

The quality of some of the items of information about Deaths held in the Vital Events statistical database

1. Introduction

1.1 This note describes the results of a comparison of the information held in nine of the fields in the National Records of Scotland (NRS) Vital Events (VE) database with the details that were recorded on the relevant Medical Certificates of Cause of Death (MCCDs) held by the Registration Office. The work was done in February 2011, using photocopies of MCCDs for a sample of 364 randomly-chosen deaths, which were supplied by 62 of the larger Registration Offices.

1.2 The information in the database was assumed to be correct if it was the same as what appears on the MCCD. Where there were differences, it was assumed that there was an error in the information in the database unless:

- (a) the Informant could provide more accurate information than the MCCD (in the case of the spelling of the deceased's surname); or
- (b) what was in the database could take account of other information that may become available after the MCCD was completed (e.g. the 'post-mortem may be available later' code will be updated when NRS is told whether or not the post-mortem was conducted).

Such errors are believed to arise when the details of the deaths are registered, and to be due to Registrars mis-keying data or selecting the wrong entries from drop-down menus (NRS's computer systems would not change slightly the information recorded by the Registrar, and eight of the nine fields are not ones that are amended by NRS staff).

1.3 For this sample:

- there were no errors in the recording of the data for 'Deceased's Surname', 'Date of Death' or 'Time of Death';
- it seemed that Registration Offices failed to key the name of the consultant in about 15% of the cases for which a consultant's name had been entered on the MCCD;
- otherwise, the error rate (for mis-keying / mis-spelling of the name) appeared to be around ½ - 1% for 'Certifying Doctor' and 'Consultant' (for the latter, this is the error rate for the cases where the name was keyed into the computer);
- the wrong code had been recorded in about 1-2% of cases for 'Post-mortem', 'Attendance on Deceased', 'Extra Information for Statistical Purposes' and 'Procurator Fiscal'; but
- the Extra Information for Statistical Purposes box had not been ticked for about a sixth of the cases which had 'Y' in the corresponding field in the database.

1.4 In the light of these results, NRS is taking steps to improve the quality of the data. When they scrutinise the information that has been recorded for deaths, the Examiners will start to look at two specific points:

Vital Events – Deaths – Background Information

- if a death occurred in hospital - has the consultant's name been recorded?
- if a record has 'Y' in 'Extra Information for Statistical Purposes' - was that box on the MCCD actually ticked?

In addition, when other work permits, NRS IT staff will consider whether it might be possible to improve the computer system's 'user interface' to reduce the likelihood of the inadvertent selection of the wrong option from a 'drop down' menu.

- 1.5 Section 2 describes the main findings in more detail, [section 3](#) provides more detailed points on the comparison for some of the individual fields, and the [Annex](#) describes briefly the design and processing of the sample.

2. Main findings

- 2.1 On the basis of this sample, it appeared that the error rate was zero, or a small percentage, for seven of the nine fields that were examined. There appeared to be:

- no errors in three of the fields: 'Deceased's Surname', 'Date of Death', and 'Time of Death';
- two errors in 'Certifying Doctor's Name', representing an error rate of 0.5%;
- three errors in 'Post-mortem', representing an error rate of 0.8%;
- three errors in 'Attendance on Deceased' - again, an error rate of 0.8%;
- eight errors in 'Procurator Fiscal' - an error rate of 2.2%;

- 2.2 However, there were proportionately many more errors in the other two fields that were examined. In the case of 'Consultant's Name':

- there were 26 cases where it seems that the Registrar overlooked the need to enter the consultant's name; and
- two cases where it was mis-keyed.

Because the consultant's name should be blank for deaths outwith hospital:

- the error rate for the mis-keying of 'Consultant's Name' is based on the number of deaths for which the name was keyed - this was 149, so the error rate is 1.3%;
- the error rate for overlooking the consultant's name is based on the number of deaths for which a consultant's name should have been keyed - this was 175 (i.e. the 149 cases where it was keyed, plus the 26 cases where it wasn't), so this error rate is 14.9%.

- 2.3 There were also many errors in 'Extra Information for Statistical Purposes':

- the doctor had not put anything in the box on the MCCD in 28 of the 180 cases for which the database had 'Y' - so 15.6% of those entries were wrong; and
- the doctor had ticked the box in one of the 184 cases for which there was 'N' - so 0.5% of those entries were wrong.

Because the sample was designed to be split '50:50' between 'Y' and 'N' values for 'Extra Information ...' ([Annex](#)), one cannot calculate an overall error rate for this field purely from the sample's results. However, [Paragraph 3.6](#) provides an estimate of the likely overall error rate for this field.

