

Scotland's Census 2022

Resolve Multiple Responses – Methodology

January 2021

Contents

1. Plain English Abstract	3
2. Abstract	4
3. Background and Introduction	5
4. The 2011 Resolve Multiple Responses Method	8
4.1 2011 Questionnaire Types	9
4.2 Identifying Records to Resolve in 2011	9
4.3 Resolving Records in 2011	9
4.4 Resolving Records in 2011	10
4.5 Prioritisation in 2011	13
5. 2022 Resolve Multiple Response Method	15
5.1 Background	15
5.2 Overview of RMR for 2022	16
5.3 Changes for 2022 Made or Considered	18
6. Strengths and Limitations	23
7. Conclusion	24
8. References	26
9. Glossary	27

1. Plain English Abstract

Sometimes, people respond to more than one Census questionnaire. This can be for an appropriate reason - for example, when someone elects to fill out an individual questionnaire, they may still be included in the main household of an online response. Such situations are accounted for, and do not pose a problem for data accuracy or quality.

However, there are also times where there is more than one response submitted, which results in duplicate people and/or duplicate households in the census dataset. If left in the dataset, issues with data quality arise and cause problems during processing. This is primarily in the form of what we call *overcount*, which leads to an over-estimation of the number of people in the Census population.

The Resolve Multiple Responses (RMR) step in census data processing works to resolve these issues. The process identifies cases where a duplicate response was submitted, and resolves them into one response, ensuring that people are accounted for but not duplicated. This paper is primarily concerned with the application of rules that are used to resolve these responses together; a separate paper on the identification of duplicates was presented earlier.¹ It details the methodology used in Scotland's 2011 Census, and presents a proposal for the changes that should be made to this methodology to bring it in line with other 2022 Census statistical processing methods.

¹ The Resolve Multiple Responses: Identification Methodology paper can be found here: [PMP014: Resolve multiple responses - identify duplicates | Scotland's Census \(scotlandscensus.gov.uk\)](#)

2. Abstract

People who respond to the Census can do so more than once. In some cases this is by design, but in most this is in error. Multiple person records associated with one person and likewise multiple household records for the same household, are duplicate responses. Duplicate responses contribute to an over-estimate of the population and households.

The Resolve Multiple Responses (RMR) step in Census processing identifies cases of multiple response error associated with duplicate responses within a small area, and resolves those errors by combining a group of duplicate responses into a single response which retains as much data as possible from the group. This paper is concerned with the rules which underpin the resolving of such errors.

This paper provides background on how multiple response error was resolved in Census 2011, and proposed changes to the associated rules in light of changes to Census design for 2022 and the additional options afforded by identifying duplicate person responses at postcode level.

The RMR process is intended to identify and resolve duplicate response within small areas (either address or postcode); duplicate response across different postcodes are dealt with later in Census processing as overcount correction.

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

Sometimes, people respond to more than one Census questionnaire. This can be for an appropriate reason - for example, when someone elects to fill out an individual questionnaire, they may still be included in the main household of an online response. Such situations are accounted for, and do not pose a problem for data accuracy or quality.

However, there are also cases where people respond to the Census more than once, which can result in duplicate people or households (or both) persisting in the census dataset. A number of situations can arise where more than one response is submitted, such as:

1. Where more than one member of a household completes a questionnaire on the assumption that they are the only one to do so;
2. Where a respondent begins filling in the census using one method (paper or online), but changes their mind and decides to fill it in using another. This most often occurs when the respondent runs into issues online, and switches

to paper. In such cases, the information on the online return would be collected as an un-submitted return²;

3. A respondent begins filling in the census return online, but forgets their login details before completing it. They are required to request a new Internet Access Code (IAC), and begin a new return³. The information on the first would be collected as an un-submitted return;
4. When someone changes their mind about what they want to include in their response after it has been submitted, and sends a new one;
5. Where a person gets confused about how to answer a paper response, and provides answers to the individual questions for themselves more than once in a questionnaire
6. Where one household is enumerated more than once (for example, if an error in the address frame leads to the same household being enumerated more than once).

