The ethnicity of the deceased person: the apparent quality of the data that are collected when deaths are registered

Introduction

With effect from 2012, when a death is registered in Scotland, information about the ethnicity of the deceased person may be collected, on a voluntary basis, for use in research to improve health and healthcare services. As there was uncertainty about the reliability of the resulting data, National Records of Scotland (NRS) has assessed the apparent quality of the information for 2012 to 2014, mainly by comparing it with any ethnicity information for that person from the 2011 Census. (This paper uses the term ‘apparent quality’ because, for example, there may be errors in the census data, or in the linkage between the Death Registration and census records.)

The overall percentage agreement between the Death Registration and census information is high, but this is due to a particularly high percentage for ‘White - Scottish’: almost every other ethnicity has a much lower percentage agreement.

The key conclusions are that:

- the data on the ethnicity of the deceased person are not (at present) suitable for calculating reliable mortality rates for most ethnicities; and
- NRS may be able to improve the quality of future data.

Contents

1. Summary

1.1 When a death is registered in Scotland, information about the ethnicity of the deceased person is sought, for use in research to improve health and healthcare services. The provision of ethnicity data is voluntary (refer to paragraph 2.1). The first results showed an unexpectedly low Black, Asian and Minority Ethnic (BAME) percentage, and that the BAME numbers would double if, say, a sixth of the 'ethnicity not known' deaths were of BAME people (refer to paragraph 2.2).

1.2 NRS has assessed the apparent quality of the Death Registration ethnicity data by comparing them with what was recorded for the same people in the 2011 Census (refer to paragraph 2.6). NRS believes that, where there are differences, the information that was collected in the census will, in general, be more reliable, because it would usually have been provided by the person him/herself (refer to paragraph 2.3), whereas the Death Registration ethnicity is provided by someone who may not even be a close relative or friend. Data from the census and Death Registration records were linked using a probabilistic approach (refer to Annex A), and NRS assessed the quality of the data linkage in order to identify which pairs of records were 'high quality' matches.

1.3 The structure of the ethnicity classification used in the 2011 Census and the Death Registration system is shown in Table 1. There are six ethnic groups (for example 'White', 'African …'), most of which are sub-divided into a number of ethnicities (for example the 'White' ethnic group includes six ethnicities: 'White – Scottish', 'White – Other British', and so on). For the deaths which were registered from 2012 to 2014, one can calculate (from the numbers given in Table 1) that no ethnicity was provided in 4.1 % of cases, 95.4 % were counted as 'White', and 0.6 % were counted as 'BAME'; 84.3 % of all the deaths were counted as 'White – Scottish'.

1.4 At first sight, there appears to be good agreement between the information about ethnicity that was given when a death was registered and that which was reported in the 2011 Census:

- the ethnicity (for example 'White - Scottish' or 'African, African Scottish or African British') was the same in 93.8 % of the cases for which a person's ethnicity was available from both sources (refer to paragraph 3.3);
- the ethnic group (for example 'White' or 'African') was the same in 99.7 % of the cases for which a person's ethnic group was available from both sources (refer to paragraph 3.4);
- the 'White'/'BAME' classification was the same in 99.7 % of the cases for which the person's ethnicity was available from both sources (refer to paragraph 3.5).

1.5 Slightly higher percentages may be obtained when the comparison is limited to those cases for which it was thought that it should be the most reliable (for example 'high' quality matches, and the person who registered the death was the spouse/partner, child, sibling or parent of the deceased):

- the ethnicity was the same in 94.7 % of the cases for which a person's ethnicity was available from both sources (refer to paragraph 4.2);
- the ethnic group was the same in 99.8 % of the cases for which a person's ethnic group was available from both sources (refer to paragraph 4.3);
- the 'White'/'BAME' classification was the same in 99.8 % of the cases for which the person's ethnicity was available from both sources (refer to paragraph 4.3).

1.6 However, more detailed analysis (of the cases for which it was thought that the comparison should be the most reliable) reveals that:

- the overall 'agreement' percentages are so high because the 'agreement' percentages are very high for 'White - Scottish' (refer to paragraph 5.2);
- the 'agreement' percentages are much lower for some (but not all) of the other ethnicities – for example:
  - only 64 % of those with 'White - Other British' in the Death Registration record had that ethnicity in the census (refer to paragraph 5.3);
  - only 78 % of those with 'Indian, Indian Scottish or Indian British' in the Death Registration record had that ethnicity in the census (refer to paragraph 5.8).
- so the ethnicity recorded when a death is registered was reported differently to the census (and so may be wrong) in many of the cases for which the ethnicity was said to be something other than 'White - Scottish' (refer to paragraph 5.9)
- many of the differences may be due to the Death Registration ethnicity being implicitly based on the deceased's country of birth. NRS could try to reduce such errors by (a) providing further guidance to registrars on this point and (b) changing the order of the questions, so that the registrar asks about the ethnicity of the deceased well before asking about the country of birth (refer to paragraph 5.10).

1.7 Based on the comparison with the information from the census, most of the ethnicities (other than 'White - Scottish') are over-represented among those for whom no ethnicity information was provided when the death was registered (refer to paragraph 6.3).

1.8 The quality of the ethnicity information provided when a death is registered has varied:

- over time, as:
  - the percentage of deaths with no ethnicity information at all has increased from 3.4 % in 2012 to 5.3 % in 2015 (refer to paragraph 7.1), and
  - the percentage agreement between the Death Registration and the census ethnicity information tended to be slightly lower in 2014 than in the previous two years (refer to paragraph 7.5)
- with the age of the deceased, as the percentage of deaths for which ethnicity information is provided is lowest for child deaths, and rises with age (refer to paragraph 8.2).

NRS could consider, with registrars, how to encourage greater provision of this information, but people would not be pressed for it, given the circumstances (refer to paragraph 8.4).

1.9 Based on the additional information about the deceased's ethnicity that was provided in respect of those who were said (when the death was registered) to have an 'other ….' ethnicity, some White people were clearly mis-classified as 'Other White …' (refer to paragraph 9.2). NRS can check such cases and – where appropriate - correct the ethnicity code that is used for the production of its statistics when it finalises its statistical data for each year (refer to paragraph 9.3).

1.10 Conclusions

- the high overall percentage agreement between the Death Registration and census information is due to a particularly high percentage for 'White - Scottish';
- most of the other ethnicities have a much lower percentage agreement;
- therefore, the data on the ethnicity of the deceased person are not (at present) suitable for calculating reliable mortality rates for most ethnicities (Section 10 explains in more detail why this is so); and
- NRS could try to improve the quality of the data by:
  - providing further guidance;
  - changing the order of the questions, so that the registrar asks about the deceased's ethnicity well before asking about the country of birth;
  - considering how to encourage people to provide the information; and
  - correcting 'other …' ethnicity codes that are clearly wrong, on the basis of the text description of the ethnicity of the deceased.

2. Background

2.1 In Scotland, information about the ethnicity of the deceased person has been sought, for all deaths registered since the start of 2012, for use in research into the links between ethnic group and causes of death to help improve health and healthcare services. The procedure that should be followed is described on the How the information should be collected page of the NRS website. The key points to note are as follows:

- while the registration of deaths is compulsory, the provision of ethnicity information is completely voluntary;
- the registrar should ask whether the 'informant' (the relative or other person who goes to the registration office to register the death) is willing to provide the Registrar General with information about the deceased person's ethnic group, and describe briefly the purposes for which the information might be used;
- if the informant consents to provide the information, the registrar should ask him/her to pick the most appropriate ethnic group from a list of categories used in 2011 Census outputs. In cases where the informant picks an 'other' group (for example 'Other White', 'Other Asian'), he/she should be asked to describe it in more detail, and the registrar should record the response (for example if the informant says 'German', the registrar should key that in);

- otherwise, the registrar should record either 'not willing to provide' or 'ethnic group not known' (whichever is appropriate);
- the information is held in the registration computer system, and may be used to produce statistics. However, it does not appear in the so-called 'death certificate' (this is the Extract of the entry in the Register of Deaths).

2.2 At a conference in Edinburgh in November 2012, NRS's Vital Events Statistician presented a summary of the results of the first six months' collection of the data. The full presentation can be found on The statistics that have been published page of the NRS website. This showed that:

- 27,920 deaths were registered from January to June 2012;
- an ethnic group was recorded in 26,910 cases (96.4 %);
- the ethnic group was a 'White' one in 26,750 cases (95.8 %) and a 'BAME' one in 160 cases (0.6 %);
- the informant was not willing to provide the information in 867 cases (3.1 %); and
- the informant was unable to provide the information, because he/she did not know the deceased's ethnic group, in 143 cases (0.5 %).

The Statistician commented that the BAME percentage appeared unexpectedly low (in comparison to, for example, a BAME figure of 3.2 % of the total population, from the 2011 Scottish Household Survey's results). He wondered whether a 'not willing to provide' or 'not known' response could be more likely in the case of the death of a BAME person. Any estimates of death rates would have to be used carefully, because the 'BAME' numbers would double if, say, a sixth of the 'not willing / unable to provide' deaths were actually deaths of BAME people. The Statistician added that, in the longer-term, NRS might try to compare individuals' ethnic groups in the census and Death Registration records, to refer to (for example) if the same code was given in both cases, and what information the census had in the case of the deaths for which no ethnic group had been provided. (However, NRS had no way to assess the quality of the data until it did the work which is described in this paper.)

