

Scotland's Census 2022

Filter Rules: Methodology

February 2023

Contents

1. Plain English Abstract	3
2. Abstract	5
3. Background and Introduction.....	7
4. 2011 Census Filter Rules Methodology	9
4.1 2011 Filter Rules.....	10
4.2 Additional Filter Rules (“CANCEIS Pre-Processing” Steps).....	15
4.3 2011 Harmonisation with Other UK Offices (ONS, NISRA).....	16
5. 2022 Filter Rules Methodology	17
5.1 Differences Between Paper and Online Questionnaires	18
5.1.1 Age-Based Routing Differences in Paper vs Online Questionnaires	19
5.2 Processing Software	21
5.3 Methodology Changes Needed or Considered for 2022.....	22
5.4 2022 Filter Rule Methodology	23
6. Strengths and Limitations	63
7. Conclusion.....	63
8. References	65

1. Plain English Abstract

Every 10 years or so, the people of Scotland (as well as the UK) are required to fill out a Census, which contains questions about their everyday lives, and gives the people of the country a snapshot into daily life at a certain point in time. This picture assists governments and other agencies on making decisions for planning and services.

Some questions in Scotland's Census are more applicable to a respondent than others, and when this occurs the respondent is often asked to skip certain questions - usually based on a previous answer, or on criteria such as age. However, if not done correctly, this can create contradictory responses. For example, this can occur if the respondent misses the guidance to skip a question (or a group of questions) and subsequently provide answers they are not meant to or do not have to. Since these questions do not apply to them, the answers tend to be misinterpreted and of poorer quality, creating conflicts within a respondent's census record.

Generally, these types of conflicts are resolved in a process called Edit and Imputation. However, the number of conflicts which need to be resolved can put unnecessary strain on computing resources, and some of the responses simply require reinforcement of the routing structure on the questionnaire. As such, a process called the **Filter Rules** is run prior to Edit and Imputation, to reduce this burden - firstly, by resolving inconsistencies (where a response can reasonably be assumed), and secondly, by signalling to Edit and Imputation that a response is not required for a particular question.

For 2022, the Filter Rules primarily reinforces or applies routing for *paper* questionnaires, as the online questionnaire automatically applies routing during the course of a response. To ensure we treat data in a consistent manner however, online responses will also be processed through the Filter Rules.

This paper outlines the background to the Filter Rules, its place in census data processing, and how it was used in 2011, before outlining the changes to the process for Scotland's Census 2022, and detailing the methodology to be used.

2. Abstract

A census is taken every 10 years or so in Scotland and the UK, which asks questions about a person's daily life. Collectively, the responses are processed, compiled and the aggregate answers released so that the country has a picture of what life in Scotland at that particular point in time looks like.

However, not all questions are applicable to every person, and so respondents are asked to skip questions that do not pertain to them (often referred to as skip logic). On paper questionnaires in particular, this can present a problem, as respondents have the opportunity to see and answer any question. By answering questions that aren't applicable to them, respondents can inadvertently provide conflicting answers.

Equally, where respondents have answered questions that were appropriately skipped, it is necessary to distinguish these at the item (question) response level from those questions which were skipped, but should have been answered.

In processing census data for outputs, the step where inconsistencies and missing responses are resolved is a statistical process called Edit and Imputation, or E&I. The process uses hot-deck (within dataset) donor methodology which identifies and resolves missing responses and inconsistencies to create a whole, plausible dataset which reflects the population in an accurate manner.

However, the number of conflicts that would need to be resolved in Edit and Imputation can be innumerable, which can put a strain on processing. The Edit and Imputation system also does not distinguish between a missing response which requires resolution, and a missing response which does not because it was appropriately skipped. Further, the donor methodology that Edit and Imputation uses benefits from as many "complete" questions as possible, maximising the pool of donor responses that can be used.

As such, a step run prior to Edit and Imputation called **Filter Rules** is used to reinforce the questionnaire routing on the Census questionnaire, by re-coding appropriately skipped responses from "missing" to "no code required", signalling a

complete answer to Edit and Imputation. Additionally, it resolves basic inconsistencies by deterministically applying rules in situations where the response can be reasonably assumed. This in turn relieves Edit and Imputation of some burden, and maximises the donor pool.

This paper will take a look at the Filter Rules methodology that was used in 2011, and outlines the changes to the process to be used for Scotland's Census in 2022.

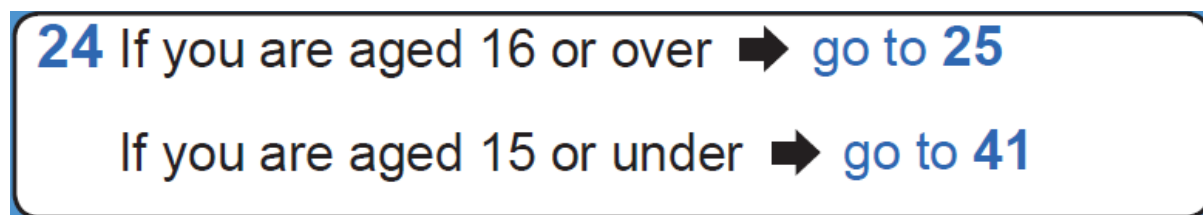
Note: On 17 July 2020 Scottish Government announced the decision to move Scotland's Census to 2022 following the impact of the COVID-19 pandemic. The information included in this report reflects the methodology intended, at the time of publication, to be used in the 2022 Census. It is not expected that there will be any major differences between the methodology presented here and that used. However, some detail may change or be completed before or during census processing. Any major changes to the intended methodology will be described in an update here.

3. Background and Introduction

Traditionally in Scotland (and elsewhere in the UK), a census questionnaire has been issued every 10 years to every person and household in the country, asking respondents questions about themselves (and the household). Primarily, this was done in a paper format, but 2022 will mark the first census in Scotland where the majority of responses will be conducted online.

Generally, a respondent answering a questionnaire is not required to provide an answer for every question, depending on what the questions are and the respondent's circumstances. The same holds true for Scotland's Census - for example, those under the age of 16 are asked to skip subsequent questions about their employment activity (Figure 1). When guidance to skip subsequent questions is not followed (or misinterpreted), the potential for these answers to be inconsistent with earlier answers on the questionnaire exists. This creates issues in data processing, particularly for the Edit and Imputation (E&I) function - a process where missing and/or contradictory responses are resolved.¹

Figure 1: Example of Routing Guidance in the Census Questionnaire



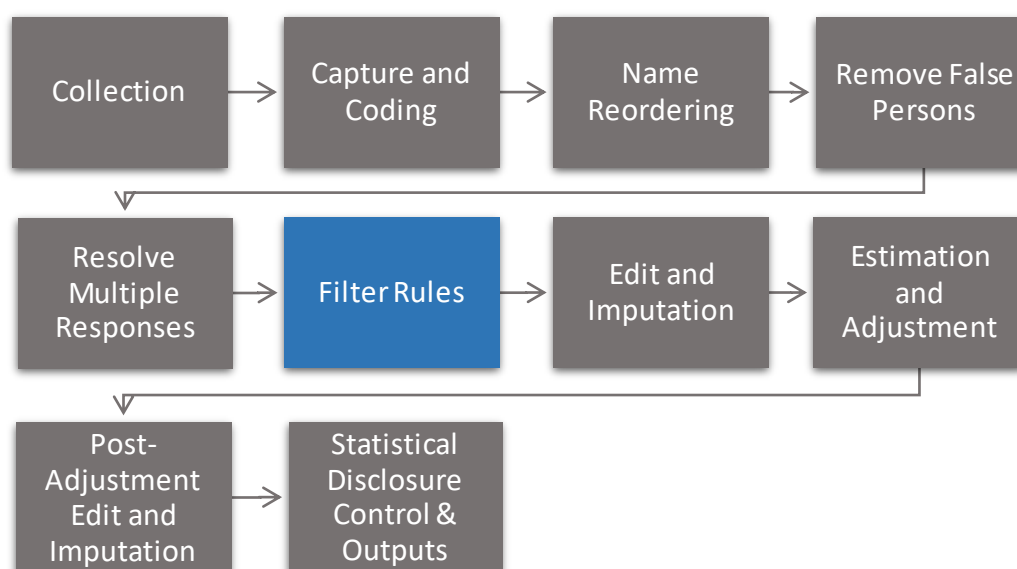
During census processing, there are a number of pre-determined rules which are implemented to assist in addressing this issue. They cover two overall purposes - firstly, to resolve inconsistencies **where the answer can reasonably be assumed**, and secondly to reinforce routing by turning *appropriately* skipped responses from “missing” to “no code required”. Marking these values as such allow a more accurate measurement of item (question) level non-response.

¹ Please see the Edit and Imputation Overview for more information on this process, located at: <https://www.scotlandscensus.gov.uk/media/cmepkqpp/scotland-s-census-2022-emaps-pmp012-overview-of-edit-and-imputation-for-2022-pdf-61.pdf>

Collectively, these pre-determined rules are referred to as the **Filter Rules**.

