Scotland's Census
Shaping our future
A' dealbhadh ar n-àm ri teachd

Scotland's Census 2021

Census Coverage Survey Sample Methodology

May 2020

National Records of Scotland

Contents

1. Introduction

1.1 High-level Summary

The Census Coverage Survey (CCS) is a voluntary survey that takes place 6 weeks after census day. It collects information from around 1.5% of people in Scotland. The CCS is used, along with Census data, to help estimate the total population of Scotland. This paper explains how many households will be asked to take part in the CCS in 2021 and how these households will be chosen.

1.2 Overview of the CCS

The Census Coverage Survey (CCS) is a voluntary, interviewer led, follow-up survey that takes place 6 weeks after census day. The CCS samples approximately 1.5-2% of the population in Scotland and collects information at an individual and household level. The primary aim of the CCS is to gather age-sex data which can be used in conjunction with census data to provide population estimates. The CCS data undergoes a matching process to the collected census data and the resulting output allows us to identify the persons and households enumerated in both the census and CCS or those captured in one but not the other.

Estimation & Adjustment (E&A) applies Dual System Estimation which uses the matched CCS and Census data to estimate the number of persons or households that have been missed overall in the Census. The initial census data is then adjusted to account for these missed individuals and provide a more accurate estimation of the true population in Scotland. The CCS is therefore a crucial factor in ensuring a complete, well-rounded population count.

1.3 Review of 2011 CCS Strategy

In 2011, 1.5% of all households in Scotland were sampled by the CCS (~ 45,000 households and 400 Communal Establishments (CEs)) with an overall return rate of 87%.

The CCS sample design was a two stage cluster sample, stratified by Local Authority (LA) and Hard to Count (HtC) index. The HtC index is a scale of 1 (easiest to count) to 5 (hardest to count) which was created to indicate how difficult it may be to enumerate a particular geographical area based on certain demographic features. The 40% easiest to count areas are assigned as HtC 1, with the next 40% to HtC 2, 10% to HtC 3, 8% to HtC 4 and the hardest 2% to count assigned to HtC 5.

Stage 1: Selection of the Primary Sampling Unit in 2011

The first stage of sampling used Datazones[1] as the Primary Sampling Unit (PSU) with 4% of the total selected using optimal (Neyman) allocation. This sampling strategy was used to allocate the overall sample among each LA/HtC strata in proportion to the size and variance of the stratum.

In 2011, HtC levels were collapsed where there were less than 20 Datazones per HtC level. As such, when an HtC level contained less than 20 Datazones, they were moved to the next available HtC level to ensure adequate sample size for the E&A process; this was the case in all but one of the processing units in 2011.

Stage 2: Selection of Secondary Sampling Unit in 2011

Once the Datazones were selected, a set proportion of Secondary Sampling Units (SSU) were sampled from each PSU. Postcodes were utilised as the SSUs in 2011. The second sampling stage involved the selection of 50% of the postcodes within each Datazone by a method of simple random selection.

[1] The data zone geography covers the whole of Scotland and nests within local authority boundaries.

1.4 Purpose of Document

To ensure precise and unbiased estimates, the CCS sample must provide an adequate representative sample of the population to enable accurate estimation. This sample must be distributed across the population so as to minimise variation in the population estimates. This is achieved through statistically efficient sampling techniques while utilising a suitable sample size. There will always be a balance between increasing the sample size for improved statistical precision and the costs to conduct the survey.

This paper will serve to investigate the sample selection to be utilised in the 2021 CCS, including the sample clustering and size. This investigation is based on previous research conducted by the Office of National Statistics (ONS) on the impact of changing the sampling fractions of each sampling stage to achieve better statistical estimates (Castaldo & Nikolakis, 2018).

The precision associated with different sample clustering and size strategies will be compared in terms of Relative Standard Error (RSE) derived from the variability of PSUs and SSUs within the sample frame, as well as through simulated runs of the Estimation process with different samples. The effect of different clustering on field force worker travel times will also be presented, as well as exploring the effects of lower response rates on different sample sizes and clustering strategies.