2.3 NRS believes that, where there are differences, the information about a person's ethnicity that was collected in the census will, in general, be more reliable than that provided when the person's death was registered. One would expect that the census information would usually have been provided by the person him/herself, and so be his/her own view of his/her ethnicity (rather than someone else's view). Also, NRS has assessed the quality of the census data. Summary information on the accuracy and quality of the data collected in the 2011 Census in Scotland was provided in Section 9 of NRS's General Report on the Scotland's Census website. That reports a high level of completeness in the responses provided to the ethnic group question in the returned census questionnaires, with missing or invalid responses contributing to an item non-response rate for that question of just 2.1 % (Table 9.4 in that report). It also records a high level of consistency between the responses to the ethnic group question obtained in the Census Quality Survey and the corresponding responses obtained in the census itself, reflected in a gross agreement rate of 96.8 % (Table 9.5

in that report; also, this paper's Annex C provides information about the Census Quality Survey, its main findings regarding the ethnicity question and how they compare to the main differences between the ethnicity recorded in the census and in the Death Registration records). However, it should be noted that a person's sense of ethnic identity may not fit well into the census question's categories: Simpson et al (JRSSA vol. 179 part 4, October 2016) found that the recorded ethnicity changed between 2001 and 2011 for 4.0 % of cases in a linked sample of those Censuses' data for England and Wales. They wrote "… it is reasonable to conclude that about a third to a half of all instability [by which they mean such changes] was due to people for whom more than one response [to the census question] is suitable, and that this amounts to about 2 % of the whole population …".

2.4 In June 2013, NRS was informed of occasions on which registrars had not followed the correct procedure. Therefore, in July 2013, NRS's Head of Registration Policy wrote to all Chief/Senior Registrars, seeking information about what the local practice had been and asking them to remind their staff to follow the specified process. The responses to NRS's questions are summarised on The extent to which the procedures were followed in 2012 page of the NRS website.

2.5 A month or so later, in August 2013, NRS published a summary of all the information that was collected in 2012, in its Vital Events Reference Table 5.4 (which is available on the NRS website). The numbers were roughly double those of the first six months, and showed a similar pattern in terms of the percentages of deaths for which an ethnic group was recorded, for which 'White' and 'BAME' ethnicities were recorded, and for which the informant was not willing, or unable, to provide the information. Given the uncertainty about the reliability of the figures (mentioned in paragraph 2.2), NRS put them on its website (along with many other statistics for 2012 that it published at the same time) but did not publicise their availability because it could not assess their quality (that could only be done when NRS did the work described in this paper).

2.6 The 2011 Census breakdown of the population of Scotland by ethnic group, age and sex was published in February 2014. Soon afterwards, in April 2014, the Director of Public Health Science at NHS Scotland, and two Professors of Public Health, asked NRS to match the ethnic information that was provided when deaths were registered in 2012 with what was recorded in the 2011 Census for the same people, in order to assess the accuracy of the information that was recorded at Death Registration (and also to carry out similar work in respect of later years' deaths, in order to obtain an on-going measure of the quality of the data). This became a 'pilot' project for a new data linkage service, so took much longer than should now be the case, because of the need to rigorously establish various sets of governance, administrative arrangements and technical procedures for the linking of the different sets of data records. It was May 2016 before NRS Vital Events could finally bring together the Death Registration records for 2012, 2013 and 2014 and the ethnicity information from the 2011 Census records (if any) for those people. Annex A describes how the links

were made, and explains that this approach ensures that those involved in a data linkage project have access only to the data which they need for their part in the project (this is for the data which are not 'theirs', they see only the personal identifiers or the additional variables which are of interest for their analysis - not both together).

2.7 It was originally intended to:

- assess the apparent quality of such information for 2012 to 2014, mainly by comparing it with the information (if any) that was obtained from the same person in the census;
- look at the extent to which the two sets of information agree, and whether this varies with (for example) the ethnicity that was reported in the census;
- set out how reliable (or not) the data on the ethnicity of the deceased person appear to be, for Scotland as a whole and for some sub-groups of the population; and
- emphasise points which should be kept in mind when using the statistics that have been published in the annual editions of Vital Events Reference Table 5.4, for example whether there appeared to be significant under-recording of certain ethnic groups and, if so, the implications for ethnic group-specific mortality rates calculated from the published figures.

However, the relatively low percentages agreeing for most ethnicities mean that the data are not at present suitable for calculating reliable mortality rates for most ethnicities (refer to Section 10). Therefore, there was little point in analysing in detail the figures for some sub-groups of the population, or suggesting any points to be kept in mind when using the statistics for particular sub-groups.

2.8 In the meantime, NRS had published summaries of the information that had been collected for 2013 and 2014 in the versions of Vital Events Reference Table 5.4 for those years. NRS subsequently published the corresponding figures for 2015 in the same way in August 2016. All the versions of Table 5.4 are available in the Vital Event Reference Tables section of the NRS website.

3. Overall comparison with the ethnicity information from the 2011 Census: for all the cases for which this can be done

3.1 In total, 163,876 deaths were registered in Scotland in the three years 2012, 2013 and 2014 taken together.

3.2 Table 1 shows their numbers, cross-classified by the ethnicity (if any) that was given when the death was registered (in the columns of the table) and the ethnicity that was reported in the census, in those cases for which a matching Census record was found (in the rows of the table). To make this table (and those that follow) easier to read, values of 0, 1, 2 and 3 have been replaced by a single dot. It seems likely that most such 'odd' cases are errors, due to (for example) the wrong records being matched, or a keying error in the underlying data. 'Hiding' such values should make it

easier to see the true patterns in the data. Values of 0, 1, 2 and 3 are all included in the row and column totals, and in the figures that are given in this report, so there may be some apparent slight discrepancies between (for example) the total that is shown for a particular column and the sum of the figures that appear in that column.

3.3 The first row of the table shows that the 163,876 deaths comprised 6,644 for which the deceased's ethnicity was not provided when the death was registered, 138,131 for which it was said to be 'White - Scottish', 14,668 for which it was said to be 'White – Other British', and so on. The first column of the table shows that there were 17,609 deaths for which the person's ethnicity was not available from a census record (usually because there was no matching census record), 132,376 for which the person was described as 'White – Scottish' in the census, 10,091 for which the census had 'White – Other British', and so on. The figures on the diagonal show the numbers of cases for which the ethnicity given when the death was registered was the same as the ethnicity reported in the census. For example, there were 122,602 people who were described as 'White – Scottish' in both the census and the Death Registration record, 7,527 people described as 'White – Other British' in both records, and so on. The 'off-diagonal' numbers show that, for example, 1,362 people who were reported to be 'White – Other British' in the census were counted as 'White – Scottish' in the Death Registration records, and 4,635 people who were recorded as 'White – Scottish' in the census were described as 'White – Other British' when their deaths were registered. The numbers for the other categories are generally much smaller. The figures from this table may be summarised as follows:

- 163,876 deaths registered in the three years;
  of which
- 23,070 cases for which:
  - ethnicity was not provided when the death was registered (6,644 cases); or
  - ethnicity was not available from the census (17,609 cases), usually because no matching census record could be found.

From their total, 1,183 is subtracted, to avoid double-counting cases for which ethnicity was not available from both sources (as they are included in both the 6,644 cases and the 17,609 cases). The result, 23,070, is the total of all the numbers which are in the cells (which are highlighted in light grey) on the 'not provided' row or in the 'not provided' column (including all the values of 0, 1, 2 and 3 which have been replaced by dots in the table);
leaving

- 140,806 cases for which ethnicity information is available from both the Death Registration record and from the census (this is the rest of the table);

which consist of:

- 132,057 cases for which the ethnicity given when the death was registered was the same as the ethnicity reported in the census – this is the total of the numbers in the diagonal part of the rest of the table, in the cells which have a border and are white in colour (including values of 0, 1, 2 and 3 which have been replaced by dots);

and

- 8,749 cases for which there were different ethnicities in the Death Registration and census records – this is the total of the numbers in the remainder of the table, in the cells which are highlighted in dark grey (again, including values which have been replaced by dots).

On this basis, the ethnicity given when a death was registered was the same as the person's ethnicity reported in the census in 93.8 % of the cases for which his/her ethnicity was available from both sources (132,057 out of 140,806).

3.4 In some cases, there is what might be regarded, for some purposes, as only a 'slight' difference between the ethnicity recorded when a death was registered and how the person was counted in the census. For example, one source of data might have 'White – Scottish' as the ethnicity, and the other might have 'White – Other British'. (There are, of course, times when that would definitely not be a 'slight' difference. For example, when comparing 'White Scottish' and 'White Other British' mortality rates, a lot of misclassification of 'White – Scottish' as 'White – Other British', or vice versa, could make unreliable any conclusions from analyses of the Death Registration data.) Therefore, NRS used the data which underlie Table 1 to produce a cross-tabulation by ethnic group, for example with all the 'White' ethnicities combined into a single 'White' ethnic group (Note: NRS will not publish that table because it could be used to deduce some of the values that have been replaced by '.'s in Table 1). Doing this increases the percentage for which the Death Registration and census data agree because, for example, when a person is counted as 'White – Scottish' in one and as 'White – Other British' in the other, both sources of data have the person in the 'White' ethnic group. Its main results are as follows:

- 140,806 cases for which an ethnic group is available from both the Death Registration record and the census record;

of which

- 140,398 cases for which the ethnic group is the same in both records;

and

- 408 cases for which the two records have different ethnic groups.