Filter Rules is one of the pre-processing, or “data cleansing” steps in census data processing. Data cleansing is a collection of preliminary processes which are largely deterministic, responsible for cleaning and preparing the data for further statistical processing. In 2022, data cleansing covers the Name Reordering, Remove False Persons, Resolve Multiple Responses and Filter Rules processes². Filter Rules is the last of these pre-processes, but closely linked with Edit and Imputation, which follows directly after (Figure 2, below).

Figure 2: Simplified Overview of the 2022 Census Data Journey



It is important to note that the Filter Rules process *primarily* applies to responses collected via paper questionnaires. Inconsistent answers or incorrect routing is a much smaller issue for responses made through the census questionnaire online. As a feature of the online questionnaire, respondents are simply brought to the next question, based on the answer(s) given in preceding questions. This prevents the respondent from seeing unnecessary questions, which reduces respondent burden

² For more details about these processes, please see the relevant published methodologies: <https://www.scotlandscensus.gov.uk/external-methodology-assurance-panels-emaps-0>

and minimises the ability to provide inconsistent answers. On paper however, a respondent is able to see all of the questions at once, providing the opportunity to create these inconsistencies.

Questionnaire routing, and therefore Filter Rules, are applied to Census Household, Household Person and Individual questionnaires, and this paper covers the methodology used in resolving inconsistencies or routing errors in the Census dataset. A similar filter rules methodology will be used for the Census Coverage Survey (CCS), with the difference being in the number of filter rules needed for the CCS due to it being a smaller questionnaire.

4. 2011 Census Filter Rules Methodology

In 2011, the Filter Rules step was an automated process written in SAS, a statistical/analytical programming language that is used to make data transformations and conduct data analysis. The Filter Rules were pre-determined, based on questionnaire content (routing and guidance). These rules essentially:

- a. Located records where the answering path was incorrect; for example, if individuals under 16 chose to answer the “Employment Activity”³ questions. Individuals under the age of 16 were asked to skip these questions. Where appropriate, Filter Rules changed the subsequent Employment Activity answers to a code which indicated the provision of these answers were unnecessary (known as “no code required”, or NCR).

Located records where an answer was inconsistent with subsequent answers on the questionnaire. Where appropriate, Filter Rules changed either the answer to the routing question OR subsequent answers to “no code required” (or, in a few circumstances, “missing”). For example, where a person answered that they owned their home, they were asked to skip the landlord question. If the landlord question was answered, it created an inconsistent

³ The group of questions on the census that related to a person’s economic activity

response, and the filter rule changed the landlord answer to “no code required” as we prioritised the ownership question (ie, making answering the landlord question is unnecessary).

- b. Located records where an answer was inconsistent with subsequent answers on the questionnaire, but, based on the information provided in these answers, a resolution could not be reasonably or logically determined. In these cases, no action was taken and these were left for Edit and Imputation steps to resolve. An example of this was where a person’s age was stated as younger than 16, but the respondent subsequently filled all of the economic activity questions. Here, we could not reasonably assume which was filled correctly (age or economic activity), so no changes are made and this falls to E&I to resolve.

Where inconsistencies exist, the filter rules may take into consideration the information which has been provided (even when it shouldn’t have been), rather than rely purely on the answer of the routing question. Answers where combinations cannot be reasonably assumed are left for the Edit and Imputation steps to resolve, as the donor methodology used is more appropriate for making these resolutions. Filter rules aim to reduce the burden on E&I however by processing these logical, pre-determined steps upfront. Reinforcing routing (ie by recoding values from “missing” to “no code required”) allows the donor pool for E&I to be maximised.

4.1 2011 Filter Rules

In 2011, the Filter Rules concerned the following questions⁴:

1. Tenure (H10) and Landlord (H11) Questions (Household Section)

⁴ What follows is a summary and/or general description of the rule as it was in 2011 - however, there may be further details not discussed. Each rule contained a number of sub-rules which determined specific actions (or noted potential actions).

5. 0 Year Olds (2) and Address One Year Ago (10)

10 One year ago, what was your usual address?
◆ If you had no usual address one year ago, state the address where you were staying.

The address on the front of this questionnaire

Student term-time / boarding school address in the UK, please write in below

Another address in the UK, please write in

Outside the UK, please write in country

Postcode

For the purposes of children under the age of 1 (“0 Year Olds”), it was found that sometimes householders would sometimes write in an address for them - presumably the address that the parents were in at that time - as there was no guidance on the questionnaire suggesting that they skip the question. However, for the purposes of processing this was unnecessary, and so this rule existed to “re-code” the answered Address One Year Ago variables to No Code Required.

6. Working Age (22 - Guidance)

22 If you are aged 16 or over → Go to **23**
If you are aged 15 or under → Go to **38**

Questions 23 - 37 on the 2011 Census questionnaire were about a person’s Economic Activity (qualifications or employment). We do not require this information from those individuals under the age of 16, so guidance (22) directed them to skip these questions. However, they were sometimes answered by under 16s regardless, and in those cases this Filter Rule turned the employment questions to No Code Required to follow routing guidelines.

This rule also routed past the qualifications question - the rule concerning qualifications corrected a different type of mistake.

7. Qualifications (23)

23 Which of these qualifications do you have?

◆ Tick all that apply.

- O Grade, Standard Grade, Access 3 Cluster, Intermediate 1 or 2, GCSE, CSE, Senior Certificate or equivalent
- SCE Higher Grade, Higher, Advanced Higher, CSYS, A Level, AS Level, Advanced Senior Certificate or equivalent
- GSVQ Foundation or Intermediate, SVQ level 1 or 2, SCOTVEC Module, City and Guilds Craft or equivalent
- GSVQ Advanced, SVQ level 3, ONC, OND, SCOTVEC National Diploma, City and Guilds Advanced Craft or equivalent
- HNC, HND, SVQ level 4 or equivalent
- Degree, Postgraduate qualifications, Masters, PhD, SVQ level 5 or equivalent
- Professional qualifications (for example, teaching, nursing, accountancy)
- Other school qualifications not already mentioned (including foreign qualifications)
- Other post-school but pre-Higher Education qualifications not already mentioned (including foreign qualifications)
- Other Higher Education qualifications not already mentioned (including foreign qualifications)
- No qualifications

The *Qualifications (23)* rule looked at situations where there may have been discrepancy within the qualifications answer itself. As an example, if a respondent had ticked that they had any of the listed qualifications (such as school or post-secondary) but also went on to tick the last box, “no qualifications”, the Filter Rule subsequently set the tickbox answer for no qualifications to No Code Required.

8. Work Last Week (24), Ever Worked (29) (Economic Activity)

24 Last week were you:

◆ Tick all that apply.

◆ Include any paid work, including casual or temporary work, even if only for one hour.

- working as an employee? → Go to **30**
- on a Government sponsored training scheme? → Go to **30**
- self-employed or freelance? → Go to **30**
- working paid or unpaid for your own or your family's business? → Go to **30**
- away from work ill, on maternity leave, on holiday or temporarily laid off? → Go to **30**
- doing any other kind of paid work? → Go to **30**
- none of the above

29 Have you ever worked?

- Yes, please write in the year you last worked
 → Go to **30**
- No, have never worked → Go to **38**

The Filter Rule which looked at the Economic Activity questions reviewed questions 24 to 37 and potentially made a number of changes by balancing groups of answers - if the respondent indicated they were working, if they have indicated they ever worked (and the last year they were in work), and two groups of employment variables. This rule was the most complex of all the rules in 2011, partly because of the combination tick/write in box on question 29, and had the largest number of sub-rules.

9. Rules on the Placeholder Questionnaire

In addition to these, there were two filter rules which concerned Placeholder Forms (forms which were filled out by enumerators and returned in place of a response from a householder). However, routing for placeholder forms will be unnecessary for 2022 and so the 2011 equivalents will not be discussed here.

4.2 Additional Filter Rules (“CANCEIS Pre-Processing” Steps)

During the live running of Census Data Processing in 2011, Imputation found some data issues in the first processing unit, thought likely to be prevalent throughout the rest of the data. These issues were addressed by creating five extra Filter Rules to assist with the donor process in Edit and Imputation, primarily to maximise the donor pool or re-code responses so that the Edit and Imputation software, CANCEIS, functioned as intended. Four of the rules were addendums to already existing rules (Students, Qualifications, Travel to Work and 0 Year Olds). A rule concerning the Long Term Health question was the only entirely new rule added.

The additional filter rules concerned the following questions:

1. Student and Term Time Address
2. Qualifications (for Under 16s)
3. Long-Term Health Conditions
4. Travel to Work (Working from Home)
5. 0 Year Olds and Address One Year Ago

For 2022, these rules will be incorporated into their corresponding existing Filter Rules, if they remain applicable to the 2022 dataset.

4.3 2011 Harmonisation with Other UK Offices (ONS, NISRA)

In 2011, there were three offices responsible for carrying out their respective Census. In England and Wales this was headed by the Office for National Statistics (ONS) and the Census in Northern Ireland was led by Northern Ireland Statistics and Research Agency (NISRA). Alongside this, the National Records of Scotland (NRS) conducted the Census for Scotland (then known as the General Register Office of Scotland, or GROS).