2. Sample Design for 2021

In 2021 we aim to improve upon the design of the CCS and quality of the estimates produced in 2011; the aim of this investigation is to determine the sample design and sample size which meets the target Key Performance Indicators (KPIs) for precision with maximum efficiency. A list of KPIs related to statistical quality are given in Appendix B. The 2011 CCS aimed to sample between 1-2% of the population. This has been used as a rule of thumb by a number of statistical agencies (ONS, Statistics Canada, Australian Bureau of Statistics) in order to maintain consistency between the balance of overall sample size and associated cost.

2.1 Methodology for Analysis

Two methods were primarily used for this paper to compare different sampling strategies, each of which are detailed more thoroughly in Appendix A.

2.1.1 Design variable analysis

A design variable was calculated which measures how much the adjustment applied in each postcode differed from the average adjustment across the entire population in Scotland's Census 2011. The expected RSE value was derived via statistical analyses for each sampling strategy, taking into account the variance of the design variable both within and between PSUs.

2.1.2 Estimation Simulation Methods

A simplified simulation of 2011 methodology was used in order to examine the effects of the CCS sample design on the estimates produced. The adjusted 2011 census was used to create a synthetic CCS sampling frame. The response rate within the synthetic CCS could be varied by selecting the desired proportion of households to be included.

Under the strategy to be tested, 500 different samples are drawn, with age-sex group estimates calculated for each sample. The variance and corresponding RSE for the average estimate can then be calculated to compare the different strategies.

2.2 Clustering Analysis

The level of clustering in a two-stage clustered sample depends on the size of the clusters; this is given by the proportion of both PSUs and SSUs selected. If a larger proportion of PSUs are selected with smaller proportion of SSUs selected in each, this gives smaller clusters and less clustering. Therefore there is an apparent trade off; with larger clusters (a smaller fraction of PSUs, with a larger fraction of SSUs) the areas should be easier to enumerate for interviewers executing the survey, but the sample could choose large areas of homogeneity which would increase the error.

As mentioned previously, the sampling proportions in 2011 were 4% of PSUs and 50% of SSUs. To improve upon the previous strategy, we have conducted a precision analysis using person and household design variables created from 2011 data to investigate improvement in the design estimates by varying the existing cluster proportions. The PSU sampling fraction was varied from 4% up to 10%, while varying the SSU sampling fraction to maintain the number of postcodes in the sample at around 2000. The number of postcodes in the sample was kept at around 2000 to avoid the sample being too large and because 2000 was around the number of postcodes samples in 2011.

There is a slight change in the sampling units from 2011. Instead of Datazones the PSUs in 2021 will be Planning Areas while the SSUs will remain as postcodes. The change in the geographical aggregation from Datazones to Planning Areas was to facilitate easier enumeration and travel within the area for field force workers; this is because the postcodes are more homogenised within Planning Areas compared to Datazones where they are more spread out and can be split across geographic barriers, such as rivers. The difference in size for the average Planning Area and Datazone are shown in the table below.

	Household count	**Area (km^2)**
Planning Area	302 ± 0.72	8.82 ± 0.51
Datazone	397 ± 1.3	11.37 ± 0.68

2.2.1 Results of Design variable analysis

The results of the analysis were compared to the RSE value of the 2011 cluster proportions to evaluate the variability. To show an improvement in design, the RSE of the design variables (DV) for persons and households would need to be lower in comparison to the RSE values for the 2011 proportions of 4% PSU, 50% SSU, which were 0.161 and 0.188 respectively.

To examine the impact of clustering on precision, 8 different combinations of PSU/SSU were analysed (Table 1). The equation of variability from Brown et al. (2011) was used to determine the RSE of the selected proportions. A calculation was made for a 2% simple random sample of all postcodes, and a 2% random sample stratified by LA and HtC for comparison.

Table 1: Table showing the PSU and SSU Sampling Rates and RSE values of the Design Variable Analysis

PSU %	SSU %	Population %	RSE % Household DV	RSE % Population DV
4	40	1.6	0.190	0.164
4	50	2	0.188	0.161
5	30	1.5	0.178	0.154
6	25	1.5	0.171	0.149
7	25	1.75	0.161	0.140
8	20	1.6	0.168	0.147
9	20	1.8	0.162	0.142
10	15	1.5	0.187	0.167
2% simple random			0.200	0.175
2% stratified simple random			0.180	0.161

The analysis shows that using the same PSU/SSU proportions as in 2011 results in a relatively high RSE value for both the household and person level design variable. The results further show that by increasing the PSU proportion while altering the SSU sampling fraction results in a decreased RSE.