On this basis, the ethnic group given when a death was registered was the same as the person's ethnic group in the census in 99.7 % of the cases for which his/her ethnic group was available from both sources (140,398 out of 140,806).

3.5 The figures which were used to produce Table 1 can be combined further to produce a cross-tabulation using a simple 'White' / 'BAME' split (which, again, will not be published). Its main results are as follows:

- 140,806 cases for which whether the person was counted as 'White' or as 'BAME' is available from both the Death Registration record and the Census record;

of which

- 140,426 cases for which the person was counted the same way in both records;

and

- 380 cases for which the person was counted as ‘White’ in one record and as ‘BAME’ in the other.

On this basis, 99.7 % of the cases for which the person’s ethnicity was available from both sources (140,426 out of 140,806) had the same ‘White’/’BAME’ classification from the Death Registration record and from the census.

4. Overall comparison with the ethnicity information from the 2011 Census: for those cases for which it was thought that the comparison should be the most reliable

4.1 The comparisons in the previous section used all the cases for which a match was found between a Death Registration record and a census record, and then concentrated on those for which the person’s ethnicity was available from both sources. However, while one would usually be confident that a pair of matching records relate to the same person, there are cases where this is not so certain. Therefore, it may be better to limit the comparison to those cases for which the quality of the matching was ‘high’ (refer to Annex A). In addition, in a small proportion of cases, the ethnicity in the census had to be imputed (for example because the relevant question had not been answered – information about how the census data were processed is available in the Census methodology section of the Scotland’s Census website), and therefore might not be correct – so it is best to exclude such cases from the comparison. One might also expect that, in general, information about the deceased’s ethnicity should be more reliable when the informant was the deceased’s spouse or partner, child, sibling or parent, than when the informant was a more distant relative (for example a niece or a grandchild) or someone who is not a relation (for example a friend, a solicitor, a funeral director or a care home worker). So the comparison is limited further, by considering only cases where the informant was the deceased’s spouse/partner or a child, sibling or parent. (Towards the end of the project, the data were examined to see whether and, if so, how the apparent quality of the information varied with the relationship between the informant and the deceased: Annex B provides more information about this.)

4.2 Restricting the numbers in these ways produces Table 2, which shows that there were 100,760 deaths (registered in the three years) for which the quality of matching was ‘high’, the census provided information about ethnicity which was not imputed, and the informant was the spouse/partner or a child, sibling or parent. These are the cases for which one might expect the least disagreement between the ethnicities recorded in the census and when the death was registered. In 3,294 of these cases, the deceased’s ethnicity was not provided when the death was registered, leaving 97,466 cases for which ethnicity information was available from both records. These comprised

- 92,303 cases for which the ethnicity given when the death was registered was the same as the ethnicity reported in the census (this is the diagonal part of the rest of the table);

and

- 5,163 cases for which different ethnicity codes were allocated in the Death Registration and census records.

On this basis, the ethnicity given when a death was registered was the same as the person's ethnicity reported in the census in 94.7 % of cases (92,303 out of 97,466). This is slightly higher than the 93.8 % obtained using all the matching records, presumably because (for example) some of the 'low' and 'medium' quality matches involved records which did not relate to the same person, and some incorrect ethnicity information was imputed or provided by someone who was not a close relative.

4.3 The figures which were used to produce Table 2 can be combined to produce cross-classifications by ethnic group and by 'White'/'BAME' (which NRS will not publish – refer to paragraph 3.4). The main points from these tables are as follows:

- 97,466 cases for which ethnicity information was available from both records;
- 97,243 or 99.8 % had the same ethnic group in both records;
- 97,263 or 99.8 % had the same 'White'/'BAME' classification in both records.

5. Comparisons with the information for certain ethnicities from the 2011 Census: for the types of case for which it was thought that the comparison should be the most reliable

5.1 The percentages given above suggest very good agreement between the census and the Death Registration data. However, that is not the case when one considers the figures for some of the ethnicity categories. The analysis in this section is based on the 97,466 cases for which the quality of matching was 'high', the census provided information which was not imputed, and the informant was the spouse/partner or a child, sibling or parent who provided information about the ethnicity of the deceased, as these are the cases for which it was thought that the ethnicity of the deceased should be the most reliable.

5.2 Table 3 summarises the main figures from Table 2. It shows only the ethnicities and ethnic groups for which the 97,466 cases included at least 20 from the census (for the purpose of the first three columns of the table) and at least 20 from the Death Registration data (for the purpose of the final three columns). It seems that the overall 'agreement' percentages are so high because the 'agreement' percentages are very high for the 'White - Scottish' ethnicity, and that is the ethnicity of a very large proportion of the population of Scotland (and of the deaths registered in Scotland). Table 3 shows that the 97,466 cases included:

- 88,856 with 'White – Scottish' in the census record;
- of which 85,538 (96.3 %) had 'White – Scottish' in the matching Death Registration record;

and

- 86,712 with 'White – Scottish' in the Death Registration record;
- of which 85,538 (98.6 %) had 'White – Scottish' in the matching Census record.

Both the ‘agreement’ percentages for ‘White – Scottish’ are clearly above the overall 94.7 % ‘agreement’ (which was calculated from the data for all 97,466 cases), so some ethnicities’ ‘agreement’ percentages must be well below 94.7 %.

5.3 The ‘White – Other British’ ethnicity is an example of this. Table 3 shows that the 97,466 cases included:

- 6,549 with ‘White – Other British’ in the census record;
- of which only 5,591 (85 %) had ‘White – Other British’ in the matching Death Registration record;
- with Table 2 revealing that this was mainly because 660 (10 %) had ‘White – Scottish’ in the matching death record, and a further 251 (4 %) had ‘Other White ethnic group’ in the matching death record;

and

- 8,728 with ‘White – Other British’ in the Death Registration record;
- of which only 5,591 or 64 % had ‘White – Other British’ in the matching census record;
- with Table 2 revealing that this was mainly because 2,859 (33 %) had ‘White – Scottish’ in the matching census record;

5.4 Some of the difference between the ethnicities recorded in the Death Registration and census records may be ‘explained’ by the person’s country of birth, which is also recorded when a death is registered. While a person’s ethnicity is not necessarily determined by where he or she was born, an informant who does not know the deceased’s own view of his/her ethnicity could well say that the deceased’s ethnicity was whatever seemed most appropriate on the basis of where the deceased was born (or where the informant, perhaps incorrectly, thought that the deceased was born). So, for example, in the absence of any knowledge of the deceased’s own view of his/her ethnicity, someone registering the death of a White person who was born in England might well describe the deceased as ‘White – Other British’.

5.5 The largest number of differences between the ethnicities in the census and the Death Registration records was the 2,859 cases where the census had ‘White – Scottish’ and the Death Registration record had ‘White – Other British’. More detailed analysis of NRS’s data for these cases (for which there is no table in this report) shows that half (1,417, or 50 %) of those concerned had been born in England, so it is understandable why they were described as ‘White – Other British’ by informants who (presumably) did not know that the deceased people had (presumably) ticked the ‘White – Scottish’ box in the census (perhaps because the deceased people had been born in England to Scottish parents, but their families had returned to Scotland a few years later, and so the deceased people had lived in Scotland for many years). However, most of the rest of the 2,859 (1,140, or 40 %) had been born in Scotland, so it is not obvious why they should have been described as ‘White – Other British’ when their deaths were registered.

5.6 The second largest number of differences was the 660 cases where the census had 'White – Other British' and the Death Registration record had 'White – Scottish'. More detailed analysis of NRS's data (for which there is no table) shows that over three-fifths (408, or 62 %) of those concerned had been born in Scotland, so it is understandable why they were described as 'White – Scottish' by informants who (presumably) did not know that they had (presumably) ticked the 'White – Other British' box in the census. However, most of the rest of the 660 (206, or 31 %) had been born in England, so it is not obvious why they should have been described as 'White – Scottish' when their deaths were registered.

5.7 The BAME ethnicities with the largest numbers in the three years' data were 'Pakistani, Pakistani Scottish or Pakistani British' and 'Indian, Indian Scottish or Indian British'. Table 3 shows that the 97,466 cases included:

- 153 with 'Pakistani, Pakistani Scottish or Pakistani British' in the census record;
- of which only 104 (68 %) had 'Pakistani, Pakistani Scottish or Pakistani British' in the matching Death Registration record.
- with Table 2 revealing that the remaining 49 mainly consisted of 30 (20 % of the 153) who had 'White Scottish' in the matching Death record, and a further 11 (7 %) who had 'Indian, Indian Scottish or Indian British' in the matching Death record – which fits in (apart from the description of 'White') with more detailed analysis of NRS's data finding that 32 of the 49 had been born in Scotland and 12 had been born in India. So a reason for the discrepancy seems to be that, in some cases, the Death Registration ethnicity has been based on the country of birth. It should be noted that some of the people whose country of birth was said to be India may have been born in places that were part of India at the time of their birth, but are now part of Pakistan or Bangladesh.

and

- 108 with 'Pakistani, Pakistani Scottish or Pakistani British' in the Death Registration record;
- of which 104 or 96 % had 'Pakistani, Pakistani Scottish or Pakistani British' in the matching census record, so there is no need for any further analyses of these cases.