Scotland's 2011 Filter Rules process was originally inherited from the Office of National Statistics, and adapted to the Scottish Census to account for the differences between the Census questionnaires and routing. The goal of the process remained the same between all Census offices - to resolve inconsistencies (where the response could reasonably be assumed), and to enforce questionnaire routing (by changing the appropriate "missing" responses to "no code required"), thereby reducing load and burden on subsequent Edit and Imputation processes.

However, the application of these rules may differ somewhat between the offices, as it is entirely dependent upon the questions and structure of each country's Census. This being said, the questionnaires remain similar in many ways, as comparability across countries (and over time) is a significant consideration when the questionnaires are developed.

5. 2022 Filter Rules Methodology

Overall, the context for using Filter Rules remains the same for 2022 as it did in 2011. As a process, it is required as a precursor to Edit and Imputation in order to ensure consistency in the application of questionnaire routing, relieve some of the burden from Edit and Imputation processing, and maximise the E&I donor pool.

Thus, the objectives of Filter Rules in 2022 are to:

1. Ensure the application of questionnaire routing in responses

Primarily, this will apply to paper responses as online responses should skip questions that do not apply, and automatically assign the skipped questions with “no code required” (-5) rather than “missing” (-9). This distinguishes appropriate non-response at the item (question) level from those questions which were skipped but should have been answered, and signals to Edit and Imputation processes that the remaining “missing” items require imputation.

2. Resolve (certain, limited) inconsistencies

The Filter Rules will also resolve certain inconsistencies, where we can reasonably assume the response. This is done through creating pre-determined rules which follow the questionnaire routing and take into account the intended meaning of which the question was written.

For example, where a person has answered that they own their home, they are then asked to skip the landlord question. If the landlord question is answered, creating an inconsistent response, it is changed to “no code required” as we prioritise the ownership question - thus making answering the landlord question unnecessary (this example is the same in 2022 as in 2011).

It is important to note that while these specified changes are a necessary part of data preparation, in cases where ambiguity exists between responses and we cannot reasonably assume a response, changes are **not** made. These types of conflicts are evaluated through the Edit and Imputation methodology⁵, which is more appropriate to resolve such contradictions. The Filter Rules only make certain pre-determined corrections, based on the intended routing of the questionnaire. In those cases, the specification should read a “no action”, to indicate that we have considered the scenario better suited to be resolved by Edit and Imputation’s donor method (rather than the deterministic method employed by the Filter Rules).

3. Create a flag dataset

The 2011 Filter Rules produced a flag dataset at the variable level where, if a particular variable was changed, a flag would be set to indicate it had done so. However, because several rules had the ability to make changes to a particular variable, it was difficult to untangle which rule(s) caused the change - making evaluation more cumbersome to do. For 2022, a larger range of flags will be produced based on the itemised step of each filter rule, allowing the ability to trace back not only to the variable which was changed, but the step in the rule which initiated the change. This should help ensure that the changes made were correctly applied, and assist in tracking the scale at which Filter Rules makes these changes.

5.1 Differences Between Paper and Online Questionnaires

Although Scotland’s Census in 2011 was delivered both online and on paper, the majority of returns in 2011 were made through paper - approximately 80%. In 2022, household respondents will first be provided with an Internet Access Code (IAC) and encouraged to provide their census return online, with paper questionnaires available for those who request one. Accordingly, we expect many more online

⁵ The Edit and Imputation methodology paper can be found: <https://www.scotlandscensus.gov.uk/media/cmepkqpp/scotland-s-census-2022-emaps-pmp012-overview-of-edit-and-imputation-for-2022-pdf-61.pdf>

responses in 2022 than in 2011 (supported by the 2019 rehearsal), and fewer paper responses. The programme is aiming for a completion rate of 80% for *online* responses, inverse to 2011 - leaving 20% on paper.

When respondents complete the census questionnaire online in 2022, the Online Collection Instrument (OCI) will automatically route respondents past questions that they do not have to answer, based on the answer they have provided (the “routing question”). However, on paper responses, this will be unavailable to respondents and the opportunity to view all questions - and hence the ability to make mistakes and/or provide inconsistent responses - will exist. To minimise the potential bias between the two modes of completion, rules are developed as consistently as possible with the routing of the questionnaire, and all responses (both online and paper) will be processed through the filter rules. This may also have the added benefit of catching potential issues in the routing of online responses.

5.1.1 Age-Based Routing Differences in Paper vs Online Questionnaires

Age Routing is where respondents under a specified age are not required to answer certain questions. For example, in 2011, there was guidance provided to indicate that under 16s did not have to answer the qualifications or employment questions (regardless of the question applied/could be answered by the respondent or not).

22 If you are aged 16 or over → Go to **23**
If you are aged 15 or under → Go to **38**

In 2011, this “Under 16s” rule, and an unwritten rule about the address of babies under the age of one, were the only situations to which this applied.

The Census questionnaire in 2022 has changed so that age-based routing will apply to other questions as well:

Questions asked of people aged 1 and over:

- The question on address 1 year ago (question 11 in the person section)

Questions asked of people aged 3 and over:

- All language questions, which include English Skills (13), Gaelic and Scots Skills (14), the BSL question (15) and the main language question (16)
- The question on the provision of unpaid care (12)

Questions asked of people aged 4 and over:

- The question on whether you are a school child or full-time student (6)
- The questions on travel to place of work or study (42 and 43)

Questions asked of people aged 16 or over:

- The new (voluntary) question on trans status or history (4)
- The new (voluntary) question on sexual orientation (8)
- The new question on Ex Service status (26)
- The question on legal marital and civil partnership status (5)
- The qualifications held question (25)
- The questions on employment and activity last week (27 - 40)

However, the routing concerning these questions will be applied differently between online and paper questionnaires. Online, guidance and automatic routing will direct respondents through the appropriate answering path as outlined above. While paper questionnaires by nature give respondents the opportunity to answer any and all questions (mistakenly or otherwise), the 2022 paper questionnaire will *not* be providing guidance for age-based routing⁶, meaning it is likely that we will see responses to questions that we would have routed around online.

However, the routing will apply to all respondents regardless of if the questionnaire is answered online or through paper, so the routing for these responses will need to be implemented during processing, partially through the Filter Rules. In such cases, we

⁶ Aside from on the two new voluntary questions regarding trans status and sexual orientation, the marital status question, and the under 16s qualifications and employment questions (as in 2011).

may find that the rules for age-based routing will result in constant and numerous changes for paper questionnaires.

This brings up some uncertainty in the event where a respondent's age is incorrect or ambiguous, as there have been known quality issues surrounding the date of birth answer (capture issues, or respondent mistake). Normally, as with the Under 16s rule, the Filter Rules would take into account if there is contradictory evidence between the stated age and subsequent questions that respondents are asked to skip if the questions did not apply to them. With some age routing however the number of questions that an "under 1" or "under 3" would be asked to answer is quite small, and basing changes on such information may risk introducing bias into the dataset. For these cases, it is proposed that the only changes that the Filter Rules would make would be where the age agrees with the routing; thus the change would be from (-9) "missing" to (-5) "no code required", with the rest left to Edit and Imputation to resolve. Please see the specification in section 5.4 for details.

5.2 Processing Software

In 2011, the Filter Rules were coded and processed using SAS, a statistical software package which was used for most of Census data processing by the National Records of Scotland (NRS).

For 2022, a decision was made to process the Filter Rules in specialist Edit and Imputation software called CANCEIS (the CANadian Edit and Imputation System)⁷. This was due to the number of errors which were found in the original SAS code and the lack of versioning to inform how it related to the specification. Since the rules themselves require review and re-write (due to changes in variable structure, formats, questionnaire layout, etc), writing and modifying the rules into CANCEIS was tested with 2011 Census data and it was decided that development time for the rules in CANCEIS is acceptable. Further, since the Edit and Imputation process will be run in CANCEIS, integration solutions will be required regardless of the software

⁷ For more information on CANCEIS, please refer to the following page:
<https://www.istat.it/en/methods-and-tools/methods-and-it-tools/process/processing-tools/canceis>

selected for the use of Filter Rules. Since it is a related process to Edit and Imputation, CANCEIS seemed to also be the best fit for the software to process Filter Rules.

CANCEIS contains two engines to process data - a deterministic (rules-based) engine, and the donor engine that Edit and Imputation primarily uses. Filter Rules will solely use the deterministic engine.

SAS will still be used in other parts of data processing, as well as to facilitate the load and extraction of text files which are processed by CANCEIS.

5.3 Methodology Changes Needed or Considered for 2022

As many aspects of the 2022 Census questionnaire are similar to the questionnaire in 2011, as a starting point many of the rules (and sub-rules) from 2011 will be carried over to be used in 2022. These act as a template for areas where the questionnaire didn't change (or didn't change much). From there, the methodology reviewed and changes are implemented where required. The major areas for change consist of:

1. Incorporation of new routing from the questionnaire - including age-based routing (as described in section 5.1.1).
2. Removal of routing that no longer applies
3. Re-work of the Economic Activity Routing Question
In 2011, the "Year Last Worked" question.
4. Updating the Dependent Variable Groups (aka Dependent Fields)

To determine whether questionnaire routing has been followed, Filter Rules checks whether a certain number of questions which should be skipped in the routing have been answered. These are known as 'dependent fields'.

