on this discussion, presenting the results of a simulation study of alternative sampling designs using individual-level data from the 2006 Census of Population. In total, 27 different sampling designs are compared, spanning small (5,200-6,000 households), medium-sized (20,000-26,000 households) and large (38,600 households) surveys. Three different measures of precision are used to compare these designs and their strengths and weaknesses are discussed in detail.

Sample size and clustering

Following sample size, the most powerful influence on the statistical efficiency of survey estimates is the degree of clustering. The simulation study compares four different scenarios including (1) no clustering; (2) a second-stage sample size of 4; (3)

a second-stage sample size of 15; (4) a second-stage sample size of 20 households. A simple random sample (SRS) of individual households produces the most precise estimates, although carrying out face-to-face interviews with a geographically-dispersed sample is generally not feasible due to cost considerations. Even where it is feasible to carry out repeat interviews using telephone interviewing techniques, initial interviews for CSO surveys will continue to involve face-to-face interviews for the foreseeable future, and therefore require a multi-stage sampling design.

There is, however, a considerably decline in efficiency when moving from a simple random sample to a two-stage sampling design. With a second-stage sample of 15 households, there is an increase in the Relative Standard Errors of between 15 and 60 per cent, depending on the degree of spatial autocorrelation of the variable concerned (the stronger the spatial structure, the greater the effect of clustering). Increasing the second-stage sample size from 15 to 20 households leads to a further (but smaller) loss in efficiency. A compromise thus needs to be achieved when balancing the potential cost savings that would result from an increase in the cluster size, on the one hand, and the resulting loss in efficiency, on the other.

Stratification

As clustering tends to reduce precision, this aspect of sampling design is typically combined with stratification, which generally improves statistical efficiency. Stratification for CSO household surveys currently relies on a classification by "area type", and this is broadly in line with international practice. One of the objectives of this study is to assess the effectiveness of this approach as compared with stratification by the Haase-Pratschke Index of Relative Affluence and Deprivation or a combination of area typology and deprivation scores. The results are striking: stratification by area type has an almost negligible effect on statistical efficiency, whilst stratification by deprivation scores yields significant improvements in precision, partly compensating for the negative effects of clustering.

Recommendations for Household Surveys in Ireland

The simulation results lead to several recommendations for the design of household surveys in Ireland. The most important of these is that stratification by deprivation score, using a measure such as the Haase-Pratschke Index of Relative Affluence and Deprivation, leads to significant and consistent gains in precision for all estimators. Estimates are therefore likely to be less volatile, to have smaller standard errors and to be more sensitive to temporal trends and inter-regional differences.

This is an exciting result, which may be applied to both the SILC and the HBS, and to any similar survey which relies on a two-stage cluster sample design. As the SILC has the goal of estimating EU poverty indicators (which are most strongly correlated with affluence and deprivation), the efficiency gains associated with this stratification method would be particularly pronounced.

Stratification by deprivation score would also lead to significant improvements in the precision of labour force-related estimators based on the QNHS. The simulation study suggests that the best approach, as far as this survey is concerned, would be to combine deprivation score with either area type or region to create a multi-criteria stratification variable. When compared with stratification by area type alone, the efficiency gains yielded by this method would facilitate an increase in the degree of clustering from 15 to 20 households per PSU whilst simultaneously achieving substantial improvements in precision. A more detailed summary of the recommendations flowing from this study are summarised below.

Theory-based Recommendations

The first set of recommendations are derived from the discussion in Part One of this study and are based primarily on theoretical considerations. In certain cases, these coincide with innovations that the CSO has identified as attractive or necessary.

No.	***Page***	***Domain***	***Description***
1	6	PSUs	*The desired population size of PSUs should be near the maximum value that allows for acceptable travel times between pairs of dwellings, which may imply aggregating Small Areas (SAs)*
2	10	Sampling of PSUs	*PSUs should be selected randomly but with probability of selection proportional to size (PPS), where the latter is measured by the number of households or persons (as appropriate)*
3	17	Selection of households	*A fixed number of private households should be selected within each PSU included in the sample*
4	16	Sampling of non-private households	*Where it is necessary to include non-private households, use a separate sampling frame, e.g. estimate the resident population in non-private households from Census data and select using PPS*
5	9	Stratification	*If deemed necessary, create a stratum for "surprises" encountered during data collection; this can be used to re-weight PSUs found to contain a larger population than expected*
6	19	Method of data collection	*If feasible, use telephone interviewing for follow-up questionnaires for the QNHS*
7	22	Rotation patterns	*Replace PSUs only when all final-stage units have been sampled; SAs should be aggregated to form PSUs of the required size for the rotation pattern, bearing in mind recommendation 1 above*
8	23	Monthly survey	*Move towards a monthly national household survey; this would facilitate a more precise approach to modelling underlying trends and provide valuable additional estimates at national level*
9	24	Missing values	*Adopt a comprehensive strategy to reduce item non-response and imputation techniques to deal with missing values*

Recommendations Based on the Simulation Study

The second set of recommendations derives from the empirical results presented in Part Two of this report and centres on how sampling methodology for CSO household surveys might be enhanced using innovative stratification techniques.

No.	***Page***	***Domain***	***Description***
10	13	Stratification	*Use the Haase-Pratschke Index of Relative Affluence and Deprivation for stratification, using arbitrary cut-off points or deciles; construct a multi-criteria stratification measure by crossing this variable with a measure of area type or region*
11	57	Stratification	*For the SILC, stratification based on a two-way intersection of NUTS3 (8 categories) and 10 deprivation categories is the strongest design and would yield significant and consistent gains in precision for all estimators*
12	57	Stratification	*For the QNHS, either the M10 or the M11 sampling design would result in significant improvements vis-a-vis the current design: estimates would be less volatile and have smaller standard errors*

Recommendations for Future Research

The final set of recommendations relates to key research questions with the potential to further optimise sampling and estimation methodologies, which require further evaluation and study.

No.	***Page***	***Domain***	***Description***
13	37	Simulation studies	*Repeat the present study using 2011 census data, extending the scope of the simulation to include rotation patterns, change scores and estimation/weighting/calibration techniques*
14	8	Small area population estimation	*Assess the potential role of census and non-census data in the estimation of population change for small areas*
15	22	Rotation patterns	*Carry out a comparative analysis of rotation designs using simulation techniques to assess accuracy of change estimates*
16	25	Estimation	*Evaluate the use of generalised regression estimators to estimate national and regional totals etc. with the relevant standard errors, using auxiliary information and benchmark data*

Implementation

It is important to note that at least some of these recommendations could have quite far-reaching consequences for survey estimates and field operations, implying that they should be assessed carefully prior to implementation. The potential gains offered by each innovation should be evaluated not only in its own terms, but also in relation to its impact on field operations and on the data series as a whole.

To the extent that the adoption of new PSUs, selection techniques or rotation patterns, a shorter interval between waves or new data collection techniques is likely to have an impact on survey estimates (e.g. by improving accuracy), it may make sense to introduce them together as a combined package, allowing for both an overlap between the old and new sampling designs and a gradual phasing-in of the new strategy.

This would enable field operators to adapt to the new procedures and to identify the most appropriate way of organising their work, as well as enabling the CSO to conduct a comparative analysis of the overall impact of the new sampling design and to take steps to maintain data integrity and comparability. On the other hand, changes that primarily influence the precision of estimates - such as stratifying by a new variable or applying new estimation techniques - could be implemented relatively easily, as they pose few risks, have low costs and minimal consequences for field operations.

Part One: Principal Considerations in the Design of Household Surveys

1 Introduction

This report analyses the implications of alternative sampling designs for existing CSO household surveys (including the Quarterly National Household Survey, the annual Survey on Income and Living Conditions, the Household Budget Survey), as well as considering the design of a possible future General Household Survey. The aim of the report is to provide recommendations in relation to the most appropriate sampling methodology to be used for these surveys, including considerations in relation to overall sample size, the size of clusters, the type of stratification to be used, criteria and procedures for selecting areas, households and individuals, the most appropriate forms of estimation and calibration, management of the population frame, rotation patterns for longitudinal designs and the precision of estimates. Whilst exploring these different issues, we will take into account their implications for field operations and the overall cost of survey operations.

The first part of this report is divided into 10 sections, which deal with analytically distinct aspects of sampling design for large-scale household surveys in Ireland. Whilst discussing each issue individually, we will also highlight the ways in which they are linked and the need to develop a comprehensive strategy that optimises simultaneously across various factors.

We will not provide a detailed or exhaustive treatment of any of these issues, but instead summarise the key considerations which we feel are most important when thinking about alternative sampling designs and are of potential relevance to the simulation study. Where appropriate, references are made to relevant scientific publications, which may be consulted for further information on specific techniques or procedures.

In Part Two of this report, we report on a simulation study designed to assess the impact of different sampling designs on the precision and accuracy of estimates. The study uses Monte Carlo techniques to estimate the sampling error and bias of various estimators of the population total for key socio-economic variables which represent the substantive areas of interest of existing household surveys. Individual-level census data were used to represent the target population and a series of samples were drawn from this population in line with the sampling designs to be tested.

This enables us to quantify the impact of complex survey design on the sample estimates, taking account of (a) the size of the overall sample and the degree of clustering; (b) the actual distribution of the variables of interest, including their degree of spatial auto-correlation; (c) the effects of using different stratification variables; (d) the impact of using different kinds of selection procedures. The design of the simulation study can be extended to include the effects of different estimation procedures, rotation patterns, forms of non-response and estimates of change, although these issues fall beyond the scope of the present study. By identifying the standard cost of various elements of survey operations, it is possible to construct a model that relates the efficiency of different sampling designs to their cost, with a view to identifying the optimal design for each survey.

2 Description of Household Sample Surveys in Ireland

The **Quarterly National Household Survey** (QNHS) is the main source of data on the Irish labour market. Estimates are derived from data collected in face-to-face interviews with individuals in selected households. The country is geographically stratified into 8 area type strata based on the Census of Population. Within strata, blocks of 75 households are identified as Primary Sampling Units (PSUs). A sample of 1 in 6 blocks is selected, and within each selected block 5 clusters of approximately 15 households are identified using systematic sampling. The first cluster consists of the 1st, 6th, 11th ... household, the second cluster consists of the 2nd, 7th, 12th ..., and so on. Initially one of the clusters is selected at random and when the households in that cluster are due for rotation they are replaced by households from the next cluster. An "in-for-5" rotation pattern is used, whereby one-fifth of the sample is replaced by a new household every quarter, ensuring that the quarterly overlap of selected households is 80 per cent and the annual overlap is 20 per cent. This design leads to a 1 in 30 overall sample; for the population of approximately 1.15 million households in Ireland, there are just under 15,000 blocks of 75 households. A one-in-six sample of these blocks results in about 2,550 blocks being selected, each block having 5 clusters of 15 households.

The **Annual Survey on Income and Living Conditions (SILC)** in Ireland is a household survey covering a broad range of issues in relation to income and living conditions. It is the official source of data on household and individual income and provides a number of key national poverty indicators, such as the "at risk of poverty" rate, the "consistent poverty" rate and rates of "enforced deprivation". SILC was conducted by the Central Statistics Office (CSO) for the first time in 2003 under EU legislation (Council regulation No. 1177/2003) and is currently being conducted on an annual basis. Each year, a different topic relating to poverty or social exclusion is included. A two-stage sample design is used, comprising a first-stage sample of 1,690 "blocks" selected to proportionately represent eight strata. The second stage of sampling involves the random selection of a sample of households (including two substitute households) from each block. In cases where interviewers cannot secure an interview from a household, they approach the two substitute households in the selected order. One quarter of households interviewed in a given year remain in the sample for the following three waves of data collection. The overall annual sample size is in the region of 5,000 households. In accordance with Eurostat recommendations, CALMAR is used to calculate the household cross-sectional weights, following the application of design weights (defined as the inverse of the probability of selection) (cf. Deville & Sarndal, 1992). Benchmark data derived from the Census of Population and the QNHS are used to gross up the data to population estimates (age by sex, region and household composition).

The **Household Budget Survey** is conducted every five years with a view to collecting detailed household income and expenditure for the purposes of updating the Consumer Price Index. It has been undertaken at regular intervals since 1951, and relies on a detailed diary of household expenditure over a two-week period. Detailed information on all sources of household income is collected as part of the survey. A two-stage sample design is used, the first-stage units being stratified by area type (as in the case of the QNHS and SILC). Each survey area contains up to 6 blocks of households and each block contains about 75 households. 2,600 blocks are selected by systematic sampling from a geographical ordering of the blocks within each stratum. The number of blocks selected from a stratum depends on the number of blocks in the stratum so that each block has equal probability of selection. At stage two, 4 original households and 4 substitute households are randomly selected from each of the 2,600 blocks. A total of four households are interviewed in each block, and interviewers approach as many substitute households as necessary to reach this quota, yielding a sample of 10,400.

3 Theoretical Considerations

3.1 Overall sampling strategy

It is generally not feasible to sample individuals directly, in the absence of high-quality register data[1]. For this reason, individuals are typically sampled within households, using a multi-stage sample design. Although it is possible to obtain more efficient statistical estimates by sampling specific individuals within households, rather than including all household members, these efficiency gains are generally small and far out-weighted by the economic cost (Clark & Steel, 2002). This is due primarily to the great heterogeneity of households, which reduces the negative effects of household clustering on the sampling error of the estimates (Pascal Ardilly & Tillé, 2006). For this reason, it has become common practice in survey sampling to collect data on all members of households using a cluster sample at the second stage with households as units. This also tends to improve response rates, whilst leaving sampling error unaffected (Groves, 1989).

Households may be sampled directly from a list or identified directly within the geographical areas which are used in the first stage of sampling. Even in the absence of an accurate population register, it is sometimes possible to construct a sampling frame of residential addresses[2], opening up the possibility of sampling households directly in a single-stage design. Whether households are selected at random or sampled systematically, the resulting designs are highly efficient; this is the sampling design typically used for CATI-based telephone surveys.

If face-to-face interviewing is used, however, it is generally not cost-effective to sample households directly in a single stage, as this would increase travel times to each appointment. In large countries (the US, Australia etc.), it is often not possible to maintain a team of professional interviewers within reasonable travel times of all areas, implying a need to use a multi-stage design where the initial sample of counties or regions remains stable for months or even years. In Ireland, by contrast, it is feasible for a large survey organisation to maintain a network of interviewers, who work on various surveys and have the capacity to reach all areas of the country (with the exception of the islands). This implies that the selection of PSUs and their population size are not constrained by considerations relating to the location of interviewers, and may be defined in relation to operational requirements.

1 The only countries that rely on individual-level sampling from a population register are Finland, Sweden, Denmark and Switzerland (Statistics Finland, 2011).

2 The Census of Population is conducted every five years in Ireland (compared to every 10 years in most other developed countries), which facilitates the maintenance of an up-to-date list of residential addresses, as well as providing high-quality data for weighting and calibration.

3.2 The frame of Primary Sampling Units (PSUs)

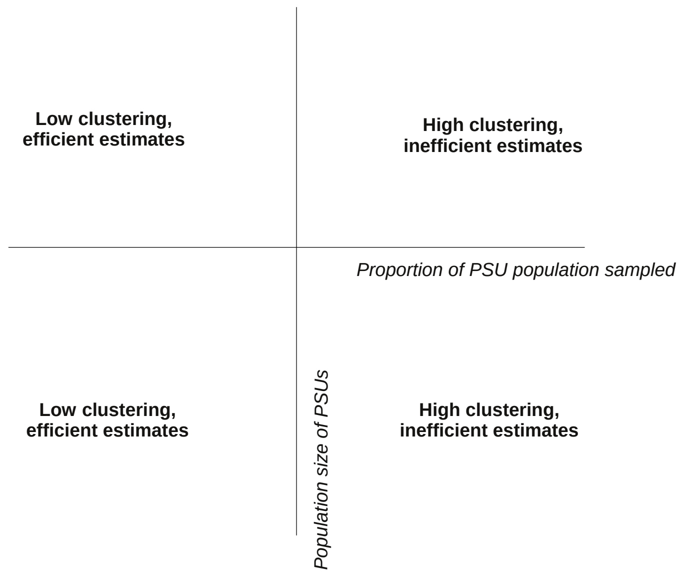
As we have seen, there are compelling motives for sampling households and including all household members in the sample and (where face-to-face interviewing is used) for sampling households within geographical areas. Operational criteria and cost considerations lead to this conclusion even if an appropriate sampling frame of dwellings exists. If it were possible to use telephone interviewing for all data collection for a given survey, then a simple random sample of dwellings drawn from an appropriate sampling frame would be more efficient in statistical terms.

As is widely known, the sampling design is strongly influenced by the degree of clustering of interviews, and this is one of the main considerations when evaluating alternative sampling designs (Pascal Ardilly & Tillé, 2006)[3]. The strength of this effect depends on the degree of spatial autocorrelation of the variable(s) of interest, as clustering will have no impact on the efficiency of estimates if a characteristic is randomly distributed throughout the population. By contrast, if a characteristic is relatively homogeneous at local level, clustering will have a particularly strong impact on the statistical efficiency of sample estimates (even a relatively low intra-class correlation coefficient can have a large effect on the variance of the estimator if the sub-sample cluster is large). This is intuitively understandable, as homogeneity within clusters implies a reduction in the informational content of successive interviews, reducing the effective sample size.

It is therefore possible to imagine a typology of cluster samples formed by the intersection of two axes relating to the population size of the PSU, on the one hand, and the relative size of the cluster, on the other (Figure 3.1).

3 "Among diverse selection procedures, clustering often has be far the greatest effect on both the variance and the cost. ... The effects of clustering on the variance come from two sources. First, the selection consists of actual clusters of the physical distribution of the population. Second, the distribution of the population in those clusters is generally not random. Instead, it is characterized by some homogeneity that tends to increase the variance of the sample" (Kish, 1965, p. 162).

Figure 3.1 Typology of Cluster Sample Designs



Considering the X-axis (proportion of the population of the Primary Sampling Unit included in the sample), at one extreme we might imagine selecting a single element within each PSU, whilst at the other extreme we might conduct a census of all local residents. As we move from the left to the right along this axis, assuming constant sample size, the cost of the survey decreases, but there is also a reduction in the efficiency of the estimates. There is therefore a need to identify an optimal compromise between these competing constraints. Considering the Y-axis (population size of the PSU), at one extreme we have PSUs with a single resident, whilst at the other extreme we have a single PSU which encompasses the entire national territory. As we move from the bottom to the top along the vertical axis, the cost of the survey initially decreases, as the degree of clustering rises, before subsequently increasing, as the growing size of the PSUs reduces clustering once again, and the efficiency of the estimates follows a similar trajectory, initially declining then rising. For this reason, we must adopt an optimal compromise based on the available resources and the precision requirements of a given survey.

This illustrates a very basic point: PSUs should be small enough to permit the cost savings associated with the geographical clustering of interviews (short travel times between all pairs of dwellings within the area), but large enough to facilitate survey operations (enough dwellings to provide interviews for a period long enough to enable the interviewer to return several times to non-responding households). The simulation study will examine sample sizes of 15 and 20 households within each PSU to allow the CSO to evaluate the relative cost-effectiveness and statistical efficiency associated with these two second-stage sample sizes.

It should be noted that the overall population size of PSUs should be set to the maximum value that allows for acceptable travel times between pairs of dwellings. This means that it may be necessary to pre-aggregate Small Areas (SAs) to obtain a higher minimum population than the current value of 65 households (which was determined with a view to protecting confidentiality)[4]. The reason for this is once again related to spatial structure, namely the tendency for the members of neighbouring households to be more similar to one another than the members of households that are far apart. By increasing the population size of PSUs, we obtain more heterogeneous PSUs[5]. As spatial autocorrelation in socio-economic variables does not decline in linear fashion over distance, there is a threshold point at which the variability of households within PSUs begins to increase more rapidly (equivalent to passing from a single street to a more mixed neighbourhood), before subsequently levelling off. If we must choose between sampling a large proportion of a small, homogeneous PSU, or a smaller proportion of a larger, more heterogeneous PSU, then there is a clear preference for the latter.

Given the costs involved in defining the boundaries of thousands of PSUs, we cannot freely choose how these are constructed. If we could, then we would certainly seek to optimise the aforementioned criteria, obtaining heterogeneous but compact areal units with relatively large populations and short average travel times between pairs of dwellings. In practice, we are forced to rely on existing spatial units such as Census Enumeration Areas or Small Areas, which are generally constructed with different aims in mind (maximum internal homogeneity and minimal size).

An optimal definition might be determined empirically by calculating the intra-class coefficients for various aggregations of small areas with increasing population (i.e. moving from a minimum of 65 households to 120, 180, 240 etc.) for a range of possible variables of interest. Although this falls beyond the scope of the present analysis, and would involve the use of complex, iterative spatial aggregation routines, it is technically feasible and of undeniable interest. As far as the present analysis is concerned, the simulation study will simply assess the impact of a sample size of 15 and 20 households at the second stage of sampling, taking the PSU for the moment as coinciding with Small Area census units.

Separate consideration should be given to the question of how auxiliary data for the PSUs should be obtained, including the current population (i.e. at the moment of each survey) and the variables needed for stratification and estimation/calibration procedures. We believe that the Census of Population is the most appropriate source of this information. Carried out every five years in Ireland, the census represents a

4 It is worth noting, in this regard, that roughly 13% of SAs (representing approximately 7% of the population) had fewer than 65 households in 2006.

5 "Larger clusters tend to have smaller roh, and this is one reason for subsampling, rather than confining the sample into compact clusters. ... Second, the sampler may try to assemble dissimilar units to form the sample clusters. ... For example, with systematic selection a sample of four dwellings can be spread around the block; this should produce less homogeneity than a cluster of four adjacent dwellings" (Kish, 1965, p. 164).

(relatively) timely source of disaggregate data, which will continue to be available at SA level[6]. If the population of a SA increases dramatically, it will be sub-divided, resolving a potential problem in the maintenance of the sampling frame. However, appreciable changes are possible, at local level, between successive census waves, most crucially in relation to the total population.

In small areas, the construction of a new apartment block or housing estate, or a period of intense emigration, can have an appreciable impact on the total population (Haase, 2009). If these changes are not correctly estimated, then the areas in question risk being either under- or over-represented in the sample, leading to potential bias.