5.8 In the case of the 'Indian, Indian Scottish or Indian British' ethnicity, Table 3 shows that the 97,466 cases included:

- 87 with 'Indian, Indian Scottish or Indian British' in the census record;
- of which only 63 (72 %) had 'Indian, Indian Scottish or Indian British' in the matching Death Registration record.
- with Table 2 revealing that the remaining 24 consisted mainly of 13 (15 % of the 87) who had 'White Scottish' in the matching Death record, and a further six (7 %) who had 'White - Other British' in the matching Death record – which kind of fits in (apart from the description of 'White') with more detailed analysis of NRS's data finding that 11 of them had been born in Scotland and three had been born in England and Wales. Again, a reason for the discrepancy seems to be that, in some cases, the Death Registration ethnicity has been based on the country of birth.

and

- 81 with ‘Indian, Indian Scottish or Indian British’ in the Death Registration record;
- of which only 63 or 78 % had ‘Indian, Indian Scottish or Indian British’ in the matching census record;
- with Table 2 revealing that the remaining 18 mainly consisted of 11 (14 % of the 81) who had ‘Pakistani, Pakistani Scottish or Pakistani British’ in the matching census record. More detailed analysis of NRS’s data found that all 18 such cases were born in India. Yet again, the main reason for the discrepancy seems to be that, in some cases, the Death Registration ethnicity has been based on the country of birth. Again, some of the people whose country of birth was said to be India may have been born in places that were part of India at the time of their birth, but are now parts of Pakistan or Bangladesh.

5.9 It is clear that the ‘agreement’ percentages for the ‘White – Other British’, ‘Pakistani …’ and ‘Indian …’ ethnicities are generally well below the ‘overall’ 94.7 %. Indeed, Table 3 shows that that is the case for almost all the other ethnicities for which the numbers were large enough to warrant inclusion in the table. Where there are differences, one would expect that the information that was collected in the census would normally be the more reliable (for reasons which were given in paragraph 2.3). So it appears that the ethnicity recorded when a death is registered may be wrong in a large proportion of the cases for which the ethnicity was said to be something other than ‘White – Scottish’.

5.10 The analyses described in paragraphs 5.5 to 5.8 (and of some other NRS data) suggest that, in many of the cases where the ethnicity reported when a death was registered differs from the ethnicity shown in a person’s census record, the Death Registration ethnicity was (wrongly, presumably) based on the deceased’s country of birth. NRS could try to reduce the number of such errors in future by:

(a) providing further guidance to registrars on this point – for example, by including a statement like ‘the ethnic group of the deceased person should describe the national or cultural tradition(s) or background(s) he/she most strongly identified with, and may have no connection to where he/she was born’ in the material that is provided for the informant and the registrar;

(b) changing the order of the questions, so that the registrar asks about the ethnicity of the deceased well before asking about the country of birth, as this should reduce the likelihood of people thinking that the deceased’s ethnic group should be based on his/her place of birth.

6. Ethnicity information from the 2011 Census: for those cases for which the deceased's ethnicity was not provided when the death was registered

6.1 Table 1 shows that 163,876 deaths were registered in Scotland in the three years 2012, 2013 and 2014, taken together, and that there were 17,609 deaths for which no information could be obtained from a census record for the person (in most cases, this would be because it was not possible to find a census record that matched the Death Registration record). It follows that ethnicity information from the census was available for 146,267 of the people whose deaths were registered in the three years.

6.2 Table 4 shows those 146,267 cases cross-classified by (in the rows) the ethnicity information that was obtained from the census and (in the columns) whether or not ethnicity information was provided when the death was registered. The second figure in the first row shows that no ethnicity information was obtained when 5,461 of those deaths were registered. The second column provides a breakdown of those deaths by the ethnicity that was recorded in the census; the third column provides a similar breakdown for the other 140,806 deaths for which ethnicity information was obtained from the census. The next three columns show the percentage that each ethnicity, as recorded in the census, represents of the total number in each of the first three columns.

6.3 It is clear that the breakdown by ethnicity (as recorded in the census) differs between those for whom no ethnicity was provided when a death was registered, and those for whom the ethnicity was provided when the death was registered. Just 78 % of the cases for which no ethnicity was provided when the death were registered had 'White – Scottish' as the ethnicity from the census, compared with 91 % of the cases for which ethnicity was provided when the death was registered. Because a lower proportion of the cases for which no ethnicity was provided when the death was registered have 'White – Scottish' as the ethnicity from the census, higher proportions of such cases have other ethnicities from the census. For example:

- 'White – Other British' accounts for 13.8 % of those for whom ethnicity was not provided when the death was registered, compared with 6.6 % of those for whom ethnicity was provided when the death was registered;
- for 'White – Irish', the corresponding figures are 2.3 % and 1.1 %; for 'Other White ethnic group', they are 2.7 % and 0.5 %;
- for 'Pakistani …', the corresponding figures are 0.8 % and 0.2 %; for 'Indian ….' 0.5 % and 0.1 %; for 'Chinese …' 0.4 % and 0.1 %.

It follows that most of the ethnicities (other than 'White – Scottish') are over-represented among those for whom no ethnicity information was provided when the death was registered.

6.4 The effect that this has on the overall statistics will be less than that of the other problems. Because ethnicity is not provided for only (very roughly) 1-in-20 of all deaths, the breakdown by ethnicity (recorded in the census) of the 146,267 deaths for which ethnicity is available from the census is very similar to that of the 140,806 of those deaths for which ethnicity was also provided when the death was registered. For example:

- ‘White – Scottish’ accounts for 90.5 % of the 146,267 compared with 91.0 % of the 140,806;
- for ‘White – Other British’ the corresponding percentages are 6.9 % and 6.6 %;
- for ‘Pakistani …’ the corresponding figures (when expressed to two decimal places) are 0.23 % and 0.21 %, for ‘Indian …’ they are 0.12 % and 0.10 %, and for ‘Chinese …’ they are 0.08 % and 0.07 %;

The final two columns in Table 4 show the scale of the increase in the number of deaths for each ethnicity (as recorded in the census) when account is taken of those for whom ethnicity was not provided when the death was registered. Including the ‘not provided’ cases increases the overall number by 3.9 %, the ‘White – Other British’ figure by 8.1 %, the ‘Pakistani …’ and ‘Indian ….’ numbers by 14 % and 17 % respectively, and the ‘Chinese …’ figure by 23 %. However, the scale of these effects tends to be less than that of the problems which were identified in Section 5 (for example, paragraph 5.3 stated that only 64 % of cases with ‘White – Other British’ in the Death Registration record had been counted as ‘White – Other British’ in the matching census record).

7. Any change over time in the apparent quality of the ethnicity information provided when a death is registered

7.1 The figures which NRS has published in each year’s version of Vital Events Reference Table 5.4 (available on the NRS website), and which are summarised in Table 5, show that the proportion of deaths for which ethnicity information is provided has been falling fairly steadily, from 96.6 % in 2012 to 94.7 % in 2015. It follows that there has been a slight decline in this aspect of the quality of the ethnicity information, because the percentage of deaths with no ethnicity information at all has increased from 3.4 % in 2012 to 5.3 % in 2015 (it appears that, in most such cases, people were not willing to provide the information: relatively few said that they did not know the deceased’s ethnic group).

7.2 The extent of agreement between the census and the Death Registration data is also an indicator of the quality of the latter. The chart shows how the percentage agreement has changed, month by month, over the three years covered by the linked data for deaths registered from 2012 to 2014. As in Section 5, the figures that it shows are based on the 97,466 cases for which the quality of matching was ‘high’, the census provided information which was not imputed, and the informant was the deceased’s spouse/partner or a child, sibling or parent who provided information about the ethnicity of the deceased, as these are the cases for which the ethnicity of the deceased should be the most reliable.

7.3 The grey dashed line towards the foot of the chart shows the percentage of such cases for which the ethnicity recorded in the Death Registration record was the same as the ethnicity recorded in the census. This fluctuates between slightly over 93 % and a little less than 96 %. The thick black line is a moving average of the monthly values, which has been calculated from the figures for five month periods and is shown against the

middle month of each such period. The moving average 'smoothes out' much of the month-to-month fluctuation in the values, and shows what may be thought of as the underlying level of this aspect of the quality of the data. From this, it appears that percentage of cases for which the ethnicity was the same was initially around 95.5 %, declined to roughly 94.75 % in autumn 2012, rose and was about 95 % in the first half of 2013, then declined to roughly 94 % towards the end of 2014. The fall from about 95 % in mid-2013 to roughly 94.5 % at the end of 2013 started at around the time (July 2013) that NRS's Head of Registration wrote to all Chief/Senior Registrars (refer to paragraph 2.4).