For each rule which uses 'dependent fields', the groups have been updated appropriately with variables new to the census for 2022. . For example, the working age routing filter rule takes account of responses to new questions which are only applicable to those aged 16 or over. These are the armed forces veteran, sexual orientation and gender identity questions (note, while some of these are voluntary questions and respondents may choose not to answer them regardless of age, a response would still contribute to the overall "score" as we consider it evidence that perhaps the age stated on the questionnaire is incorrect).

For detail on each of the 2022 rules, the specification details each proposed rule below (Section 5.4).

5.4 2022 Filter Rule Methodology

Household Rules

- A. Filter Rule A - Tenure and Landlord

Person Rules

- B. Filter Rule B - Marital Status for Under 16s
- C. Filter Rule C - Full Time Education for 0 - 3 Year Olds
- D. Filter Rule D – Student and Term Time Address
- E. Filter Rule E – Country of Birth and Year of Arrival
- F. Filter Rule F – 0 Year Olds and Address One Year Ago
- G. Filter Rule G – Carers Aged 0 - 2
- H. Filter Rule H – Language Variables for 0 - 2 Year Olds
- I. Filter Rule I – Long Term Health
- J. Filter Rule J – Under 16s and Economic Activity
- K. Filter Rule K - Economic Activity
- L. Filter Rule L - Travel to Work or Study
- M. Filter Rule M - Method of Travel
- N. Filter Rule N - Travel or Work from Home

Notes:

-5 means “No Code Required”

-7 means “Invalid”

-9 means “Missing”

Filter Rule A – Tenure (H12) and Landlord (H13)

Filter Rule A is the only filter rule that is included in the Household section of the census questionnaire.

Item	Variable A	And	Variable B	Action (Then...)	Notes
	(If <i>tenure</i> is...)		(If <i>landlord</i> is...)		
1	Ticked, either: 1 (Owns outright), or 2 (Mortgage/Loan)		Ticked: 1 (Council/LA), or 2 (Private/Agency)	No Action	This is (essentially) the same as 2011.
2	Ticked, either: 1 (Owns outright), or 2 (Mortgage/Loan)		Ticked: 3 (Other)	Set: <i>landlord</i> to -5 (NCR)	This is (essentially) the same as 2011.
3	Ticked, either: 1 (Owns outright), or 2 (Mortgage/Loan)		-7 (Invalid), or -9 (Missing)	Set: <i>landlord</i> to -5 (NCR)	This is (essentially) the same as 2011.
4	Ticked, either: 3 (Owns w/ Shared Equity), or 5 (Part Owns/Part Rents)		Ticked: 1 (Council/LA), or 2 (Private/Agency)	No Action	In 2011, we did not route around the Landlord question for partial ownership answers to the Tenure question. However, we have elected to do so for 2022, so these three sub-rules reflect this change.
5	Ticked, either: 3 (Owns w/ Shared Equity), or 5 (Part Owns/Part Rents)		Ticked: 3 (Other)	Set: <i>landlord</i> to -5 (NCR)	

6	Ticked, either: 3 (Owns w/ Shared Equity), or 5 (Part Owns/Part Rents)		-7 (Invalid), or -9 (Missing)	Set: <i>landlord</i> to -5 (NCR)	
7	Ticked, either: 4 (Rents), or 6 (Lives Here Rent Free), or -7 (Invalid) -9 (Missing)	And	Any Value (ticked or unticked)	No Action	This is (essentially) the same as 2011.

Filter Rule B – Marital Status (5) for Under 16s

This is the first of several new age-based routing questions, and one where the guidance on paper questionnaires does not match the online routing. As such, a deterministic edit (in Filter Rules) will only occur where responses match routing appropriately - meaning any contradiction will be resolved in E&I.

To qualify as “Filled”, two or more questions in the Dependent Fields must be answered. See below for a list of the variables in the Dependent Fields group (generally questions that apply to 16+).

Item	Variable A (If <i>age</i> is...)		Variable B (If <i>mar_stat</i> is...)		Dependent Fields	Action (Then...)	Notes
1	0 - 15	&	Ticked: 1 (Never)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
2	0 - 15	&	Ticked: 1 (Never)	&	Not Filled	Set: <i>mar_stat</i> to -5 (No Code Required)	Generally expected behaviour; since there is no routing guidance for this question on paper, a response of “never” should be sufficient to route
3	0 - 15	&	Ticked: ANY <i>except</i> 1 (Never)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
4	0 - 15	&	Ticked: ANY <i>except</i> 1 (Never))	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
5	0 - 15	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
Item	Variable A		Variable B			Action (Then...)	Notes

	(If age is...)		(If <i>mar_stat</i> is...)		Dependent Fields		
6	0 - 15	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>mar_stat</i> to -5 (No Code Required)	Expected behaviour
7	16+	&	Ticked: ANY	&	Filled	No Action	Expected behaviour
8	16+	&	Ticked: ANY	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
9	16+	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
10	16+	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
11	-7 (Invalid), or -9 (Missing)	&	Ticked: ANY	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
12	-7 (Invalid), or -9 (Missing)	&	Ticked: ANY	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
13	-7 (Invalid), or -9 (Missing)	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve

14	-7 (Invalid), or -9 (Missing)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
----	----------------------------------	---	----------------------------------	---	-------------------	------------------	--

Dependent Fields
trans_stat
sex_orient
quals_school
quals_apprentice
quals_further
quals_higher
ex_service
emplyd_last_week
emplyd_last_week_grp2
other_act
look_work
avail_work
wait_work
ever_worked
emp_stat
employer
occupation
industry
supervisor
hours_worked

Filter Rule C – Full Time Education (6) and 0 - 3 Year Olds (2)

This is another (new) age-based filter rule where the guidance on paper questionnaires does not match the online routing. As such, a deterministic edit (in Filter Rules) will only occur where responses match routing appropriately - meaning any contradiction will be resolved in E&I.

To qualify as "Filled", two or more questions in the Dependent Fields must be answered. See below for a list of the variables in the Dependent Fields group (generally questions that apply to 16+).

Item	Variable A (If age is...)		Variable B (If student is...)		Dependent Fields	Action (Then...)	Notes
1	0 - 3	&	Ticked: 1 (Yes), or 2 (No)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
2	0 - 3	&	Ticked: 1 (Yes)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
3	0 - 3	&	Ticked: 1 (Yes), or 2 (No)	&	Not Filled	If 2 (No), Set: <i>student</i> to -5 (No Code Required)	Expected behaviour without guidance; in line with routing
4	0 - 3	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
5	0 - 3	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>student</i> to -5 (No Code Required)	Expected behaviour in line with routing

Item	Variable A (If age is...)		Variable B (If student is...)		Dependent Fields	Action (Then...)	Notes
6	4+	&	Ticked: 1 (Yes), or 2 (No)	&	Filled	No Action	Expected behaviour
7	4+	&	Ticked: 1 (Yes), or 2 (No)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
8	4+	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
9	4+	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
10	-7 (Invalid), or -9 (Missing)	&	Ticked: 1 (Yes), or 2 (No)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
11	-7 (Invalid), or -9 (Missing)	&	Ticked: 1 (Yes), or 2 (No)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
12	-7 (Invalid), or -9 (Missing)	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
13	-7 (Invalid), or -9 (Missing)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve

Dependent Fields
trans_stat
mar_stat
sex_orient
quals_school
quals_apprentice
quals_further
quals_higher
ex_service
emplyd_last_week
emplyd_last_week_grp2
other_act
look_work
avail_work
wait_work
ever_worked
emp_stat
employer
occupation
industry
supervisor
hours_worked

Filter Rule D – Student (6) and Term Time Address (7)

For Variable Groups A and B: Variable Groups are made up of several different questions (variables), noted below. In order to be considered “Filled”, there needs to be a valid response to at least three or more questions in Group A. “Not Filled” is a response to two or fewer questions.