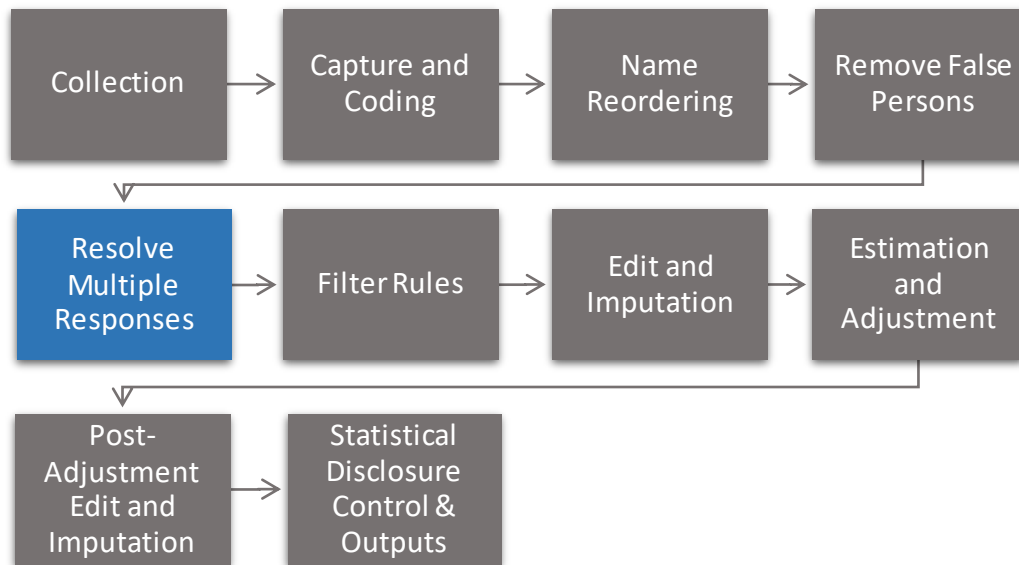
This type of unintended duplication causes quality issues – the most notable referred to as an overcount. Overcount causes the Census population to be falsely inflated. Later statistical processing can estimate and account for small amounts of an overcount, but the Dual System Estimation methodology used in the Estimation and Adjustment procedure is better at compensating for undercount⁴, and so most duplicates will need to be identified and resolved before reaching this point.

² An unsubmitted return is an online questionnaire which is started but left unfinished by the respondent, and collected at the end of the Census collection period to ensure that all responses are accounted for. Please see glossary for definition.

³ For data security reasons, individuals cannot be given access to partially completed census returns over the phone, hence the need to start a new return.

⁴ For more information on Dual System Estimation and how it is used, please see: [PMP001: Estimation and adjustment methodology | Scotland's Census \(scotlandscensus.gov.uk\)](https://www.scotlandscensus.gov.uk/PMP001:Estimation-and-adjustment-methodology)

Figure 1: Simplified Overview of the 2022 Census Data Journey



Resolve Multiple Responses (RMR) is the process by which responses are reviewed in order to identify duplicate person records and duplicate household or communal establishment records, and resolves them by combining duplicate responses in a way which retains information where possible (rather than discarding it). This occurs for duplicates found in the same location - it does not deal with issues of placement, where duplicates exist at different locations (for example, parents who live separately both recording their child as living with them). Resolving such duplicates by combining them would require making a judgement on which location is the 'correct' one. Such overcount issues are left for subsequent Estimation processes.

This paper covers the general methodology used in *resolving* duplicates in the Census dataset during the Resolve Multiple Response stage. The methodology which is used in order to identify (locate, or link together) potential duplicates is new for Scotland's Census 2022, and is detailed in a separately published paper⁵.

⁵ [PMP014: Resolve multiple responses - identify duplicates | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk/pmp014-resolve-multiple-responses-identify-duplicates)

4. The 2011 Resolve Multiple Responses Method

In 2011, the Resolve Multiple Response step:

- a. Identified census records which are potentially from the same communal establishment, household or person (through a matching algorithm written in SAS). Households and communal establishment records were matched on address; person records within those addresses were matched on a combination of name, date of birth and sex (accounting for cases where some of those were missing).
- b. Selected a priority record based on a set of criteria, which becomes the base record to be carried into further processing.
- c. Resolves the identified duplicates into one response by merging the records together, using the priority record as a base. If there were any missing or invalid variables on the priority record, RMR took values from the non-priority record and wrote them to the priority record, forming a singular, 'new' response. Where more than one household or communal establishment records are merged, the associated person records are associated to the priority household or communal establishment.