There are various ways of producing small-area population estimates, which may be applied in isolation or in combination: (1) use data from other sources, such as the GeoDirectory[7], or administrative databases to estimate population change between census waves; (2) build a predictive model using, for example, small area estimation techniques with auxiliary data; (3) adopt a sampling design based on the selection of a fixed *proportion* rather than a fixed *number* of dwellings within each PSU, combined with a random or systematic sample of PSUs. As manual enumeration is costly, both in terms of the additional tasks that must be carried out and the difficulties it creates in organising the workload of interviewers, some combination of (1) and (2) would be preferable.

This implies a need to estimate population change at various spatial scales, most likely on the basis of a model of demographic change that applies estimated change parameters to distinct demographic groups, whilst taking account of both natural change (births and deaths) and migration flows. We recommend that the CSO assesses the capacity of existing techniques to provide satisfactory estimates of population and demographic change between census waves. We recommend carrying out an assessment of the potential role of census and non-census data in the estimation of population change at the level of small areas, perhaps using recent advances in small area estimation techniques and taking advantage of new administrative databases.

The distribution of the resident population between local areas, strata and regions is of great relevance to sample surveys, as it has an impact on stratification, selection probabilities and estimation. The changes between successive census waves may be considerable, particularly at higher levels of disaggregation and during periods of rapid economic change. In general terms, the most dramatic changes result from

6 Due to the temporal stability of spatial patterns of relative affluence and deprivation, it would be perfectly feasible to use decennial census data.

7 The availability of new databases of residential and non-residential addresses (including the GeoDirectory and the register of dwellings for the 2011 Census) has the potential to overcome the need for on-the-ground enumeration of addresses prior to selection. If found to be sufficiently accurate and up-to-date, this tool could also alter the nature of the trade-off between costs and benefits when considering the dimensions of PSUs.

rapid growth (new developments) rather than decline in the resident population. The only way to identify small areas undergoing rapid growth during the intercensal period is by using auxiliary data. Where a fixed sample size is drawn within each PSU, the latter being selected with probability proportional to size, revised estimates should be obtained for all PSUs prior to sampling (using not only a demographic model, but also the percentage of new addresses registered within the GeoDirectory, for example). This would permit the inclusion of new dwellings within the sampling frame. If this is not possible, then visual enumeration should be carried out within each of the sampled PSUs, in order to identify anomalies, which could be assigned to a separate stratum. We recommend creating a specific stratum for "surprises" encountered during data collection, following Kish's well-known recommendation (Kish, 1965). This stratum could be used to reweight PSUs found to contain a larger population than expected. A weighting factor would be applied to observations from PSUs undergoing rapid population growth in order to ensure that they have the appropriate influence on estimates[8].

3.3 Selection of PSUs

If PSUs are based on the new census geography of Small Areas (SAs), they will have relatively uniform populations, as this is one of the characteristics of these new spatial units. This is a beneficial characteristic, as the negative effects of clustering are greatest when the size of clusters varies considerably (Pascal Ardilly & Tillé, 2006). The Small Areas should have a minimum population size in order to ensure that all units can be included in the selection process. However, urban SAs are likely to have systematically larger populations than rural ones. This difference has the potential to give rise to bias if a fixed second-stage sample size is used; if we were to collect a simple random sample of SAs, urban residents would be under-represented in the overall sample, to the extent that urban SAs have a larger average population than rural SAs. The same applies to any other factor associated with above-average or below-average SA populations, such as residential expansion, demographic decline or population density.

These two considerations (the attractiveness of a fixed second-stage sample size, and the correlation between the population of the PSU and variables of potential interest) lead us towards a preference for PPS sampling, using the estimated current resident population of the PSU (Rao, 2006, p. 119). This has the attractive side effect of fixing the sample size at the outset. If all household members are included in the sample, then this ensures equal probability of selection for all elements of the target population, a desirable property for sample surveys which simplifies estimation/weighting procedures. Sampling should be carried out without replacement, in line with common practice, which is not only intuitively compelling,

8 This weighting factor could be defined as the inverse of the average ratio between the expected and the observed population of the PSUs, in order to compensate for the underestimation of the probability of selection during initial sampling of PSUs. The PSUs allocated to this stratum could be identified using auxiliary data or on-the-ground enumeration.

but also produces estimates with a lower variance (Fellegi, 1963, p. 183)[9]. We therefore recommend that the PSUs be selected randomly but with probability of selection proportional to size (PPS), where the latter is measured by the number of households:

In two-stage cluster sampling designs, PPS sampling is sometimes used for the first-stage units, that is, the clusters (for example regional units from the population of regions, enterprises from a business register, etc.). An equal probability or self-weighting sampling design is obtained if the elements are sampled from the sampled clusters with an equal sample size. (Eurostat, 2008, p. 23)

3.4 Stratification of PSUs

As we noted above, the spatial clustering of survey respondents yields cost savings but tends to reduce the statistical efficiency of the estimates. The latter can be mitigated, however, by stratifying the sample of PSUs prior to selection, using auxiliary information. The rationale for this procedure is straightforward (albeit somewhat counter-intuitive). To the extent that the variables of interest are correlated with a set of auxiliary variables, the efficiency of the estimates can be increased if we stratify the sampling units in accordance with their scores on these variables (Pascal Ardilly & Tillé, 2006). Stratification essentially removes differences between stratum means from the sampling error (Kish, 1965). In practical terms, each stratum is treated as a distinct sample, and the size of each stratum should be proportionate to its incidence in the target population of units[10].

This ensures that random selection does not lead to the inclusion of a disproportionate number of PSUs with extreme values for the variables of interest and, by reducing the sampling error of estimates within strata, yields efficiency gains. Stratification is widely used in sampling designs, as cluster sampling enhances the gains that it yields. The benefits produced by stratification at the aggregate level do not necessarily apply to all sub-classes, and are typically lost when estimating small classes. Kish provides the following overview of its role:

Success in choosing strata with great heterogeneity among their means determines the gains made by proportionate sampling. ... Generally we obtain only small or moderate gains from proportionate sampling of elements, because the variables available for stratification, such as age and sex, do not separate the population into very homogeneous strata. Variables with the high relationships necessary for large gains are rarely available for stratification. (Kish, 1965, p. 88)

9 The term "replacement" refers here to the sampling process itself (i.e. units can only be selected once for inclusion in any given sample). In longitudinal surveys such as the SILC, a different form of replacement is used to substitute households that drop out due to sample attrition, seeking to ensure that the number of substitutions is as low as possible.

10 Alternative procedures exist, but are less attractive in the present context.

The considerations at stake here are, in a certain sense, the opposite of those set out earlier in relation to the definition of PSUs. Whilst PSUs should, ideally, be as heterogeneous as possible in composition, strata should be internally homogeneous. Stratification is often carried out using regional identifiers and/or a typology of areas (metropolitan region, small town, rural, etc.). The Swedish Labour Force Survey is unusual in using individual-level data from a population register, and the sample is stratified by 192 categories formed from the intersection of sex, nationality, employment status and county (Bergman, Kristiansson, Olofsson, & Safstrom, 1994).

On theoretical grounds, it is possible to show that the optimal stratification variable for a sampling design that aims to estimate the population value of a specific variable coincides with that outcome variable itself. In other words, if we are interested in estimating the number of employed people in the country, the best stratification variable would be the proportion of employed people in each PSU. This extreme example helps to illustrate the potential of stratification as well as the challenges and difficulties that it poses during the development of a sampling design.

If we can identify a single stratification variable (or vector of distinct variables) that is strongly associated with the variables of interest, then we can achieve efficiency gains (smaller standard errors). The first difficulty is that any single variable is likely to be efficient when estimating similar variables, but much less so when estimating others; given the large range of applications of household sample surveys, this is not a trivial issue. The second difficulty is that stratification variables must measure relatively stable attributes of PSUs, which are not subject to rapid change. The third difficulty is that the stratification variable must be available for all PSUs, implying that it must be estimated with an acceptable degree of precision at the small area level, and thus cannot be derived from previous sample surveys.

In general household surveys, health surveys and labour force surveys, many of the variables of interest are correlated with the socio-economic status of households. Empirical studies at both individual and aggregate level have consistently found a powerful "social gradient" in responses to key outcome variables (Wilkinson, 1996). Individual-level analyses using aggregate-level measures of socio-economic status have replicated these findings. This provides a rationale for seeking to identify a broadly-based measure of aggregate socio-economic status for use in stratification.

Several British household surveys use the proportion of families in various socio-economic status groups to stratify the sampling frame of PSUs (which are often defined as Super Output Areas or Postcode areas), along with one or more other variables (geographical region, area type, population density, the unemployment rate, household tenure or car ownership). This technique is used, for example, in the Family Expenditure Survey (from 1996-7), the Family Resource Survey, the General Lifestyle Survey, the Living Costs and Food Survey and the Time Use Survey.

These stratification designs are moving in the right direction, in our opinion, although the simple combination of census variables may not be the best solution, as these are prone to measurement error and may not lead to the identification of homogeneous categories. Rather than adding more and more variables to the stratification, we believe that these should be analysed in an appropriate way before being deployed.

The classification used in the aforementioned (British) surveys is derived from the National Statistics Socio-economic Classification (NS-SEC), which is based on the Goldthorpe social class schema (Rose & Harrison, n.d.; Rose, Pevalin, & O'Reilly, 2005). This classification is not particularly suitable as a stratification variable, for various reasons: (1) as a categorical *individual-level* variable, when aggregated this yields the percentage of households in each category, which raises the question of how these categories should be combined[11]; (2) it is an occupational classification and cannot distinguish between the conditions of people within categories such as the long-term unemployed, those who have never worked, students and those who provide incomplete occupational descriptions; (3) it is based on a single member of each household (principally the household member who is responsible for owning or renting the accommodation), which means that it may understate the degree of disadvantage and fail to discriminate between households with multiple earners and those with a single earner.

In applied sociological research, a number of aggregate measures of socio-economic status have been developed in recent years. For example, Haase and Pratschke conceptualise this as a latent, multidimensional construct based on a set of aggregate-level indicators (Haase & Pratschke, 2005, 2008; Haase, Pratschke, & Gleeson, 2012; Pratschke & Haase, 2007). The resulting set of scores reflect the overall affluence/deprivation of local areas and are readily available for Small Areas. The scale is continuous, extending from the most affluent to the most deprived areas, and the distribution of the index follows that of the component variables (i.e. it is not "normalised" or transformed). The scores are measured using a Confirmatory Factor Analysis model which yields a measure that is highly stable over time and has the attractive property of capturing both individual-level compositional effects and the influence of neighbourhood concentrations of deprivation.

There is only one precedent, as far as we are aware, involving stratification by a deprivation index, namely the Scottish Health Survey, which is stratified by Health Board and by the 2006 Scottish Index of Multiple Deprivation. The reason for adopting this approach was the perceived need to over-sample deprived areas. The Scottish Index of Multiple Deprivation relies primarily on administrative data, and is therefore not applicable in the Irish context[12].

11 In many cases, a single measure is used, such as the proportion of households with reference person in a non-manual occupation (e.g. British Crime Survey, Health Survey for England, General Lifestyle Survey).

12 It is also important to note that the Scottish Health Survey does not use geographical clustering, as addresses are extracted directly from the Postcode Address File. These addresses are stratified primarily by Health Board (12 geographical regions), distinguishing only between "deprived" and "non-deprived" areas within these regions, the former being defined as "datazones" (areas) within the most deprived 15% of areas. The deprived areas were over-sampled.

To summarise, socio-economic classifications and deprivation indices are sometimes used as stratification variables for health and social surveys, although they are generally cast in a subordinate role, with a view to identifying deprived areas or those with high (or low) proportions of non-manual employees. The potential value of deprivation indices in stratification for sample surveys is not widely understood, which is due, in many cases, to the nature of many of the existing indices. As these are often unstable over time, data-driven and primarily concerned with the identification of highly-disadvantaged areas, they cannot provide an over-arching framework for stratification. Methodological studies of stratification variables are generally limited to individual census variables, yielding inconclusive results, with certain census variables having strong associations with certain survey questions but not others (Bruce, 1993).

We recommend using the Haase-Pratschke Index of Relative Affluence and Deprivation, applying arbitrary cut-off points[13] or deciles. This innovation would be without bias and would not influence the expectation of the estimators, which means that no interruption would occur in the data series. Because affluent rural areas may have different characteristics to affluent urban areas, and disadvantaged rural areas may differ from disadvantaged urban areas, it may be advisable to construct a multi-criteria stratification measure by crossing this variable with a measure of area type or region[14]. This would ensure that the main urban areas and towns are correctly represented in the final sample, as well as taking into account the potential differences between areas[15]. Tables 3.1 and 3.2 show the means for a range of potential variables of interest, derived from the 2006 Census of Population, by area type and deprivation category. In Part Two, we will report simulation results that show the impact of these different types of stratification on the precision and accuracy of sampling estimators in Ireland.

In larger surveys, it may be worth considering a third criterion (alongside deprivation score and area type), such as the 8 NUTS 3 regions, for a theoretical total of 5x5x8 = 200 strata. Although the number of strata would, in practice, be lower than 200 (as some regions do not have the full range of population density strata), this approach would only be feasible in the context of a large survey such as the QNHS. In smaller surveys, it would be advisable to limit stratification to deprivation score and area type, resulting in a maximum of 5x5 = 25 strata[16]. If a more precise approach to the definition of strata boundaries is required, Lavallée and Hidiroglou (1988) suggested

13 Such as +/- 0.25 standard deviations; +/- 0.25-1.0 standard deviations; +/- 1.0 standard deviations to maximum value.

14 e.g. (1) Dublin County Borough or suburbs of Dublin; (2) Other County Boroughs or suburbs; (3) large towns with a population of more than 5,000; (4) mixed urban/rural areas bordering larger town, town with a population between 1,000 and 5,000, mixed urban/rural areas bordering small town or town with a population below 1,000; (5) rural areas.

15 “Generally, more gain accrues from the use of coarser divisions of several variables than from the finer divisions of one. ... When using several variables, there is no need for completeness and symmetry in forming the cells; smaller and less important cells may be combined. ... Stratifying variables unrelated to each other (but related to the survey variables) should be preferred." (Kish, 1965, p. 101)

16 To facilitate comparisons, the simulation study will employ a standard stratification using 5x5 = 25 strata.

an iterative algorithm that yields optimal boundaries for a given variable. Their algorithm can encounter difficulties during convergence, but alternative methods have been proposed (Baillargeon & Rivest, 2011).

Table 3.1 Strata Means (%) for Area (8) Typology

	Dublin + environs	Other cities + environs	Towns over 10,000	Towns 5,000 to 9,999	Towns 1,500 to 4,999	Towns 1,000 to 1,499	Towns under 1,000	Open country-side
p_employed	59.1	54.3	59.4	58.8	59.8	54.0	58.0	57.3
p_unemployed	5.9	6.0	6.6	6.8	6.2	7.6	6.4	4.2
p_student	11.4	15.1	10.2	8.3	7.9	7.5	7.5	9.8
p_homefam	9.4	10.1	10.4	11.5	11.4	13.0	12.9	13.4
p_retired	10.5	9.6	9.3	9.9	10.3	12.2	10.5	11.4
p_disability	3.3	4.5	3.9	4.3	4.2	5.3	4.4	3.7
p_other_pstat	.3	.3	.3	.3	.3	.3	.3	.3
p_owner_occ	67.9	63.2	69.8	71.7	70.6	70.6	73.1	86.7
p_priv_rented	15.3	18.5	13.5	12.2	12.6	11.1	8.3	6.0
p_LA_VB_rented	12.8	14.7	13.0	13.1	13.3	15.1	15.6	4.8
p_LL_illness	9.5	9.8	9.1	9.5	9.0	10.3	9.7	8.4
p_Irish	82.6	84.6	83.5	83.8	84.0	87.1	89.7	91.5
p_Traveller	.2	.4	.6	.8	.6	1.1	.2	.3
p_other_white	10.0	10.2	10.5	11.2	11.2	9.3	7.6	6.3
p_non_white	5.4	3.4	3.7	2.7	2.5	1.4	1.3	.9
p_lo_educ	15.2	13.3	14.5	15.6	16.9	21.9	18.9	19.3
p_hi_educ	35.3	33.0	29.0	25.7	25.6	19.5	23.0	24.4
p_single	49.3	51.1	44.8	42.4	43.1	42.4	41.7	37.0
p_married	40.6	39.2	44.2	46.0	45.1	44.7	46.0	53.3
p_sep_div_w	10.2	9.8	11.0	11.6	11.8	12.9	12.3	9.7
p_age1529	34.0	37.6	33.0	30.8	30.5	29.0	28.8	25.5
p_age3044	28.4	26.5	31.1	31.7	32.1	28.7	32.6	28.9
p_age4559	20.3	19.5	20.7	21.0	20.2	21.7	20.8	25.0
p_age6074	12.2	11.9	11.1	12.0	12.2	14.1	12.5	14.4
p_age75ov	5.1	4.5	4.1	4.5	5.0	6.5	5.3	6.3
p_male	48.4	48.4	49.0	49.6	49.4	50.1	49.3	51.2
p_female	51.6	51.6	51.0	50.4	50.6	49.9	50.7	48.8

Table 3.2 Strata Means (%) by Deprivation (5) Typology

	Below – 1 STD	-1 STD to -.25 STD	-.25 STD to .25 STD	.25 STD to 1 STD	Above 1 STD
p_employed	45.7	53.9	58.5	63.4	65.8
p_unemployed	10.9	6.2	4.6	3.8	2.2
p_student	8.8	9.5	10.2	11.0	13.4
p_homefam	13.6	12.8	11.9	10.3	9.3
p_retired	13.2	12.5	11.0	8.8	7.7
p_disability	7.4	4.7	3.5	2.4	1.5
p_other_pstat	.4	.3	.3	.2	.2
p_owner_occ	64.7	76.6	80.8	78.5	76.2
p_priv_rented	5.8	9.2	10.4	13.2	16.2
p_LA_VB_rented	25.8	11.0	5.8	5.2	4.8
p_LL_illness	13.6	10.2	8.6	7.2	6.1
p_Irish	90.8	88.2	87.5	84.6	84.0
p_Traveller	1.2	.5	.2	.1	.0
p_other_white	4.8	7.8	8.7	10.3	10.5
p_non_white	1.6	2.1	2.3	3.6	4.3
p_lo_educ	32.1	21.8	16.3	10.5	6.0
p_hi_educ	12.3	20.5	26.0	35.6	47.9
p_single	45.9	41.7	40.5	42.4	45.6
p_married	39.6	46.6	49.3	49.0	47.4
p_sep_div_w	14.5	11.7	10.1	8.5	7.0
p_age1529	29.2	28.3	28.6	31.4	34.2
p_age3044	25.3	26.7	28.6	31.8	32.8
p_age4559	23.1	23.3	23.3	21.5	20.0
p_age6074	15.9	15.0	13.7	11.1	9.4
p_age75ov	6.5	6.7	5.8	4.3	3.6
p_male	48.6	50.3	50.5	49.8	49.0
p_female	51.4	49.7	49.5	50.2	51.0

3.5 Selection of Secondary Sampling Units (SSUs)

The simulation study evaluates a range of different second-stage sample sizes, including 15 and 20 households within each PSU. The latter figure is higher than the 15 households recommended by Prof. Steel in his 1997 report, but well below the figure of 45 previously used in the Annual Labour Force Survey. These figures are also roughly in line with international practice[17]. Given the data and techniques presently available, some degree of clustering is essential in order to reduce the cost and burden associated with official household surveys[18].

For the purposes of sampling design, the household is typically assumed to constitute a residential group associated with a single dwelling. Although this is a reasonable assumption, some people do not live in private households and therefore risk having a selection probability of zero. This group comprises groups which are difficult to sample (such as the homeless and Travellers who do not live in halting sites or permanent accommodation) as well as members of residential communities (such as convents and nursing homes), those who are in an institutional setting (such as hospitals and prisons) and hotel guests. As Prof. Steel observes in his 1997 report, the rotation of a very large non-private dwelling in or out of the sample may contribute to volatility over time, and these dwellings range greatly in size[19].

For this reason, if it is deemed necessary to include them, non-private dwellings should ideally be considered separately, as they do not form part of the main target population of private households, and their identification within PSUs is problematic. A more satisfactory approach would rely on a separate sampling frame of non-private households, to be sampled using appropriate procedures (e.g. estimation of the resident population in non-private households from Census data, selection with probability proportional to size and systematic selection of a fixed number of individuals from a list of residents). We recommend that an assessment of the viability of this approach be carried out by the CSO.

17 For example, the French Labour Force Survey is based on a cluster sample of "around 20 dwellings", drawn from areas which are selected by SRS within "sectors" containing 120-240 dwellings, which are chosen by PPS within each of the PSUs, which are selected using PPS. Sectors are only divided into areas once they have been sampled. The PSUs are large units, equivalent to Municipalities in many cases (P. Ardilly & Osier, 2007).

18 Recent developments in multi-level modelling techniques draw attention to an unexpected benefit of cluster designs, which has to do with the analysis of contextual influences in relation to social phenomena. These techniques (which have come to a position of dominance in a large number of applications in the field of spatial, sociological and epidemiological modelling) use clustered data to partition the variance of a variable (or set of variables) between various levels from the individual to the household, school and neighbourhood. In this way, it is possible to explore the influence of the spatial and temporal context on individual behaviours and outcomes.

19 Current household surveys in Ireland, including the QNHS, SILC and HBS are only concerned with individuals living in private households.

As far as the selection of private households is concerned, the simplest approach relies on random selection[20]. Although a systematic sample of dwellings may produce estimates with a slightly lower variance, at least in the presence of specific patterns of auto-correlation (Chang & Huang, 2000; Iachan, 1982), the periodicity of systematic sampling can give rise to bias and selection errors can be difficult to detect[21]. Random selection has the added advantage of simplifying estimation procedures following data collection and can be reproduced in the context of simulation studies, allowing a closer alignment between experimental studies and real-life surveys (Levy & Lemeshow, 1999).