3. Detailed points on the results for some of the fields

- 3.1 **Deceased's Surname:** There were 8 cases for which the deceased's surname recorded in the database differed slightly from that on the MCCD (e.g. 'Cruickshank' vs 'Cruikshank'; 'Elliott' vs. 'Elliot'; etc). In such cases, it was assumed that the surname recorded in the database is correct, because the doctor who completed the MCCD might not know the precise spelling of the deceased's surname. The Informant, who is usually a member of the family, is more likely to know the correct spelling; and the name is something that the Informant and the Registrar should check, when they look over the draft entry before the registration is finalised.
- 3.2 **Certifying Doctor's name:** There appeared to be two errors in the database: in one case, it appears that the Registrar mis-keyed the doctor's name (which had been clearly written on the MCCD); in the other, it appears that the Registrar took the name that was in the 'consultant' box.
- 3.3 **Consultant's name:** There appeared to be 28 errors in the database:
- 17 cases where the MCCD gave the name of the consultant, which was not the same as that of the certifying doctor, but there was no consultant's name in the database. It was assumed that, in these cases, the Registrar simply overlooked the need to enter the name of the consultant.
 - 9 cases where the MCCD gave the name of the consultant, who was the certifying doctor as well, and there was no consultant's name in the database. Again, we assume that the Registrar has overlooked the need to enter it.
 - 2 cases where the Registrar appears to have mis-keyed a consultant's name that was clearly written on the MCCD (e.g. 'David' whereas the MCCD had 'Davie').
- 3.4 **Post-mortem:** There appeared to be 3 errors in the database - for example, one case where the database had code 2 'post-mortem information may be available later' but, on the MCCD, box 3 'no post-mortem is being done' was clearly chosen. There were several other cases where the code in the database differs from what appears on the MCCD, but these were likely to be due to the later provision of more information. For example, in cases where MCCD box 2 'post-mortem information may be available later' was ticked, the database will have a different code if NRS was later told (e.g.) that the proposed post-mortem had not actually been performed.
- 3.5 **Attendance on Deceased:** There appeared to be 3 errors in the database - for example, two cases where the MCCD had a tick in box 2 'I was not in attendance ...' but the database had code 1 for "I was in attendance...".

Vital Events – Deaths – Background Information

3.6 **Extra information for statistical purposes:** The doctor makes a mark (e.g. a tick, or a cross) in this box on the MCCD if he/she wants to indicate that 'I may later be able to supply the Registrar General with additional information'.

However:

- in 28 of the 180 cases for which the database had 'Y', the doctor had not put anything into the box on the MCCD, so in all those cases the doctor did not have any additional information to supply
- in one of the 184 cases for which the database had 'N', the doctor had ticked the box, so in that case the doctor did have additional information to supply.

The errors mean that the code that was recorded is wrong in:

- 28 of the 180 cases where 'Y' is in the database - i.e. 15.6% of those entries are wrong;
- one of the 184 cases where 'N' is in the database - i.e. 0.5% of those entries are wrong;

Because the sample was designed to be split '50:50' between 'Y' and 'N' values for 'Extra Information ...', one cannot calculate an overall error rate for this field purely from the sample's results. Instead, the overall error rate for the field is a weighted average of the above error rates. As only about 3½% of records for deaths registered in 2010 had 'Y' in the field, the overall error rate must be much closer to the error rate for the cases which have 'N' (it will be roughly 1%, calculated thus: $.035 \times 15.6\% + .965 \times 0.5\%$). Note that the error percentage for cases which have 'Y' is much higher than 1% because the mis-recording of a small proportion of the very large percentage of cases which should have had 'N' represents a significant fraction of all the ones that have 'Y'.

3.7 **Procurator Fiscal:** A tick in the box on the MCCD indicates that 'this death has been reported to the Procurator Fiscal'. There appeared to be 8 errors in the database: (e.g.) two cases where the database had 'blank' (indicating that the death had not been reported to the PF), but there was a tick in the box.

Annex - Sample design and processing

A1 The sample was picked from the statistical records for the deaths which were registered in 2010 at offices which registered 200+ deaths in that year. Each such office received a letter which identified up to six cases for which photocopies of the MCCDs were requested (using the Year of Registration, the Registration District code, Entry Number and Surname of the Deceased). Because there was particular interest in the accuracy of the 'additional information may be available later' indicator, each office's sample had (a) up to three deaths randomly-chosen from those for which that indicator's value (in the VE database) was 'Y' plus (b) up to three deaths randomly-chosen from those for which the indicator's value was 'N'. There were six cases in the sample for almost all the offices: the exceptions were offices which had fewer than three deaths registered in 2010 for which the indicator's value was 'Y' (as their samples could not include three such deaths). Also, a few offices failed to provide a photocopy of the MCCD for one of the deaths which had been included in the sample.

A2 Such a sample is not completely representative - for example:

- deaths registered at the largest offices were particularly under-represented, because no office had more than six cases in the sample;
- deaths registered at the smallest offices were completely unrepresented, because only offices with 200+ deaths were asked to provide cases; and
- deaths for which the 'additional ...' indicator was 'Y' were considerably over-represented (about half the sample compared with only 3½% of all deaths).

However, it should provide a reasonable general indication of the accuracy of the data taken from the MCCD, as the cases were chosen at random and are spread across most of Scotland.

A3 An NRS VE clerk compared what was in the MCCD photocopies and the database: for some fields, this was done by eye; for other fields, by keying information from the MCCD into a spreadsheet that VE used for the work. Differences were 'flagged', and checked by a more senior person, who excluded any which were, or seemed likely to be, due (e.g.): to the clerk misreading the photocopy of the MCCD or mis-keying something into the spreadsheet that VE used for the work; to further information becoming available to NRS after the MCCD had been completed; or to the Registrar or the Informant having better information than the certifying doctor. Therefore, the remaining differences were all cases where the information held in the VE database was wrong.