Interestingly there appears to be a point at which the gains from increasing the proportion of PSUs selected level off.

2.2.2 Results of Simulation Analysis

The RSE values for the estimates produced under each CCS sample clustering proportion are shown in Table 2. The synthetic CCS for these simulations has a 100% response rate.

Table 2: Table showing RSE values from Estimation simulations of different CCS sample cluster proportions – 100% response rate

PSU %	SSU %	Estimate RSE %	Sample household count
4	40	0.203	42786
4	50	0.189	51330
5	30	0.185	41678
6	25	0.189	41789
7	25	0.173	48475
8	20	0.171	46234
9	20	0.158	52029
10	15	0.170	47290
2% simple random		0.165	49460
2% stratified simple random		0.156	50550

In general, the less clustering in the sample, the lower the RSE for the estimates produced, which is in agreement with the expected RSE values calculated in Table 1. While the aim was to maintain approximately the same number of households and persons captured in the sample in each clustering rate, there is some variation in sample size. As the sample size increases, the RSE also decreases, but in addition samples with larger PSU values tend to have lower RSEs.

The Key Performance Indicator for the Estimation system is to produce confidence intervals on the estimate within ±0.4% at national level. The required RSE to achieve this target precision is 0.204%. This is only just achieved with a PSU rate of 4%, while, when clustering is decreased by using PSU rates of 5% or more, RSE falls comfortably below the required target. It was agreed by NRS Statisticians that the RSE should be less than or equal to 0.19% to increase the likelihood that the national level KPI target is achieved.

However, these results do not take into account variations in the CCS response rate; the RSE calculations using the design variable equation assume a 100% response rate to the CCS and the synthetic CCS used in the simulations is a sample of the complete adjusted Census without any omission due to non-response. To examine how RSE values change with varying levels of CCS non-response, further simulations were run using synthetic CCS samples with omission rates equivalent to non-response. To better replicate the expected pattern of non-response, in one scenario the omission rate was varied between each HtC stratum according to response rate projections while holding the overall omission rate at 80%.
These simulations are shown in the table below:

Table 3: RSE values from Estimation simulations with varying response rates in the synthetic CCS

Response Rate	Sampling Strategy	RSE (%)	Households in sample
100%	4% PSU 50% SSU	0.189	51330
	4% PSU 40% SSU	0.203	42786
	7% PSU 25% SSU	0.173	48752
87%	4% PSU 50% SSU	0.193	51330
	4% PSU 40% SSU	0.212	42786
	7% PSU 25% SSU	0.183	51330
80%	4% PSU 50% SSU	0.202	51330
	4% PSU 40% SSU	0.215	42786
	7% PSU 25% SSU	0.185	48752
80% Stratified	4% PSU 50% SSU	0.201	51330
	4% PSU 40% SSU	0.213	42786
	7% PSU 25% SSU	0.181	48752