This approach would be relatively straightforward to apply if a listing of residential addresses can be constructed that is sufficiently accurate and up-to-date to permit the selection of Secondary Sampling Units (SSUs) to be carried out centrally. Up until now, interviewers have carried out a visual enumeration of dwellings prior to selecting households for interview, which imposes a cost that is roughly proportional to the geographical size and population of the PSU. By using the GeoDirectory or a similar register of residential addresses, it may be possible to select the dwellings to be included in the sample as well as calculating, in advance, an optimal interview order and route for interviewers. This would be advantageous in economic terms as well as facilitating improvements in the coordination of survey operations.

Although second homes, uninhabitable dwellings and certain kinds of commercial or mixed-use buildings may be difficult to identify using registers of addresses, this difficulty could be accommodated within the sampling design. The order of visits could be optimised using GIS techniques to reduce interviewer travel time, perhaps using a mobile device to assign workloads and provide directions to the next household to be visited. This would enable greater control over return visits to households where it was not possible to conduct an interview on previous occasions, perhaps using an interview management system that schedules and assigns cases.

In the case of the QNHS, the number of interviews to be carried out within each PSU should ideally be completed within one week, and should use the same reference week. Follow-up interviews should similarly be organised within a specific week of the subsequent month or quarter, to ensure that the required time interval between successive interviews is respected[22].

---

20 “Departures from simplicity must be justified by strong considerations” (Kish, 1965, p. 23).

21 Moreover, there is no known analytical variance estimator for the design variance of a systematic sample, which implies the need to use approximations or non-parametric estimation techniques.

22 The current requirement for CSO sample surveys is that 80% of interviews should be performed during the week following the reference week, whilst the remainder should be carried out within two weeks. Operational changes with the potential to increase the percentage of interviews in the former category should be evaluated positively.

3.6 Effect of proxy interviewing and mixed mode data collection

Studies of the impact of proxy interviewing on data quality have been carried out and the results of these studies are reported in the international statistical literature (Thomsen & Villund, 2011). The evidence suggests that proxy response is an acceptable technique for gathering data on family members who are not present at the moment of the interview, although it tends to increase overall error. An exception to this overall finding relates, somewhat predictably, to sensitive information which is particularly subject to social desirability effects (e.g. substance use), where proxy response is generally not satisfactory.

A related issue is the mode of data collection, which has the potential to generate bias and to influence the variance of estimates. National statistical agencies began experimenting with techniques for remote data collection during the 1980s, and the statistical literature once again provides an extensive treatment of its influence on estimates:

Whilst the face-to-face interview was the gold standard in the fifties and sixties of the twentieth century, the telephone survey quickly became popular during the seventies and soon became the predominant mode in the U.S.A. (see Nathan 2001). The popularity of telephone surveys led to a new mixed-mode approach as mixes of face-to-face and telephone surveys were implemented. For instance, beginning in 1984, the British Labour Force Survey used telephone interviews in a quarterly panel design... (De Leeuw, 2005, p. 234)

The main modes of data collection are face-to-face interviewing, self-administration and telephone interviewing (De Leeuw, Hox, & Huisman, 2003), and each of these has a computer-assisted variant (CAPI, CASI, CATI). In recent years, national statistics agencies have experimented with web-based interviewing techniques (CAWI), and more than eight million families completed the 2011 Italian Census of Population using a dedicated web portal.

Although official household surveys in Ireland have traditionally been carried out using face-to-face interviews (computer-assisted in recent years), there is a strong case for considering the role of computer-assisted telephone interviewing in the context of the Quarterly National Household Survey, where households remain in the sample for several waves before being replaced[23]. Multiple-mode surveys are now widely-used at international level (including the Swedish Labour Force Survey, in which all but a small percentage of interviews are conducted by telephone), and have several advantages.

The first advantage stems from the use of distance interviewing, and is of an economic nature. The second relates to CATI as a mode of data collection, which produces additional cost savings, ensures effective case management and a high

23 The SILC also involves repeated interviews, one year apart, but the complexity and length of the questionnaire would effectively exclude telephone interviewing.

level of standardisation of interactions between interviewer and participant. Methodological studies during the 1980s (cf. Weeks, 1992) showed that CATI can improve data quality when compared with paper-and-pencil survey techniques. Thirdly, a larger number of contacts at different times of day is possible in order to obtain an interview, compared with face-to-face interviewing. Fourthly, telephone interviews are typically less intrusive and may be more willingly accepted by participants (e.g. elderly people), particularly if they have already participated in at least one interview and provided contact details. Having collected socio-demographic information at first interview, subsequent interviews are likely to be relatively brief and focused on the changes that have occurred since the last interview (this strategy is used for Labour Force Surveys in several countries). Finally, it may be easier to contact persons who have moved out of an area, permitting greater control over this potential source of non-response and bias[24].

The CSO has decided to implement this mode of interviewing for follow-up interviews, and a detailed plan to achieve this objective has already been established. This is a sound decision, in our view, and De Leeuw (1992) reports that only small differences have been found between face-to-face and telephone surveys, which are limited to item-level non-response and do not concern validity or social desirability bias (cf. Dixon, 2001). If the savings derived from this innovation are used to increase the overall sample size, to reduce degree of clustering or to increase the frequency of surveys, the net gains would be substantial. This does not exclude the possibility of using face-to-face interviewing in a targeted fashion (i.e. within a sequential multi-mode system), with the aim of reducing unit non-response (De Leeuw, 2005).

3.7 Rotation patterns for panel designs and frequency of repeated surveys

An important aspect of many official sample surveys relates to the estimation not only of the population values of various quantities, but also of change over time in those quantities (increase in the population at work, change in the number of households in poverty etc.). Change over time can be studied using various sampling designs (fresh sample for each survey, partial overlap of sample across different surveys, full panel design), and most national statistics agencies incorporate a panel element in repeated surveys.

When developing sampling designs for repeated surveys, excessive emphasis is often placed on cross-sectional estimates, with the assumption that good measures of *level* translate automatically into good measures of *change*. There is, however, a difference between these two kinds of estimators, and there are differences between

24 Where sample attrition is due to unobserved characteristics that are related to the phenomenon of interest (e.g. finding a job, in the context of a Labour Force Survey), the resulting missing data pattern is "non-ignorable" in statistical terms. Dorsett (2010) shows that re-weighting based on variables available in the sample frame does not resolve this problem. However, if weights can be constructed to reflect outcomes since the time of sampling, bias can be minimised.

short-term and long-run changes, with alternative sampling designs producing more or less precise measures of one or the other (P. A. Bell, 1998).

This is closely related to the question of rotation patterns. The most important reason for using a rotation pattern, within a sample survey, aside from considerations of cost[25], has to do with the efficiency of estimates of change: by maintaining all or part of the sample elements constant between waves, the standard error of the estimate of change between waves is lower than when distinct samples are used (Kish, 1965, p. 474; Kolenikov & Angeles, 2011). The panel design reduces variations in scores and produces a smoother data series. For this reason, many surveys (including the QNHS and SILC) adopt this procedure.

However, there are also drawbacks to this kind of design, starting with the well-known phenomenon of "rotation group bias", whereby survey participants' responses are influenced by the mere fact of having responded to a similar questionnaire on a previous occasion (Bailar, 1975). The levels of certain characteristics have been found to vary depending on the length of time households remain within a panel survey. Panel attrition can exacerbate problems deriving from unit non-response and the integration of panel and non-panel data components can increase the complexity of estimation procedures (but see Hox, 2000). Finally, previous work suggests that wave-to-wave sample overlap can reduce the precision of trend estimation:

Tallis (1995) suggested that high overlap between successive surveys for the LFS reduces the ability to detect turning points in the economy. This and work by Sutcliffe and Lee (1995) suggest that a sample rotation pattern with no month to month overlap would provide better estimates of the underlying direction of the series. (P. A. Bell, 1998, p. 1)

However, a rotation pattern with no overlap gives poorer estimates of wave-to-wave change (what is termed "lag one movement"):

The sampling error on this estimate depends not just on the sampling error on the level estimates but also on the correlation between estimates from the two months. The best estimates of movement will result from a high correlation - this can often be obtained by retaining a large portion of the sample common to the two months. (P. A. Bell, 1998, p. 3)

The challenge, when developing rotation designs, is therefore to anticipate the requirements of the users and to identify the most appropriate compromise between these two aims. As Bell (1998) points out, Labour Force Surveys are typically used to estimate trends, and sophisticated users are typically wary of reading too much into

25 Finally, the panel design can lead to substantial cost savings, as a shorter questionnaire can typically be administered during second and subsequent waves, as consent is often easier to obtain and it may be possible to use distance interviewing techniques, given that contact has already been established with the survey participant (e.g. CATI).

lag one movement. It therefore makes sense to focus on rotation designs that yield efficient estimates of both short-term and longer-term trends.

Letters are typically used to represent different rotation designs in periodic surveys, where the degree of "lag one overlap" is given by *P*[26]. We will discuss this issue with reference to the Quarterly National Household Survey, which currently relies on an "in-for-5" (or "5-in") rotation pattern, whereby one-fifth of the sample is replaced each quarter, leading to a lag one overlap of 80 per cent and a lag four (or annual) overlap of 20 per cent. Many labour force surveys adopt a similar design, including a "six-in" design in the case of the monthly Canadian Labour Force Survey (Fuller & Rao, 2001, p. 45; Gambino, Singh, Dufour, Kennedy, & Lindeyer, 1998), an "8-in" design in the case of the Australian LFS (Bell, 1998) and a "4-in-8-out" design in the case of the American Current Population Survey, where households participate in 4 monthly surveys, are excluded from the sample for 8 months, and subsequently interviewed on 4 more occasions (Bailar, 1975).

The Swedish Labour Force Survey - which is sampled using individual data from a register of the total population - involves a rotation pattern of "1-in-2-out" for eight interviews, based on a monthly survey (Bergman et al., 1994, p. 183). The Finnish Labour Force Survey also relies on a population register, and involves a monthly sample with a complex rotation pattern (Statistics Finland, 2011). Statistics Canada uses a "6-in" pattern, and the Japanese statistics agency uses a "2-in-10-out" (where households participate in 2 monthly surveys, are excluded for 10 months and are then included again for 2 months), both of which increase overlap between samples one year apart, improving estimates of annual change.

As these examples illustrate, it is possible to design rotation patterns that reduce the degree of overlap between successive quarters and increase the annual overlap, with the aim of achieving an optimal balance between estimates of change over different periods of time. For example, a “2-in-2-out" rotation pattern, with a total of 6 interviews per rotation group, would ensure that half the sample remains stable between successive quarterly surveys and one-third between surveys that are a year apart. Each household would remain in the survey for a total of 30 months (assuming a quarterly survey). One new rotation group, comprising one-sixth of the sample, would be added each quarter[27].

---

26 e.g.
$P = 1/2$: ab - bc - cd - de - ef ... ("2-in")
$P = 2/3$: abc - bcd - cde - def - efg ... ("3-in")
$P_1 = 1/3$; $P_4 = 1/3$: eaf - fbg - gch - hdi - | - iej - jfk - kgl - lhm - |- min - njo ... ("2-in-3-out")

27 In a given quarter, one-sixth of the sample would be included for the first time, one-sixth would be included for the second time, one-sixth would be included for the third time (following a gap of two quarters), one-sixth would be included for the fourth time, one-sixth would be included for the fifth time (following another gap of two quarters), and one-sixth would be included for the last time.

This "2-in-2-out" rotation pattern, with a total of six interviews, would yield an acceptable overlap between quarterly and annual estimates. Whenever a new rotation group must be added to the sample, the same PSUs could be used, where possible[28]. Assuming a minimum PSU size of 65 households, and a second-stage sample size of 20, it would be possible to draw three successive samples from each PSU. As each rotation group would remain in the sample for three years on the basis of the "2-in-2-out" quarterly design (six interviews), the PSU would only need to be replaced after nine years (excluding transition to the new sampling design).

This rotation pattern is recommended by Bell (1998), who describes it as yielding good long-term estimates as well as good lag-one movement estimates. We recommend carrying out a comparative analysis of rotation designs using simulation techniques, with a view to identifying the most appropriate pattern for the Irish context. It would be possible to assess the accuracy of the estimates of change that would be produced by this design, and to compare alternative rotation patterns using simulation techniques like those described in Part Two of this report. This could be achieved by simulating change in the target population and then sampling repeatedly from the resulting populations on the basis of two or three alternative rotation patterns[29].

It is worth briefly considering how rotation patterns relate to first-stage and second-stage sampling procedures. As we noted at the beginning of this report, the current sampling design for the QNHS divides each block of 75 households into five groups of 15 households, which are introduced successively into the sample[30]. This reveals the inter-relations between different elements of the sampling design: the size of the block (PSU) is related to both the size of the cluster and to the rotation pattern. It also illustrates a second form of rotation, involving overlap between blocks. Because of the influence of spatial autocorrelation, this will have the effect of further smoothing the temporal data series over a longer period of time and reducing the sampling error of the estimates of level and change. It is cost-efficient, as block enumeration and other preparatory work is exploited to the full.

For these reasons, a common criterion in the rotation of PSUs is the exhaustion of households within a given PSU (areas are replaced when all final-stage units have been sampled). This is particularly cost-effective where manual enumeration or checking of addresses must be carried out within each PSU. It may also lead to a more trusting relationship between the survey respondents and the interviewer, who would work for a longer period of time in the local area. As we have seen, it has positive effects in relation to the variance of longer-run change scores and the

28 At the moment of transition to the new design, participants in each PSU would have to be randomly assigned to one of six rotation groups, and all but one of these groups would leave the survey before completing a full set of 6 interviews.

29 It would be feasible, for example, to apply a mathematical function to the individual-level census data (e.g. conditionally-random changes in the employment state, some degree of mixing between PSUs, population ageing, etc.).

30 The SILC also relies on a rotation pattern ("in-for-4"), as recommended by Eurostat.

smoothness of the data series. It may therefore make sense to draw multiple samples from PSUs, replacing each rotation group with another from the same PSU until all final-stage units have been exhausted.

The introduction of a new rotation pattern is a complex and costly step, which must be managed carefully in order to ensure that this transition does not provoke undue difficulties for field operations or for the maintenance of important data series. Eurostat is currently considering alternative rotation patterns for European Labour Force Surveys and is likely to publish its recommendations in the near future. It therefore makes most sense for the CSO to postpone any decisions about rotation patterns until new European guidelines have been formulated, whilst continuing to explore and discuss the properties of alternative approaches.

The consultants were also asked to evaluate the possibility of conducting a *monthly* rather than a *quarterly* household survey. The rotation pattern described above would be well-suited to this kind of design, as the "2-in-2-out" rule would reduce month-to-month correlations. If a longer duration is required, each household could remain in the sample for more waves (8, for example). The required precision of monthly estimates (and the relevant geographical units) would need to be decided in advance, with a view to determining the appropriate sample size. Quarterly and national estimates would then be obtained by pooling or averaging the monthly estimates. The selection procedures described above would facilitate this step, ensuring that monthly, quarterly and national estimates are representative of the same target population.

In broad terms, a monthly data series is an attractive option, as it would facilitate a more precise approach to modelling underlying trends (both seasonal and secular). Although most labour force surveys, at international level, are quarterly (with the exception of large developed countries like the US, Canada and Australia, but also Finland and others), the advantage of a monthly survey are self-evident. With a small increase in the cost of sampling, estimation and data dissemination, it may be possible to maintain an identical quarterly sample size, whilst obtaining additional monthly estimates of key variables at national level. If a decision is made in the future to adopt a monthly survey design for the QNHS, this might also be an appropriate moment in which to consider alternative rotation patterns.

3.8 Controlling for non-response

Once a survey has been carried out, it is necessary to deal with missing data. Eurostat guidelines indicate the broad approach to be adopted in relation to (a) unit non-response, and (b) item non-response. As far as the first is concerned, the most appropriate solution is provided by weighting and estimation techniques following data collection (see Section 3.9). The best way to reduce bias, in the presence of non-random unit non-response, is to use auxiliary information (to the extent that such information is available at an individual or aggregate level) for respondents and non-respondents (Haziza, Thompson, & Yung, 2010, p. 35). As far as item non-response is concerned, it is generally considered to be important to implement strategies to reduce this, as well as using some form of imputation when it does occur (cf. De Leeuw et al., 2003). The CAPI software currently used for household surveys in Ireland reduces item non-response due to interviewer error (skipped questions) and mistakes during data entry[31].

3.9 Weighting and estimation techniques

Statisticians frequently stress that estimation procedures form an integral part of any sampling design and should be assessed within the context of the objectives of each survey. This is due to the complex inter-dependence that exists between different parts of the design and the need to identify optimal trade-offs between the different requirements (cost, efficiency, precision of estimates of level and change, operational simplicity, consistency with benchmark data and other surveys, etc.). The complexity of the sampling design of large-scale official household surveys is matched by the growing sophistication and complexity of estimation procedures.

Estimation is intimately related to the question of bias, and there are a number of reasons for giving careful consideration to this issue when designing sample surveys. Various aspects of sampling design have the potential to generate bias, due to (a) the fact that PSUs (and households) are selected as integers; (b) inaccuracies in census data, in registers of dwellings as well as changes over time in relation to the stratification base; (c) non-sampling error; (d) the temporal clustering of interviews; (d) selective non-response.

For these reasons, estimation techniques are typically employed following data collection to ensure that sample data are in line with (known) population characteristics, particularly age, gender composition and region:

In modern survey sampling practice, auxiliary information is often used to improve the efficiency of estimation for a given sample, by using model-assisted estimation techniques. Thus, in addition to the sampling design, an estimation design enters on the scene. The concept of estimation strategy is sometimes used referring to a combination of a sampling design and an estimation design. (Eurostat, 2008, p. 28)

These techniques lead to a differential weighting of cases, which raises some complex issues. There is great variation between national statistics agencies in how they deal with this, although there appears to be broad agreement about the attractive properties of a broad family of estimators known as generalised regression

31 Item non-response can lead to the exclusion of units during the analysis phase, although it is true that the number of cases concerned is extremely low in the case of the QNHS - less than 0.005% of individuals each quarter, or 1 in 20,000. Regardless of the number of cases involved, however, we would advise imputing these values rather than excluding the units, as the latter decision relies on the particularly strong (and generally untenable) assumption that data are "missing completely at random".

estimators (GREG)[32], which may be used for composite estimation, calibration and variance estimation[33]. Estimates from previous waves can be integrated with other kinds of predictors into a regression-based estimating equation, ensuring coherent age-sex and regional control totals and consistent estimates:

The modified regression method is another way to provide composite estimates that can be obtained as weighted aggregates of the current survey dataset. The method targets a predetermined set of key items, for which it achieves particularly low sampling errors. ... The modified regression technique uses generalised regression on the current month's dataset after attaching new auxiliary variables z_{ti} to each unit i at time t. Here z_{ti} is a row vector with an element for each of the key items. Corresponding to these we have "pseudo-benchmarks" Z_t based on the previous month's estimates for the key items. The modified regression estimator is then given by a generalised regression step applying both the demographic benchmarks and the pseudo-benchmarks. (P. Bell, 2001, p. 55; cf. Singh, Kennedy, & Wu, 2001)

We recommend using some form of generalised regression estimator to estimate national and regional totals, means and ratios, with the relevant standard errors, making full use of auxiliary information and calibrating to benchmark data derived from the Census of Population (cf. Deville & Sarndal, 1992). Key considerations here relate to: (a) the choice between design-based, model-based or model-assisted estimators, relative to their efficiency, cost and complexity; (b) the number of revisions that estimates relating to previous waves are allowed to undergo; (c)

32 "Model-assisted estimation procedures substitute for post-stratification, and assume continuous variables, relying on weighted least squares (WLS). If the study variable is binary or polytomous, a non-linear assisting model may be appropriate ... These are all special cases of generalised regression estimators (GREG)" (Eurostat, 2008). "The GREG estimator ... is design-consistent as well as model-unbiased under the working model ... Moreover, it is nearly "optimal" in the sense of minimizing the asymptotic anticipated MSE (model expectation of the design MSE) under the working model, provided the inclusion probability, πi, is proportional to the model standard deviation σi. However, in surveys with multiple variables of interest, the model variance may vary across variables." (Rao, 2006, p. 123)

33 Composite estimates take advantage of the large positive correlation over time in responses by the same individuals to provide estimates with smaller variances (Bailar, 1975; Fuller & Rao, 2001), and have been adopted (typically using a regression estimator) in many Labour Force Surveys (Gambino, Kennedy, & Singh, 2001; Statistics Finland, 2011).
Suppose that x_i is a row vector of auxiliary variables, and that X is a corresponding vector of benchmark values.
Let w_i^A be input weights which give estimates $\hat{y}^A \text{ and } \hat{x}^A = \sum_i w_i^A x_i$.
The generalised regression estimator is given by:

$$\hat{y}^{GR} = \hat{y}^A + (X - x^A)\hat{\beta}^{GR} \quad \text{where}$$
$$\hat{\beta}^{GR} = \left(\sum_i (w_i^A / c_i) x'_i x_i\right)^{-1} \sum_i (w_i^A / c_i) x'_i y_i$$

The $\hat{\beta}^{GR}$ is a sample estimate of the parameter describing the regression of y_i against x_i in the whole population. Typically c_i = 1 is used (larger values increase the penalty for changing the weight of unit *i*) (P. Bell, 2000).

harmonisation of estimates across multiple data sources (different surveys, etc.) (cf. Knottnerus & van Duin, 2006).

Gambino *et al.* (2001) observe that the use of composite estimation in the context of the Canadian Labour Force Survey made seasonal adjustment more effective and meant that three-month moving averages could be dropped in favour of the more detailed monthly estimates (p. 68). Proprietary software applications are available to perform these kinds of calculations (POULPE, g-CALIB, CLAN, CALMAR, GREGWT etc.), although the open source programme *R* also provides a range of routines for complex survey designs. Some of these applications are capable of calibrating simultaneously at individual and household level, and this procedure, when correctly implemented, has the potential to further improve the precision of estimates (Lemaître & Dufour, 1986). We recommend that the CSO evaluate the effectiveness of these composite estimation techniques in the Irish context by integrating the appropriate sub-routines within an extended simulation study of rotation patterns and estimates of change.