Item	Variable A		Variable B		Variable Group	Action (Then...)	Notes
	(If <i>ft_student</i> is...)		(If <i>term_time</i> is...)		A		
1	Ticked: 1 (Yes)	&	Ticked: 1 (Address on Front)	&	Filled	No Action	As in 2011 (expected outcome)
2	Ticked: 1 (Yes)	&	Ticked: 1 (Address on Front)	&	Not Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
3	Ticked: 1 (Yes)	&	Ticked: 2 (Another Address)	&	Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
4	Ticked: 1 (Yes)	&	Ticked: 2 (Another Address)	&	Not Filled	Set: Group B to -5 (No Code Required)	As in 2011 (expected outcome)
5	Ticked: 1 (Yes)	&	-7 (Invalid), or -9 (Missing)	&	Filled	Set: <i>term_time</i> to -5 (No Code Required)	As in 2011 (behaviour shows respondent simply missed <i>term_time</i> question)
6	Ticked: 1 (Yes)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>term_time</i> to -5 (NCR) Group B to -5 (NCR)	As in 2011 (behaviour shows it was likely the respondent missed ticking a response for <i>term_time</i>)

Item	Variable A		Variable B		Variable Group	Action (Then...)	Notes
	(If <i>ft_student</i> is...)		(If <i>term_time</i> is...)		A		
7	Ticked: 2 (No)	&	Ticked: 1 (Address on Front)	&	Filled	Set: <i>term_time</i> to -5 (No Code Required)	As in 2011 (respondent did not have to answer <i>term_time</i>)
8	Ticked: 2 (No)	&	Ticked: 1 (Address on Front)	&	Not Filled	Set: <i>term_time</i> to -5 (No Code Required)	As in 2011 (respondent did not have to answer <i>term_time</i> ; dependent fields will be resolved by E&I)
9	Ticked: 2 (No)	&	Ticked: 2 (Another Address)	&	Filled	Set: <i>term_time</i> to -5 (No Code Required)	As in 2011 (respondent did not have to answer <i>term_time</i>)
10	Ticked: 2 (No)	&	Ticked: 2 (Another Address)	&	Not Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
11	Ticked: 2 (No)	&	-7 (Invalid), or -9 (Missing)	&	Filled	Set: <i>term_time</i> to -5 (No Code Required)	Expected outcome. This incorporates the 2011 pre-processing rule which was later added during live.
12	Ticked: 2 (No)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>term_time</i> to -5 (No Code Required)	This incorporates the 2011 pre-processing rule which was later added during live. The dependent fields are left and should be resolved by E&I.
13	-7 (Invalid), or -9 (Missing)	&	Ticked: 1 (Address on Front)	&	Filled	No Action	As in 2011 (unable to determine; <i>ft_student</i> was missed; resolved by E&I)

Item	Variable A		Variable B		Variable Group	Action (Then...)	Notes
	(If <i>ft_student</i> is...)		(If <i>term_time</i> is...)		A		
14	-7 (Invalid), or -9 (Missing)	&	Ticked: 1 (Address on Front)	&	Not Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
15	-7 (Invalid), or -9 (Missing)	&	Ticked: 2 (Another Address)	&	Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
16	Ticked: 2 (No)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
17	Ticked: 2 (No)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	As in 2011 (contradictory response; resolved by E&I)
18	Ticked: 2 (No)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	As in 2011 (contradictory response; resolved by E&I)

The Dependent Variable Group A is the group of variables that is checked (if the respondent has filled three or more of) as evidence that the questions were to be answered. Group B is *all* the variables that are affected by the Filter Rule/routing (which will be changed to -5, No Code Required where appropriate).

In 2011, we did not check voluntary questions (of which there was only *religion*).

Dependent Group A		Dependent Group B	
	quals_higher	sex_orient	quals_higher
cob	ex_service	cob	ex_service
mig_arr_month	emplyd_last_week	mig_arr_month	emplyd_last_week
mig_arr_year		mig_arr_year	emplyd_last_week_grp2
add_1year_ago		add_1year_ago	other_act
		add_1year_ago_pc	look_work
		add_1year_ago_cty	avail_work
carer		carer	wait_work
lang_eng_understand		lang_eng_understand	ever_worked
lang_eng_speak	emp_stat	lang_eng_speak	emp_stat
lang_eng_read	employer	lang_eng_read	employer
lang_eng_write	occupation	lang_eng_write	occupation
lang_gael	industry	lang_gael	industry
	supervisor	lang_scots	supervisor
	hours_worked	lang_bsl	hours_worked
main_lang		main_lang	work_study_address
health		health	work_study_pc
lt_cond		lt_cond	work_study_cty
	work_study	lt_cond_other	work_study
disability		disability	
passport		passport	
		passport_other_1_of_2	
		passport_other_2_of_2	
		religion	
nat_ident_uk		nat_ident_uk	
		nat_ident_other	
ethnic_major		ethnic_major	
		ethnic_1_of_2	
		ethnic_2_of_2	
quals_school		quals_school	
quals_apprentice		quals_apprentice	
quals_further		quals_further	

Filter Rule E – Country of Birth (9) and Year of Arrival (10)

In 2011, where a respondent was born in a UK country - ie Scotland, England, Wales or Northern Ireland (*cob*), but filled in year of arrival (then known as *yrarr*) **AND** the year of arrival was after year of birth (then *dob*), this rule changed the country you were born to No Code Required.

However, for 2022 this Filter Rule will prioritise the Country of Birth variable, meaning the *mig_arr_month* and *mig_arr_year* variables will be turned to -5 (No Code Required) if a country within the UK is selected.

This is because it was found that the level of accuracy for the Country of Birth question was much higher than the Migration question (ONS 2011 Census Quality Survey), likely due to the ability to interpret the migration question differently than intended - for example, if a respondent thought it meant a return from a holiday outside the UK.

Item	Variable A		Variable B	Action (Then...)	Notes
	(if <i>cob</i> is...)		(if <i>mig_arr</i> is...)		
1	Ticked, either: 923 (Scotland), or 921 (England), or 922 (Northern Ireland), or 924 (Wales)	&	<i>mig_arr_month</i> and <i>mig_arr_year</i> are valid	Set: <i>mig_arr_month</i> and <i>mig_arr_year</i> to -5 (No Code Required)	This is slightly different to 2011; then, we <i>would</i> have also checked date of birth before setting the <i>mig_arr</i> variables to no code required. For 2022 we have not checked <i>dob</i> (as explained above).
2	Ticked, either: 923 (Scotland), or 921 (England), or 922 (Northern Ireland), or 924 (Wales)	&	<i>mig_arr_month</i> and <i>mig_arr_year</i> are invalid -7 (Invalid), or -9 (Missing)	Set: <i>mig_arr_month</i> and <i>mig_arr_year</i> to -5 (No Code Required)	As in 2011

Item	Variable A		Variable B	Action (Then...)	Notes
	(if <i>cob</i> is...)		(if <i>mig_arr</i> is...)		
3	Any Country Code OTHER THAN UK Countries (923, 921, 922, 924)	&	<i>mig_arr_month</i> and <i>mig_arr_year</i> are valid	No Action	As in 2011 (appropriate response, no action needed)
4	Any Country Code OTHER THAN UK Countries (923, 921, 922, 924)	&	<i>mig_arr_month</i> and <i>mig_arr_year</i> are invalid -7 (Invalid), or -9 (Missing)	No Action	As in 2011 (potentially missed response; E&I to resolve)
5	-7 (Invalid), or -9 (Missing)	&	<i>mig_arr_month</i> and <i>mig_arr_year</i> are valid	No Action	As in 2011 (E&I to resolve)
6	-7 (Invalid), or -9 (Missing)	&	<i>mig_arr_month</i> and <i>mig_arr_year</i> are invalid -7 (Invalid), or -9 (Missing)	No Action	As in 2011 (E&I to resolve)

Filter Rule F – 0 Year Olds (2) and Address One Year Ago (11)

In 2011, this rule was in two parts – the initial filter rule only set a “no code required” to responses if they were under 0 AND had filled out an address. However, during live processing it was found that a larger donor pool was required and a secondary rule was required which covered the rest of respondents younger than 1. For 2022 the rules have been combined.

Item	Variable A (if age is...)		Address 1 Year Ago	Action (Then...)	Notes
1	0	&	Filled - response to one or more variable: <i>add_1year_ago</i> <i>add_1year_ago_pc</i> <i>add_1year_ago_cty</i>	Set: All Address 1 Year Ago Fields to -5 (No Code Required)	As in 2011 (respondent does not have to answer the Address 1 Year Ago question if under the age of 1)
2	0	&	Not Filled -7 Invalid, or -9 Missing	Set: All Address 1 Year Ago Fields to -5 (No Code Required)	As in 2011 (appropriate response; incorporates the pre-processing rule implemented during 2011 live)
3	1+	&	Filled - response to one or more variable: <i>add_1year_ago</i> <i>add_1year_ago_pc</i> <i>add_1year_ago_cty</i>	No Action	As in 2011 (appropriate response)
4	1 +	&	Not Filled -7 Invalid, or -9 Missing	No Action	As in 2011 (resolved by E&I)

Item	Variable A (if age is...)		Address 1 Year Ago	Action (Then...)	Notes
5	-7 Invalid, or -9 Missing	&	Filled - response to one or more variable: <i>add_1year_ago</i> <i>add_1year_ago_pc</i> <i>add_1year_ago_cty</i>	No Action	As in 2011 (resolved by E&I)
6	-7 Invalid, or -9 Missing	&	Not Filled -7 Invalid, or -9 Missing	No Action	As in 2011 (resolved by E&I)

Address One Year Ago Fields
<i>add_1year_ago</i>
<i>add_1year_ago_pc</i>
<i>add_1year_ago_cty</i>

Filter Rule G – Carers (12) Aged 0 - 2 (2)

This is another (new) age-based filter rule where the guidance on paper questionnaires does not match the online routing. As such, a deterministic edit (in Filter Rules) will only occur where responses match routing appropriately - meaning any contradiction will be resolved in E&I.