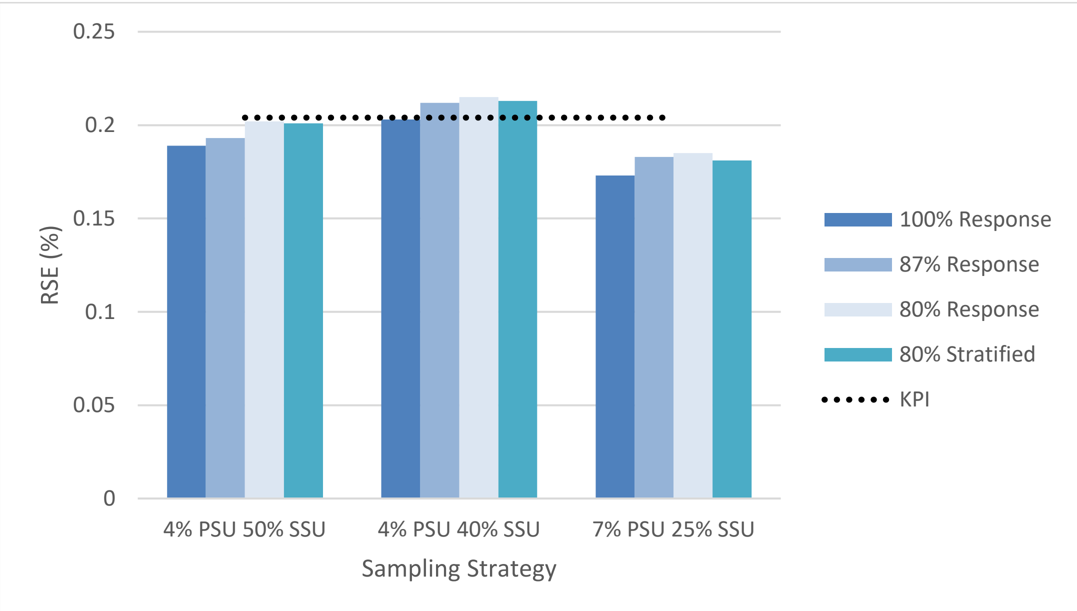

Figure 1: Graph showing RSE values for three different sample strategies under different response rates. The dotted line shows the target KPI of 0.204%

As expected, the RSE value increases as the response rate decreases. For the smaller 4% PSU sample, this was found to push all samples with less than 100% response rate over the acceptable value to meet the KPI.

The RSE decreased slightly for the sample stratified by HtC at overall 80% response rate. This may be due to HtC 1 and 2 having a response rate above 80% in the stratified sample, while the smaller HtC strata had a lower response rate.

The simulation analysis was run again, this time using an 80% response rate for different PSU and SSU. The RSE values for the estimates produced under each CCS sample clustering proportion are shown in Table 4.

Table 4: Table showing RSE values from Estimation simulations of different CCS sample cluster proportions – 80% response rate

PSU %	SSU %	Estimate RSE %	Sample household count
4	44.5	0.200	46643
5	35.5	0.190	46989
6	28.5	0.192	45808
7	25	0.185	46888
8	21.5	0.186	49163
9	18.5	0.170	48856
10	16	0.177	47655

These results also showed that where there is less clustering in the sample, the RSE for the estimates produced is lower and there is also some variation in sample size.

2.2.3 Field Force Modelling and Travel Times

It should be noted that while these estimates produce less clustering, there is expected to be a greater associated geographic distance between the households sampled which could have an impact on travel times for field force workers.

To explore the scale of this issue, an NRS model for Field Force visits to the CCS was adapted, in order to estimate the differences in travel time between different clustering strategies.

The proportion of operational time taken up by travelling was calculated for 3 different samples of around the same size. There is some variation in the sample size, and correspondingly in the number of interviewers needed, so the total number of hours worked between each sample is not exactly the same.

Table 5: Travel times as a proportion of total time worked under different sampling clustering.

Sample	Time travelling (% of total)	Postcodes in sample
4% PSU 45% SSU	23.4	2479
7% PSU 25% SSU	23.9	2429
9% PSU 20% SSU	25.2	2478

Clustering did not appear to dramatically affect travel times. This may be due to Planning Areas (the clusters) being so small that interviewers had to travel between different clusters irrespective of how many SSUs (post codes) are selected in each.

2.3 Sample Size Analysis

Using the sampling proportions from the above cluster proportion analysis we investigated the effect of varying the sample sizes by altering the SSU sampling fraction. This analysis involved using 3 different PSU values (4%, 7% & 9%) each with a range of 3 different SSUs.

RSE values were calculated both theoretically using the design variable analysis, and with estimation simulations.

2.3.1 Results of Sample Size Design Variable Analysis

The RSE values for these 9 combinations of different PSU and SSU rate are presented in Table 6 below.

Table 6: Table showing RSE values for PSU values of 4%, 7% & 9% with varied SSU values using design variable analysis

PSU%	SSU%	RSE % Household DV	RSE % Population DV
4	0.40	0.190	0.164
4	0.50	0.188	0.161
4	0.60	0.186	0.160
7	0.15	0.203	0.179
7	0.25	0.161	0.140
7	0.35	0.148	0.128
9	0.10	0.259	0.233
9	0.20	0.162	0.142
9	0.30	0.138	0.120

This analysis shows that, for each of the PSU values, the RSE for both the household and individual level design variables decrease as the SSU increases, as expected for a larger sample size. The best RSE values are obtained for the highest PSU sample of 9% and improves as the SSU is increased for this value.