7.4 The lines near the top of the chart show that there has been relatively little month-to-month change in the percentages of such cases which have the same ethnic group (the thin black line) or the same White/BAME category (black dashed line): over the period, they have both remained within the ranges 99.6 % to 100.0 %. In consequence, the grey hollow line, which shows the moving average of the percentage with the same ethnic group, is almost horizontal: when expressed to one decimal place, it has had only two values (99.7 % and 99.8 %).

7.5 In conclusion, the percentage agreement between the Death Registration record and the 2011 Census record ethnicity information varied over the period from 2012 to 2014, and tended to be slightly lower in 2014 than in most of the previous two years.

8. Variation with the age of the deceased in the apparent quality of the ethnicity information provided when a death is registered

8.1 The figures which NRS has published in each year's version of Vital Events Reference Table 5.4 (available on the NRS website) can be used to see how the proportion of deaths for which ethnicity information is provided varies with the age of the deceased. The top part of Table 6 contains the relevant numbers (which were copied from the versions of Table 5.4 for each of the four years for which it has been published), and the final few lines give the percentages of deaths in each year, and for all four years combined. Because the underlying numbers are smaller, the percentages for the younger age-groups are shown without decimal places. As was noted in paragraph 7.1, the overall proportion of deaths (of all ages) for which ethnicity information is provided has fallen fairly steadily, from 96.6 % in 2012 to 94.7 % in 2015.

8.2 The final row of the table shows that the percentage of deaths for which ethnicity information is provided is lowest for child deaths, and rises with age. Taking the four years together, ethnicity information was provided for only 61 % of deaths of 0-4 year olds, 68 % of 5-9 year olds and 77 % of 10-14 year olds. It is 88 % for 15-19 year olds, then rises slowly and steadily (by one percentage point per age-group) to 95 % for the 50-54 age-group and the two that follow, and 96 % (when rounded) for each age group from 65-69 onwards.

8.3 Detailed year by year analysis of the figures for the younger age-groups is not recommended because of the relatively small numbers involved, which could lead to large percentage year-to-year fluctuations. The figures for each of the older age-groups generally show broadly the same pattern as for the 'all ages' percentages: a slight fall each year from 2012 to 2015 (the only exception being the 70-74 age-group, for which a slight rise between 2014 and 2015 is assumed to be a 'random' fluctuation).

8.4 NRS could consider, with registrars, how to encourage more informants to provide their view of the ethnicity of the deceased. For example, the explanatory material that should be shown to informants could say more about the potential uses of the information, and give examples of the kinds of benefits to public health that might eventually result from the analysis of such data. However, registrars could not try to press informants who have said that they are unwilling to provide such information, lest they cause distress. And, of course, it would not be possible to seek the information in the minority of cases where it is not provided because the informant says that he/she does not know the ethnicity of the deceased.

8.5 It is assumed that the percentage for which ethnicity information is provided is lowest for child deaths because there may be particularly distressing circumstances, and that it would not be sensible to try to urge informants to provide their view of a child's ethnicity in such cases. Therefore, it is very unlikely that NRS could collect ethnicity information for all child deaths. However, that may not matter: Child Death Reviews (more information can be found on the Scottish Government website) may be able to obtain more comprehensive information about child deaths than NRS can, and the proposed National Resource Centre (which should have a data analyst) could produce any statistics that might be required.

9. Apparent quality of 'other' ethnicity information: consistency of text description of ethnicity of deceased and ethnicity codes used in Death Registration and 2011 Census records

9.1 In cases where the informant picks an 'other' ethnicity (for example 'Other White', 'Other Asian'), the registrar should ask the informant to describe it in more detail, and record the response. Table 7 shows the main results, by listing all the descriptions which were given in the case of at least 10 'other' ethnicity deaths for which the informant was the deceased's spouse/partner, child, sibling or parent. (Because all the information is taken from the Death Registration record, no data linkage was required for this part of the analysis and all the relevant records can be considered: there is no need to restrict the selection on the basis of the quality of matching.) Apart from converting all the text to upper case (so that, say, 'British', 'british' and 'BRITISH' will be counted together), it has not been edited in any way, because to do so could have taken a long time and would have been of little benefit. The purpose of this table is just to show the main descriptions that have been recorded and the extent to which they are consistent with the ethnicity that was provided. For example, 'WHITE ENGLISH' does not include cases where 'WHITE, ENGLISH' (with a

comma) was recorded; ‘GERMAN’ and ‘GERMANY’ are counted separately, and ‘AMERICAN’ does not include a case where ‘AMERICAN CITIZEN’ was entered.

9.2 Comparing the descriptions and the ethnicity codes, it is clear that they are consistent in many cases – for example, it is credible (although not certain to be accurate) that 65 of the 67 Americans are counted in ‘Other White ethnic group’; and likewise for all 19 Australians, 72 of the 73 Canadians and all 11 Danes. However, assuming that the descriptions are correct, some White people were clearly mis-classified as ‘Other White ethnic group’, including:

- 114 described as ‘British’ - any who were Scottish should have been counted as ‘White – Scottish’, and those who were English, Welsh or Northern Irish should have been counted as ‘White – Other British’;
- 160 described as ‘English’, who should have been counted in ‘White – Other British’;
- 40 described as ‘Shetland’, who should have been counted in ‘White – Scottish’;
- 34 described as ‘Welsh’, who should have been counted in ‘White – Other British’;
- 69 described as ‘White British’ - any who were Scottish should have been counted in ‘White – Scottish’, and those who were English, Welsh or Northern Irish should have been counted in ‘White – Other British’;
- 19 described as ‘White English’, who should have been counted in ‘White – Other British’;

9.3 In theory, such errors could be prevented by providing more training and guidance for registrars. However, given the number of registrars across Scotland, and the relatively small number of such cases that each will (on average) be likely to deal with, it would probably be better if NRS were to check such cases and – where it seems appropriate - correct the ethnicity code that is used for the production of its statistics. This would probably be best done when NRS finalises its statistical data for each year, so that all such cases can be dealt with together (as that should be more efficient than trying to correct them as and when they are received by NRS). It should be noted that NRS will not correct any of the data for the years for which it has already ‘frozen’ its statistical database.

9.4 Tables 8 and 8b provide information about the consistency (or not) of the ‘other’ ethnic group information from the Death Registration record and the ethnicity that was recorded in the census. As with some of the earlier tables, the comparison is restricted to the cases for which it was thought that it should be the most reliable. Table 8 shows that the census had ‘Other White ethnic group’ for only 256 of the 856 people who were described as ‘Other White ethnic group’ when their deaths were registered, with 320 of the 856 having described themselves as ‘White – Scottish’ in the census, and 251 having said that they were ‘White – Other British’.

9.5 Table 8b compares the description of the deceased’s ethnicity and the ethnicity recorded in the census, in those cases where a particular form of

text description was used for at least 10 such deaths. It will be seen that there were many who were counted as 'White - Scottish' in the census, but who were not described as such when their deaths were registered, including 28 who were said to be American, 44 said to be British, 36 said to be Canadian, 23 said to be German and 23 said to be Italian. There were also people counted as 'White - Other British' in the census but described as Canadian, German or Italian when their deaths were registered. Such inconsistencies could be due to misunderstandings on the part of informants, who may have assumed that the deceased's ethnicity was determined by his/her country of birth. These could be cases where it was not realised that someone who had been born elsewhere (perhaps to parents from Scotland) considered him/herself to be Scottish (perhaps having lived elsewhere for only the first few years of life). Of course, there might also be cases where the census information was wrong (for example due to the person ticking the wrong box on the census form).

9.6 Finally, Table 9 provides the same kind of information as Table 7, but without any restriction on the type of person who registered the death. Again, it shows only descriptions which were given in respect of at least 10 'other' ethnicity deaths. The numbers are a little larger than in Table 7, because there is no restriction based on the type of informant, and a few more descriptions appear (for example 'Austrian', which was not shown in Table 7 because it was given for only 9 of the cases that were counted in that table). The same kinds of points as were made in paragraph 9.2 apply here, but based on slightly larger numbers.

10. Are the data (at present) suitable for calculating reliable ethnic group-specific mortality rates?

10.1 As was explained in Section 5, the high overall percentage agreement between the Death Registration and census information is due to a particularly high percentage for 'White - Scottish': most of the other ethnicities have much lower agreement percentages. It follows that the data on the ethnicity of the deceased person are not (at present) suitable for calculating reliable mortality rates for most ethnicities. This can be seen from some examples, which use the figures for deaths registered in 2012, 2013 and 2014 (which are given in Table 1) to produce estimates of the number of deaths of people of a particular ethnicity (mainly) from the Death Registration data, and compare them with (better) estimates produced (mainly) using the census information about the person's ethnicity. The greater the difference between the two estimates, the less reliable is the estimate which is based on the Death Registration data.

10.2 The example that will be described in the next few paragraphs uses the figures for 'White - Scottish', since that is the ethnicity with the largest numbers. Basing the estimate mainly on the information from the Death Registration record, there were:

- 138,131 cases where the deceased was said to be 'White – Scottish' plus

- 4,271 cases where the ethnicity of the deceased was not provided when the death was registered, but one can assume that the person was 'White – Scottish' because that is the ethnicity that was recorded on the census record that was found to match the Death Registration record;

Adding those two figures together gives an estimate of 142,402 deaths of people who might be counted as 'White – Scottish'.