To qualify as “Filled”, two or more questions in the Dependent Fields must be answered. See below for a list of the variables in the Dependent Fields group (generally questions that apply to 16+).

Item	Variable A		Variable B		Dependent Fields	Action (Then...)	Notes
	(If age is...)		(If carer is...)				
1	0 - 2	&	Ticked: 1 (No)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
2	0 - 2	&	Ticked: 1 (No)	&	Not Filled	Set: <i>carer</i> to -5 (No Code Required)	Generally expected behaviour; since there is no routing guidance for this question on paper, a response of “no” should be sufficient to route
3	0 - 2	&	Ticked: Any other than 1 (No)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
4	0 - 2	&	Ticked: Any other than 1 (No)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
5	0 - 2	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
6	0 - 2	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>carer</i> to -5 (No Code Required)	Expected behaviour

Item	Variable A		Variable B		Dependent Fields	Action (Then...)	Notes
	(If age is...)		(If student is...)				
7	3+	&	Ticked: 1 (Yes), or 2 (No)	&	Filled	No Action	Expected behaviour
8	3+	&	Ticked: 1 (Yes), or 2 (No)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
9	3+	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
10	3+	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
11	-7 (Invalid), or -9 (Missing)	&	Ticked: 1 (Yes), or 2 (No)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
12	-7 (Invalid), or -9 (Missing)	&	Ticked: 1 (Yes), or 2 (No)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
9	-7 (Invalid), or -9 (Missing)	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve
10	-7 (Invalid), or -9 (Missing)	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve

Dependent Fields
trans_stat
mar_stat
sex_orient
quals_school
quals_apprentice
quals_further
quals_higher
ex_service
emplyd_last_week
emplyd_last_week_grp2
other_act
look_work
avail_work
wait_work
ever_worked
emp_stat
employer
occupation
industry
supervisor
hours_worked

Filter Rule H – Language Variables (13 - 16) for 0 - 2 Year Olds (2)

This is another (new) age-based filter rule where the guidance on paper questionnaires does not match the online routing. As such, a deterministic edit (in Filter Rules) will only occur where responses match routing appropriately - meaning any contradiction will be resolved in E&I.

For the *language* variable group to qualify as “filled”, one or more questions must be answered. See below for a list of variables in the *language* variable group.

For the Dependent Fields to qualify as “Filled”, two or more questions must be answered. See below for a list of the variables in the Dependent Fields group (generally questions that apply to 16+).

Item	Variable A (If <i>age</i> is...)		Variable(s) B (If <i>language</i> is...)		Dependent Fields	Action (Then...)	Notes
1	0 - 2	&	Filled	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
2	0 - 2	&	Filled	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
3	0 - 2	&	Filled: <i>all</i> language variables are “no skills” and <i>main_lang</i> is -7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>language</i> group to -5 (No Code Required)	Expected behaviour without guidance; in line with routing
4	0 - 2	&	Not Filled, ie -7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.

Item	Variable A (If age is...)		Variable(s) B (If language is...)		Dependent Fields	Action (Then...)	Notes
5	0 - 2	&	Not Filled , ie -7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>language</i> group to -5 (No Code Required)	Expected behaviour.
6	3+	&	Filled	&	Filled	No Action	Expected behaviour.
7	3+	&	Filled	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
8	3+	&	Not Filled , ie -7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
9	3+	&	Not Filled , ie -7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
10	-7 (Invalid), or -9 (Missing)	&	Filled	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
11	-7 (Invalid), or -9 (Missing)	&	Filled	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.
12	-7 (Invalid), or -9 (Missing)	&	Not Filled , ie -7 (Invalid), or -9 (Missing)	&	Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.

Item	Variable A (If age is...)		Variable(s) B (If language is...)		Dependent Fields	Action (Then...)	Notes
13	-7 (Invalid), or -9 (Missing)	&	Not Filled , ie -7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	Response cannot be reasonably assumed - E&I to resolve.

Language Variable Group	Dependent Fields
lang_eng_understand	trans_stat
lang_eng_speak	mar_stat
lang_eng_read	sex_orient
lang_eng_write	quals_school
lang_gael	quals_apprentice
lang_scots	quals_further
lang_bsl	quals_higher
main_lang	ex_service
	emplyd_last_week
	emplyd_last_week_grp2
	other_act
	look_work
	avail_work
	wait_work
	ever_worked
	emp_stat
	employer
	occupation
	industry
	supervisor
	hours_worked

Filter Rule I – Long Term Health (18)

In 2011, this rule was put into CANCEIS pre-processing during live because it was found a large proportion of people did not respond to the question - presumably because those in good health may have missed the “no conditions” option at the bottom of the list. For 2022, this may not be as big a problem as online responses (which make up a greater proportion of submissions) should assist with data quality, but for paper, the list layout is nearly the same in 2022 as in 2011 - so this rule is incorporated here to maximise the donor pool.

Item	Variable A	Variable B	Variable C	Action (Then...)	Notes
	(If <i>lt_cond_other</i> is...)	(If <i>health</i> is...)	(If <i>disability</i> is...)		
1	Filled (Any)	Any	Any	No Action	As 2011.
2	Not Filled , either: -7 (Invalid) -9 (Missing)	Ticked: 1 Very Good 2 Good	Ticked: 3 No	Set: <i>lt_cond</i> to 0000000001	As 2011.

Filter Rule J – Under 16s (2) and Economic Activity (25 - 40)

Note that while this rule primarily deals with the Economic Activity questions, for 2022 it will also route the voluntary under 16 questions (question 4, *trans_stat* and question 8, *sex_orient*). Additionally, unlike 2011, the *qualifications* variables will also be routed here (in 2011 this was dealt with in a pre-processing step).

For the Dependent Fields to qualify as “Filled”, two or more questions must be answered. See below for a list of the variables in the Dependent Fields group (generally questions that apply to 16+).

Item	Variable A		Variable B		Dependent Fields	Action (Then...)	Notes
	(If age is...)		(If <i>emplyd_last_week_grp2</i> is...)				
1	0 - 15	&	1 (Working)	&	Filled	No Action	As 2011. Response cannot be reasonably assumed - E&I to resolve.
2	0 - 15	&	1 (Working)	&	Not Filled	Set: <i>Dependent Fields</i> to -5 (NCR)	As 2011
3	0 - 15	&	2 (Not Working)	&	Filled	Set: <i>Dependent Fields</i> to -5 (NCR)	As 2011 - conceivable behaviour IF under 16s answer these questions.
4	0 - 15	&	2 (Not Working)	&	Not Filled	Set: <i>Dependent Fields</i> to -5 (NCR)	As 2011. Generally in line with expected behaviour; just answered a question they were not required to.
5	0 - 15	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	As 2011. Response cannot be reasonably assumed - E&I to resolve

Item	Variable A		Variable B		Dependent Fields	Action (Then...)	Notes
	(If age is...)		(If <i>emplyd_last_week_grp2</i> is...)				
6	0 - 15	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	Set: <i>Dependent Fields</i> to -5 (NCR)	As 2011 - expected behaviour.
7	16+	&	1 (Working)	&	Filled	No Action	As 2011 - expected behaviour.
8	16+	&	1 (Working)	&	Not Filled	No Action	As 2011 - E&I to resolve.
9	16+	&	2 (Not Working)	&	Filled	No Action	As 2011.
10	16+	&	2 (Not Working)	&	Not Filled	No Action	As 2011 - E&I to resolve.
11	16+	&	-7 (Invalid), or -9 (Missing)	&	Filled	No Action	As 2011 - E&I to resolve.
12	16+	&	-7 (Invalid), or -9 (Missing)	&	Not Filled	No Action	As 2011 - E&I to resolve.

Dependent Fields			
trans_stat	ex_service	wait_work	supervisor
sex_orient	emplyd_last_week	ever_worked	hours_worked
quals_school	emplyd_last_week_grp2	emp_stat	
quals_apprentice	other_act	employer	
quals_further	look_work	occupation	
quals_higher	avail_work	industry	

Filter Rule K – Economic Activity (Working Last Week)

In 2011, the question, “Have you ever done any paid work?” was covered by two variables, *ever worked* and *last year worked*. *Ever worked* covered the tick response while a space was provided to write in a year for the *last year worked* variable. If the year “2011” was filled in the *last year worked* variable, the assumption generally was that the respondent was referring to current employment.

For 2022, the options for the *ever worked* variable have changed. There is no longer a *last year worked* variable and to account for this, the options have been split into “Yes, in the last 12 months” and “Yes but not in the last 12 months”. This has added some uncertainty to changing the *ever_worked* variable in the 2022 filter rule; as the census is issued in March of 2022, the equivalent to writing “2011” in *last year worked* is not the same as “Yes, in the last 12 months”. Thus, many changes that require choice between the two *ever worked* options have been left for E&I to resolve where we may originally (in 2011) have made changes.