4.1 2011 Questionnaire Types

In 2011, household questionnaires could be completed either by sending in paper questionnaires, or completing a questionnaire online. Census in 2022 will also be available for completion both online and on paper. However, while the majority of people will have completed the Census on paper in 2011, 2022 will aim to have the majority completed online.

Generally, there are two parts to a Census response - the household section, where there are questions asked about household characteristics, and the person section, which asks about the individuals living within the household. The whole questionnaire is sometimes simply referred to as a household questionnaire (with household and person sections).

Communal establishments, or CEs, are a type of managed accommodation. The questionnaire for these were similar in structure, although the person questions are asked on a questionnaire given to individuals, while the questions about the establishment are asked to the CE manager on a separate form. All communal establishment questionnaires in 2011 were paper.

4.2 Identifying Records to Resolve in 2011

In 2011, person records within an address were matched on a combination of name, date of birth and sex (accounting for cases where some of those were missing). All person matches within that address were flagged to be resolved as duplicates. In addition to this, if a person match was identified between two or more households at an address, those households were also flagged to be resolved as duplicates.

4.3 Resolving Records in 2011

RMR 'resolved' identified duplicates by merging them together. Within a group of duplicate records, a priority record was chosen to be retained, and any missing or invalid variables on the priority record were taken from the non-priority records, and merged onto the priority. Note that valid data on the record chosen as priority was

not overwritten with information from non-priority records – only missing or invalid values. Here 'invalid' can be taken to mean a value that is out of scope (e.g. an impossible date of birth) or could not be coded. At this stage the valid/invalid distinction only applies to variables in isolation; validity across variables or questions is dealt with in a later processing stage, Filter Rules. The religion question was the only voluntary (i.e. non-response is considered a valid response) question in 2011, but was treated the same as other questions. If a group of person records were resolved, and any contained a response to the religion question, then that response would overwrite the non-response of any other record.

4.4 Resolving Records in 2011

Once records were matched and a priority record determined, RMR proceeded to “resolve” the duplicate records (i.e. merge them together). The record chosen as priority was used as a base, and any missing or invalid variables on the priority record is taken from the non-priority records, and merged onto the priority. Note that valid data on the record chosen as priority was not overwritten with information from non-priority records – only missing or invalid values.

Merging of response variables occurs regardless of if the question “needed” to be answered or not. If the variable is not needed (i.e. if the respondent would have been routed past the question), this was fixed in the later Filter Rules process⁶. Subsequently, this means that the variable may have been changed back to a missing or ‘no code required’ type response. Similarly, any inconsistencies created were identified and resolved in Edit & Imputation⁷.

Resolving person records

When resolving Person Records, the relationship variables were not merged or overwritten, and instead all instances were changed to missing. Relationship

⁶ More information on Filter Rules can be found here: [Data cleansing | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk)

⁷ For details on Edit and Imputation, please refer to the published methodology here: [PMP012: Overview of edit and imputation for Scotland's Census 2022 | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk)

variables on the priority record were kept as-is. This was done to avoid breaking the appropriate relationship links within a household, and the 'blanks' were later imputed in later statistical processes.

Resolving households

In general, it was not assumed that multiple household responses at one address were in fact the same household or should be combined. It was consistent with the definition of a household that two or more may share an address. For example, consider a family who live in a house and an unrelated lodger in an outbuilding on the same property – they share an address but may correctly consider themselves to be two households and provide two household returns. Combining them into one household would contradict that, and potentially under-count the number of distinct households.

Where they were resolved, this was for one of two reasons. Either the presence of a duplicate person record across households indicated that these households are a case of multiple response, or another characteristic of the household response suggested it represented part of a household.