A key task to be tackled during the estimation phase is estimation of the variance of the estimates. As we have seen, household surveys are often carried out using a multi-stage sample design that first selects a stratified set of geographical areas and then a sample of dwellings, interviewing all household members, applying corrections to compensate for non-response or sampling error, and calibrating/weighting to known population characteristics and using auxiliary information. Variance estimation in the presence of these kinds of complex designs and estimation techniques is not straightforward, and must take account of the clustering, the stratification, the rotation pattern and different forms of weighting (Wolter, 1985).

Whilst we can estimate the variance of our estimators empirically in the simulation study reported in Part Two of this report by sampling repeatedly from the target population, in normal circumstances we do not have this option. Techniques similar to simulation, which rely on re-sampling are frequently used in academic research and by national statistics agencies (such as the bootstrap, see Hinkley, 1988), although the latter often prefer to use analytical variance formulae, operationalised using software routines that use numerical approximations or linearisation routines to simplify computation[34]. For example, POULPE, used by INSEE in France, transforms the variance of multi-stage sampling designs into a sum of variance terms representing the contribution of each stage, and then combines these terms using efficient algorithms (P. Ardilly & Osier, 2007).

34 It may be worth observing, however, that the Australian Bureau of Statistics is moving towards the adoption of the group jackknife approach to variance estimation, with a generalised regression weighting approach. These techniques are accessed using SAS macros written by the ABS itself (P. Bell, 2000).

3.10 Precision requirements and overall sample size

In ideal terms, one would fix the desired precision for a given sample estimate and then determine the design and sample size that yield this precision for a minimum cost. In practice, it is more common to start with considerations relating to cost and operational procedures, with the subsequent aim of maximising precision within these constraints (Kish, 1965). Horvitz & Thompson provide a useful summary of the issues at stake:

Research in the theory of sampling for surveys has been concerned with the development of more efficient sampling systems, the system including both the sample design and the method of estimation. One sampling system is said to be more efficient than another if the variance or mean square error of the estimate with the first system is less than that of the second, provided the cost of obtaining the data and results is the same for both. The development of stratified, multi-stage, multiphase, cluster, systematic, and other sample designs beyond simple or unrestricted random sampling, as well as alternative methods of estimation, have all resulted in increased efficiency in specific circumstances. (Horvitz & Thompson, 1952, p. 664)

The precision of the sampling designs discussed in the preceding sections are assessed in this report using a simulation study, the results of which are reported in Part Two. Sample size will therefore be evaluated alongside (a) the efficiency of different sampling designs; (b) the required precision; (c) the variance of key study variables; (d) the geographical units for which estimates must be provided; (e) the anticipated rate of non-response; and (f) cost and operational constraints.

The key challenge of sampling design is to identify unbiased estimators which are efficient, which have the lowest possible sampling variance and which are consistent with the required survey operations.

When interpreting the results of our simulations, it is important to bear in mind the regulations imposed by the European Council (No. 577/1998, No. 430/2005) in relation to Labour Force Survey sampling. As the European Commission needs comparable information on the level and trend of employment and unemployment in member states, these regulations aim to achieve data comparability. In line with Regulation No. 577/1998, Member States must:

- if possible, conduct a LFS each year, on a continual basis, yielding quarterly and annual results
- gather data on the situation during the "reference week" (generally the week preceding the interview), without exceeding a 5 week interval between the two
- gather data for reference weeks spread uniformly throughout the whole year
- calculate quarterly data by combining data for 13 consecutive reference weeks and annual data by combining data for 52 consecutive reference weeks.

The survey itself must:

- be a sample of households or persons
- be (mainly) targeted at persons residing in private households within the economic territory of the state (supplemented by persons living in collective households if possible)
- be conducted via interviews with the person concerned, or another household member (or be obtained from sources of equivalent quality)
- include all individuals in a given household (or comply with further regulations)
- use an appropriate method of statistical imputation to deal with item non-response
- use weighting factors that are calculated from (a) the probability of selection; (b) external data relating to the distribution of the population being surveyed by sex, age (five-year age groups) and region (NUTS II level)

As far as the statistical efficiency of the estimates is concerned, the regulations include the following criterion:

For a group of unemployed people representing 5% of the working age population the relative standard error for the estimation of annual averages ... at NUTS II level shall not exceed 8% of the sub-population in question. ... Regions with less than 300 000 inhabitants shall be exempt from this requirement. (European Communities, 1998)

It is possible to analytically derive the sample size required to satisfy this condition for each region, assuming (initially) a simple random sample of the target population. Eurostat (2008) shows that this is roughly 3,000 individuals, which should be inflated to approximately 3,500 to account for non-response. In Ireland, the smallest NUTS 2 region is the BMW Region, which had a population of 1,166,500 in 2007, which implies that the sample size for the SE Region would be approximately 9,500 individuals. In other words, the (theoretical) minimum sample size required to meet the Eurostat requirements would be 13,000 people per annum (i.e. roughly 6,000 households). Although stratification has the potential to compensate, at least in part, for the clustering element of the two-stage designs currently used in household surveys in Ireland, clustering generally produces a DEFF that is greater than 1, implying that the effective sample size would be (considerably) larger than 13,000.

A further rule for statistical efficiency is provided by Regulation No. 577/1998:

In the case of a continuous survey, for sub-populations which constitute 5% of the working age population the relative standard error at national level for the estimate of changes between two successive quarters, shall not exceed 2% of the sub-population in question. ... For Member States with a population of between one million and twenty million inhabitants, this requirement is relaxed so that the relative standard error for the estimate of quarterly changes shall not exceed 3% of the sub-population in question. (European Communities, 1998, p. L77/4)

In addition, Regulation No. 430/2005 stipulates that the relative standard error of any yearly estimate of "structural" variables, representing 1 per cent or more of the working-age population, should not exceed 9 per cent for a country with a population between 1 million and 20 million, whilst consistency between annual sub-sample totals and full sample annual averages should be ensured for employment, unemployment and the inactive population by sex and age group (ten-year intervals between 15 and 54).

Many national statistics agencies set standards in relation to the precision of sample-based estimates as part of their quality-control and reporting mechanisms (accuracy is rarely referred to, as it can only be reliably estimated using simulation techniques). The CSO follows this practice by requiring a maximum Relative Standard Error of 8 per cent for the measurement of an unemployment rate of 5 per cent at the level of NUTS 3 regions rather than NUTS 2 (excluding regions with a population below 300,000 - in practice, only the Midlands Region).

Part Two: A Simulation Study of Alternative Sampling Designs

4 Introduction

Although many of the issues mentioned above can be formalised mathematically and assessed in isolation, the evaluation of their combined impact within a realistic setting poses formidable difficulties. To achieve the research aims, it was necessary to use simulation techniques. Simulation involves random sampling from probability distributions with a view to determining the properties of sampling designs or estimators, including their bias, consistency, sampling variance and relative efficiency. The overall approach involves the following steps:

1. Generate S independent datasets under the conditions of interest
2. Compute the numerical value of the estimator (e.g. the estimated total) for each dataset (T_1, ..., T_S)
3. Calculate summary statistics across T_1, ..., T_S, which can be used (if S is large enough) as an approximation of the true sampling properties of the estimator under the conditions of interest
4. When comparing estimators, the relative efficiency of estimator 2 to estimator 1 is defined as the ratio of the variance of the first estimator to the variance of the second.

We will assess the alternative sampling designs by comparing the estimated mean of the sampling distribution with the true population total and by analysing the dispersion of the sampling distribution itself, using three different statistics: (1) the Relative Standard Error (RSE), (2) the Mean Square Error (MSE) and (3) 95% Confidence Intervals (CI). All three use the sampling variation of the estimates generated by the 600 repetitions carried out during each simulation. We use the Relative Standard Error rather than the Standard Error as the latter depends on the scale of the variable of interest and therefore makes it difficult to compare results across different variables.

In the presence of bias, the RSE only provides part of the story, as it measures precision but not accuracy. For this reason, we will also report the MSE for each estimator. This expresses the difference between the estimates and the true population value on the basis of a quadratic loss function. It therefore incorporates both the variance of the estimator and its bias, with a greater penalisation of extreme values than the RSE. Squared error loss is one of the most widely used loss functions within statistics precisely for this reason. When evaluating alternative sampling designs for household surveys, this is a particularly apt criterion. Finally, we will report the 95% Confidence Intervals, as these have an intuitively-compelling interpretation (if a large number of samples are drawn using an identical design, 95 samples out of 100 will yield an estimate that falls within the interval)[35].

35 See the Glossary of Terms for additional information and mathematical formulae.

The most effective and appropriate means of conducting these simulations is to use individual-level census data from the 2006 Census of Population. Using individual-level data, it is possible to simulate various combinations of procedures for stratification, selection, clustering and estimation, and to quantify their effects on the precision and standard error of key estimates. Estimates of the standard errors obtained when gathering data on a given characteristic can facilitate informed decisions about sampling design, sample size and the most appropriate way of reporting survey findings. As mentioned earlier, this analysis could be combined, in future, with an estimate of costs to develop a decision-support system that can estimate the cost-efficiency trade-off for different surveys.

The analysis is, of necessity, limited to the information provided by the Census of Population, although the available information relating to employment, education, housing quality and ownership, family composition and long-term limiting illness is rich enough to provide accurate insights into the properties of the sampling designs considered, particularly if we consider each of these variables to represent a broader set of covariates.

Access to individual-level census data was provided by the CSO, for the purposes of the present analysis, at its facility in Swords. A secure laboratory, following rigid criteria regarding data access, aggregation and the protection of anonymity, ensured that existing norms were respected. All calculations were carried out using syntax batch files in SPSS 18.0.3. The only data taken out of the secure laboratory and included in this report are the aggregate regional and national estimates and measures of their sampling error. For each simulation, 600 repetitions were completed in order to obtain accurate estimates of the sampling distribution of the estimates[36].

The simulations assume a 100 per cent response rate with no measurement error. Current QNHS response rates are in the region of 80-85 per cent, and for this reason we conducted two additional simulations, reducing the overall sample size from 26,000 to 23,000 (88.5%) and to 20,000 (76.9%). The precision estimates for these reduced sample sizes are merely indicative of the effects of non-response, as they are based on the unrealistic assumption that non-response is "ignorable" (i.e. missing completely at random) (De Leeuw et al., 2003). It is worth noting, in this context, that the CSO plans to introduce the goal, for field staff, of achieving 100 per cent coverage of the sample each quarter. The current non-response rate of 85 per cent is due to vacant dwellings as well as refusals, which could be considerably reduced in the future by identifying occupancy status from census returns.

36 A simulation is an experiment, which implies that statistical principles can be applied to the results. The number of repetitions is therefore chosen in order to achieve the required precision. The mean of the sampling distribution is the expected value of the estimate, and the sampling distribution itself converges on a normal bell-shaped curve in moderate to large samples (Kish, 1965, p. 13). The difference between this value and the population value is the "sampling bias", and the standard error coincides with the standard deviation of the sampling distribution: $\mathrm{SE}(\bar{y}) = \sqrt{\mathrm{Var}(\bar{y})}$.

All estimates reported in this section were weighted by the inverse of the sampling fraction in order to "gross up" to the national population, and by the regional distribution of the population (NUTS 3). No additional attempt was made to calibrate by gender, age or household structure. Given the beneficial effects of modern estimation techniques on the standard error of the estimates, this implies that further improvements in precision and accuracy are possible. Further simulations, including alternative estimation techniques, with a view to identifying the most appropriate approach and quantifying its impact on the variance of the estimates, are beyond the scope of the present study.

As we use proportionate stratified random samples, the samples are self-weighting and the estimates can be calculated without sorting cases into strata (Kish, 1965). Sub-class totals (the number of women, or young people who are unemployed) could also be calculated using standard formula, based on stratum weights which are specific to the sub-class, although these are not reported or analysed in this report. The sampling designs analysed in this report are relatively simple and yield (largely) self-weighting samples, although it should be noted that they are not identical to those currently used by the CSO. The Horvitz-Thompson estimator of the total is valid, namely:

$$\hat{Y}_{\pi} = \sum_{k \in S} \frac{y_k}{\pi_k}$$ (Horvitz & Thompson, 1952)

This formula can be further simplified, as the inclusion probabilities are a constant across all elements and equal to the sampling fraction (n/N). The estimated total can be reliably used to calculate the mean as the total population is known (or can be estimated accurately) (Pascal Ardilly & Tillé, 2006).

5 Description of Sample designs

Two sets of simulations were carried out: (1) a set of simulations to gain insights into the effects of individual aspects of the sampling design; (2) a set of simulations which compare a limited number of comprehensive competing designs.

Sample size

We started by considering three different sample sizes, the first of which comprises 6,000 households, which represents smaller household surveys like the SILC (currently in the region of 5,600 households) and the Household Budget Survey (8,000). A future Wealth Survey would be likely to have an annual sample size of this order of magnitude.

The second sample size included 26,000 households, broadly in line with the sample size of each Quarterly National Household Survey. Precision estimates for this survey are particularly important, as it is used to estimate labour force statistics and must conform to EU requirements. For this reason, we also considered the effects of reducing the quarterly sample size to 23,000 and 20,000, primarily with a view to estimating the impact of non-response (under simplifying assumptions, as noted earlier).

Finally, we considered a sample size of 38,600, which coincides with the maximum sample size that has been considered for the QNHS (Steel, 1997). A small number of simulations were conducted using this sample size, enabling us to compare our simulation results with those derived on theoretical grounds and reported elsewhere (Steel, 1997) or by applying GREG estimation techniques to sample data (CSO, 2003).

In light of the results of these three sets of simulations, we obtained additional results for a sample size of 5,200 for the SILC and Wealth surveys and 26,000 for the QNHS.

Stratification

Within the first set of simulations, we considered three different stratification designs, starting with no stratification (i.e. a simple two-stage cluster sample (2SCS) of household members), which provides a baseline against which the other designs can be assessed. The first stage of sampling is based on a PPS sample of PSUs (Small Areas), and the second is based on an SRS of a fixed number of households. All household members are included in the sample via a combination of direct or proxy interviewing.

The second stratification design reproduces the current CSO area classification using eight categories whilst the third is based upon the Haase-Pratschke Index of Relative Affluence and Deprivation, divided into five categories, as described in Section 3.4.

The final design uses 25 categories formed from the intersection between area type (reduced to 5 categories, as described earlier) and deprivation index. These two designs are identified as "two-stage stratified cluster samples" (2SSCS) in the tables. A simple random sample (SRS) of households was also simulated for the three sample sizes (6,000; 26,000; 38,600)[37].

The final set of simulations comprised three alternative stratification designs: (1) an intersection of the 34 Local Authority Areas (NUTS 4) with the 8 current area types, resulting in 92 strata; (2) an intersection of 8 Regional Authorities (NUTS 3) with 5 area types and 5 deprivation categories, resulting in 126 strata; (3) an intersection of 8 Regional Authorities (NUTS 3) with 10 deprivation categories, resulting in 80 strata. In the first two stratifications, strata were combined in such a way as to include a minimum of 20 Small Areas. The precision estimates for the three alternative stratification designs were evaluated against those obtained for the SRS and non-stratified 2SCS sample designs.

Cluster size

Within the first set of simulations, we studied two different levels of clustering: 15 (the current value for the QNHS) and 20 households per PSU. This provides information on the current approach to sampling for the QNHS, whilst comparing it with an alternative design that implies a higher degree of clustering. The aim is to assess the efficiency loss associated with a design based on a fixed sample size at the small area level which would yield cost savings and improvements in survey organisation for the CSO. The final set of simulations adopts a cluster size of 4 for the SILC survey and 20 for the QNHS.

Variables of interest

In contrast to previous studies, which focus exclusively on labour force variables, we include a broader range of measures in our analysis, in order to evaluate the impact of different sampling designs on the socio-economic variables that are at the centre of the SILC and HBS, as well as the QNHS, notably the calculation of poverty rates and estimates of well-being.

The variables of interest are all derived from the 2006 Census of Population and include the following: (1) the employed population; (2) the unemployed population; (3) the student population; (4) people on home duties; (5) retired people; (6) people in owner-occupied dwellings; (7) people in privately-rented dwellings; (8) people in local-authority rented dwellings; (9) people with a long-term limiting condition or disability; (10) Irish people; (11) other white people; (12) non-white people; (13) Travellers; (14) people with low educational attainments; (15) people with high

37 Although it would have been interesting to compare this with a simple random sample of individuals, with a view to calculating precise design effects for each of the sampling designs, and estimating the impact of clustering within households, this was not feasible within the present study.

educational attainments; (16) single people; (17) married people; (18) separated/divorced/widowed people.

Not all of the results can be shown in this report due to space constraints. We have therefore selected four variables which exemplify the differences between the sampling designs. These are the number of people employed and unemployed, the number of people with a long-term limiting illness and the number of people with primary education only. We also report the average number of households and adults in each sample and variations in these quantities. All analyses are based on the adult population, defined as persons aged 15 years and over.

There are various reasons for including these variables. Estimation of the number of **employed** and **unemployed** people are central to the QNHS and relate to the precision requirements set by Eurostat. They also permit comparisons with previous CSO studies. **Long-lasting limiting Illness** is a variable not included in the construction of the Haase-Pratschke Index of Relative Affluence and Deprivation and allows us to test the effects of the stratification designs on a seemingly unrelated variable. **Low education** is included because it has a higher value, more akin to the unemployment rates prevailing at the time of the 2011 Census, whilst nevertheless being strongly correlated with unemployment. Low education is also the census variable most strongly correlated with deprivation, which should provide a good idea of how effective the different sampling designs are at estimating poverty-related outcomes[38].

Table 5.1 below summarises the 27 sample designs included in the simulation study. The first three sample designs (SRS, 2SCS and 2SSCS with stratification by area type) were implemented for all three survey sizes (6,000, 26,000, 38,600). The first two designs are the most relevant to the present study, and this is why subsequent simulations focus specifically on these. The simulations presented in the final five rows of the table report the results of the second set of simulations, which compare the most relevant sampling designs.

38 The simulation results for the latter variable should, however, be interpreted with care, as the percentage of people with a low education is included in the calculation of the Haase-Pratschke Index. As we showed earlier, optimal results are obtained when stratifying by the variable of interest itself, although this is practically impossible. The precision of the estimated total number of people with low education is therefore likely to be relatively high, representing an upper threshold for the precision levels that can be achieved when stratifying by the Haase-Pratschke Index and estimating poverty-related variables.

Table 5.1 Description of Sampling Designs used in Simulations

Design / Stratification	Small Sample: 6,000 or 5,200 Households	Medium Sample: 26,000 or less Households	Large Sample: 38,600 Households
Simple Random Sample	**S1** SRS (6,000 HH)	**M1** SRS (26k)	**L1** SRS (38k)
PSU15 - 2 Stage Cluster Sample (2SCS)	**S2** 2SCS_15	**M2** 2SCS_15	**L2** 2SCS_15
2SSCS 15 - stratified by Area(8)	**S3** 2SSCS_15_A8	**M3** 2SSCS_15_A8	**L3** 2SSCS_15_A8
2SSCS 15 - stratified by Deprivation(5)	**S4** 2SSCS_15_D5	**M4** 2SSCS_15_D5	
2SSCS 15 - strat. by Area(5) x Depr(5)	**S5** 2SSCS_15_AD	**M5** 2SSCS_15_AD	
2SSCS 20 - strat. by Area(5) x Depr(5)	**S6** 2SSCS_20	**M6** 2SSCS_20	
26,000 2SSCS 20 - Deprivation(5)		**M726** 2SSCS_20 (26k)	
23,000 2SSCS 20 - Deprivation(5)		**M723** 2SSCS_20 (23k)	
20,000 2SCS 20 - Deprivation(5)		**M720** 2SSCS_20 (20k)	
Simple Random Sample (SRS)	**S7** SRS (5,200 HH)		
2 Stage Cluster Sample (2SCS)	**S8** 2SCS_4	**M8** 2SCS_20 (26k)	
2SSCS - NUTS4 x Area(8)	**S9** 2SCS_4_N4_A8	**M9** 2SCS_20_N4_A8	
2SSCS - NUTS3 x Area(5) x Deprivation(5)	**S10** 2SCS_4_N3_A5_D5	**M10** 2SCS_20_N3_A5_D5	
2SSCS - NUTS3 x Deprivation(10)	**S11** 2SCS_4_N3_D10	**M11** 2SCS_20_N3_D10	

The simulations based on 6,000 households include results for six designs, starting with a simple random sample of households (S1), a non-stratified two-stage cluster sample of households (S2) and a two-stage cluster sample that is stratified by area type (S3). The sampling designs labelled S4 and S5 are based on alternative forms of stratification, using the Haase-Pratschke Index (S4) and a combination of area type and the Haase-Pratschke Index (S5). Finally, S6 is based on a higher degree of clustering (20 households within each PSU, rather than the 15 used in the S1-S5), although this kind of design is unlikely to be utilised in the foreseeable future, as household surveys of roughly 6,000 dwellings typically rely on smaller "clusters".

Following an evaluation of the first set of simulations, the second set was based on the exact number of interviews currently carried out for the SILC (5,200). Again, we start with a simple random sample (S7), followed by the non-stratified two-stage cluster design (S8). In contrast to the first set of simulations, however, the second-stage sample size is reduced to 4 households, which represents current practice for this survey. This is followed by three alternative stratifications: NUTS4 by Area8 (S9), NUTS3 by Area5 by Deprivation5 (S10) and a final stratification of NUTS3 by Deprivation10. The variable Deprivation10 is based on a decile ranking of the Haase-Pratschke Deprivation Index.

The "medium-sized" sample of 26,000 households was subject to the most intensive analysis, the first six sampling designs, labelled M1-M6, being analogous to S1-S6. In contrast to the smaller survey design, the number of households sampled per PSU is of considerable operational importance for the QNHS. For this reason, we compare the effect of stratifying by a combination of area type and Deprivation Index (M6), on the one hand, and stratifying by the Haase-Pratschke Index alone (design M726), increasing the second-stage sample size from 15 to 20 households. Finally, we explore the effect of reducing the overall sample size (first to 23,000 households, then to 20,000) on the precision of estimates.

In the final simulations for the medium-sized sample, we did not have to repeat the SRS, as this is equivalent to simulation M1. The designs M8 to M11 mirror the alternative stratifications of the designs S8 to S11 for the small sample.