For the Dependent Fields to qualify as “Filled”, 2 or more questions must be answered. See below for a list of the variables in the Dependent Fields group (generally questions that apply to 16+).

Item	Variable A	Variable B	Dependent Fields		Action (Then...)	Notes
	(<i>emptyd_last_week_grp2</i>)	(<i>ever_worked</i>)	A	B		
1	1 (Working)	1 (Last 12 Months)	Filled	Filled	No Action	Unable to tell if answers pertain to current or past employment. E&I to resolve. This is similar to what was done in 2011.
2	1 (Working)	1 (Last 12 Months)	Filled	Not Filled	No Action	Unable to tell if answers pertain to current or past employment, and respondent should have filled Fields B according to routing question <i>ever_worked</i> . E&I to resolve. This is a different to what was done in 2011, but due to question change.
3	1 (Working)	1 (Last 12 Months)	Not Filled	Filled	Set: <i>ever_worked</i> to -5 <i>Fields A</i> to -5	A respondent currently working who followed routing, except they also answered <i>ever_worked</i> .. This is the same as 2011 (except there is no <i>last year worked</i> to change).
4	1 (Working)	1 (Last 12 Months)	Not Filled	Not Filled	Set: <i>ever_worked</i> to -5 <i>Fields A</i> to -5	A respondent who did not fill Fields A or B, so followed at least part of the routing. They <i>should have</i> filled B (but not A) so leave B as missing and enforce routing by setting Fields A to -5. This is the same as 2011 (except for <i>lastyrwrk</i> , as it does not exist in 2022).
5	1 (Working)	2 (Not Last 12 Months)	Filled	Filled	Set: <i>emptyd_last_week</i> to 000001 and <i>emptyd_last_week_grp2</i> to 2 Not Working	A respondent who said they are both currently working but haven't worked in the last 12 months; however, routing is followed except for the Working Last Week question. This is similar to the rule in 2011.

6	1 (Working)	2 (Not Last 12 Months)	Filled	Not Filled	Set: <i>emptyd_last_week</i> to 000001 and <i>emptyd_last_week_grp2</i> to 2 Not Working	A respondent who said they are both currently working but haven't worked in the last 12 months, however, routing is followed except for the Working Last Week question. They should have filled B, so leave these as missing. This is similar to the rule in 2011.
7	1 (Working)	2 (Not Last 12 Months)	Not Filled	Filled	No Action	A respondent who said they are both currently working but haven't worked in the last 12 months. Unable to follow routing/determine which answers pertain to, so no action.
8	1 (Working)	2 (Not Last 12 Months)	Not Filled	Not Filled	No Action	As above.
9	1 (Working)	3 (Never Worked)	Filled	Filled	No Action	A contradictory response where the respondent answers they are both working and have never worked. However, they provided all Fields A and Fields B, so there is an indication that they have at least worked in the past. In 2011, we changed the <i>emptyd_last_week</i> variables (to not working) and <i>lastyrwrkd</i> to missing (to allow E&I to resolve). This question isn't the same for 2022 so we will allow E&I to resolve. In a way, this is similar to Item 1 (2022).

10	1 (Working)	3 (Never Worked)	Filled	Not Filled	Set: <i>emptyd_last_week</i> to 000001 <i>emptyd_last_week_grp2</i> to 2 Not Working and <i>Fields B</i> to -5 No Code Required	A contradictory response where the respondent answers that they are working and have never worked. Since they filled Fields A but not B, there is heavier indication that they have never worked (routing followed except for emptyd variables). As such, we can resolve the inconsistency by changing the emptyd_last_week variables (and setting B to NCR). This is similar to what we did in 2011.
11	1 (Working)	3 (Never Worked)	Not Filled	Filled	No Action	A contradictory response where the respondent answers they are both working and have never worked. We don't have evidence whether the respondent is currently working or has not worked in the last 12 months so leave this for E&I to resolve (this is also what we did in 2011).
12	1 (Working)	3 (Never Worked)	Not Filled	Not Filled	No Action	A contradictory response where the respondent answers they are working and have never worked. Can't tell if routing was followed so leave for E&I to resolve.
13	1 (Working)	-7 (Invalid), or -9 (Missing)	Filled	Filled	Set: <i>ever_worked</i> to -5 <i>Fields A</i> to -5	Similar to 2011
14	1 (Working)	-7 (Invalid), or -9 (Missing)	Filled	Not Filled	No Action	In 2011, we changed the <i>emptyd</i> variables to Not Working and <i>ever_worked</i> to missing to allow E&I to resolve. This isn't quite the same in 2022, so No Action to allow E&I to resolve.

15	1 (Working)	-7 (Invalid), or -9 (Missing)	Not Filled	Filled	Set: <i>ever_worked</i> to -5 <i>Fields A</i> to -5	Appropriate response - enforce routing.
16	1 (Working)	-7 (Invalid), or -9 (Missing)	Not Filled	Not Filled	No Action	Not enough information - E&I to resolve.
17	2 (Not Working)	1 (Last 12 Months)	Filled	Filled	No Action	Appropriate response, no actions needed.
18	2 (Not Working)	1 (Last 12 Months)	Filled	Not Filled	No Action	Respondent should have filled B - No action, allow E&I to resolve.
19	2 (Not Working)	1 (Last 12 Months)	Not Filled	Filled	No Action	Respondent should have filled A - No action, allow E&I to resolve.
20	2 (Not Working)	1 (Last 12 Months)	Not Filled	Not Filled	No Action	Respondent should have filled A and B - No Action, allow E&I to resolve.
21	2 (Not Working)	2 (Not Last 12 Months)	Filled	Filled	No Action	Appropriate response, no actions needed.
22	2 (Not Working)	2 (Not Last 12 Months)	Filled	Not Filled	No Action	Respondent should have filled B - No action, allow E&I to resolve.

23	2 (Not Working)	2 (Not Last 12 Months)	Not Filled	Filled	No Action	Respondent should have filled A - No action, allow E&I to resolve.
24	2 (Not Working)	2 (Not Last 12 Months)	Not Filled	Not Filled	No Action	Respondent should have filled A and B - No Action, allow E&I to resolve.
25	2 (Not Working)	3 (Never Worked)	Filled	Filled	No Action	Respondent has ticked they have never worked, but by filling B there is a strong indication they have done some work previously. In 2011, we would have changed the <i>ever_worked</i> variable to No, but this isn't the same in 2022 so leave for E&I to resolve.
26	2 (Not Working)	3 (Never Worked)	Filled	Not Filled	No Action	Appropriate response - no action needed.
27	2 (Not Working)	3 (Never Worked)	Not Filled	Filled	No Action	In 2011, we changed <i>ever_worked</i> to missing to allow E&I to resolve. This isn't quite the same in 2022, so No Action to allow E&I to resolve.
28	2 (Not Working)	3 (Never Worked)	Not Filled	Not Filled	Set: <i>Fields B</i> to -5 No Code Required	Respondent has largely followed routing, except they should have filled Fields A. Leaving these unchanged will be resolved in E&I, but routing can be enforced here by setting B to NCR. This is similar to what was done in 2011.

29	2 (Not Working)	-7 (Invalid), or -9 (Missing)	Filled	Filled	No Action	In 2011, we changed <i>ever_worked</i> to Yes and allowed E&I to resolve <i>lastyrwrkd</i> . However, this isn't quite the same in 2022, so No Action allows E&I to resolve.
30	2 (Not Working)	-7 (Invalid), or -9 (Missing)	Filled	Not Filled	No Action	E&I to resolve - this is the same as 2011.
31	2 (Not Working)	-7 (Invalid), or -9 (Missing)	Not Filled	Filled	No Action	In 2011, we changed <i>ever_worked</i> to Yes and allowed E&I to resolve <i>lastyrwrkd</i> . However, this isn't quite the same in 2022, so No Action allows E&I to resolve.
32	2 (Not Working)	-7 (Invalid), or -9 (Missing)	Not Filled	Not Filled	No Action	Not enough information - E&I to resolve.
33	-7 (Invalid), or -9 (Missing)	1 (Last 12 Months)	Filled	Filled	Set: <i>emplyd_last_week</i> to 000001 <i>emplyd_last_week_grp2</i> to 2 Not Working	Aligns with routing; similar to 2011.
34	-7 (Invalid), or -9 (Missing)	1 (Last 12 Months)	Filled	Not Filled	Set: <i>emplyd_last_week</i> to 000001 <i>emplyd_last_week_grp2</i> to 2 Not Working	Aligns with routing, although respondent should have filled Fields B - E&I to resolve this.
35	-7 (Invalid), or -9 (Missing)	1 (Last 12 Months)	Not Filled	Filled	No Action	E&I to resolve - this is also what we did in 2011.
36	-7 (Invalid), or -9 (Missing)	1 (Last 12 Months)	Not Filled	Not Filled	No Action	Not enough information - E&I to resolve.