- Where multiple household records at the same address had a person match across households, these households were merged.
- Where multiple household records at the same address did not have a person match across households, but one or more households indicated via the 'number of usual residents' question that they had 6 or more usual residents, and the total number of person records at this address was less than the maximum number of usual residents for any one household, these households were merged together.
- Empty household records (i.e. with no associated person records) at the same address were merged together and then to any non-empty household record at that address.

Where household records were merged, person records which appeared on the priority household, had their 'position' in the original household (in the form of a

household person number) retained, and subsequent person records from the non-priority records were then added to the household. Any relationship information on records that were on the non-priority household were set to missing. Any visitors on a non-priority household were moved to the priority household.

Resolving Communal Establishments

Where multiple communal establishment records shared an address, these were resolved by picking a priority communal establishment record and moving all people associated with the non-priority communal establishment record.

In cases where there are several paper household returns, RMR also checked to ensure that this is not because the 'household' is actually a communal establishment. In 2011, a list of known communal establishments was produced for this purpose. This list allowed Household matching to check cases where there were a number of duplicate paper Household forms (3 or more) against the known addresses and postcodes.

Where this occurred, RMR created a communal establishment record and moved all people on the household forms to the communal establishment. Corresponding household records were subsequently removed from the dataset.

Resolving Non-response Returns

A non-response return is a return associated with non-response; it is generated by the Census field force where they believe a property to exist but there is no associated Census return. Non-response returns record the characteristics of the property and whether it is likely to be occupied or unoccupied. These returns are used in later processing. RMR resolved duplication within non-response returns (i.e. more than one return at the same address) and duplication between non-response returns and household or communal establishment returns (where a household or communal establishment return was preferred).

Flags and Tidying Up after RMR

All actions to merge or discard actions within RMR were recorded. Merged records were recorded in one table and discarded records in another. Additionally, records were assigned flags to indicate their status as the result of an RMR resolution or (in the case of person records) their having been moved to a priority household. This was done to ensure that any changes made to the data in RMR were properly recorded. Additionally, person records were renumbered within the household or communal establishment to be sequential, and relationship variables updated in line with this, as not to disrupt later processes.

Identifying duplicate person records within a postcode

An additional step was developed to extend person matching to the level of postcode, and so identify additional cases of duplicate person or household records at distinct but similar addresses. Such matches indicate duplication (and so overcount) that would not otherwise have been found. However, in practice these matches were not resolved due to time constraints in processing the Census data.

4.5 Prioritisation in 2011

Once either household, communal establishment or person records were flagged to be resolved, one of the duplicates was selected to be the priority record. This creates the basis of a record to carry forward to further processing (while non-priority records were not). Additionally, for household records, a second priority rule is required to determine how person records associated with those household records are ordered within the resulting combined household.

In 2011, selection of a priority household record was done as follows, where if more than one record meets a condition then records are checked against the next condition:

- Prioritise household records with associated people over those without;
- Then within this, prioritise completed online records over pre-addressed paper records over unaddressed paper records and un-submitted online records;

- Then within this, records with a higher number of valid fields (i.e. for a subset of variables, records will vary in terms of how complete those variables are – more complete records are preferred over less complete records);
- Otherwise select a priority record at random.

The second priority rule, which determines in which order associated person records are moved to the combined household, superseded other priority rules and was used to ensure that households with adults associated (i.e. persons aged 16 or older) were taken before those without adults associated.

Communal establishment priority records were selected in the same way, with the exception that there were no online communal establishment records.

Person records were prioritised as follows:

- Prioritise records associated with individual returns over all other records;
- Then within this, prioritise records with a higher number of valid fields as before (although the subset of variables is specific to person records).

In 2011, Individual Forms were always taken as the priority when resolving Person Records in the Individuals step of RMR. Where a person provides a separate individual questionnaire, the associated record is deemed to provide information which is more accurate (where it differs from the same person's household record) and best respects the wishes of that person to be counted in the way of their choosing. However, missing or invalid values on the individual record were overwritten where possible with values from the non-priority records.