A sample size of 38,600 households is unlikely to be used in household surveys in Ireland in the foreseeable future. This sample size is included primarily in order to facilitate comparisons with estimates reported elsewhere using different techniques.

It is important to stress that sampling designs are typically evaluated using individual estimators based on individual variables. This means that the precision and accuracy of the estimates depend on the nature of the variable chosen, and will also vary over time as the population mean and standard deviation of these variables change. Where rapid changes occur in, for example, the number of people unemployed, the results of previous simulation studies may quickly become outdated. For this reason, we recommend repeating the present study as soon as 2011 census data are available. More generally, a simulation study using individual-level census data could be conducted every five years, adapting the structure of the study to the design considerations that are considered most relevant, and gradually extending the scope of the simulation to include rotation patterns, structured non-response, change over time and estimation/calibration techniques.

6 Parameter Estimates, Sampling Distributions, Precision and Accuracy

In this Section, we will report the estimates generated by the 27 sampling designs described in the previous section, as applied to four variables of interest, including the mean of the sampling distribution, the Relative Standard Errors (RSE), Mean Square Error (MSE) and Confidence Intervals (CI). For each design and variable, we will report estimates for Ireland as a whole (NUTS 1), the Southern & Eastern Region and the Border, Midlands and Western Region (NUTS 2) and the eight Regional Authorities (NUTS 3), collapsing the latter into two columns (the minimum and maximum values). The eight Regional Authority areas differ considerably in size; as precision/accuracy is influenced primarily by sample size, the Dublin Region will generally have the highest precision and the Midlands Region the lowest.

The last column of each table includes a ranking of the three best sampling designs, based on the precision/accuracy of estimates for Ireland as a whole. This ranking is confined to the comparison between alternative designs for a "small" household survey (similar to the SILC), and for a "medium-sized" survey (like the QNHS), and excluding the simple random sample of households. Although the latter is always the most efficient sampling design, it is not currently feasible to use this kind of design for household surveys in Ireland due to its cost and the need for a high-quality population register. Table 6.13 draws together the results to facilitate an overall assessment of these alternative designs, and a detailed discussion of the results is provided in Section 7.

6.1 Relative Standard Errors (RSE)

Table 6.1 Relative Standard Error of Estimates of Employed Population

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	1,896	1,409	487	557	110	
S1 - SRS	6,000	1.11	1.26	2.13	1.97	4.16	
S2 – 2SCS	400x15	1.33	1.61	2.61	2.54	5.29	
S3 – 2SSCS area 8	400x15	1.37	1.64	2.66	2.59	5.31	
S4 – 2SSCS depr 5	400x15	1.21	1.47	2.59	2.60	5.02	1
S5 – 2SSCS a5 x d5	400x15	1.28	1.57	2.56	2.49	5.30	2
S6 – 2SSCS a5 x d5	300x20	1.30	1.60	2.67	2.84	5.69	3
M1 - SRS	26,000	.51	.57	1.04	.97	2.08	
M2 – 2SCS	1,733x15	.71	.79	1.29	1.23	2.42	
M3 – 2SSCS area 8	1,733x15	.63	.77	1.21	1.31	2.37	
M4 – 2SSCS depr 5	1,733x15	.61	.73	1.15	1.20	2.46	3
M5 – 2SSCS a5 x d5	1,733x15	.59	.73	1.13	1.18	2.50	1
M6 – 2SSCS a5 x d5	1,300x20	.59	.72	1.23	1.25	2.63	1
M726 – 2SSCS depr 5	1,300x20	.62	.76	1.37	1.34	2.66	
M723 – 2SSCS depr 5	1,150x20	.71	.83	1.42	1.44	2.71	
M720 – 2SSCS depr 5	1,000x20	.71	.89	1.47	1.54	3.08	
L1 - SRS	2,573x15	.42	.46	.86	.79	1.71	
L2 – 2SCS	2,573x15	.54	.63	.98	.98	1.88	
L3 – 2SSCS area 8	2,573x15	.53	.62	1.01	.96	1.99	
S7 - SRS	5,200	1.16	1.30	2.31	2.04	4.47	
S8 – 2SCS	1,300x4	1.20	1.39	2.39	2.27	4.76	3
S9 – 2SSCS N4A8	1,300x4	1.20	1.38	2.37	2.24	5.15	3
S10 – 2SSCS N3A5D5	1,300x4	1.13	1.34	2.31	2.29	4.54	2
S11 – 2SSCS N3D10	1,300x4	1.12	1.34	2.21	2.15	4.53	1
M1- SRS	26,000	.51	.57	1.04	.97	2.08	
M8 – 2SCS	1,300x20	.68	.81	1.37	1.35	2.56	
M9 – 2SSCS N4A8	1,300x20	.67	.80	1.27	1.32	2.60	3
M10 – 2SSCS N3A5D5	1,300x20	.64	.75	1.16	1.27	2.23	2
M11 – 2SSCS N3D10	1,300x20	.63	.74	1.15	1.26	2.27	1

Table 6.2 Relative Standard Error of Estimates of Unemployed Population

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	175	126	49	54	10	
S1 - SRS	6,000	3.92	4.78	7.82	8.10	17.48	
S2 – 2SCS	400x15	5.24	6.05	9.24	9.76	21.60	
S3 – 2SSCS area 8	400x15	5.30	6.44	9.56	10.15	21.20	
S4 – 2SSCS depr 5	400x15	4.77	5.91	9.13	10.26	22.28	2
S5 – 2SSCS a5 x d5	400x15	4.68	5.85	9.09	8.74	20.83	1
S6 – 2SSCS a5 x d5	300x20	4.93	6.12	9.85	9.62	23.23	3
M1 - SRS	26,000	2.00	2.40	3.61	3.70	8.11	
M2 – 2SCS	1,733x15	2.64	3.07	4.67	5.00	10.76	
M3 – 2SSCS area 8	1,733x15	2.32	2.75	4.36	4.50	9.75	
M4 – 2SSCS depr 5	1,733x15	2.24	2.82	4.38	4.64	9.83	2
M5 – 2SSCS a5 x d5	1,733x15	2.20	2.67	4.19	4.29	9.90	1
M6 – 2SSCS a5 x d5	1,300x20	2.27	2.75	4.65	4.56	10.92	
M726 – 2SSCS depr 5	1,300x20	2.24	2.77	4.75	4.86	10.75	2
M723 – 2SSCS depr 5	1,150x20	2.51	3.15	5.20	5.18	10.81	
M720 – 2SSCS depr 5	1,000x20	2.62	3.22	5.22	5.36	11.86	
L1 - SRS	2,573x15	1.53	1.93	3.02	3.06	6.73	
L2 – 2SCS	2,573x15	1.96	2.31	3.51	3.82	7.90	
L3 – 2SSCS area 8	2,573x15	1.96	2.39	3.54	3.92	7.80	
S7 - SRS	5,200	4.32	5.24	8.66	8.64	18.51	
S8 – 2SCS	1,300x4	4.74	5.53	8.68	8.62	19.16	3
S9 – 2SSCS N4A8	1,300x4	5.10	5.77	9.39	9.00	19.53	
S10 – 2SSCS N3A5D5	1,300x4	4.70	5.49	8.61	8.57	20.02	2
S11 – 2SSCS N3D10	1,300x4	4.47	5.13	8.62	8.36	20.42	1
M1- SRS	26,000	2.00	2.40	3.61	3.70	8.11	
M8 – 2SCS	1,300x20	2.63	3.17	4.59	4.94	10.46	3
M9 – 2SSCS N4A8	1,300x20	2.70	3.19	4.56	5.18	10.12	
M10 – 2SSCS N3A5D5	1,300x20	2.28	2.65	4.26	4.36	9.13	1
M11 – 2SSCS N3D10	1,300x20	2.28	2.61	4.36	4.22	9.69	1

Table 6.3 Relative Standard Error of Estimates of Population with Long-term Limiting Illness

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	294	215	79	85	17	
S1 - SRS	6,000	2.86	3.48	6.05	5.56	12.35	
S2 – 2SCS	400x15	3.40	4.16	6.30	6.86	14.19	3
S3 – 2SSCS area 8	400x15	3.52	4.15	6.18	6.80	13.44	
S4 – 2SSCS depr 5	400x15	3.29	4.07	6.10	6.75	14.44	2
S5 – 2SSCS a5 x d5	400x15	3.04	3.77	6.20	6.21	13.50	1
S6 – 2SSCS a5 x d5	300x20	3.46	4.00	7.17	6.75	13.85	
M1 - SRS	26,000	1.43	1.68	2.92	2.74	6.16	
M2 – 2SCS	1,733x15	1.61	1.95	2.99	3.18	6.79	
M3 – 2SSCS area 8	1,733x15	1.66	2.06	2.87	3.29	6.86	
M4 – 2SSCS depr 5	1,733x15	1.54	1.86	3.03	3.21	6.60	2
M5 – 2SSCS a5 x d5	1,733x15	1.52	1.85	3.12	2.86	6.43	1
M6 – 2SSCS a5 x d5	1,300x20	1.59	1.92	2.92	3.05	6.49	
M726 – 2SSCS depr 5	1,300x20	1.55	1.88	2.86	2.96	6.93	3
M723 – 2SSCS depr 5	1,150x20	1.72	2.00	3.42	3.41	7.28	
M720 – 2SSCS depr 5	1,000x20	1.85	2.18	3.52	3.69	7.86	
L1 - SRS	2,573x15	1.17	1.38	2.35	2.21	4.97	
L2 – 2SCS	2,573x15	1.34	1.60	2.40	2.54	5.03	
L3 – 2SSCS area 8	2,573x15	1.32	1.57	2.47	2.52	5.37	
S7 - SRS	5,200	3.06	3.71	6.33	5.89	13.31	
S8 – 2SCS	1,300x4	3.37	4.00	6.26	6.41	14.54	3
S9 – 2SSCS N4A8	1,300x4	3.29	3.79	6.71	6.14	14.51	2
S10 – 2SSCS N3A5D5	1,300x4	3.50	3.97	6.60	6.13	14.57	
S11 – 2SSCS N3D10	1,300x4	3.28	3.86	6.78	6.03	13.98	1
M1- SRS	26,000	1.43	1.68	2.92	2.74	6.16	
M8 – 2SCS	1,300x20	1.71	2.09	3.01	3.56	6.58	
M9 – 2SSCS N4A8	1,300x20	1.66	1.98	3.17	3.37	6.98	3
M10 – 2SSCS N3A5D5	1,300x20	1.52	1.74	3.07	2.96	6.81	1
M11 – 2SSCS N3D10	1,300x20	1.57	1.89	2.92	3.00	6.72	2

Table 6.4 Relative Standard Error of Estimates of People with Low Education

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	551	376	176	137	36	
S1 - SRS	6,000	2.32	2.93	4.15	4.85	9.38	
S2 – 2SCS	400x15	3.57	4.31	5.93	8.13	12.46	
S3 – 2SSCS area 8	400x15	3.48	4.29	5.67	8.46	12.34	
S4 – 2SSCS depr 5	400x15	2.74	3.76	5.20	7.87	11.99	1
S5 – 2SSCS a5 x d5	400x15	2.75	3.58	5.32	6.01	12.94	2
S6 – 2SSCS a5 x d5	300x20	2.75	3.58	5.77	5.99	12.94	2
M1 - SRS	26,000	1.08	1.33	1.89	2.25	4.40	
M2 – 2SCS	1,733x15	1.72	2.09	2.85	3.84	5.72	
M3 – 2SSCS area 8	1,733x15	1.59	2.08	2.62	3.98	5.37	
M4 – 2SSCS depr 5	1,733x15	1.32	1.80	2.61	3.63	5.43	3
M5 – 2SSCS a5 x d5	1,733x15	1.17	1.54	2.29	2.59	5.60	1
M6 – 2SSCS a5 x d5	1,300x20	1.28	1.68	2.61	2.94	6.06	2
M726 – 2SSCS depr 5	1,300x20	1.36	1.94	2.72	4.01	6.04	
M723 – 2SSCS depr 5	1,150x20	1.44	2.01	2.90	4.22	6.25	
M720 – 2SSCS depr 5	1,000x20	1.66	2.22	3.37	4.73	7.32	
L1 - SRS	2,573x15	.87	1.10	1.59	1.82	3.50	
L2 – 2SCS	2,573x15	1.37	1.64	2.21	3.11	4.51	
L3 – 2SSCS area 8	2,573x15	1.28	1.64	2.00	3.06	4.46	
S7 - SRS	5,200	2.45	3.09	4.46	5.27	10.07	
S8 – 2SCS	1,300x4	2.87	3.55	4.77	6.33	10.06	
S9 – 2SSCS N4A8	1,300x4	2.85	3.48	4.63	6.41	10.72	3
S10 – 2SSCS N3A5D5	1,300x4	2.60	3.00	4.77	5.17	9.99	1
S11 – 2SSCS N3D10	1,300x4	2.66	3.16	4.38	5.18	10.49	2
M1- SRS	26,000	1.08	1.33	1.89	2.25	4.40	
M8 – 2SCS	1,300x20	1.89	2.27	3.04	4.16	6.08	
M9 – 2SSCS N4A8	1,300x20	1.83	2.31	2.87	4.08	6.21	3
M10 – 2SSCS N3A5D5	1,300x20	1.39	1.67	2.35	2.71	5.03	2
M11 – 2SSCS N3D10	1,300x20	1.37	1.73	2.21	2.89	5.32	1

6.2 Mean Square Errors (MSE)

Table 6.5 Mean Square Error (/10^6) of Estimates of Employed Population

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	1,896	1,409	487	557	110	
S1 - SRS	6,000	440.23	315.95	107.68	21.40	120.41	
S2 – 2SCS	400x15	637.34	516.86	161.48	34.70	201.28	
S3 – 2SSCS area 8	400x15	680.00	532.42	167.09	34.52	208.97	
S4 – 2SSCS depr 5	400x15	523.40	431.36	160.33	30.48	209.79	1
S5 – 2SSCS a5 x d5	400x15	590.02	496.80	154.52	34.30	194.60	2
S6 – 2SSCS a5 x d5	300x20	644.90	535.29	170.87	39.86	254.61	3
M1 - SRS	26,000	92.51	63.72	25.69	5.33	28.91	
M2 – 2SCS	1,733x15	179.35	124.28	39.19	7.19	46.87	
M3 – 2SSCS area 8	1,733x15	144.13	116.74	34.52	6.90	53.05	
M4 – 2SSCS depr 5	1,733x15	133.55	106.49	31.18	7.40	45.03	3
M5 – 2SSCS a5 x d5	1,733x15	125.55	104.94	30.37	7.67	43.45	2
M6 – 2SSCS a5 x d5	1,300x20	125.35	103.50	35.59	8.51	48.34	1
M726 – 2SSCS depr 5	1,300x20	138.37	115.29	44.64	8.66	55.73	
M723 – 2SSCS depr 5	1,150x20	178.99	136.61	47.32	9.02	64.11	
M720 – 2SSCS depr 5	1,000x20	180.45	156.60	50.80	11.64	73.86	
L1 - SRS	2,573x15	63.76	42.52	17.63	3.59	19.46	
L2 – 2SCS	2,573x15	103.92	78.67	22.55	4.32	30.04	
L3 – 2SSCS area 8	2,573x15	100.32	76.08	24.05	4.86	28.36	
S7 - SRS	5,200	483.89	337.99	126.01	24.78	129.73	
S8 – 2SCS	1,300x4	517.73	385.88	135.41	27.65	159.61	3
S9 – 2SSCS N4A8	1,300x4	521.62	379.72	133.74	33.10	156.05	
S10 – 2SSCS N3A5D5	1,300x4	460.22	358.52	127.38	25.32	162.05	2
S11 – 2SSCS N3D10	1,300x4	451.52	358.74	116.64	25.11	143.45	1
M1- SRS	26,000	92.51	63.72	25.69	5.33	28.91	
M8 – 2SCS	1,300x20	166.17	132.19	44.52	7.44	57.17	
M9 – 2SSCS N4A8	1,300x20	164.17	127.31	38.68	8.36	54.44	3
M10 – 2SSCS N3A5D5	1,300x20	148.16	112.59	33.31	6.17	49.85	2
M11 – 2SSCS N3D10	1,300x20	141.67	108.03	32.34	6.33	49.32	1

Table 6.6 Mean Square Error (/10^6) of Estimates of Unemployed Population

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	175	126	49	54	10	
S1 - SRS	6,000	47.26	36.37	14.84	3.25	18.94	
S2 – 2SCS	400x15	85.40	59.10	20.52	5.03	27.72	
S3 – 2SSCS area 8	400x15	86.62	66.50	22.06	4.90	29.91	
S4 – 2SSCS depr 5	400x15	69.67	55.56	19.89	5.20	30.79	2
S5 – 2SSCS a5 x d5	400x15	67.43	54.33	20.05	4.60	22.13	1
S6 – 2SSCS a5 x d5	300x20	74.38	59.31	23.06	5.82	27.39	3
M1 - SRS	26,000	12.36	9.23	3.16	.70	3.97	
M2 – 2SCS	1,733x15	21.30	14.99	5.23	1.22	7.23	
M3 – 2SSCS area 8	1,733x15	16.55	12.04	4.57	1.00	5.92	
M4 – 2SSCS depr 5	1,733x15	15.38	12.68	4.61	1.03	6.22	2
M5 – 2SSCS a5 x d5	1,733x15	14.83	11.44	4.20	1.04	5.35	1
M6 – 2SSCS a5 x d5	1,300x20	15.80	11.99	5.19	1.26	6.03	
M726 – 2SSCS depr 5	1,300x20	15.40	12.18	5.43	1.22	6.85	3
M723 – 2SSCS depr 5	1,150x20	19.49	15.85	6.53	1.24	7.90	
M720 – 2SSCS depr 5	1,000x20	21.11	16.50	6.52	1.48	8.32	
L1 - SRS	2,573x15	7.22	5.92	2.20	.48	2.72	
L2 – 2SCS	2,573x15	11.77	8.46	2.95	.66	4.25	
L3 – 2SSCS area 8	2,573x15	11.81	9.10	3.00	.64	4.46	
S7 - SRS	5,200	57.76	43.89	18.30	3.67	21.61	
S8 – 2SCS	1,300x4	68.61	48.50	18.19	3.91	21.44	3
S9 – 2SSCS N4A8	1,300x4	79.61	52.78	21.01	3.91	23.57	
S10 – 2SSCS N3A5D5	1,300x4	67.77	48.00	17.89	4.11	21.39	2
S11 – 2SSCS N3D10	1,300x4	62.31	42.20	18.30	4.50	20.57	1
M1- SRS	26,000	12.36	9.23	3.16	.70	3.97	
M8 – 2SCS	1,300x20	21.23	15.91	5.05	1.16	7.07	3
M9 – 2SSCS N4A8	1,300x20	22.53	16.14	5.06	1.09	7.85	
M10 – 2SSCS N3A5D5	1,300x20	15.97	11.14	4.36	.93	5.55	1
M11 – 2SSCS N3D10	1,300x20	16.12	10.97	4.61	1.02	5.30	2

Table 6.7 Mean Square Error (/10^6) of Estimates for People with Long-term Limiting Illness

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	294	215	79	85	17	
S1 - SRS	6,000	70.80	55.96	22.61	4.36	22.40	
S2 – 2SCS	400x15	100.44	80.48	24.80	5.72	34.06	3
S3 – 2SSCS area 8	400x15	106.81	79.74	23.62	5.20	33.46	
S4 – 2SSCS depr 5	400x15	94.26	77.49	23.04	6.05	34.10	2
S5 – 2SSCS a5 x d5	400x15	79.79	65.59	23.89	5.20	27.88	1
S6 – 2SSCS a5 x d5	300x20	104.42	74.85	31.71	5.56	32.95	
M1 - SRS	26,000	17.79	13.14	5.28	1.08	5.47	
M2 – 2SCS	1,733x15	22.48	17.58	5.53	1.32	7.36	
M3 – 2SSCS area 8	1,733x15	23.81	19.83	5.10	1.34	7.94	
M4 – 2SSCS depr 5	1,733x15	20.55	16.08	5.70	1.24	7.47	2
M5 – 2SSCS a5 x d5	1,733x15	19.91	15.78	6.01	1.19	5.96	1
M6 – 2SSCS a5 x d5	1,300x20	21.95	17.15	5.29	1.20	6.75	
M726 – 2SSCS depr 5	1,300x20	20.68	16.32	5.08	1.38	6.38	3
M723 – 2SSCS depr 5	1,150x20	25.57	18.53	7.24	1.53	8.45	
M720 – 2SSCS depr 5	1,000x20	29.67	22.01	7.66	1.76	9.85	
L1 - SRS	2,573x15	11.88	8.90	3.41	.71	3.54	
L2 – 2SCS	2,573x15	15.62	12.01	3.58	.72	4.75	
L3 – 2SSCS area 8	2,573x15	14.98	11.40	3.76	.82	4.64	
S7 - SRS	5,200	80.98	63.96	24.81	5.06	25.14	
S8 – 2SCS	1,300x4	97.81	73.87	24.36	6.04	29.76	3
S9 – 2SSCS N4A8	1,300x4	93.41	66.74	27.73	5.96	27.84	2
S10 – 2SSCS N3A5D5	1,300x4	105.44	72.75	27.34	6.03	27.17	
S11 – 2SSCS N3D10	1,300x4	93.40	69.23	28.74	5.67	26.57	1
M1- SRS	26,000	17.79	13.14	5.28	1.08	5.47	
M8 – 2SCS	1,300x20	25.22	20.25	5.61	1.24	9.18	
M9 – 2SSCS N4A8	1,300x20	23.84	18.30	6.23	1.39	8.38	3
M10 – 2SSCS N3A5D5	1,300x20	20.15	14.01	5.94	1.32	6.37	1
M11 – 2SSCS N3D10	1,300x20	21.30	16.58	5.30	1.29	6.59	2