2.3.2 Results of Simulation Analysis

Estimation simulations were run for samples maintaining the same PSU sampling fraction as in the clustering analysis, and differing SSU sampling fractions. Again, a 100% response rate was used. The results of these simulations are shown in Table 7.

Table 7: Table showing Simulation RSE values and average sample household counts for a range of SSU rates while fixing PSU rates – 100% response rate

PSU %	SSU %	Estimate RSE %	Sample household count
4	40	0.203	42786
4	50	0.189	51330
4	60	0.177	62512
7	15	0.203	33106
7	25	0.173	48475
7	35	0.143	67626
9	10	0.206	30195
9	20	0.158	52029
9	30	0.137	75105

Within each of the values of PSU rate, the increasing SSU rates decrease the RSE. As shown in Figure 2, the difference between 7% and 9% seemed much less notable, potentially indicating that the difference results more from absolute sample size than the clustering proportion above a certain threshold of PSU rate. The RSE for the 4% samples appeared to follow a trend higher than the other two.

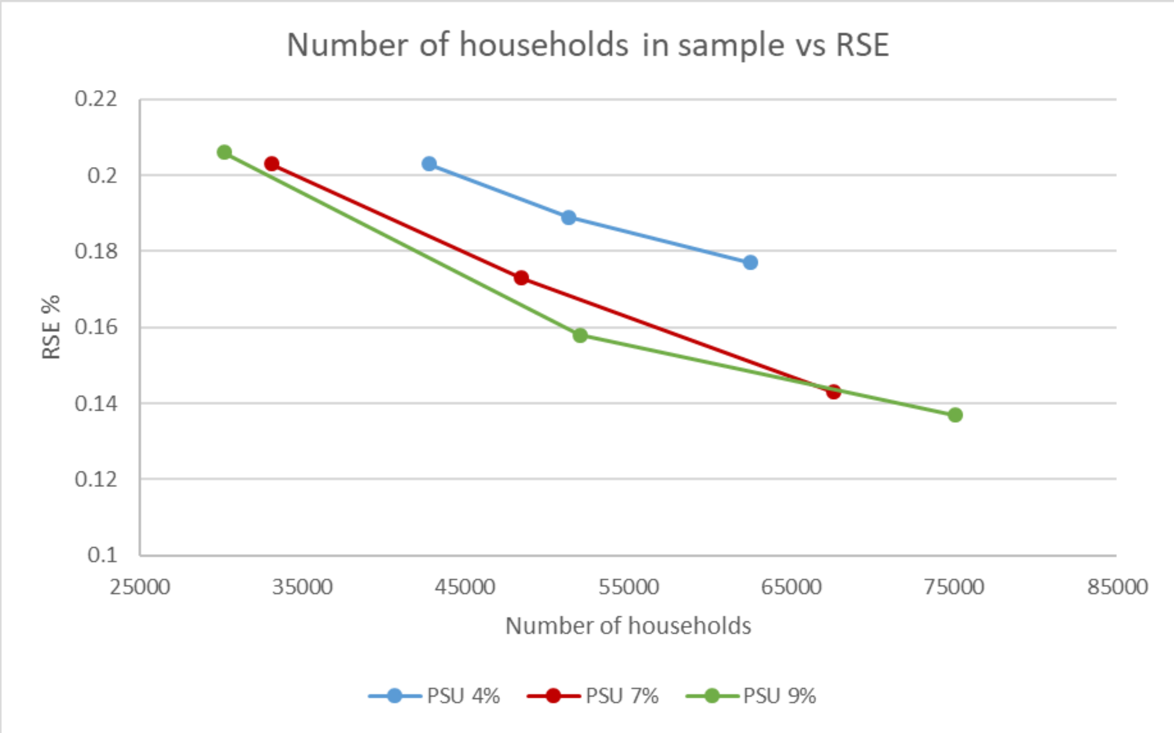

Figure 2: Graph showing RSE values against average sample household counts for ranges with fixed PSU rates

Estimation simulations were run again, this time using a 80% response rate, for samples maintaining the same PSU sampling fraction at 7% and 9%, and differing SSU sampling fractions within each PSU. The results of these simulations are shown in Table 8.