37	-7 (Invalid), or -9 (Missing)	2 (Not Last 12 Months)	Filled	Filled	Set: <i>emptyd_last_week</i> to 000001 <i>emptyd_last_week_grp2</i> to 2 Not Working	As 2011.
38	-7 (Invalid), or -9 (Missing)	2 (Not Last 12 Months)	Filled	Not Filled	Set: <i>emptyd_last_week</i> to 000001 <i>emptyd_last_week_grp2</i> to 2 Not Working	Aligns with routing, although respondent should have filled Fields B - E&I to resolve this.
39	-7 (Invalid), or -9 (Missing)	2 (Not Last 12 Months)	Not Filled	Filled	No Action	E&I to resolve. This is what was done in 2011.
40	-7 (Invalid), or -9 (Missing)	2 (Not Last 12 Months)	Not Filled	Not Filled	No Action	Not enough information - E&I to resolve.
41	-7 (Invalid), or -9 (Missing)	3 (Never Worked)	Filled	Filled	No Action	Contradictory information - respondent indicates they have never worked, but filled in Fields B (which indicates they have worked). E&I to resolve.
42	-7 (Invalid), or -9 (Missing)	3 (Never Worked;)	Filled	Not Filled	Set: <i>emptyd_last_week</i> to 000001 <i>emptyd_last_week_grp2</i> to 2 Not Working and <i>Fields B</i> to -5 No Code Required	Generally aligns with appropriate routing.
43	-7 (Invalid), or -9 (Missing)	3 (Never Worked)	Not Filled	Filled	No Action	Contradictory information - respondent indicates they have never worked, but filled in Fields B (which indicates they have worked). E&I to resolve.

44	-7 (Invalid), or -9 (Missing)	3 (Never Worked)	Not Filled	Not Filled	Set:3 <i>emplyd_last_week</i> to 000001 <i>emplyd_last_week_grp2</i> to 2 Not Working and <i>Fields B</i> to -5 No Code Required	Generally aligns with appropriate routing, though E&I will resolve Fields A.
45	-7 (Invalid), or -9 (Missing)	-7 (Invalid), or -9 (Missing)	Filled	Filled	Set: <i>emplyd_last_week</i> to 000001 <i>emplyd_lastflag_week_grp2</i> to 2 Not Working	Filling Fields B indicates that there is a history of work; however, leaving the <i>ever_worked</i> variable for E&I to resolve.
46	-7 (Invalid), or -9 (Missing)	-7 (Invalid), or -9 (Missing)	Filled	Not Filled	No Action	Not enough information - E&I to resolve.
47	-7 (Invalid), or -9 (Missing)	-7 (Invalid), or -9 (Missing)	Not Filled	Filled	No Action	E&I to resolve.
48	-7 (Invalid), or -9 (Missing)	-7 (Invalid), or -9 (Missing)	Not Filled	Not Filled	No Action	Not enough information - E&I to resolve.

Dependent Fields A	Dependent Fields B
other_act	emp_stat
look_work	employer
avail_work	occupation
wait_work	industry
	supervisor
	hours_worked

Filter Rule L - Work or Study More

The online questionnaire has a series of sub-questions around working and studying which determines which activity the respondent does more if they have indicated they both work and study. This is because the "Travel to Work" address response should apply to the place where the respondent spends the most time. However, there is no such indicator on a paper questionnaire. For the most part, the Filter Rules can't make this determination, except in a couple of straightforward cases, detailed here.

Item	Variable A	Variable B	Variable C	Action (Then...)	Notes
	(If <i>ft_student</i> is...)	(If <i>other_act</i> is...)	<i>emplyd_last_week_grp2</i>		
1	1 (Yes)	2 (Studying)	1 (Working)	No Action	Unable to determine if respondent works or studies more.
2	1 (Yes)	2 (Studying)	2 (Not Working)	Set: <i>work_study</i> to 2 (Study)	
3	1 (Yes)	Anything <i>other</i> than 2 (Studying) ⁸	1 (Working)	No Action	Unable to determine if respondent works or studies more.
4	1 (Yes)	Anything <i>other</i> than 2 (Studying)	2 (Not Working)	No Action	Unable to determine if respondent works or studies more.
5	2 (No)	2 (Studying)	1 (Working)	No Action	Unable to determine if respondent works or studies more.
6	2 (No)	2 (Studying)	2 (Not Working)	No Action	Unable to determine if respondent works or studies more.
7	2 (No)	Anything <i>other</i> than 2 (Studying)	1 (Working)	Set: <i>work_study</i> to 1 (Work)	

⁸ Including Invalid or Missing

8	2 (No)	Anything <i>other</i> than 2 (Studying)	2 (Not Working)	No Action	Not enough information
---	--------	--	-----------------	------------------	------------------------

Filter Rule M - Method of Travel

This rule is essentially the same as 2011, with the exception that travel options for “outside of the UK” are now set to “no code required”.

Item	Variable A	Variable B	Action (Then...)	Notes
	(If <i>work_study_address</i> is...)	(If <i>method_travel</i> is...)		
1	1 Work from Home	Any Value	Set: <i>method_travel</i> to -5 No Code Required	As in 2011
2	2 Equivalent	Any Value	Set: <i>method_travel</i> to -5 No Code Required	This is similar to 2011, but we did not have the <i>work_study_address</i> option (2 Equivalent)
3	3 No Fixed Place	Any Value	No Action	As in 2011
4	4 Offshore	Any Value	No Action	As in 2011
5	5 Another Address	Any Value	No Action	As in 2011
6	6 Outside UK	Any Value	Set: <i>method_travel</i> to -5 No Code Required	This is new for 2022 and will mirror the routing on OCI
7	-7 Invalid, or -9 Missing	Any Value	No Action	As in 2011

Filter Rule N - Travel or Work from Home (Transportation Question Recode Option)

There is one situation where the Filter Rules will create a new code for an (inherent) answer option in order to prevent an inconsistency. The "Travel to Work or Study" question (42) gives options for work or studying from home, however, the "Method of Travel" question (43) does not give a tick option for this. In these circumstances, rather than setting a -5 (no code required), the related filter rule (N) will create the code "11" (Work or Study from home). This was done in 2011 as well for E&I purposes.

Item	Variable A	Variable B	Action (Then...)	Notes
	(If <i>work_study_address</i> is...)	(If <i>method_travel</i> is...)		
1	1 Work from Home	Any Value	Set: <i>method_travel</i> to 11 (Work or Study from Home)	As in 2011
2	2 Equivalent	Any Value	Set: <i>method_travel</i> to 11 (Work or Study from Home)	This is similar to 2011, but we did not have the <i>work_study_address</i> option (2 Equivalent)
3	3 Any other option	Any Value	No Action	As in 2011
4	-7 Invalid, or -9 Missing	Any Value	No Action	As in 2011

6. Strengths and Limitations

The main strength of the filter rules process is that it reduces the burden on the following processing step, Edit and Imputation. It does this by reinforcing the questionnaire routing on the Census questionnaire, by re-coding appropriately skipped responses from “missing” to “no code required”, signalling to Edit and Imputation that a response is not required for that particular question. Additionally, it resolves basic inconsistencies by deterministically applying rules in situations where the response can be reasonably assumed. This in turn relieves Edit and Imputation of some burden, and maximises the donor pool.

A limitation of the filter rules could be that the rules are pre-set before we receive the data, therefore there may be routing issues or inconsistencies that had not been accounted for. To try and mitigate this, an empty filter rule has been set up for both the Household and Individual data set to allow the results to be looked at and a further filter rule created if needs be.

Without the filter rules process, the burden on Edit and Imputation would be increased, meaning that the process could run slower or less effectively. Having a less accurate donor pool means that imputation would not be as accurate.

7. Conclusion

Filter Rules is one of the “data cleansing” steps in census data processing and are applied to Census Household, Household Person and Individual questionnaires.

Filter Rules is most important for responses on paper forms. As a feature of the online questionnaire, respondents are simply brought to the next question, based on the answer(s) given in preceding questions. This prevents the respondent from seeing unnecessary questions, which reduces respondent burden and minimises the ability to provide inconsistent answers. On paper however, a respondent is able to see all of the questions at once, providing the opportunity to create these

inconsistencies. To minimise the potential bias between the two modes of completion, all responses (both online and paper) will be processed through the filter rules.

Filter Rules is required as a precursor to Edit and Imputation. This is to ensure consistency in the application of questionnaire routing and resolve inconsistencies in responses, which in turn relieve some of the burden from Edit and Imputation processing, and maximise the E&I donor pool.

8. References

National Records of Scotland: Overview Of Edit and Imputation for Scotland's Census 2022, (online) available at:

<https://www.scotlandscensus.gov.uk/media/cmepkqpp/scotland-s-census-2022-emaps-pmp012-overview-of-edit-and-imputation-for-2022-pdf-61.pdf>

National Records of Scotland: Statistical methodology papers for Census 2022, (online) available at:

<https://www.scotlandscensus.gov.uk/external-methodology-assurance-panels-emaps-0>

Information on CANCEIS, (online) available at:

<https://www.istat.it/en/methods-and-tools/methods-and-it-tools/process/processing-tools/canceis>