Table 6.8 Mean Square Error of Estimates of People with Low Education

Model	Sample Design	Ireland	SE Region	BMW Region	Min NUTS3 (Dublin)	Max NUTS3 (Midlands)	Rank
Actual Count 2006	('000)	551	376	176	137	36	
S1 - SRS	6,000	163.19	120.87	53.13	11.38	44.69	
S2 – 2SCS	400x15	388.22	262.67	108.82	20.04	124.31	
S3 – 2SSCS area 8	400x15	369.36	259.43	100.43	17.30	136.69	
S4 – 2SSCS depr 5	400x15	227.09	199.21	83.24	18.82	116.15	1
S5 – 2SSCS a5 x d5	400x15	228.77	180.79	88.85	22.06	67.97	2
S6 – 2SSCS a5 x d5	300x20	230.28	180.38	104.39	21.84	66.94	3
M1 - SRS	26,000	35.43	25.04	11.00	2.52	9.64	
M2 – 2SCS	1,733x15	89.90	61.67	24.98	3.95	30.41	
M3 – 2SSCS area 8	1,733x15	76.59	60.91	21.22	3.76	31.21	
M4 – 2SSCS depr 5	1,733x15	52.68	45.98	21.15	3.85	24.65	3
M5 – 2SSCS a 5 x d 5	1,733x15	41.66	33.57	16.16	3.60	12.62	1
M6 – 2SSCS a5 x d5	1,300x20	49.95	39.98	21.13	4.84	16.23	2
M726 – 2SSCS depr 5	1,300x20	56.01	52.90	22.83	4.68	30.17	
M723 – 2SSCS depr 5	1,150x20	62.99	56.92	25.84	4.95	33.41	
M720 – 2SSCS depr 5	1,000x20	84.09	69.36	35.34	6.99	41.83	
L1 - SRS	2,573x15	23.21	16.94	7.75	1.59	6.32	
L2 – 2SCS	2,573x15	57.01	37.96	15.08	2.64	18.93	
L3 – 2SSCS area 8	2,573x15	50.13	38.34	12.30	2.59	18.68	
S7 - SRS	5,200	182.80	135.11	61.29	13.09	53.09	
S8 – 2SCS	1,300x4	251.76	179.39	70.17	13.10	75.24	
S9 – 2SSCS N4A8	1,300x4	247.96	172.03	66.12	14.87	78.04	3
S10 – 2SSCS N3A5D5	1,300x4	208.46	126.84	76.94	13.02	50.01	1
S11 – 2SSCS N3D10	1,300x4	216.50	141.28	61.10	14.71	50.65	2
M1- SRS	26,000	35.43	25.04	11.00	2.52	9.64	
M8 – 2SCS	1,300x20	107.90	72.88	28.50	4.84	38.04	
M9 – 2SSCS N4A8	1,300x20	101.94	75.45	25.85	5.02	40.18	3
M10 – 2SSCS N3A5D5	1,300x20	62.80	39.40	23.56	3.32	13.84	2
M11 – 2SSCS N3D10	1,300x20	57.00	42.62	15.84	3.72	15.73	1

6.3 Confidence Intervals (CI)

Table 6.9 Estimates and 95% Confidence Intervals for Employed Population

Model	Sample Design	Estimate for Ireland	Lower Bound	Upper Bound	Confidence Interval	Rank
Actual Count 2006		1,895,987				
S1 - SRS	6,000	1,895,825	1,894,141	1,897,509	3,367	
S2 – 2SCS	400x15	1,897,108	1,895,084	1,899,132	4,048	
S3 – 2SSCS area 8	400x15	1,897,003	1,894,912	1,899,094	4,182	
S4 – 2SSCS depr 5	400x15	1,895,916	1,894,080	1,897,751	3,672	1
S5 – 2SSCS a5 x d5	400x15	1,898,224	1,896,283	1,900,165	3,882	2
S6 – 2SSCS a5 x d5	300x20	1,901,868	1,899,885	1,903,850	3,965	3
M1 - SRS	26,000	1,896,630	1,895,860	1,897,401	1,540	
M2 – 2SCS	1,733x15	1,895,819	1,894,745	1,896,894	2,149	
M3 – 2SSCS area 8	1,733x15	1,895,787	1,894,824	1,896,750	1,926	
M4 – 2SSCS depr 5	1,733x15	1,895,852	1,894,925	1,896,779	1,855	3
M5 – 2SSCS a5 x d5	1,733x15	1,896,024	1,895,125	1,896,923	1,798	2
M6 – 2SSCS a5 x d5	1,300x20	1,896,300	1,895,402	1,897,198	1,796	1
M726 – 2SSCS depr 5	1,300x20	1,895,599	1,894,656	1,896,543	1,887	
M723 – 2SSCS depr 5	1,150x20	1,896,132	1,895,059	1,897,206	2,147	
M720 – 2SSCS depr 5	1,000x20	1,896,320	1,895,243	1,897,398	2,155	
L1 - SRS	2,573x15	1,896,855	1,896,218	1,897,492	1,274	
L2 – 2SCS	2,573x15	1,896,093	1,895,275	1,896,911	1,636	
L3 – 2SSCS area 8	2,573x15	1,896,132	1,895,328	1,896,936	1,607	
S7 - SRS	5,200	1,895,324	1,893,560	1,897,089	3,529	
S8 – 2SCS	1,300x4	1,895,534	1,893,708	1,897,359	3,651	
S9 – 2SSCS N4A8	1,300x4	1,898,174	1,896,350	1,899,998	3,649	3
S10 – 2SSCS N3A5D5	1,300x4	1,895,738	1,894,017	1,897,460	3,443	2
S11 – 2SSCS N3D10	1,300x4	1,894,065	1,892,367	1,895,763	3,396	1
M1- SRS	26,000	1,896,630	1,895,860	1,897,401	1,540	
M8 – 2SCS	1,300x20	1,896,836	1,895,803	1,897,868	2,064	
M9 – 2SSCS N4A8	1,300x20	1,897,369	1,896,346	1,898,391	2,044	3
M10 – 2SSCS N3A5D5	1,300x20	1,896,086	1,895,110	1,897,063	1,953	2
M11 – 2SSCS N3D10	1,300x20	1,895,363	1,894,409	1,896,317	1,908	1

Table 6.10 Estimates and 95% Confidence Intervals for Unemployed Population

Model	Sample Design	Estimate for Ireland	Lower Bound	Upper Bound	Confidence Interval	Rank
Actual Count 2006		175,199				
S1 - SRS	6,000	175,453	174,902	176,004	1,103	
S2 – 2SCS	400x15	175,901	175,162	176,641	1,479	
S3 – 2SSCS area 8	400x15	175,642	174,896	176,388	1,492	
S4 – 2SSCS depr 5	400x15	175,155	174,485	175,825	1,340	2
S5 – 2SSCS a5 x d5	400x15	175,393	174,734	176,052	1,318	1
S6 – 2SSCS a5 x d5	300x20	174,434	173,745	175,124	1,379	3
M1 - SRS	26,000	175,507	175,226	175,788	562	
M2 – 2SCS	1,733x15	174,950	174,580	175,319	740	
M3 – 2SSCS area 8	1,733x15	175,318	174,992	175,644	653	
M4 – 2SSCS depr 5	1,733x15	174,949	174,635	175,263	628	2
M5 – 2SSCS a5 x d5	1,733x15	175,317	175,008	175,626	618	1
M6 – 2SSCS a5 x d5	1,300x20	174,915	174,597	175,233	636	
M726 – 2SSCS depr 5	1,300x20	175,089	174,775	175,404	630	3
M723 – 2SSCS depr 5	1,150x20	175,500	175,147	175,854	707	
M720 – 2SSCS depr 5	1,000x20	174,868	174,500	175,235	735	
L1 - SRS	2,573x15	175,346	175,130	175,561	431	
L2 – 2SCS	2,573x15	175,069	174,794	175,345	550	
L3 – 2SSCS area 8	2,573x15	175,236	174,960	175,512	552	
S7 - SRS	5,200	175,733	175,125	176,341	1,217	
S8 – 2SCS	1,300x4	175,035	174,371	175,700	1,329	3
S9 – 2SSCS N4A8	1,300x4	174,268	173,556	174,980	1,424	
S10 – 2SSCS N3A5D5	1,300x4	175,411	174,751	176,072	1,321	2
S11 – 2SSCS N3D10	1,300x4	175,888	175,257	176,519	1,262	1
M1- SRS	26,000	175,507	175,226	175,788	562	
M8 – 2SCS	1,300x20	175,004	174,635	175,374	739	3
M9 – 2SSCS N4A8	1,300x20	174,689	174,311	175,068	757	
M10 – 2SSCS N3A5D5	1,300x20	175,079	174,758	175,399	641	1
M11 – 2SSCS N3D10	1,300x20	175,531	175,210	175,852	642	2

Table 6.11 Estimates and 95% Confidence Intervals for People with Long-term Limiting Illness

Model	Sample Design	Estimate for Ireland	Lower Bound	Upper Bound	Confidence Interval	Rank
Actual Count 2006		294,118				
S1 - SRS	6,000	293,963	293,288	294,638	1,350	
S2 – 2SCS	400x15	294,545	293,742	295,348	1,607	3
S3 – 2SSCS area 8	400x15	293,596	292,767	294,424	1,657	
S4 – 2SSCS depr 5	400x15	294,648	293,870	295,426	1,556	2
S5 – 2SSCS a5 x d5	400x15	293,917	293,200	294,634	1,433	1
S6 – 2SSCS a5 x d5	300x20	292,606	291,795	293,417	1,622	
M1 - SRS	26,000	294,235	293,897	294,574	677	
M2 – 2SCS	1,733x15	293,997	293,617	294,378	761	
M3 – 2SSCS area 8	1,733x15	294,397	294,006	294,788	782	
M4 – 2SSCS depr 5	1,733x15	294,050	293,686	294,413	727	2
M5 – 2SSCS a5 x d5	1,733x15	294,125	293,767	294,483	716	1
M6 – 2SSCS a5 x d5	1,300x20	293,771	293,397	294,146	750	
M726 – 2SSCS depr 5	1,300x20	294,155	293,790	294,520	730	3
M723 – 2SSCS depr 5	1,150x20	294,079	293,673	294,484	812	
M720 – 2SSCS depr 5	1,000x20	293,763	293,327	294,199	872	
L1 - SRS	2,573x15	294,225	293,948	294,501	553	
L2 – 2SCS	2,573x15	294,314	293,997	294,631	634	
L3 – 2SSCS area 8	2,573x15	294,137	293,827	294,448	621	
S7 - SRS	5,200	294,183	293,461	294,905	1,444	
S8 – 2SCS	1,300x4	293,845	293,052	294,638	1,587	3
S9 – 2SSCS N4A8	1,300x4	294,002	293,226	294,777	1,551	2
S10 – 2SSCS N3A5D5	1,300x4	294,004	293,180	294,828	1,648	
S11 – 2SSCS N3D10	1,300x4	294,524	293,749	295,298	1,550	1
M1- SRS	26,000	294,235	293,897	294,574	677	
M8 – 2SCS	1,300x20	293,908	293,505	294,310	805	
M9 – 2SSCS N4A8	1,300x20	294,394	294,003	294,785	782	3
M10 – 2SSCS N3A5D5	1,300x20	294,317	293,957	294,677	720	1
M11 – 2SSCS N3D10	1,300x20	294,249	293,879	294,619	740	2

Table 6.12 Estimates and 95% Confidence Intervals for People with Low Education

Model	Sample Design	Estimate for Ireland	Lower Bound	Upper Bound	Confidence Interval	Rank
Actual Count 2006		551,400				
S1 - SRS	6,000	551,506	550,481	552,531	2,050	
S2 – 2SCS	400x15	552,163	550,583	553,742	3,160	
S3 – 2SSCS area 8	400x15	551,973	550,431	553,514	3,083	
S4 – 2SSCS depr 5	400x15	551,204	549,995	552,413	2,418	1
S5 – 2SSCS a5 x d5	400x15	551,119	549,906	552,333	2,427	2
S6 – 2SSCS a5 x d5	300x20	551,715	550,498	552,933	2,435	3
M1 - SRS	26,000	551,397	550,919	551,874	955	
M2 – 2SCS	1,733x15	551,047	550,286	551,807	1,521	
M3 – 2SSCS area 8	1,733x15	551,523	550,821	552,226	1,404	
M4 – 2SSCS depr 5	1,733x15	551,253	550,671	551,836	1,165	3
M5 – 2SSCS a5 x d5	1,733x15	551,952	551,436	552,468	1,032	1
M6 – 2SSCS a5 x d5	1,300x20	551,344	550,776	551,911	1,134	2
M726 – 2SSCS depr 5	1,300x20	551,438	550,838	552,039	1,201	
M723 – 2SSCS depr 5	1,150x20	551,110	550,474	551,747	1,273	
M720 – 2SSCS depr 5	1,000x20	551,541	550,805	552,277	1,472	
L1 - SRS	2,573x15	551,484	551,098	551,871	773	
L2 – 2SCS	2,573x15	551,591	550,985	552,196	1,211	
L3 – 2SSCS area 8	2,573x15	551,972	551,406	552,539	1,133	
S7 - SRS	5,200	551,743	550,659	552,828	2,169	
S8 – 2SCS	1,300x4	552,079	550,807	553,351	2,544	
S9 – 2SSCS N4A8	1,300x4	551,906	550,643	553,169	2,526	3
S10 – 2SSCS N3A5D5	1,300x4	552,972	551,820	554,123	2,303	1
S11 – 2SSCS N3D10	1,300x4	552,046	550,866	553,225	2,359	2
M1- SRS	26,000	551,397	550,919	551,874	955	
M8 – 2SCS	1,300x20	551,348	550,514	552,181	1,667	
M9 – 2SSCS N4A8	1,300x20	551,694	550,885	552,504	1,620	3
M10 – 2SSCS N3A5D5	1,300x20	553,228	552,609	553,847	1,237	2
M11 – 2SSCS N3D10	1,300x20	550,914	550,309	551,519	1,209	1

Table 6.13 Comparison of Sampling Designs

Model	Sample Design	Relative Standard Error				Mean Square Error				95% Confidence Interval			
		E	UE	LLI	ED	E	UE	LLI	ED	E	UE	LLI	ED
Actual Count 2006													
S2 – 2SCS	400x15			3				3				3	
S3 – 2SSCS area 8	400x15												
S4 – 2SSCS depr 5	400x15	1	2	2	1	1	2	2	1	1	2	2	1
S5 – 2SSCS a5 x d5	400x15	2	1	1	2	2	1	1	2	2	1	1	2
S6 – 2SSCS a5 x d5	300x20	3	3		2	3	3		3	3	3		3
M2 – 2SCS	1,733x15												
M3 – 2SSCS area 8	1,733x15												
M4 – 2SSCS depr 5	1,733x15	3	2	2	3	3	2	2	3	3	2	2	3
M5 – 2SSCS a5 x d5	1,733x15	1	1	1	1	2	1	1	1	2	1	1	1
M6 – 2SSCS a5 x d5	1,300x20	1			2	1			2	1			2
M726 – 2SSCS depr 5	1,300x20		2	3			3	3			3	3	
M723 – 2SSCS depr 5	1,150x20												
M720 – 2SSCS depr 5	1,000x20												
S8 – 2SCS	1,300x4	3	3	3		3	3	3			3	3	
S9 – 2SSCS N4A8	1,300x4	3		2	3			2	3	3		2	3
S10 – 2SSCS N3A5D5	1,300x4	2	2		1	2	2		1	2	2		1
S11 – 2SSCS N3D10	1,300x4	1	1	1	2	1	1	1	2	1	1	1	2
M8 – 2SCS	1,300x20		3				3				3		
M9 – 2SSCS N4A8	1,300x20	3		3	3	3		3	3	3		3	3
M10 – 2SSCS N3A5D5	1,300x20	2	1	1	2	2	1	1	2	2	1	1	2
M11 – 2SSCS N3D10	1,300x20	1	1	2	1	1	2	2	1	1	2	2	1

7 Results and Recommendations

The tables presented in Section 6 illustrate the results of the first simulation study to be carried out in Ireland using census data to estimate the statistical efficiency of sampling designs. The designs were derived from a review of the international literature and from a detailed discussion of the role of sample size, clustering, selection algorithms, stratification variables, area definitions and the consequences of combining these in various ways.

We will now summarise the key findings with a view to providing concrete recommendations on how the design of household surveys in Ireland might be improved in the future. These recommendations represent provisional hypotheses based on the results of a single study, necessarily limited to certain key attributes.

7.1 Effect of Overall Sample Size

The absolute size of the category to be estimated has a great influence on the precision of the estimates, which is an effect of both the sample size and the incidence of the category. When calculating the number of people in employment, for example, the Relative Standard Error of the estimate produced by a simple random sample of households increases from 0.42 (sample of 38,600) to 0.51 (sample of 26,000) reaching 1.11 (sample of 6,000). Comparing the two best designs for a "medium-sized" sample (26,000) and a "small" sample (6,000), the RSE increases from 0.59 (M5 and M6) to 1.21 (S4), which implies an increase in the 95% Confidence Interval from 1,796 (M6) to 3,672 (S4).

The unemployed population, which included roughly 175,000 people in 2006, represents a much smaller proportion of the total, and is consequently estimated less accurately. The Relative Standard Error of the estimate produced by a simple random sample of households increases from 1.53 (sample of 38,600) to 2.00 (sample of 26,000), reaching 3.92 (sample of 6,000). Comparing the two best designs for "medium-sized" and "small" samples, the RSE increases from 2.20 (M5) to 4.68 (S5) as the sample increases from 6,000 to 26,000, which implies an increase in the 95% Confidence Interval from 636 (M6) to 1,318 (S5)[39].

We also considered those with a long-term limiting illness or disability, which applied to 294,000 people in 2006. For a simple random sample of households, the RSE ranges from 1.17 (sample of 38,600) to 1.43 (sample of 26,000) and reaches 2.86 (sample of 6,000). Comparing the two best designs for a "medium-sized" sample and a "small" sample, the RSE increases from 1.52 (M5) to 3.04 (S5), with an increase in the 95% Confidence Interval from 716 to 1,433.

39 The smallest Confidence Interval for the sample of 26,000 was 618 (M5).

Finally, we have the number of people with no more than a primary school education - roughly 551,000 in 2006. As this involves a much larger population than in the case of the unemployed or those with a limiting condition, the Relative Standard Errors are lower (but not quite as low as for the employed population); for a simple random sample of households, the RSE ranges from 0.87 (sample of 38,600) to 1.08 (sample of 26,000) and reaches 2.32 (sample of 6,000). Comparing the two best designs for a "medium-sized" sample and a "small" sample, the RSE increases from 1.17 (M5) to 2.74 (S4), as the 95% Confidence Interval increases from 1,032 to 2,418. This provides an accurate impression of the impact of sample size on the precision of estimates.

7.2 Degree of Clustering

Following sample size, the next most powerful influence on the statistical efficiency of the estimates is the degree of clustering. In the first set of simulations, we studied this by comparing (1) no clustering; (2) a second-stage sample size of 15; (3) a second-stage sample size of 20.

A simple random sample of households clearly produces more precise estimates for all designs, although this is not currently feasible as a sampling design due to the costs involved in carrying out face-to-face interviews with a highly geographically-dispersed sample. As data availability improves and new technological paradigms transform data collection procedures, direct sampling of individuals or households may become an option.

Starting with "small" samples, the RSE for estimates of the employed population goes from 1.11 (no clustering) to 1.33 (second-stage sample of 15), whilst the RSE for estimation of the unemployed population increases from 3.92 to 5.24. The equivalent figures for people with a long-term limiting condition are 2.86 and 3.40, whilst the RSEs for the population with no more than a primary school education are 2.32 (no clustering) and 3.57 (second-stage sample size of 15). In other words, the RSE increases by between one fifth and a half as we introduce this relatively moderate degree of geographical clustering.

Moving on to "medium-sized" samples, the RSE for estimation of the employed population goes from 0.51 (no clustering) to 0.71 (second-stage sample of 15), whilst the RSE for estimation of the unemployed population increases from 2.00 to 2.64. The equivalent figures for people with a long-term limiting condition are 1.43 and 1.61, whilst the RSEs for the population with no more than a primary school education are 1.08 (no clustering) and 1.72 (second-stage sample size of 15).

Finally, for "large" samples, the RSE for estimation of the employed population goes from 0.42 (no clustering) to 0.54 (second-stage sample of 15), whilst the RSE for estimation of the unemployed population increases from 1.53 to 1.96. The equivalent figures for people with a long-term limiting condition are 1.17 and 1.34, whilst the RSEs for the population with no more than a primary school education are 0.87 (no clustering) and 1.37 (second-stage sample size of 15). The results therefore confirm that the effect of clustering is largely independent of overall sample size. It is greatest in relation to the population with low education, and lowest in relation to long-term limiting conditions, which is due to the higher spatial autocorrelation of the former variable. In other words, the greater the similarities, on average, between individuals living within the same Small Area, the higher that the "price" of clustering will be, in terms of statistical efficiency.

7.3 Effects of Stratification

Although clustering tends to reduce precision, as we have seen, this aspect of the sampling design is typically combined with stratification, which can improve statistical efficiency. This beneficial effect is rarely sufficient to compensate for the impact of clustering; nevertheless, the *economic* efficiency of the latter often justifies this loss in statistical efficiency. We dedicated considerable attention to this issue in Part One, where we showed that there is a strong *prima facie* case for using a multidimensional, multivariate index of social conditions when stratifying for sample surveys such as the QNHS, the SILC and the HBS. This case is reinforced by inspection of the differences in the mean of a range of socio-economic variables across the categories of alternative stratification variables (Tables 3.1 and 3.2).

Starting with the first set of simulations, stratification by area type has no beneficial effects on estimates of the employed population in "small" samples, and actually leads to a slightly higher RSE (1.37) when compared with a simple two-stage cluster sample (1.33)[40]. In "medium-sized" samples, the RSE declines slightly, from 0.71 to 0.63, and in "large" samples, stratification by area type leads to only a marginal improvement (the RSE drops from 0.54 to 0.53).

As far as estimation of the unemployed population is concerned, stratification by area type is once again associated with a slightly higher RSE in "small" samples (5.30 compared to 5.24). In "medium-sized" samples, the RSE declines from 2.64 (no stratification) to 2.32 (stratification by area type), whilst in "large" samples, there is no improvement (the RSE remaining unchanged at 1.96).