Table 8: Table showing Simulation RSE values and average sample household counts for a range of SSU rates while fixing PSU rates – 80% response rate

PSU%	SSU%	Estimate RSE%	Sample household count
7	25	0.185	46888
7	24	0.185	46754
7	23	0.188	45138
9	18.5	0.170	48856
9	17.5	0.178	46583
9	16.5	0.187	43085

3. Summary of Findings

3.1.1 Sample Cluster Proportions

The analysis of both the design variable and simulations shows that decreasing the level of clustering (using smaller clusters) improves precision. Higher PSU percentages tended to give lower predicted RSEs for a fixed sample size. However, the results of the sample size analysis suggest that beyond 7% the RSEs appear to be more dependent on sample size than on the clustering.

While reducing clustering may result in improved statistical estimates, there may be slightly increased travel times for field force workers. However, the increase in travel time observed in the model were not large relative to the total travel time. This may be because the clusters (planning areas) are too small to accommodate an entire interviewer's workload even when the SSU percentage is high. Because of this, some travel between clusters is always necessary, meaning that adding additional planning areas to interviewer's workloads does not greatly affect travel times.

3.1.2 Sample Size Selection

The analysis of the sample size showed that as the sample size increases, the RSE decreases.

The sample proportions used in 2011 (PSU/SSU 4/50) resulted in a sample size of approximately 51,000 and RSE values fall below the national target of 0.204% (the desired RSE to give a confidence interval of 0.4%) at 0.189% for the simulation analysis, however, they do not leave much room for error. In simulations accounting for an assumed CCS response rate of 80% the RSE values increase to 0.202%. This is only just inside the national target.

Decreasing the level of clustering increased the statistical efficiency, giving RSE values well within the target with a lower sample size. The sample with PSU/SSU values of 7% and 25% achieved an RSE of 0.172%, well within the national target levels, with a sample size of 48,500. In simulations with an 80% response rate this sample design was still comfortably within the national target (RSE 0.185%).

3.1.3 Conclusion

The CCS is a vital component in ensuring an accurate estimation of the entire population of Scotland is achieved in 2021. This paper evaluates the sample design in terms of clustering and sample size, making a recommendation on the appropriate level of clustering and sample size.

Analysis of different PSUs and SSUs was carried out under the assumption that the 2021 census response rate will be the same as in 2011 (94% response rate) and the 2021 CCS response rate will be 80%. This is 7% lower than in 2011. The reason for this is based on the voluntary nature of the CCS and a known decline in public response to surveys.

The results of this analysis suggest that greater statistical efficiency is achieved through decreasing the level of clustering of the sample with minimal increase in travel times. Selecting more planning areas (PSUs) and less postcodes (SSUs) in each planning area results in an improvement in expected precision of the estimates as show in the reduced RSE values. It was agreed within NRS that the most statistically efficient sample would be that shown in Table 9. This is because the precision of the estimates (lower RSE value) is acceptable given the expected response rates, and there is also a suitable sample size (a large sample size would lead to higher costs).

Table 9: Table showing recommended PSU and SSU rates for CCS sample, with RSE value average sample household count

PSU %	SSU %	Estimate RSE %	Sample household count
9	17.5	0.1778	46,583

Changes in sample size had an effect on the RSE, with higher sample sizes resulting in lower RSEs. Similarly, decreasing the response rate to the CCS increased RSE in simulations.

Based on a 9% selection of PSUs, with 17.5% of SSUs in each PSU and a sample size of 46,583 households, this gives a RSE of 0.1778% which is sufficient to meet national targets with some contingency. It is anticipated that this will not dramatically affect travel times.

4. Appendix A

4.1 Calculation of Design Variable

To look at the effectiveness of the sample design, there are two main design variables used in this analysis: one for household response rates (Z_h) and one for the individual level (Z_i) (1). Where the design variable *Z* is the difference between the 2011 post-adjusted census counts (*Y*) and the product of the initial, unadjusted, 2011 counts (X) and the ratio (*R*); R is a ratio of the summed pre-adjusted and post-adjusted counts across all postcodes (2). The script *i* is the total count of individuals and *h* is the household count across all postcodes respectively (Brown, 2011).