Two points should be noted here:

- anyone who wanted to produce a 'White – Scottish' estimate for (say) a single calendar year or for sub-groups of the population (for example age-groups) would not have the relevant figures cross-classified as in Table 1, and so could not produce any estimates in the way that has just been described. Instead, one would have to take the relevant 'White – Scottish' Death Registration figures from Vital Events Reference Table 5.4, and then increase them slightly to take account of the fact that the ethnicity was not provided for some Death Registration records. For the overall total number of 'White – Scottish' deaths, the increase should be about three % (based on the information which is given in Section 6 and Table 4). However, a different percentage increase might well be appropriate for each year or each sub-group of the population, but one would not know what those figures should be – refer to paragraph 10.9.
- this estimate does not include a share of the 1,183 deaths (registered in the three years) for which no ethnicity was provided and no matching census record was found. However, that should not matter much, because those 1,183 deaths were only 0.7 % of all the 163,876 deaths registered in the three years.

10.3 As mentioned in paragraph 2.3, the ethnicity recorded in the census should be more reliable than that provided when the death was registered. It follows that a better estimate of the number of deaths of 'White – Scottish' people would be mainly based on the census data for the people whose deaths were registered in the three years. This is produced as follows:

- 132,376 deaths of people for whom the ethnicity recorded in the matching census record was 'White – Scottish';

plus

- 13,309 deaths of people for whom no matching Census record was found, but who were described as 'White – Scottish' when the deaths were registered

Adding those two figures together gives a better estimate of 145,685 deaths of people who might be counted as 'White – Scottish' (again, this does not include a share of the 1,183 deaths for which no ethnicity was provided and no matching census record was found).

10.4 The estimate of 142,402 'White – Scottish' deaths (based mainly on data from the Death Registration records) represents 98 % of the better estimate of 145,685 deaths (based mainly on the census data for the people whose deaths were registered in the three years). So, any 'White – Scottish' death rates that were produced using the data from the Death Registration records would tend to be slightly lower than the (unknown) true values.

10.5 Table X shows those estimates, and the corresponding figures for each of the other ethnicities for which at least 100 deaths were registered in the three years (there seems little point in producing such estimates for ethnicities with smaller numbers, as they could be affected greatly, in percentage terms, by 'random' factors). As will be seen, the 'mainly Death Registration-based' estimates represent the following percentages of the (better) 'mainly census-based' estimates for those ethnicities:

- 98 % for 'White – Scottish'
- 127 % for 'White – Other British';
- 80 % for 'White – Irish'
- 92 % for 'White – Polish'
- 157 % for 'Other White'
- 83 % for 'Pakistani, Pakistani Scottish or Pakistani British'
- 106 % for 'Indian, Indian Scottish or Indian British'
- 89 % for 'Chinese, Chinese Scottish or Chinese British'

10.6 The greater the difference from 100 %, the less reliable would be any estimate of an ethnicity's death rate that was calculated from the Death Registration data. It appears that 'White – Scottish' is the only ethnicity (of those for which estimates were produced) for which the estimate, and hence any estimate of an ethnic group-specific death rate, is likely to be within a couple of percent of the (unknown) true value. The estimates for 'Indian ….' and 'White – Polish' might be within around six % to eight % of the (unknown) true values, but the estimates for the other five ethnicities would all appear to differ by more than 10 % from the (unknown) true value (including over-estimation of 27 % for 'White – Other British' and 57 % for 'Other White').

10.7 The above percentages were calculated from the data for deaths of all ages, of both sexes and registered in all three years, taken together. Had similar estimates been produced using data for (say) individual age-groups for each sex for single years, one would expect that the scale of the difference would vary from one age/sex/year 'cell' to another – so one would expect to see much larger percentage differences for some of the cells (and also some cells with much smaller percentage differences). For example, while the overall figure for 'White – Polish' might be eight % too low, one might find under-estimation by (say) 18 % for one age-group and over-estimation of (say) two % for another age-group. It follows that any age-and-sex-specific death rates that were produced from the Death Registration data could be subject to large percentage errors, because of the apparent unreliability of some of the information about the ethnicity of the deceased. In addition, 'random' factors may lead to large percentage fluctuations in figures that are based on relatively small numbers of deaths (more information about this is available from the Fluctuations in and possible unreliability of death statistics for small areas, for small sub-groups of the population, or for short periods section of the NRS website.). Therefore, the Death Registration data on the ethnicity of the deceased person are not (at present) suitable for calculating reliable mortality rates for most ethnicities.

10.8 For completeness, Table X also includes similar estimates for 'White' and 'BAME' (which were produced using the data which underlie Table 1). While the two estimates for 'White' are almost the same, the 'mainly Death Registration-based' estimate for 'BAME' is only 84 % of the (better) 'mainly census-based' estimate. It follows that the data on the ethnicity of the deceased person are not (at present) suitable even for calculating reliable overall 'BAME' mortality rates.

10.9 Finally, paragraph 10.2 referred to the need to scale up the Death Registration figures from Vital Events Reference Table 5.4 to take account of the fact that the ethnicity was not provided for a proportion of the Death Registration records. Section 6 and Table 4 show the increases that would be needed for the overall total numbers of deaths for each of the 'main' ethnicities: for example, 3.3 % for 'White – Scottish', 8.1 % for 'White – Other British', and much larger figures for some ethnicities. Two points should be made on this:

- first, different percentage increases could well be needed for different sub-groups of the population. For example, as the percentage of deaths for which ethnicity is provided rises with the age of the deceased (refer to Section 8 and Table 6), one would need much smaller percentage increases for the older age-groups than for the younger age-groups. It is not possible to determine what the percentage increases should be for the youngest age-group, because the census cannot provide any information for people who had not been born then. Some sub-groups' percentage increases would have to be large, because Section 6 and Table 4 showed that the overall Death Registration figures for most of the ethnicities should be increased by at least 10 % to take account of those for whom no ethnicity information was provided when the death was registered.
- second, there are differences between the percentage increases (to take account of the fact that the ethnicity was not provided in some cases) that were used to produce the estimates that are shown in Table X and those that are shown in Table 4. Such differences arise because those percentage increases have different denominators: for Table X, the denominator is the number of Death Registration records with the given ethnicity; for Table 4, it is the number of deaths for which there was a matching census record with the given ethnicity. The two percentage increases differ most for 'Other White' (8 % in Table X, 20 % in Table 4), due to a large difference between the number of Death Registration records with 'Other White' as the ethnicity (1,745 – used in Table X) and the number of deaths for which there was a matching census record with 'Other White' as the ethnicity (866 – used in Table 4).

11. Acknowledgment

NRS is grateful for the advice provided, towards the end of this project, by Professors Raj Bhopal and Laurence Gruer, who commented upon the emerging findings and the final draft of this report, and made some helpful suggestions for the interpretation and possible further analysis of some of the results. Any errors or omissions are the responsibility of NRS.

ANNEXES

Annex A Method of linking ethnicity information from the 2011 Census to data from the Death Registration records for 2012 to 2014

A1 This Annex describes how the ethnicity information from the 2011 Census was linked to that from the Death Registration records for 2012 to 2014. The two were brought together by linking them separately to the NRS Indexing Team's 'Population Spine' (rather than by directly linking the two sets of records together).

A2 Matching: overall process

A2.1 Each dataset was linked to the NRS Linkage Population Spine, which is a list of people in the population used for data linkage. Each entry in the list consists of a few items of identifying information. The Population Spine is updated regularly. It allows linkage keys to be maintained securely over time, without the need to create standing research datasets, and may be used to link datasets collected at different times.

A2.2 The NRS Indexing Team used the following details for matching:

- first forename,
- surname,
- sex,
- date of birth (allowing separate comparison of year, month and day) and
- postcode of the person's address

from the Death Registration records and the 2011 Census records to find the relevant records on the Population Spine, and created a series of ad-hoc references for each case involved in this project (these are project-specific index keys, assigned at random so that it is not possible to infer, from the order of the records within a dataset, to whom a particular reference relates). These references defined two relationships: first, between each Death Registration record and the corresponding Population Spine record (if there was one); second, between each 2011 Census record and the corresponding Population Spine record (again, if there was one).

A2.3 NRS Indexing Team then provided NRS Census with a file which included the ad-hoc references for all the records in the 2011 Census, plus an indicator of the quality of the match between the 2011 Census and the Population Spine and the year of registration of the death for those 2011 Census records that were associated with a Death Registration record (via the Population Spine). That enabled NRS Census to provide NRS Vital Events with a file containing, for each 2011 Census record that had been associated with a Death Registration record, an ad-hoc reference plus the

'ethnicity' information from the 2011 Census, an indicator of whether that information was imputed, the indicator of the quality of the match, and the year of registration of the death. NRS Vital Events also received from NRS Indexing Team, for each Death Registration record that had been provided for matching with the Population Spine, the ad-hoc references, an indicator of the quality of any match, and the unique identifier for the record of the death. NRS Vital Events then used the ad-hoc references to bring together the ethnicity information from the 2011 Census, the indicators of quality and the data from the Death Registration record.