The highest number of valid fields check is a test of 'completeness' undertaken on a subset of questionnaire variables. RMR looks at these variables, and counts the number of fields that are "valid", assigning a score to each record (all variables have equal weighting). The record with the highest score would then be assigned as the priority record. Again, here 'invalid' can be taken to mean a value that is out of scope (e.g. an impossible date of birth) or could not be coded. At this stage the valid/invalid

distinction only applies to variables in isolation; this means that for any paper returns, we don't distinguish between acceptable non-response (e.g. a person aged under 16 who is routed away from providing a response to employment status) and 'real' non-response which will be imputed. For this reason, the choice of variables to check against for completeness tended to avoid variables for which high non-response is expected.

5. 2022 Resolve Multiple Response Method

5.1 Background

We anticipate that the sort of multiple response error which the Resolve Multiple Response process dealt with in 2011 will persist in 2022. Respondents may still submit multiple questionnaires, leading to overcount of people and households, and these should again be resolved where possible to allow for an accurate population estimate to be made. Some changes to respondent behaviour are expected, compared with 2011. This is based on changes to Census design and observations from the 2019 rehearsal.

Online First

When first contacted, household respondents will 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 responses in 2022 than in 2011 (supported by the 2019 rehearsal), and fewer paper responses. Online collection allows for some types of error to be identified before the respondent submits their return – for example, if the same name is listed twice in one return, the householder will be prompted to provide an additional name (e.g. a nickname) to distinguish them. This will require the householder to acknowledge that the same name appears twice, and so make accidentally listing the same person multiple times more difficult than on paper. If online response does prove much more popular in 2022 then we can expect fewer

cases of duplicate persons within households. However, increased uptake of online response will likely mean an increase in other types of multiple response to resolve, whether duplication between online and paper, or multiple online responses (e.g. if the respondent partially completes an online return and then forgets their password).

Validation of Online Responses:

Returns provided online are subject to additional validation which could not be applied to paper. Respondents who provide their date of birth will see their calculated age displayed for confirmation; unlikely question responses will generate validation messages, and in some cases invalid question responses will not be accepted. Paper returns are not subject to this validation, and are subject to additional sources of error (e.g. capture error). It is expected that, given an online and a paper return relating to the same person, if any question response does not match between the two then the online response is more accurate.

Individual Returns:

The addition of more sensitive voluntary questions for Census 2022 is expected to encourage an increase in the use of Individual returns (where a respondent provides a separate return in addition to that provided by their householder). Likewise, online respondents can request an Individual return and in doing so provide minimal information on their household return. In both cases we expect more Individual returns than in 2011, and so expect to resolve more cases of multiple response of this type.

However, the principles of RMR for 2022 are the same as in 2011. Duplicate person records are identified and resolved by identifying a priority record and retaining information from other records where necessary. Duplicate household and communal establishment records are identified and resolved in a way which retains the associated residents.

5.2 Overview of RMR for 2022

1. Identify Duplicate Person Records

We identify duplicate person records upfront by linking within enumeration postcode. Linked person records are grouped into clusters. These clusters are linked again to administrative data, or clerically reviewed, or both to determine whether they come from the same person or not. This provides a list of person duplicates to resolve, and by identifying duplicates across questionnaires, we can see which households have a person in common.

2. Identify Household and Communal Establishment Records to Combine

Where there are multiple household or communal establishment returns associated with the same address, we identify them and in some cases combine them according to rules consistent with those applied in 2011.

- a. Unoccupied households are grouped with occupied households at that address where possible;
- b. Household responses which indicate they represent only part of a household are grouped with other households at that address where possible;
- c. Households with one or more people in common (i.e. a person is duplicated across households) are grouped;
- d. Communal establishments at the same address are grouped

Note that households are not grouped into communal establishments; for 2022, whether a property is considered a household or part of a communal establishment is determined at enumeration and not changed. For example, small hotels and B&Bs (as determined by the number of bed spaces) will be enumerated as households and not converted to communal establishments.

3. Resolve Household and Communal Establishment Records Within an Address

For those cases in (2) which are to be resolved, a priority record is identified based on rules set out below. The priority record and its valid question responses

are retained; missing or invalid question responses take a value from the next highest priority record where possible. Any person records associated with a non-priority record are moved to the priority record, and the non-priority record discarded.