(1)

$$Z_p^i = Y_p^i - R^i X_p^i$$

$$Z_p^h = Y_p^h - R^h X_p^h$$

(2)

$$R^i = \sum_{p=1}^{P} Y_p^i / \sum_{p=1}^{P} X_p^i$$

$$R^h = \sum_{p=1}^{P} Y_p^h / \sum_{p=1}^{P} X_p^h$$

These two design variables reflect the modelled variability in the 2011 census coverage at the postcode level. Our aim for 2021 is to minimise this variance. The initial 2011 person and household data and the final post-census adjusted person and household response rates for 2011 were used in the creation of these design variables. Within the context of this study, these design variables allow for the investigation of changes to the PSUs and SSUs of the clustering models by acting as a proxy for the variation of our estimates (Brown, 2011). The design variables were also utilised to conduct the optimal allocation of the sample in the first stage of sampling. The use of design variables was a key component in ascertaining the level of improvement in the statistical design achieved through varying the sampling proportions (Brown, 2011).

4.2 Analysis of Relative Standard Error

The improvement in the estimates was determined by evaluating their expected relative standard error (RSE) for the different cluster proportions in comparison to the values obtained using the 2011 cluster values. The RSE value (3) provided a measure of the variability of population estimates and was derived via statistical analyses using the household and individual design variables. The RSE in this case was based on the modelled variability of the population divided by the total population estimate (T), a lower resulting RSE indicated an improvement in the design.

$$\%RSE = \frac{\sqrt{\{V\,(\hat{T}-T)\}}}{T} \times 100 \qquad (3)$$

The variability of the population estimate based on the 2-stage clustered sample was determined using the equation of variation outlined in Brown et al. (2011) (4):

$$V(\hat{T}^i - T^i) = \sum d \left\{ \frac{N_d^2}{n_d} \left(1 - \frac{n_d}{N_d}\right) \sigma_d^2 + \frac{N_d}{n_d} \sum o \in d \left(1 - \frac{m_{do}}{M_{do}}\right) M_{do}^2 \, \frac{\sigma_{do}^2}{m_{do}} \right\} \qquad (4)$$

Where M_{do} is the total number of postcodes in each HtC and planning area (*o*) and m_{do} is the number of postcodes sampled. Additionally N_d and n_d are the number of postcodes in the overall population and of the sample in each HtC respectively (Brown, 2011). The equation models the variation within the clusters of the sample estimates and the variation of the design variable across the population. It accomplishes this by respectively calculating the variance of the design variable for each postcode within the sample clusters (σ_{do}^2) (5) and the variability across the totals of the clusters (σ_d^2) (6). Where Z_{dop}^i is the total of the postcode design variable within each Planning Area and Z_{do}^{-i} is the mean of the design variables for the postcodes within each Planning Area (5). Further to this, Z_{do}^i is the total of the postcode design variables for each HTC and Planning Area and Z_d^{-i} is the means for the design variables for the postcodes within each HtC and Planning Area (6). The clusters are examined at HTC stratification level (*d*) (Brown 2011). This results in the estimated variances of the population, which are then used in the equation to determine the overall variance between the two, taking into account the population estimates.

$$\sigma_{do}^2 = \frac{1}{M_{do} - 1} \sum p \in od \ (Z_{dop}^i - Z_{do}^{-i})^2 \tag{5}$$

$$\sigma_d^2 = \frac{1}{N_d - 1} \sum o \in d \ (Z_{do}^i - Z_d^{-i})^2 \tag{6}$$

These variance values were calculated for each of the 8 tentative cluster proportions, the square root of these values were then used to calculate the standard error relative to the population estimates for the individual and household design variables respectively (1). In order for the design to show improvement the overall variability of the sample design variable in relation to the population should be smaller than it was in 2011.

4.3 Methodology for Estimation Simulations

A simplified simulation of 2011 methodology was used in order to examine the effects of the CCS sample selection method on the estimates produced. The adjusted 2011 census was used to create a synthetic CCS sampling frame. The response rate within the synthetic CCS could be varied by selecting the desired proportion of households to be included in the frame.