A2.4 This approach ensures that those involved in a data linkage project have access only to the data which they need for their part in the project (this is for the data which are not 'theirs', they see only the personal identifiers or the additional variables which are of interest for their analysis - not both together):

- NRS Vital Events got only one item of information that was provided in the 2011 Census: the person's ethnicity;
- NRS Indexing Team got only those Death Registration and 2011 Census data items that they needed to link the two sets of records to the Population Spine. NRS Indexing Team did not get the ethnicity information from either dataset;
- NRS Census did not get any information from the Death Registration records other than the year in which the death was registered.

A2.5 NRS Indexing Team linked the Death Registration and 2011 Census records to the Population Spine in a way which ensured that a particular Population Spine record could be matched to no more than one Death Registration record and to no more than one 2011 Census record. The matching procedure was complicated, and used specialised software which examined the five specified fields for each person and calculated 'linkage' scores which gave greater weight to matching on 'rare' values (such as unusual surnames).

A3 Matching: the procedure followed within each phase

A3.1 The matching was done in a number of phases. In each, the records to be compared were grouped into 'blocks', by bringing together all the ones with the same value for the 'blocking criterion' field (or fields). Then, within each block, all the records which matched (on the basis of the fields listed above) were identified. A diagram at the end of the Annex summarises what is done in each phase. The blocks and the criteria on which they were based were as follows:

- phase 1 - postcode – using the current (this is the latest reported) postcode, in cases for which a Population Spine record had more than one postcode
- phase 2 - postcode – using any other postcode(s) on the Population Spine
- phase 3 - postcode district – for example EH10

- phase 4 - postcode town – for example EH (Edinburgh)
- phase 5 - the first letter of the first forename and the date of birth

A3.2 For each phase, there was no need to examine all the possible pairs of records which had the same value for the blocking criterion. The linkage tool (software called Link Plus) provides a score which is an estimate of the likelihood of a pair being a true match, and also identifies any fields for which the pair's values do not match. This enabled the NRS Indexing Team to set a threshold for selecting pairs for further inspection at a level which should prevent few genuine matches being missed, and allow few false matches (based on a subjective judgement of where, in a list of pairs in order of their matching scores, there is a change from 'likely match' to 'likely non-match'). Inevitably, this approach can miss small numbers of true matches, but they might still be identified in subsequent phases.

A3.3 The selected 'likely match' pairs were analysed using software called SAS. Those which were exact matches (on the basis of the five specified fields) were extracted first. Then, pre-programmed 'fuzzy match' criteria were applied to identify 'close matches', where only one field had values which did not match exactly (for example perhaps the day and month of birth appeared to have been swapped round on one record, and all the other fields' values agreed exactly). Finally, those pairs not assigned as matches by the first two approaches were clerically reviewed, at which stage some possible matches of alternative names were assessed (for example 'Rob' in one record, 'Robert' in another; or the first forename on one record matching the middle name of another) . Again, not all the residual pairs were examined, as it was obvious that low scoring cases were so dissimilar that there was no need to review them.

A3.4 Records that matched exactly were removed from the process in the phase in which they are identified. Inexact match records were not: they were 'flagged' as possible matches, but also went forward into the next phase with the as yet unmatched records. This means that 'competing' matches could arise later: for example, a given census record might be a 'close' match for one Population Spine record in one phase, and a 'close' match for another Population Spine record in a later phase. As part of the final phase, there was a clerical review to try to capture matches which had been missed in the earlier phases. At this stage, competing matches from different phases were compared using their scores relative to other scores within the same phase, and the competing matches with the highest relative score were taken to be true matches.

A4 The quality of the matching

A4.1 The above notes describe the linkage procedure that applied at the time that this work was done. NRS Indexing Team's assessment of its results concluded that they were sufficiently reliable. For example,

matches that were based on clerical review were verified by a second person, who worked independently.

A4.2 The quality of the matching of records to the Population Spine is defined as follows:

- High quality (H) matches
  - exact match across a minimum of forename, surname and date of birth
  - usually will also match on the sex and the full postcode (rather than, say, the postal district)
- Medium quality (M) matches
  - probabilistic matches based on the match score exceeding that of the lowest-scoring one % of the exact matches. It is difficult to describe such cases precisely, as matching on unusual names will contribute to a high score (and such cases might not match well on the other variables). Usually, though, it is expected that at least one variable would be an exact match and that there would be close matches on the other variables.
- Low quality (L) matches
  - the result of clerical review and judgement such as mis-spellings, and jumbled dates of birth.
  - possible human error and lack of consistency

A4.3 An important part of this project was to act as a pilot for NRS's new approach to linkage via the Indexing Team, who needed to work with the data analyst to understand the accuracy of these quality measures in order to allow them to be applied more widely in future.

A5 Some reasons why some records could not be matched

A5.1 Some Death Registration records (and some 2011 Census records) could not be matched to the Population Spine. There were also Population Spine records for which there were no matching 2011 Census records. As a result, some Death Registration records could not be matched (via the Population Spine) to 2011 Census records.

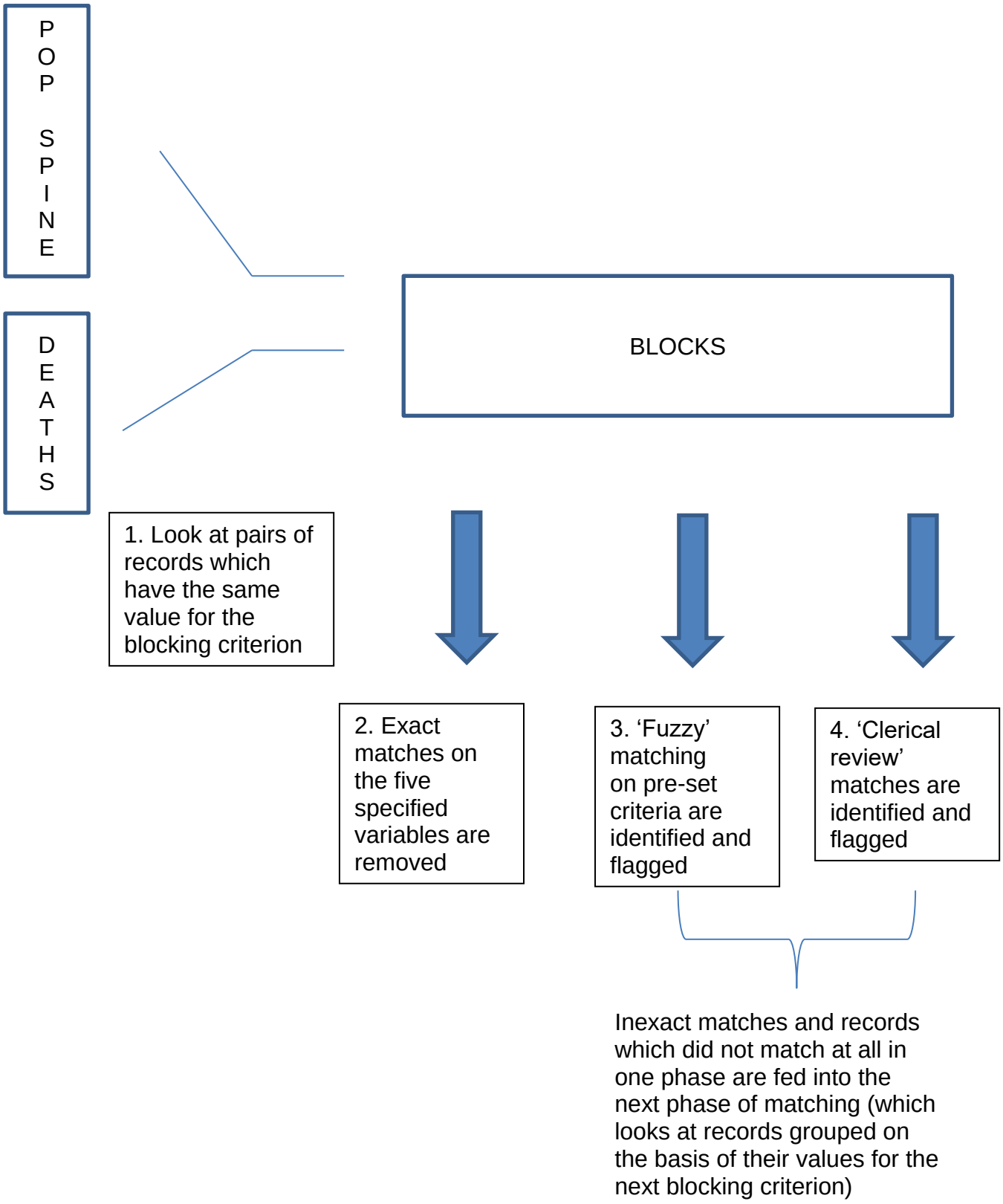
A5.2 Reasons why this might have happened include, for example:

- no Census record
  - the person might have been living outwith Scotland at the time of the 2011 Census, and so was not covered by it;
  - the person might not have been born at the time of the 2011 Census, so again would not be covered by it;
  - the person might have been in Scotland at the time of the 2011 Census, but was not enumerated for some reason – there are 'hard to count' groups within the population, and a small number of people will avoid being counted in the census.
- problems with the information that was entered on the 2011 Census form, or was recorded when the death was registered. For example, the person who registered the death might have provided an incorrect date of birth. It is important to note that

census information will normally come directly from an individual, whereas Death Registration data are provided by third parties, who may not be close relatives or friends.