4. Resolve Additional Household Records

Where we see a duplicate person across multiple household addresses within a postcode, this is taken as evidence of an error in the address frame. These households are resolved as in (3).

5. Resolve Duplicate Person Records

Duplicate person records identified in (1) are resolved in a similar way. The priority record and its valid question responses are retained; missing or invalid question responses take a value from the next highest priority record where possible. Non-priority records are discarded.

6. Resolve duplicate non-response returns so they can be used in later processes

7. Before the data is passed on for later processing, affected records (whether they have been resolved or affected by another resolution) are flagged. Person records are renumbered within their household or communal establishment, and relationships between resolved person records are recovered.

5.3 Changes for 2022 Made or Considered

Multiple households at an address

For 2022, respondents who consider their household to be distinct from another at the same address can only obtain an additional Internet Access Code or paper questionnaire by requesting them from the Contact Centre. There they will be asked to provide their address, and a distinct 'address' assigned to the additional

household so that Census operations and processing can distinguish between them. However, this does not rule out the possibility that an address will be missed, and so it is still possible that more than one distinct household may exist at an address (e.g. in the outbuilding example above). For this reason we again do not propose to combine household returns together by default, but only where there is evidence they relate to the same underlying household. Again the primary piece of evidence is the existence of a person in common across households.

Identifying Person Duplicates and the Sequence of Events

In 2011, RMR identified household and communal establishment records at the same address, and only then matched person records to determine whether those household and communal establishment records had at least one person in common. Persons in common were identified by matching on name (either an exact match or a match on Soundex of name), sex, and date of birth.

Where a person was duplicated across addresses, they were not resolved but retained as distinct person records. Depending on how similar these addresses are, the duplicate person might indicate that there is an error with the address frame (that the same 'true' address has been enumerated twice).

For 2022, using the proposed person matching method⁸ allows us to identify person duplicates within a postcode, regardless of which address they are associated with. This means person linking can be done upfront on all available person records, and the resulting matches used to identify duplicate persons. It also identifies cases where the same person appears associated with households at different addresses within the same postcode. Where the same person appears at more than one address within the same postcode, we take this as evidence that these multiple 'addresses' are actually one address, recorded twice on the address frame. We

⁸ Referring to the Resolve Multiple Responses Identification Methodology, which can be found here: [PMP014: Resolve multiple responses - identify duplicates | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk/PMP014: Resolve multiple responses - identify duplicates | Scotland's Census)

resolve these cases in the same way as other duplicate household records, retaining the address associated with the priority household.

Prioritisation

For 2022, where a group of household or communal establishment records are to be resolved, we prioritise as follows:

- Prioritise by mode of collection and return status (Online Submitted over Paper over Online Un-submitted);
- Then by number of valid fields filled, analogous to that done in 2011 although the subset of variables is tailored to the 2022 questionnaire;
- Then at random

When resolving person records, we prioritise first by questionnaire type (where individual questionnaires take priority over any other type), and then as with household or communal establishment returns.

The intention here is to retain as much accurate data as possible. Online returns are subject to more validation at collection than paper and are not subject to scanning errors, so are more accurate. Online un-submitted returns are considered an exception; they aren't subject to the same validation as a submitted return, and the presence of an un-submitted return won't stop non-response follow-up. In general un-submitted returns will duplicate complete submitted returns, and so will only be retained where follow-up did not result in a submitted return.

Consideration was given to the idea that as paper questionnaires require additional effort for the respondent to obtain and complete, the associated return could be expected to be more accurate or complete than an online submitted return. However, this does not account for errors introduced in capture or scanning. Favouring online submitted returns allows for these errors to be avoided.

By prioritising again by number of valid fields filled we favour more complete over less complete records. This means maintaining more variables associated with the priority record than we otherwise would, and so creating fewer inconsistencies by overwriting missing or invalid values.

Multi-part questions

We resolve person duplicates by using fields from lower-priority records to overwrite missing or invalid fields in the priority record. In general, this allows us to retain the most data, and any inconsistencies created between questions will be resolved in Edit & Imputation as they would be for non-duplicates. However, where one question response is spread over multiple variables, inconsistency is more of a problem and harder to account for later on.