Turning now to the population with a long-term limiting condition, stratification by area type yields a slightly higher RSE in "small" samples (3.52 compared to 3.40), and the same is true for "medium-sized" samples (1.66, compared to 1.61). In "large" samples, stratification by area type leads to a small improvement (the RSE declines slightly from 1.34 to 1.32).

Finally, for the population with low education, stratification by area type yields some improvements in statistical efficiency for all three sample sizes, with a reduction in

40 This is an anomalous result, as stratification should normally lead to an improvement in precision.

the RSE from 3.57 to 3.48 in the sample of 6,000 households, from 1.72 to 1.59 in the sample of 26,000 households and from 1.37 to 1.28 in the sample of 38,600 households.

This shows that stratification by area type alone does not yield large improvements in precision and can even reduce this in small samples. The reason for this is that area type does not coincide with sharp differentials in the values of key socio-economic variables such as employment, unemployment, poverty and education (Table 3.1). This makes it particularly important to evaluate alternative criteria, with a view to reducing the statistical efficiency cost of clustering in household surveys and in order to improve the precision of sample estimates.

The next stratification variable considered is the Haase-Pratschke Index of Relative Affluence and Deprivation. Starting with the employed population, the simulation study shows that this index is indeed much more effective than area type, yielding considerable improvements in precision in both "small" and "medium-sized" samples (this was not tested for "large" samples). In the first case, the RSE declines from 1.33 (no stratification) to 1.21 (stratification by deprivation index); the latter figure is only a little higher than the RSE for the simple random sample of households (1.11). In the second case (sample size of 26,000), the RSE declines from 0.71 to 0.61, which may be compared with an RSE of 0.51 for the SRS of households.

This pattern is repeated for the unemployed population, where the RSE falls from 5.24 (no stratification) to 4.77 (stratification by deprivation categories) in the "small" sample, and from 2.64 to 2.24 in the "medium-sized" sample. Once again, the latter value is only a little higher than that obtained for an SRS of households (2.00).

The gains are visible but less marked for long-term limiting conditions, where the RSE falls from 3.40 (no stratification) to 3.29 (stratification by deprivation categories) in the "small" sample, and from 1.61 to 1.54 in the "medium-sized" sample, which may be compared to 1.43 for an SRS of households.

As might be expected, given its stronger spatial structure and correlation with disadvantage, the RSE for estimation of the population with low educational attainments declines more dramatically, from 3.57 (no stratification) to 2.74 (stratification by deprivation categories) in the "small" sample, and from 1.72 to 1.32 in the "medium-sized" sample, which may be compared with 1.08 for a SRS of households.

Larger improvements in precision may be obtained by combining area type (using a five-way classification) with the Haase-Pratschke Index of Relative Affluence and Deprivation (five categories). Controlling for deprivation level, the measure of area type now discriminates in a more effective way between the Primary Sampling Units. This yields the lowest RSEs for the employed population when using a "medium-sized" sample (0.59, compared to 0.51 for the SRS), the unemployed population (4.68 for a sample of 6,000, where the equivalent RSE for the SRS is 3.92; 2.20 for a sample of 26,000, where the equivalent RSE is 2.00), those with a long-term limiting condition (3.04 for "small" samples, compared to an RSE for the SRS of 2.86; 1.52 for "medium-sized" samples, compared to a value of 1.43 for the SRS) and those with low education levels (1.17 in "medium-sized" samples, compared to an RSE of 1.08 for the simple random sample).

7.4 Detailed Assessment of Alternative Sampling Designs

The final two sections reported in Tables 6.1 to 6.13 show the results of the second set of simulations. Sample designs S7 to S11 contain the results for a SILC sample of 5,200 households, based on a sample of 4 households per PSU. Designs M1 and M8 to M11 contain the results for a QNHS sample of 26,000 households using a sample of 20 households per PSU. Designs S9 and M9 are therefore roughly equivalent to the sampling designs currently employed by the CSO. The last two designs apply alternative stratifications to these designs: S10/M10 are based on a three-way intersection of NUTS3 (8), Area (5) and Deprivation (5); and S11/M11 are based on a two-way intersection of NUTS3 (8) and Deprivation (10).

The results are fully in line with those of the first set of simulations, clearly showing that a stratification design that includes a measure of deprivation leads to greater precision in all cases. Further to this general finding, the last two designs allow us to investigate how this measure should be defined.

With regard to the optimal sampling design for the SILC, stratification based on a two-way intersection of NUTS3 (8 categories) and 10 deprivation categories is undoubtedly the strongest design. We believe that the weakness of the S9 design (NUTS4 x Area8) lies with the way in which it treats the Dublin area. In this design, Dublin is included in a single stratum, as it is both a county and a single area type. The presence of a large and undifferentiated stratum for Dublin in existing sampling designs was one of the reasons for commissioning this study in the first place. The S10 design sub-divides the Dublin area into 5 deprivation strata, but each of these strata is very large by comparison to the others. This helps to explain the poor results obtained for the S10 design (NUTS3 x Area5 x Deprivation5) with respect to long-term limiting illness, which is representative of a variable which has a relatively low degree of spatial auto-correlation.

Unlike the SILC, which inquires into a wider spectrum of socio-economic indicators, the QNHS is more narrowly focused on the labour market. Hence, the precision of estimates for employment and unemployment is of particular pertinence for this survey. Regardless of whether we use Relative Standard Errors, Mean Square Errors or Confidence Intervals to assess the results, no clear preference between the M10 and M11 designs emerges. Either of these designs would result in significant improvements vis-a-vis the current design. As noted earlier, these improvements would be achieved at practically no cost, as operational procedures would be unaffected and there would be no interruption in the data series.

7.5 Precision and Accuracy of Estimates

In the preceding four sections, we referred to the Relative Standard Errors, as these provide a measure of the precision of the estimates generated by the different sampling designs. The reported precision estimates can be extrapolated to different sample sizes without great difficulty. However, it is also interesting to assess the designs included in the simulation study using different measures of efficiency, including the question of whether the mean of the estimates produced by the 600 repeated samples is centred on the true population value (within sampling error). With 600 repetitions, we have high power to detect bias, and we can therefore distinguish between "precision" (the variance of the estimates) and "accuracy" (the difference between the mean of the sampling distribution and the true population value). This is one of the great strengths of simulation studies, and Kish emphasises the importance of this distinction:

Research objectives are commonly stated in terms of the precision, the inverse of the variance of survey estimates. Accuracy is the inverse of the total error, including bias as well as the variance. If important biases, especially nonsampling errors, are present and distinguishable, accuracy is a better measure of survey objectives than precision alone. (Kish, 1965, p. 24–5)

As noted earlier, the Mean Square Error (MSE) quantifies total survey error, including both the variance of the estimator and its (squared) bias. It therefore provides an overall measure of the efficiency of a sampling design. When we compare the MSEs and RSEs for the different sampling designs, we find a number of patterns. Although the rank order of the sampling designs is very similar, certain differences are more pronounced, due to the way in which the MSE penalises errors.

For the employed population, for example, the MSE (divided by 1,000,000) for "medium-sized" samples drops from 179.35 (unstratified two-stage cluster design) to 125.55 (two-stage cluster sample stratified by area and deprivation), a reduction of 30 per cent. The equivalent figures for the RSE were 0.71 and 0.59, which represents a reduction of just under 17 per cent. For the unemployed population, the MSE for "medium-sized" samples declines from 21.30 (unstratified two-stage cluster design) to 14.83 (two-stage sampling design stratified by area and deprivation index), a reduction of 30 per cent once again (compared to 17 per cent for the RSE). For long-term limiting illness, the MSE signals an improvement of just over 11 per cent for the same comparison, compared with under 6 per cent for the RSE, whilst the MSE signals an improvement of 54 per cent for the estimation of low education (32% according to the RSE). This provides further support for stratifying by deprivation or by a combination of deprivation and area type in household surveys in Ireland.

As we noted earlier, EU regulations specify various criteria in relation to the precision of official Labour Force Surveys in member states. We will limit our analysis here to the annual estimates of the unemployed population for NUTS 2 Regions, which are based on the assumption of a national unemployment rate of 5 per cent (these estimates should have an RSE of 8% or less). The unemployment rate was less than 5 per cent in Ireland in 2006, which means that we can use our simulation results for the unemployment rate to assess (in a conservative manner) whether European regulations can be satisfied using the sampling designs analysed.

The simulations reported in the previous section indicate that all sampling designs with "medium-sized" samples can satisfy Eurostat requirements without difficulties. The highest RSEs are observed for the BMW Region, with a maximum of 5.22 for a sample size of 20,000 households and 20 households per PSU (stratification by deprivation index). In other words, the Eurostat requirements can be satisfied using a single quarterly survey. As the highest RSE at NUTS 3 level is just 11.86 (assuming the same sample design as above), even the more exacting CSO requirements can be comfortably achieved using pooled annual data.

7.6 Effectiveness of Stratification by Deprivation Score

This report provides strong evidence that the precision of estimates based on CSO household surveys can be significantly improved by including an aggregate-level measure of deprivation in the stratification design for two-stage cluster samples. This empirical finding is highly original and of great interest, as it has not been considered in published simulation studies, at least as far as we are aware. These results are nevertheless fully in line with theoretical expectations, as we showed earlier. The question therefore arises why this kind of design has not been given serious consideration until now, despite the fact that deprivation indices have been available in many countries since at least the 1990s.

We believe that the explanation lies primarily with the nature of existing deprivation indices, as the Irish Deprivation Measures have a series of attributes which makes them particularly suitable for sample stratification. Firstly, other indices, such as those developed for the UK (M. Noble, Penhale, Smith, & Wright, 2000; Robson, Bradford, & Deas, 1994) aim to estimate the number of people living in poverty in a given area. This results in a skewed distribution of scores which fails to distinguish between affluent areas. The Irish indices measure relative affluence and deprivation, with a near-normal distribution of scores, which makes them particularly useful as a general covariate across the full spectrum of geographical areas.

Secondly, the Irish Deprivation Measures are based on a more comprehensive definition of deprivation than other indices, including the concept of opportunity deprivation. This is an important issue, as it leads to a more balanced consideration of the specific forms of urban and rural deprivation. Most other deprivation indices follow Townsend by assuming that deprivation can only be defined and measured at the level of the individual person. As we have argued in a lengthy series of publications, all indices which confine their attention to individual attributes tend to exhibit an urban bias, as they inevitably overlook the aggregate and social forms of deprivation observed in rural areas (see Appendix).

Thirdly, the Irish Deprivation Measures have a multidimensional structure. Whilst all deprivation indices are intrinsically multivariate, they generally fail to achieve "true" multidimensionality. Appropriate coverage of all relevant dimensions of deprivation ensures that the Irish Deprivation Measure is strongly correlated with a wide range of social and economic variables.

Finally, the Irish Deprivation Index developed by Haase and Pratschke uses a Confirmatory Factor Analysis (CFA) approach, which means that an identical structure and measurement model can be applied to successive waves of census data. The Factor Scores produced by this kind of statistical model are not only fully comparable over time, but also more reliable and precise. The Irish Deprivation Measures are the only ones at international level with these attractive properties.

All of the above are critical issues when using deprivation index scores as a stratification variable in the context of stratification for a two-stage cluster sample design and make the Irish Deprivation Measures particularly suitable for this purpose. At the same time, their method of construction can easily be replicated in other contexts (cf. Haase et al., 2012). A detailed description of the Irish Measures of Deprivation is provided in the Appendix.

7.7 Recommendations for Household Surveys in Ireland

The simulation results presented in Section 6, and discussed above, lead to several important recommendations for the design of "small" sample surveys in Ireland. The first is that stratification by deprivation scores, using a measure such as the Haase-Pratschke Index of Relative Affluence and Deprivation, leads to significant and consistent gains in precision for all estimators. Estimates would be less volatile, have a much smaller standard error and be more sensitive to temporal trends and inter-regional differences.

This is an exciting result, which may be applied to both the SILC and the HBS, and to any similar survey which relies on a two-stage cluster sample design. As the SILC has the goal of estimating EU poverty indicators, the efficiency gains associated with this stratification method would be particularly pronounced.

Stratification by deprivation score would also lead to significant improvements in the precision of estimators based on the QNHS. The simulation study suggests that the best approach would be to combine deprivation scores with either area type or region to create a multi-criteria stratification variable. When compared with stratification by area type alone, the efficiency gains yielded by this method would facilitate an increase in the degree of clustering (from 15 to 20 households per PSU), whilst simultaneously achieving substantial improvements in precision. All recommendations are summarised in the tables below.

Theory-based Recommendations

The first set of recommendations are derived from the discussion in Part One of this study and are based primarily on theoretical considerations. In certain cases, these coincide with innovations that the CSO has already identified as attractive or necessary.

No.	***Page***	***Domain***	***Description***
1	6	PSUs	*The desired population size of PSUs should be near the maximum value that allows for acceptable travel times between pairs of dwellings, which may imply aggregating Small Areas (SAs)*
2	10	Sampling of PSUs	*PSUs should be selected randomly but with probability of selection proportional to size (PPS), where the latter is measured by the number of households or persons*
3	17	Selection of households	*A fixed number of private households should be selected within each PSU included in the sample*
4	16	Sampling of non-private households	*Where it is necessary to include non-private households, use a separate sampling frame, e.g. estimate the resident population in non-private households from Census data and select using PPS*
5	9	Stratification	*If deemed necessary, create a stratum for "surprises" encountered during data collection; this can be used to re-weight PSUs found to contain a larger population than expected*
6	19	Method of data collection	*If feasible, use telephone interviewing for follow-up questionnaires for the QNHS*
7	22	Rotation patterns	*Replace PSUs only when all final-stage units have been sampled; SAs should be aggregated to form PSUs of the required size, bearing in mind travel times (recommendation no. 1)*
8	23	Monthly survey	*Move towards a monthly national household survey; this would facilitate a more precise approach to modelling underlying trends and provide valuable additional estimates at national level*
9	24	Missing values	*Adopt a comprehensive strategy to reduce item non-response and imputation techniques to deal with missing values*

Recommendations Based on the Simulation Study

The second set of recommendations derives from the empirical results presented in Part Two of this report and centres on how sampling methodology for CSO household surveys might be enhanced using innovative stratification techniques.

No.	***Page***	***Domain***	***Description***
10	13	Stratification	*Use the Haase-Pratschke Index of Relative Affluence and Deprivation for stratification, using arbitrary cut-off points or deciles; construct a multi-criteria stratification measure by crossing this variable with a measure of area type or region*
11	57	Stratification	*For the SILC, stratification based on a two-way intersection of NUTS3 (8 categories) and 10 deprivation categories is the strongest design and would yield significant and consistent gains in precision for all estimators*
12	57	Stratification	*For the QNHS, either the M10 or the M11 sampling design would result in significant improvements vis-a-vis the current design: estimates would be less volatile and have smaller standard errors*

Recommendations for Future Research

The final set of recommendations relates to key research questions with the potential to further optimise sampling and estimation methodologies, which require further evaluation and study.

No.	***Page***	***Domain***	***Description***
13	37	Simulation studies	*Repeat the present study using 2011 census data, extending the scope of the simulation to include rotation patterns, change scores and estimation/weighting/calibration techniques*
14	8	Small area population estimation	*Assess the potential role of census and non-census data in the estimation of population change for small areas*
15	22	Rotation patterns	*Carry out a comparative analysis of rotation designs using simulation techniques to assess accuracy of change estimates*
16	25	Estimation	*Evaluate the use of generalised regression estimators to estimate national and regional totals etc. with the relevant standard errors, using auxiliary information and benchmark data*

At least some of these recommendations could have far-reaching consequences for survey estimates and field operations, implying that they should be assessed carefully prior to implementation. The potential gains offered by each innovation should be evaluated not only in its own terms, but also in relation to its impact on field operations and on the data series as a whole.

To the extent that the adoption of new PSUs, selection techniques or rotation patterns, a shorter interval between waves or new data collection techniques is likely to have an impact on survey estimates (e.g. by improving accuracy), it may make sense to introduce them together as a combined package, allowing for both an overlap between the old and new sampling designs and a gradual phasing-in of the new strategy.

This would enable field operators to adapt to the new procedures and to identify the most appropriate way of organising their work, as well as enabling the CSO to conduct a comparative analysis of the overall impact of the new sampling design and to take steps to maintain data integrity and comparability. On the other hand, changes that primarily influence the precision of estimates - such as stratifying by a new variable or applying new estimation techniques - could be implemented relatively easily, as they pose few risks, have low costs and minimal consequences for field operations.

Appendix 1: Glossary of Terms

Bias

A key desideratum in survey sampling is an unbiased estimator. Bias may be defined, in mathematical terms, as the expectation of the estimator (mean of the sample estimates in a simulation study) minus the true population value:

$$Bias(\hat{t})=E(\hat{t})-T$$

Calibration estimator

This kind of estimator uses calibrated weights which are as close as possible to the original sampling design weights, while also respecting a set of constraints. For example, it is possible to use calibration on known population *x*-totals to modify the basic sampling design weights $d_k=1/\pi_k$ that appear in the Horvitz-Thompson estimator, $\hat{t_{y\pi}}=\sum_s y_k/\pi_k=\sum_s d_k y_k$. "A new estimator, $\hat{t_{yw}}=\sum_s w_k y_k$ is sought, with weights w_k as close as possible, in an average sense for a given metric, to the d_k while respecting the calibration equation $\sum_s w_k \mathbf{x_k}=\mathbf{t}_x$ " (Deville & Sarndal, 1992, p. 376).

Cluster sample

In a (one-stage) cluster sample, all elements of the sampled clusters are included in the element sample. In two-stage cluster sampling, an element-level sample is drawn from the sampled clusters by using a sampling technique such as simple random sampling or sampling with probability proportional to size.

Confidence Interval

This is the range of values which will (asymptotically) include the true population total an average of 95 times out of 100 (assuming a (1 - α) confidence level of .95), if a series of independent samples are drawn in an identical way. We assume that the estimate generated by each simulation represents an I.I.D. sample from the sampling distribution, and the 95% confidence interval is calculated in the following way:

$$\bar{\hat{t}}-1.96\frac{s.e.(\hat{t})}{\sqrt{n}} < t < \bar{\hat{t}}+1.96\frac{s.e.(\hat{t})}{\sqrt{n}}$$

Design effect (Deff)

The *design effect* is the ratio of the variance estimated with a given sampling design to the variance estimated by a random sample of the same overall size *n* and accounts for the combined effects of stratification, clustering and weighting. It may be defined as:

$$deff\,(\hat{t}) = \frac{(\hat{v}(\hat{t}))}{(v_{\hat{SRS}}(\hat{t}))}$$

If deff > 1, the effective sample size is smaller than the actual sample size, and is given by the formula:

$$n_{eff} = n / \mathrm{deff}(\hat{t})$$

Mean square error (MSE)

The MSE is a measure of total survey error (and, conversely, of the accuracy of an estimator), with reference to a parameter (the population total) *t*, given a specific sampling design, and is defined as the sum of the variance and the squared bias. One of the advantages of Monte Carlo simulation studies is that it is possible to estimate this statistic (as the mean of the squared deviations of the sample estimates from the true population value) and to break it down into the following variance components:

$$MSE(\hat{t}) = E(\hat{t} - t)^2 = Var(\hat{t}) + Bias^2(\hat{t})$$

Population total

This is a key parameter of interest in survey sampling:

$$T = \sum_{k=1}^{N} y_k$$

Under a simple random sample, this can be written as:

$$\hat{t} = N \sum_{k=1}^{n} y_k / n = N\,\bar{y} \text{ or} \sum_{k=1}^{n} w_k\, y_k$$

where w_k = *N*/*n* is the sampling weight (inverse inclusion probability). In designs based on an equal probability of selection, the sampling weights are equal for all sample elements and the above yields an unbiased estimate.

Precision

This is also referred to as "efficiency", and may be defined as the variance of an estimator $Var(\hat{t})$, conditional upon the sampling design.

Probability proportional to size

A form of sampling where the inclusion probability depends on the size of the population element, based on the assumption that the value Z_k of the auxiliary size variable *z* is known for every population element *k* (e.g. population of a PSU).

Relative standard error

A measure of the reliability of a statistical estimate which is obtained by dividing the standard error of the estimate by the estimate itself and multiplying by 100 (to obtain a %).

$RSE(\hat{t}) = s.e.(\hat{t})/\hat{t} * 100$

Standard error

The standard error (s.e.) of an estimate is defined as:

$s.e.(\hat{t}) = \sqrt{(\hat{Var}(\hat{t}))}$ where $\hat{Var}(\hat{t})$ is the estimated sampling variance of the total estimate $\hat{t}$.

Sampling variance

The sampling variance V of a given parameter is the square of the standard error (see above), and is a function of the sampling design, the size of the sample, the variability of the parameter of interest and the sampling weights. It is defined mathematically as:

$$Var(\hat{t}) = \frac{1}{(n-1)} \sum_{i=1}^{n} (\hat{t} - E(\hat{t}))^2$$

Small Areas

These are the spatial units used in the new census geography for the 2011 Census of Population. They have a minimum population size of 65 households, in order to protect anonymity, and have been constructed with a view to obtaining maximum internal homogeneity, respecting administrative boundaries and natural/ infrastructural features.

Stratified sample

This is where the target population is divided into non-overlapping sub-populations, using auxiliary information. Stratum sample sizes are typically determined by proportional allocation (i.e. the number of sample elements in stratum h is given by $n_h = n \mathrm{x} W_h$, where $W_h = N_h/N$ is the stratum weight, and n is the specified overall sample size). The inclusion probability $\pi_{hk} = \pi = n/N$ of population element k in stratum h is constant, and the sampling weight is also a constant $w_{hk} = N/n$ (this aspect of the sampling design may therefore be described as self-weighting).