- differences between the information held in the 2011 Census, Population Spine and the Death Registration records – for example,
  - different records may hold markedly different forms of a name (or an variant thereof), such as ‘Alexander’ (or ‘Alexandra’) in one record and ‘Sandy’ in another; ‘Agnes’ in one and ‘Senga’ in another.
  - someone with multiple forenames, who normally goes by the second forename, might have that name appearing first in his/her Population Spine record, but the person’s forenames in the Death Registration record might start with the one that appeared first on his/her birth certificate.

Summary of what was done in each phase of the matching process



Annex B Variation in the apparent quality of information provided when a death is registered with the informant's relationship to the deceased

B1 This Annex considers whether and, if so, how the apparent quality of the information varied with the relationship between the informant and the deceased.

B2 Paragraph 4.1 stated that one would expect that, in general, information about the deceased's ethnicity should be more reliable when the informant is the deceased's spouse/partner, or a child, sibling or parent, than when the informant is a more distant relative or not related at all. After the Death Registration and 2011 Census information about ethnicity had been brought together, and most of the analyses described earlier had been done, it was felt that it would be as well to investigate the accuracy (or otherwise) of this statement.

B3 Table 10 shows that there were 116,275 deaths (registered in the three years) for which the quality of matching was 'high', the 2011 Census provided information about ethnicity which was not imputed, and the ethnicity of the deceased was provided. From the first row, it will be seen that the ethnicity that was given when the death was registered was the same as that reported in the 2011 Census for 94.7 % of these cases. The next four rows show the corresponding figures for some groups of types of informant:

- 94.4 % when the informant was the widow, widower or partner of the deceased;
- 94.8 % when the informant was a child, sibling or parent of the deceased;
- 95.0 % when the informant was one of certain other kinds of relative of the deceased, including grandchild, stepchild, nephew/niece and certain kinds of in-law;
- 93.2 % when the informant was anything else, including a friend, cousin, executor or solicitor.

These groupings are, of course, arbitrary: for example, it could be argued that (say) stepchildren should be in the same grouping as children, or that cousins should be in the same grouping as nephews and nieces. The groupings were chosen before the analyses were done, in the expectation that they would show that the apparent 'accuracy percentage' was higher for certain types of informant whom one might consider to tend to be 'closest' to the deceased, and lower for certain types of informant who might tend to have a more distant connection (rather than to provide definitive groupings of types of informant based on the apparent accuracy of their information).

B4 It is surprising that the percentage agreement is (only just) highest (95.0 %) when the informant was in the 'certain other kinds of relative' group, and that the figure for deaths registered by a widow, widower or partner (94.4 %) is slightly below that for all the deaths covered in the table (94.7 %). The percentage is lowest (93.2 %) for the 'other' category, as would be expected.

B5 The middle of the table shows the figures for the main types of informant in each group (for this purpose, those which registered at least 100 deaths over the three years). The descriptions are as they were recorded when each death was registered, and converted to wholly upper case so that it does not matter whether (for example) initial capitals were used. However, there is no standardisation of some variation in the way that a term is recorded – for example 'step daughter' (with a space), 'step-daughter' (with a hyphen) and 'stepdaughter' (with no separator) are treated separately. It will be seen that, for some of the types of informant who registered most deaths:

- widow (94.2 %) and widower (94.4 %) have very similar percentages;
- father (93.8 %) is below the overall figure of 94.7 % agreement;
- daughter (94.8 %) and son (94.6 %) have almost identical percentages;
- brother (95.4 %) is less than sister (96.3 %);
- surprising, nephew (95.2 %) and niece (95.4 %) have clearly higher percentages than any of widow, widower, daughter, son and father. Given the numbers of deaths involved, the high percentages for nephew / niece are the main reason for the 'certain other kinds of relative' group having the highest overall percentage;

Of course, some of the apparent differences between types of informant may be due to 'random' factors. There may be no 'real-life' reason why (say) nephews and nieces should, in general, be able to provide information which is (on average) slightly more consistent with people's views of their ethnicity than that provided by closer relatives. Or, it might be that a larger proportion of the deaths which are registered by (say) nephews and nieces are of people whose ethnicity is relatively easy to 'get right', because (say) they spent all their lives in a particular part of Scotland.

B6 The foot of the table shows the corresponding figures for those for whom the census had a 'BAME' ethnicity. Overall, 55 % of them had the same ethnicity in the Death Registration and census records; again, the percentage was higher for the 'certain other kinds of relative' group (66 %) than for child/sibling/parent (54 %) or widow/widower/partner (56 %). The numbers involved are relatively small, so cannot be analysed in great detail. However, it should be noted (from the final column of the table) that only about two-thirds of these people were described as 'BAME' when their deaths were registered, with the percentages for the main groups being:

- 69 % for widow / widower / partner;
- 63 % for child, sibling or parent;
- 74 % for certain other types of relative; and
- 45 % for the 'other' group.

Annex C Some Census Quality Survey results regarding ethnicity

C1 This Annex describes the Census Quality Survey (CQS), its main findings regarding responses to the ethnicity question in the 2011 Census, and how they compare to the main differences between the ethnicity recorded in the census and the Death Registration records.

C2 Background

The CQS was conducted to obtain an indication of the reliability of the answers to each of the census questions. Its sample was chosen from among the earliest census returns in order that census data would be available in time to load onto the CQS interviewers' laptop computers. In total, 1,760 people were interviewed in 787 households in certain parts of Glasgow, Fife and Scottish Borders: areas that were chosen to reflect the diversity of Scotland's population and circumstances (for example from built-up to rural). Participation was voluntary, and the overall response rate was around 50 % (somewhat lower in Glasgow, and higher in both Fife and Scottish Borders). The CQS was not designed to provide nationally-representative results, so its findings are general indications rather than precise figures, and should be analysed only in a 'broad-brush' way.

C3 Differences between the ethnicity given in the CQS and that reported in the Census

C3.1 In total, 54 respondents to the CQS said that their ethnicity was something different from the ethnicity that had been reported for them in the census:

- in about half of these cases, the person was shown as 'White - Scottish' in the census, but described him/herself as 'White - Other British' in the CQS;
- in roughly a fifth of cases, the census had 'White - Other British' and the CQS answer was 'White - Scottish';
- the remaining 30 %-or-so were spread over a range of combinations of census and CQS responses, with the most numerous being three people for whom the 'White - Other white ethnic group' box had been ticked in the census, and 'ENGLISH' had been put in the 'please write in' space, who said that they were 'White - Other British' when they answered the CQS. These responses (and some of the others that were given by only one or two people in the CQS) suggested that they had not properly understood which ethnicity category was the most appropriate for them. (Note: the answers to census questions that were put onto the interviewers' laptops were 'raw' data that had not yet been subject to all NRS's data quality checks. Later, when NRS processed the census data, it corrected answers that contained obvious errors or inconsistencies – like the ones, mentioned above, for which 'White - Other white ethnic group' had been ticked and 'ENGLISH' had been written in. It follows that such

cases were counted as ‘White - Other British’ in the final census datasets.)

C3.2 When the census and CQS answers differed, the interviewee was asked to pick the reason for the discrepancy from a list of possible causes. The reasons most often given for the differences in ethnicity were as follows:

- ‘mistake’ – about a quarter of the 54 cases;
- ‘changed mind’ – roughly a fifth;
- ‘someone else completed form’ – around a seventh.

A few people said that they had misunderstood the question; a third picked ‘other reason’ or did not give a reason.

C4 Comparison with differences between ethnicity in the Death Registration and Census data

The two main differences between the census and the CQS are very similar to the two main differences between the census and the Death Registration ethnicity data:

- in about half of these cases, the person was shown as ‘White - Scottish’ in the census, but described him/herself as ‘White - Other British’ in the CQS – which is akin to 53 % (4,635 out of 8,749) of the ‘off-diagonal’ cases in Table 1 being ‘White – Scottish’ in the census and ‘White – Other British’ in the Death Registration record;
- in roughly a fifth of cases, the census had ‘White - Other British’ and the CQS answer was ‘White - Scottish’ – which is comparable to 16 % (1,362 out of 8,749) being ‘White – Other British’ in the census and ‘White – Scottish’ in the Death Registration record.

Those who sometimes describe themselves as ‘White - Scottish’ and sometimes as ‘White – Other British’ could be among the Scottish equivalents of the people for whom Simpson et al (refer to paragraph 2.3 of the main report) felt that more than one response to the census question was suitable, and who were thought to account for about a third to a half of all the changes of individuals’ ethnicity between the 2001 and 2011 Censuses in England and Wales.

C5 People for whom ethnicity had not been recorded on the census form

For completeness, it may be noted that the CQS included 12 people whose ethnicity had not been given in the census (and, therefore, for whom the ethnicity would be imputed at a later stage in the processing of the census data). They were asked to pick from a list of reasons for not answering the question in the census:

- five said that they had not seen the question in the census (even though it occupied half a page in the printed version of the form);
- two said that they had not thought that it applied to them (although none of the ‘routing’ instructions should cause anyone to ‘bypass’ the question);

- one ‘didn’t understand’ the census question; and
- no reason was given in the other four cases.