If we were to allow overwriting of a missing or invalid response for part of a multi-part question, we could create an inconsistency between parts of that question, or a response which is then not valid. For example, the 'date of birth' question populates variables for day, month, year of birth as well as age on Census day. By combining two partial dates of birth which may not match exactly we could create a new date of birth combination which is less accurate in data quality terms than either of the contributing partial dates. This is a problem for later processes that use date of birth (for example, any other linking or matching process – the presence of dates of birth which don't match is evidence against a match whereas a missing or partial date of birth is not).

Instead we treat specific multi-part questions as one variable; we only overwrite a response if completely missing or invalid, and then with a complete response from a lower priority record. This avoids creating combined responses which are of less value than the original partial response.

The multi-part questions treated in this way are date of birth (made up of day, month, year of birth and age) and year of arrival in the UK (made up of month and year of arrival).

Voluntary questions

Most Census questions and associated variables are mandatory – we require a valid answer and will go on to impute any missing or invalid values to produce a complete dataset. However, some particularly sensitive questions and associated variables are voluntary – respondents can choose not to provide an answer, and if they do not, this is considered a valid response and will not be imputed over.

When resolving duplicate persons, we treat non-response for these questions as valid and do not overwrite a missing or invalid value here. This is done to ensure that the response associated with an individual return is given the appropriate priority and not overwritten – even where that response is valid non-response. This is in contrast to 2011, where the only voluntary question was treated the same as other questions. As a consequence of this change, individual person returns take priority over other person returns. This means that if any voluntary question is left missing in the individual response, this missing value is treated as valid and retained rather than being overwritten. As an example, consider a case of multiple response where a respondent's householder provides a response to the religion question and no response to sexual orientation. If that respondent requests an individual questionnaire so that they can provide a response to the sexual orientation question, then the response to religion on that individual questionnaire will be retained as well, even if it is non-response.

This is an unavoidable consequence of prioritising individual response over other response – we cannot distinguish between intentional and unintentional non-response for voluntary questions, and would otherwise risk overwriting legitimate data provided by a respondent, with inaccurate data provided by their householder. However, we do note that incomplete individual questionnaires may result in

unintentional non-response being retained; this will be considered in the quality assurance checks associated with the RMR process.

6. Strengths and Limitations

Identifying and resolving duplicates across addresses

An advantage of Resolve Multiple Responses for 2022 is the identification and resolution of duplicate person and household records at different addresses in the same postcode, where there is evidence these addresses correspond with one 'true' address. This enables us to resolve the associated duplicate people and reduces known overcount. The scale of this potential overcount depends on how frequent these duplicate addresses are, but research from 2011 Census and scaling up results from the 2019 rehearsal suggest a figure of around 2,000 additional duplicate persons who would otherwise be retained.

A disadvantage of resolving such cases in RMR by combining records is that, where combining e.g. two households at different addresses, the resulting combined household can only take one address. For most subsequent processing purposes it's sufficient to know the postcode of a household. However, when Census records are linked to the Census Coverage Survey for the purpose of producing a population estimate, the exact address does matter. If the combined Census household takes an address different to that provided on the CCS, the associated person matches will need to be clerically reviewed.

However, the number of such household resolutions across Scotland is expected to be small, and the number which also appear in the CCS smaller still. The benefit in terms of reducing overcount across Scotland should outweigh the small number of additional cases requiring clerical review.

Combining communal establishments across addresses in a similar way is not proposed – given their nature as managed residences this sort of multiple response

should not occur, and if it did, could not be resolved in the same way without effectively discarding an enumerated communal establishment.

Prioritisation

Prioritising responses from individual questionnaires over other questionnaires is essential to ensure that data on sensitive questions is captured and used appropriately, as the household questionnaire response is more likely to have been answered by or visible to another household member.

Prioritising online submitted returns over paper returns, and paper returns over online un-submitted returns, allows resolved records to retain the most accurate information available.

Multi-part questions

Treating multi-part questions together results in resolved records which appear less complete than if the variables had been treated separately, but the resulting values are more likely to be accurate and useful in later processing (e.g. in matching).