The numbers of people in the pre-adjusted 2011 census and the synthetic CCS were aggregated within each postcode by age-sex group. This dataset can then be queried to pull out the postcodes selected through the different sample methods to be tested, which then are aggregated by age-sex group and hard to count index within each Processing Unit (using the same groupings as used in 2011) to calculate estimates through DSE, and produce a scaling ratio between the estimates and original census count within sample areas.

These ratios were then applied to the overall population, again stratified within each Processing Unit by age-sex group and hard to count index, giving the estimates for the overall population. Estimates were only calculated by age-sex group, and no additional correction methodologies were used in the simulation. To calculate Local Authority estimates, a synthetic estimator approach was used, applying the DSE ratios for the Processing Unit to the original census count of each Local Authority separately.

From the 500 different estimates produced for each of the 500 replicate CCS samples, the variance of the average and corresponding RSE can be calculated. In cases with lower than 100% response rate, 500 different replicates of selecting which households were responding were used.

5. Appendix B

Measures of success for Scotland's Census 2021 objectives, as at November 2019[2].

How we will achieve high quality results?	How will we measure success? (Level 1 Key Performance Indicators (KPIs)[1] and acceptance levels)
We will maximise our overall person response rate	Person response rate[2] of at least 94%
We will ensure a minimum level of response with every local authority in Scotland	Person response rate in every council area of at least 85%.
We will maximise the accuracy of our national population estimates	Variability[3]: national estimates will achieve 95% Confidence Intervals (CI) +/- 0.4%; Bias: < 0.5%
We will maximise the accuracy of our local authority population estimates	Variability[4]: council area estimates will achieve 95% CI +/- 3%
We will minimise the non-response to all mandatory questions	Achieve or exceed target non-response rates for all mandatory questions
Our data will demonstrate high agreement rates with post coverage quality surveys	Agreement rates of at least XX%[5] achieved for all questions
All national and local authority level results for each main release will be assessed by a quality assurance panel	Undertaken with no residual issues remaining
We will publish details of methods and full details of all our data quality indicators	Published on our website
We will publish the results of an independent methodology review	Positive review published.
We will maintain our National Statistics Accreditation	Accreditation maintained throughout

1. Lower-level KPIs may sit below individual Level 1 KPIs.
2. Precise measure for person response rate to be defined.
3. This target is under review.
4. This target is under review.
5. Precise measure for agreement rate to be defined.

[2] As found in Scotland's Census 2021 Statistical Quality Assurance Strategy https://www.scotlandscensus.gov.uk/documents/Statistical%20Quality%20Assurance%20Strategy.pdf

6. Appendix C

List of Acronyms

CCS	Census Coverage Survey
CE	Communal Establishment
DV	Design variables
E&A	Estimation and Adjustment
HtC	Hard to Count Index
KPI	Key Performance Indicator
LA	Local Authority
NRS	National Records of Scotland
ONS	Office of National Statistics
PSU	Primary Sampling Unit
RSE	Relative Standard Error
SSU	Secondary Sampling Unit

Geography Definitions

Data Zone	The data zone geography covers the whole of Scotland and nests within local authority boundaries.
Hard to Count Index	The Hard to Count index is a scale of 1 (easiest to count) to 5 (hardest to count) which was created to indicate how difficult it may be to enumerate a particular geographical area based on certain demographic features.
Local Authority	Local Authorities are the 32 council areas within Scotland.
Planning Areas	Planning Areas are geographic areas built from groups of postcodes and averaging between 200-400 residential addresses. They nest within Local Authorities.

7. References

Brown, J., Abbott, O., & Smith, P.A. (2011). Design of the 2001 and 2011 Census Coverage Surveys for England and Wales, *Journal of the Royal Statistical Society.* 174(*4*), pp. 881-906.

Castaldo, A (2018a). 2021 Census Coverage Survey Design Strategy. Version 4. *The Office of National Statistics.*

Castaldo, A. & Nikolakis D. (2018b). Assessing the use of an Address Based Design for the 2021 Census Coverage Survey. *The Office of National Statistics.*