In stratified sampling, an estimator of the population total is the sum of stratum total estimators, given by:

$\hat{t} = \sum_{h=1}^{H} \hat{t}_h$ where $\hat{t}_h = \sum_{k=1}^{n_h} y_{hk}/\pi_{hk} = \sum_{k=1}^{n_h} w_{hk} y_{hk}$ is the Horvitz-Thompson estimator of the stratum total T_h. Assuming simple random sampling without replacement in each stratum, the total estimator is given by:

$\hat{t} = \sum_{h=1}^{H} N_h / n_h \sum_{k=1}^{n_h} y_{kh}$ where N_h/n_h is the stratum-specific sampling weight. With proportional allocation this simplifies, because the weights N_h/n_h are equal to a constant, N_h/n_h:

$\hat{t} = \sum_{h=1}^{H} N_h / n_h \sum_{k=1}^{n_h} y_{kh} = N/n \sum_{h=1}^{H} \sum_{k=1}^{n_h} y_{kh}$. We will use this result in the simulation study to estimate the population total for each sample as the sum of the variable of interest across all households, PSUs and strata, multiplied by N/n.

Weighting

Weighting ensures that a given estimate conforms to the benchmark distribution of the population by a set of variables (typically age, gender and geographical area). A weight is allocated to each sample respondent, which may generally be defined as the inverse of the probabilities of selection, adjusted for any under-enumeration and non-response.

Appendix 2: Irish Deprivation Measures

Background

The Irish Deprivation Measures are now in their fourth generation, spanning the 1991, 1996, 2002 and 2006 censuses. Initially taking the form of a multivariate index of relative affluence and deprivation (Haase, McKeown, & Rourke, 1996), these were subsequently based on a Confirmatory Factor Analysis approach to emphasise the multidimensionality of the concept of deprivation and to achieve comparability of index scores over time (Haase & Pratschke, 2005, 2008). More recent changes include the extension of the analysis to the new census geography, resulting in the Pobal-Haase Deprivation Index for Small Areas (Haase & Pratschke, 2010) and the development of the first All-Island Deprivation Index (Haase et al., 2012), a proof-of-concept study to demonstrate the applicability of this method of index construction to multiple jurisdictions.

Since their first publication, the Irish deprivation measures have gained great popularity and support. The index is used by many government departments, including the Department of Children and Youth Affairs, Department of Education and Skills, Department of Environment, Community and Local Government, Department of Health, and Department of Social Protection. A comprehensive overview of the different applications of the index is available at www.trutzhaase.eu. The most prominent exposure is provided by the online mapping facility at www.pobal.ie. The following are some examples of how the index has been used:

- The Measures of Deprivation have been central to the designation of area-based initiatives to tackle poverty in Ireland for almost 20 years.

- The Measures of Deprivation have played an essential role in developing Resource Allocation Models (RAMs) for successive local development programmes.

- Work remains under way to extend the use of the Measures of Deprivation as a basis for an integrated Resource Allocation Model covering a multiplicity of area-based initiatives, including the LDSIP, CDP, LEADER and LES.

- Work carried out during 2009 included the fine-tuning of areas to be designated under the RAPID initiative. The RAPID designation will further benefit from the use of the new Small Area geography, as many areas are defined at the sub-ED scale.

- Through the operation of the Social Inclusion Monitoring (SIM) groups, the index has become a major reference point for many Local Authorities. An increasing number of County Development Plans now include designated chapters on the social and economic well-being of county residents, generally with reference to the Irish Measures of Deprivation.

- Since November 2009 it is possible to gain special consideration when applying to a higher education institution under the Higher Education Access Routes (HEAR) initiative. One of the six eligibility criteria is based on the relative deprivation of the student's area of residence. Until now, this was based on the ED-level Measures of Deprivation and work is currently being undertaken by NIRSA/NUI Maynooth to develop a new interface based on the new SA geography.

- The National Transportation Agency has adopted the Measures of Deprivation with the aim of modelling the social benefits of alternative transport routes.

- The Measures of Deprivation have been endorsed by the Health Information Unit under the Department of Health, and included as part of Health Atlas Ireland. The Health Atlas is a major tool for health researchers and aims to enhance the integration of health-related information with socio-economic data, as represented by the index. The deprivation index is fully documented on the HSE FactFile website and a number of major health studies are in preparation which will include the use of the deprivation index data for predictive modelling purposes. These include a major study on the spatial distribution of mortality and the development of a formal resource allocation model for Primary Health Care areas.

- During 2009, the two Regional Assemblies commissioned the development of a Gateway Development Index (Fitzpatrick & Haase, 2009). The GDI plays an important role in the longer-term evaluation of the National Spatial Strategy (NSS) under the National Development Plan (NDP). The Measures of Deprivation form an integral part of the GDI and will form part of the mid-term and final evaluations of the NSS.

Overall, the Measures of Deprivation enjoy unparalleled support across a large number of Government Departments, State Agencies, non-governmental organisations and local stakeholders. This has been achieved through ongoing contact with a wide variety of stakeholders and by applying state-of-the-art statistical methods in the construction of the index.

Conceptual Underpinnings

The definition of deprivation underlying the Irish Deprivation Measures as developed by Haase and Pratschke is derived from that of Coombes *et. al.* (1995), who state that:

The fundamental implication of the term deprivation is of an absence – of essential or desirable attributes, possessions and opportunities which are considered no more than the minimum by that society. (Coombes et al., 1995, p. 5)

It is the expressed intention of the index to include considerations of opportunity deprivation, although the operationalisation of this concept is limited by the available census data. In the absence of more complex measures of opportunity deprivation, which would require extensive studies (cf. Haase & Walsh, 2007; MacDonald, 2003), proxies are used which can be associated with opportunity deprivation. This, in turn, makes it essential to consider aspects of deprivation which do not rest within the individual person alone. As a result, the Irish Measures of Deprivation differ from all other existing deprivation indices which base themselves on Townsend's definition of poverty.

Most deprivation indices are based on a factor analytical approach which reduces a larger number of indicator variables to a smaller number of underlying dimensions or factors. This approach is taken a step further in the Measures of Deprivation developed by Haase & Pratschke (2005, 2008): rather than allowing the definition of the underlying dimensions of deprivation to be determined by data-driven techniques, the authors develop a prior conceptualisation of these dimensions. Based on earlier deprivation indices for Ireland, as well as analyses from other countries, three dimensions of affluence/disadvantage are identified: Demographic Profile, Social Class Composition and Labour Market Situation.

Demographic Profile is first and foremost a measure of rural affluence/deprivation. Whilst long-term adverse labour market conditions tend to manifest themselves in urban areas in the form of unemployment blackspots, in rural areas, by contrast, the result is typically agricultural underemployment and/or emigration. Emigration from deprived rural areas is also, and increasingly, the result of a mismatch between education and skill levels, on the one hand, and available job opportunities, on the other. Emigration is socially selective, being concentrated amongst core working-age cohorts and those with further education, leaving the communities concerned with a disproportionate concentration of economically-dependent individuals as well as those with lower levels of education. Sustained emigration leads to an erosion of the local labour force, a decreased attractiveness for commercial and industrial investment and, ultimately, a decline in the availability of services.

Demographic Profile is measured by five indicators:

- the percentage increase in population over the previous five years
- the percentage of population aged under 15 or over 64 years of age
- the percentage of population with a primary school education only
- the percentage of population with a third level education
- the percentage of households with children aged under 15 years and headed by a single parent

Social Class Composition is of equal relevance to both urban and rural areas. Social class background has a considerable impact in many areas of life, including educational achievements, health, housing, crime and economic status. Furthermore, social class is relatively stable over time and constitutes a key factor in the inter-generational transmission of economic, cultural and social assets. Areas with a weak social class profile tend to have higher unemployment rates, are more vulnerable to the effects of economic restructuring and recession and are more likely to experience low pay, poor working conditions as well as poor housing and social environments.

Social Class Composition is measured by five indicators:

- the percentage of population with a primary school education only
- the percentage of population with a third level education
- the percentage of households headed by professionals or managerial and technical employees, including farmers with 100 acres or more
- the percentage of households headed by semi-skilled or unskilled manual workers, including farmers with less than 30 acres
- the mean number of persons per room

Labour Market Situation is predominantly, but not exclusively, an urban measure. Unemployment and long-term unemployment remain the principal causes of disadvantage at national level and are responsible for the most concentrated forms of multiple disadvantage found in urban areas. In addition to the economic hardship that results from the lack of paid employment, young people living in areas with particularly high unemployment rates frequently lack positive role models. A further expression of social and economic hardship in urban unemployment blackspots is the large proportion of young families headed by a single parent.

Labour Market Situation is measured by four indicators:

- the percentage of households headed by semi-skilled or unskilled manual workers, including farmers with less than 30 acres
- the percentage of households with children aged under 15 years and headed by a single parent
- the male unemployment rate
- the female unemployment rate

Statistical Properties

A detailed discussion of the measurement issues involved in the New Measures of Deprivation have been provided in a recent article by the consultants (Pratschke & Haase, 2007). Following a critical appraisal, simple additive approaches are dismissed on account that they apply an implicit weighting according to the number of indicators or domains falling within each of the underlying dimensions of deprivation. A second approach, that of weighing each indicator variable on the basis of a specifically commissioned survey (Forrest & Gordon, 1993) is dismissed on the basis that it would require a single outcome measure against which the individual deprivation indicators could be regressed, which contradicts the very understanding of deprivation as a multi-dimensional concept.

A third approach to the weighing of indicator variables relies on Principal Components Analysis (PCA) or other forms of Exploratory Factor Analysis (EFA). These are the most commonly used techniques in the construction of deprivation indices. PCA and EFA techniques extract 'variance components' from a set of variables, which may be said to reflect the different dimensions of disadvantage. They are therefore consistent – at least in principle – with a dimensional analysis and a large number of variables can be included in the analysis without requiring specific theoretical justification. Examples of indices that use PCA include those by Carstairs and Morris (1990), Duncan and Aber (1997), Haase (1996), McIntyre and Gilson (2000), SAHRU (1997) and Salmond and Crampton (2002). Noble et al. (2000; 2003, 2005; 2001) also use a form of EFA, but only to identify the first component within each domain and not for the combination of these domain indicators into a single index.

The main weakness of EFA and PCA, as far as the measurement of disadvantage is concerned, is that all the variables included in the analysis are treated as being related to all the components or factors. This can lead to ambiguities in interpretation, as the definition of the components or factors depends on the precise pattern of the loadings. Further problems arise from the tendency for researchers to use the first component of an EFA or PCA as a unidimensional index of disadvantage. In summary:

... the main problems that must be overcome when combining indicator variables involve: (1) controlling for dimensionality and measurement error; (2) producing a stable and interpretable set of dimensions; and (3) avoiding arbitrary operational decisions that make it impossible to compare scores (Pratschke & Haase, 2007, p. 725).

A more powerful and general form of statistical analysis, known as Structural Equation Modelling, can provide acceptable solutions to all these problems, enabling one to incorporate latent variables within longitudinal analyses (Bollen, 1989; Loehlin, 1992). The dimensions of disadvantage are conceptualised on theoretical grounds and indicator variables are constructed to measure these. There are statistical tests to assess whether the dimensions specified by the model are consistent with the data. Each dimension is linked with a subset of indicator variables, which simplifies interpretation, and the latent variables control for errors of measurement in the observed variables. Weights can be created for the indicator variables according to established statistical criteria, which can be used to estimate disadvantage scores for each individual area. Above all, factor score estimates are comparable from one period of time to another, and from one country to another, as long as the model has the same structure (Meredith, 1993).

Bibliography

Ardilly, P., & Osier, G. (2007). Cross-sectional variance estimation for the French "Labour Force Survey." *Survey Research Methods*, *1*(2), 75–83.

Ardilly, Pascal, & Tillé, Y. (2006). *Sampling methods: Exercises and solutions*. New York: Springer Verlag.

Bailar, B. A. (1975). The Effects of Rotation Group Bias on Estimates from Panel Surveys. *Journal of the American Statistical Association*, *70*(349), 23–30. doi:10.2307/2285370

Baillargeon, S., & Rivest, L. P. (2011). The construction of stratified designs in R with the package stratification. *Survey Methodology*, *37*(1), 53–65.

Bell, P. (2000). Weighting and standard error estimation for ABS household surveys. *Unpublished Australian Bureau of Statistics Memorandum*.

Bell, P. (2001). Comparison of alternative labour force survey estimators. *Survey Methodology*, *27*(1), 53–64.

Bell, P. A. (1998). *The impact of sample rotation patterns and composite estimation on survey outcomes* (ABS Working Paper No. 1351.0). Citeseer.

Bergman, L. R., Kristiansson, K.-E., Olofsson, A., & Safstrom, M. (1994). Decentralised CATI versus paper and pencil interviewing: Effects on the results in the Swedish Labour Force Surveys. *Journal of Official Statistics*, *10*(2), 181–195.

Bollen, K. (1989). *Structural Equations with Latent Variables*. New York: John Wiley.

Bruce, S. (1993). Selecting stratifiers for the Family Resources Survey (FRS). *Survey Methodology Bulletin*, *32*, 20–25.

Carstairs, V., & Morris, R. (1990). Deprivation and health in Scotland. *Health Bulletin*, *48*, 162–174.

Chang, H.-J., & Huang, K.-C. (2000). Remainder Linear Systematic Sampling. *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, *62*(2), 249–256.

Clark, R. G., & Steel, D. G. (2002). The effect of using household as a sampling unit. *International Statistical Review*, *70*(2), 289–314.

Coombes, M., Raybould, S., Wong, C., & Openshaw, S. (1995). Towards an Index of Deprivation: A review of alternative approaches. *1991 Deprivation Index: A review of approaches and a matrix of results*. London: HMSO Department of the Environment.

CSO. (2003). Variance estimation for complex surveys in CSO.

De Leeuw, E. D. (1992). *Data quality in mail, telephone, and face-to-face surveys*. Amsterdam: TT-Publicaties.

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics*, *21*(5), 233–255.

De Leeuw, E. D., Hox, J. J., & Huisman, M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*, *19*(2), 153–176.

Deville, J.-C., & Sarndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, *87*(418), 376–382. doi:10.2307/2290268

Dixon, R. (2001). *Australian Labour Force Data: How Representative is the "Population Represented by the Matched Sample"?* (Research Paper). Melbourne: Department of Economics, University of Melbourne.

Dorsett, R. (2010). Adjusting for nonignorable sample attrition using survey substitutes identified by propensity score matching: An empirical investigation using labour market data. *Journal of Official Statistics*, *26*(1), 105–125.

Duncan, G., & Aber, J. L. (1997). Neighborhood models and measures. In J. Brooks-Gunn, G. Duncan, & Aber (Eds.), *Neighborhood Poverty: Context and consequences for children* (Vol. 1, pp. 62–78). New York: Russell Sage Foundation.

European Communities. (1998). *On the organisation of a labour force sample survey in the Community. EC Council Regulation 577/98*.

Eurostat. (2008). *Survey sampling reference guidelines: Introduction to sample design and estimation techniques*. Luxembourg: European Communities.

Fellegi, I. P. (1963). Sampling with Varying Probabilities Without Replacement: Rotating and Non-Rotating Samples. *Journal of the American Statistical Association*, *58*(301), 183–201. doi:10.2307/2282963

Fitzpatrick, J., & Haase, T. (2009). *Preparation of a Gateway Development Index - Report on Stages 1 & 2*. Dublin: BMW and S&E Regional Assemblies.

Forrest, R., & Gordon, D. (1993). *People and Places: A 1991 Census atlas of England*. Bristol: School for Advanced Urban Studies.

Fuller, W. A., & Rao, J. (2001). A regression composite estimator with application to the Canadian Labour Force Survey. *Survey Methodology*, *27*(1), 45–52.

Gambino, J., Kennedy, B., & Singh, M. P. (2001). Regression composite estimation for the Canadian labour force survey: Evaluation and implementation. *Survey Methodology*, *27*(1), 65–74.

Gambino, J., Singh, M., Dufour, J., Kennedy, B., & Lindeyer, J. (1998). *Methodology of the Canadian Labour Force Survey*. Ontario: Statistics Canada.

Groves, R. M. (1989). *Survey errors and survey costs*. New York: Wiley.

Haase, T. (2009). *Divided City: The Changing Face of Dublin's Inner City*. Dublin: DICP.

Haase, T., McKeown, K., & Rourke, S. (1996). *Local Development Strategies for Disadvantaged Areas - Evaluation of the Global Grant, 1992-1995*. Dublin: Area Development Management Ltd.

Haase, T., & Pratschke, J. (2005). *Deprivation and its Spatial Articulation in the Republic of Ireland: New Measures of Deprivation Based on the Census of Population, 1991, 1996 and 2002*. Dublin: ADM.

Haase, T., & Pratschke, J. (2008). *New Measures of Deprivation in the Republic of Ireland: An Inter-temporal and Spatial Analysis of Data from the Census of Population, 1991, 1996, 2002 and 2006*. Dublin: Pobal.

Haase, T., & Pratschke, J. (2010). *The Pobal-Haase Deprivation Index for Small Areas*. Dublin: Pobal.

Haase, T., Pratschke, J., & Gleeson, J. (2012). All-Island Deprivation Index: Towards the Development of Consistent Deprivation Measures for the Island of Ireland. *Borderlands: The Journal of Spatial Planning in Ireland*, *2*, 21–38.

Haase, T., & Walsh, K. (2007). *Measuring Rural Deprivation: Literature review and issues arising*. Dublin: Rural Development Advisory Committee.

Haziza, D., Thompson, K. J., & Yung, W. (2010). The effect of nonresponse adjustments on variance estimation. *Survey Methodology*, *36*(1), 35–43.

Hinkley, D. V. (1988). Bootstrap Methods. *Journal of the Royal Statistical Society. Series B (Methodological)*, *50*(3), 321–337.

Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663–685.

Hox, J. J. (2000). Multilevel analysis of grouped and longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling Longitudinal and Multilevel Data* (pp. 15–32). Mahway, NJ: Lawrence Erlbaum Associated.

Iachan, R. (1982). Systematic Sampling: A Critical Review. *International Statistical Review / Revue Internationale de Statistique*, *50*(3), 293–303. doi:10.2307/1402499

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

Knottnerus, P., & van Duin, C. (2006). Variances in repeated weighting with an application to the Dutch Labour Force Survey. *Journal of Official Statistics*, *22*(3), 565–584.

Kolenikov, S., & Angeles, G. (2011). Cost efficiency of repeated cluster surveys. *Survey Methodology*, *37*(1), 75–94.

Lavallée, P., & Hidiroglou, M. (1988). On the stratification of skewed populations. *Survey Methodology*, *14*, 33–43.

Lemaître, G., & Dufour, J. (1986). Comparative study of weighting methods for family data. *Proceedings of the Survey Methods Research Section of the American Statistical Association.*

Levy, P. S., & Lemeshow, S. (1999). *Sampling of Populations: Methods and Applications* (Third ed.). New York: John Wiley & Sons.

Loehlin, J. (1992). *Latent Variable Models: An introduction to Factor, Path and Structural Analysis*. Hillsdale, NJ: Erlbaum.

MacDonald, M. (2003). *Social Inclusion: Transport Aspects*. Leeds: Centre for Transport Studies, Imperial College.

McIntyre, D., & Gilson, L. (2000). Redressing dis-advantage: Promoting vertical equity within South Africa. *Health Care Analysis*, *8*, 235–258.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.

Noble, Michael, Barnes, H., Smith, G., McLennan, D., Dibben, C., Anttila, C., Sigala, M., et al. (2005). *Northern Ireland Multiple Deprivation Measure 2005*. Belfast: Northern Ireland Statistics and Research Agency.

Noble, Michael, Wright, G., Lloyd, M., Dibben, C., Smith, G., Ratcliffe, A., McLennan, D., et al. (2003). *Scottish Indices of Deprivation 2003*. Oxford: Social Disadvantage Research Centre.

Noble, M., Penhale, B., Smith, G., & Wright, G. (2000). *Measuring Multiple Deprivation at the Local Level: The Indices of Deprivation 2000*. London: Department of the Environment, Transport and the Regions.

Noble, M., Smith, G. A. N., Wright, G., Dibben, C., & Lloyd, M. (2001). *Measures of Deprivation in Northern Ireland*. Belfast: Northern Ireland Statistics and Research Agency (NISRA).

Noble, M., Smith, G. A. N., Wright, G., Dibben, C., Lloyd, M., & Penhale, B. (2000). *Welsh Index of Multiple Deprivation*. Cardiff: National Assembly for Wales.

Pratschke, J., & Haase, T. (2007). Measurement of social disadvantage and its spatial articulation in the Republic of Ireland. *Regional Studies*, *41*(6), 719–734.

Rao, J. N. K. (2006). Interplay between sample survey theory and practice: An appraisal. *Survey Methodology*, *31*(2), 117–138.

Robson, B., Bradford, M., & Deas, I. (1994). *Relative Deprivation in Northern Ireland*. Manchester: Centre for Urban Policy Studies, Manchester University.

Rose, D., & Harrison, E. (n.d.). *Social Class in Europe: An Introduction to the European Socio-economic Classification*. London: Routledge.

Rose, D., Pevalin, D., & O'Reilly, K. (2005). *The National Statistics Socio-economic Classification: Origins, Development and Use*. London: Office for National Statistics.

Salmond, C., & Crampton, P. (2002). Heterogeneity of deprivation within very small areas. *Journal of Epidemiology & Community Health*, *56*, 669–670. doi:10.1136/jech.56.9.669

Singh, A. C., Kennedy, B., & Wu, S. (2001). Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design. *Survey Methodology*, *27*(1), 33–44.

Small Area Health Research Unit (SAHRU). (1997). *A National Deprivation Index for Health and Health Services Research* (Technical Report No. 2). Dublin: SAHRU, Trinity College.

Statistics Finland. (2011). Quality description: Labour force survey.

Steel, D. (1997). *Annual and quarterly Labour Force Surveys: Report on sample design and estimation procedures*. Dublin: CSO.

Tallis, K. (1995). Survey design and signal extraction - meeting changing user expectations. *Proceedings of the Fiftieth Session of the International Statistical Institute*. Beijing.

Thomsen, I., & Villund, O. (2011). Using register data to evaluate the effects of proxy interviews in the Norwegian Labour Force Survey. *Journal of Official Statistics*, *27*(1), 87–98.

Weeks, M. F. (1992). Computer-assisted survey information collection: a review of CASIC methods and their implications for survey operations. *Journal of Official Statistics*, *8*(4), 445–466.

Wilkinson, R. G. (1996). *Unhealthy Societies: The afflictions of inequality*. London: Routledge.

Wolter, K. M. (1985). *Introduction to Variance Estimation*. New York: Springer Verlag.