Voluntary questions

By treating non-response in voluntary questions as valid, and so not overwriting with responses from lower-priority duplicates, we better respect respondents' wishes not to respond to these questions. The disadvantage is that, since we cannot distinguish between intentional and unintentional non-response, we may discard a valid response in favour of an unintentional non-response. This can be mitigated by monitoring the incidence of very incomplete individual responses, and if it is common, treating non-response as missing as in 2011 RMR.

7. Conclusion

The Resolve Multiple Responses process should proceed for Census 2022 using the same principles from Census 2011 –

- That duplicate person records are identified and combined in a way which retains quality data where possible;
- That duplicate households are identified on the basis of having a person in common, and are also combined in a way which retains quality data where possible;
- That residual cases of duplicate households associated with specific errors are combined even if there is no person in common.

But run on an iterative basis as data is received, rather than all at once, to support the fine-tuning of later processes and allow more time for clerical review of matches.

Replacing the more simple 2011 person matching step with the more sophisticated 2022 method allows for more duplicate person records to be identified and for those matches to be more robust (when quality assured against administrative data).

Additionally, duplicate persons identified across addresses should be treated as evidence of those addresses relating to the same 'true' address, and associated household records resolved accordingly.

8. References

National Records of Scotland: *Resolve Multiple Responses Identification Methodology*, (online) available at:

[PMP014: Resolve multiple responses - identify duplicates | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk/pmp014)

National Records of Scotland: *Estimation and Adjustment Methodology*, (online) available at:

[PMP001: Estimation and adjustment methodology | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk/pmp001)

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

[PMP012: Overview of edit and imputation for Scotland's Census 2022 | Scotland's Census \(scotlandscensus.gov.uk\)](https://scotlandscensus.gov.uk/pmp012)

9. Glossary

Term	Definition
Household	<ul style="list-style-type: none"> • One person living alone, or • A group of people (not necessarily related) living at the same address who share cooking facilities and share a living room, sitting room or dining area. <p>A household may also be:</p> <ul style="list-style-type: none"> • a person or a group of people living in sheltered housing or very sheltered housing (irrespective of whether there are other communal facilities), • a person or a group of people living in a temporary or mobile structure (for example a caravan, mobile home or boat) on any type of site that is their usual place of residence.
Communal Establishment	A communal establishment is typically managed residential accommodation where there is full-time or part-time supervision of the accommodation. For example - care homes, hospitals, holiday/leisure accommodation
Clerical Review	A process where an individual statistician manually recalls and reviews the record in question in order to make decisions on how to proceed with the record (i.e., remove it, merge it, move to next process, etc). This generally happens with records which are ambiguous in some respect - for example, there is text written in the name field, but may not actually reflect a person and rather information instead.
Link	Two records that have been connected
Match	Two records that represent the same individual

<p>Online submitted return</p>	<p>An online submitted return is an online return where the respondent has completed the online collection process by submitting their questionnaire responses to National Records of Scotland (NRS) and has been allocated a submission ID by the Data Collections Operational Management System (DCOMS).</p>
<p>Online un-submitted return</p>	<p>An online un-submitted return is an online return where the respondent has not completed the online collection process by submitting their questionnaire responses to National Records of Scotland (NRS). There is no submission ID attached to this response.</p>
<p>Non-response return</p>	<p>A non-response return is a rules-generated return based on information provided by field force during follow-up fieldwork. Non response returns will be created for every census enumeration address that has not returned a census questionnaire. In addition to the unique address identifier it will contain the perceived reason for non-response along with field force responses to key questions relating to the address.</p>
<p>Non-response follow up</p>	<p>Non-response follow up (NRFU) (sometimes referred to simply as 'follow-up') will have two forms: letter and visit. NRFU will be flexible and will depend on a number of factors, including but not limited to: area characteristics, immediate area return rate, local authority return rate and estimation area return rate. The type of NRFU (letter or visit) will also be flexible and depend on a number of factors, including but not limited to: area characteristics, previous NRFU activity and time since census day.</p>
<p>Household Individual Questionnaire</p>	<p>A questionnaire for an individual in a household offered on request.</p>