Automatic extraction of multiple underlying causes from textual death records

Graham Kirby[1], Masih Hajiarab Derkani[1], Alan Dearle[1], Jamie Carson[1], Fraser Dunlop[1], Chris Dibben[2], Lee Williamson[2]

[1] University of St Andrews, [2] University of Edinburgh

Addendum: Classification Performance Considered at Various Hierarchy Levels

The classification performance metrics shown on the main poster (precision, recall, F1) are calculated according to whether each classification extracted by the system is exactly the same as that defined in the gold standard data. Even a minor difference, for example "alcohol dependence syndrome" rather than "acute intoxication", is counted as being completely incorrect.

The primary purpose of classifying causes of death in the Digitising Scotland project is to enhance the usefulness of the data set, by enabling researchers to retrieve records relevant to particular research questions. In some cases fine-grained classification accuracy may not be essential. For example, there are around 500 ICD-10 codes specifying various forms of cancer; a researcher investigating geographical patterns in cancer incidence may be satisfied with retrieving all cancer-related records, whether or not the precise form is accurately classified.

With a hierarchical classification system such as ICD-10, and the variant used for the Kilmarnock data set, it is possible to deduce whether one classification is related to another based on sub-strings of the classification codes. For example, in ICD-10 all cancer-related codes begin with "C"; all codes related to cancer of respiratory organs begin with "C3", and so on.

Table 4 shows the performance metrics for the Kilmarnock data set after re-calculation using code sub-strings of various lengths. The first row corresponds to the original metrics; the second row shows small improvements when the last letter of the code is ignored. The final row shows micro-precision/recall of 93% when only the first letter of the code is considered. That is, for each major category in the classification system, 93% of the decisions made by the classifier as to whether that category was present or absent in a particular record were correct.

Table 4. Performance Metrics at Various Hierarchy Levels for Kilmarnock Data Set

	Macro-Precision	Macro-Recall	Macro-F1	Micro-Precision/Recall
Level 5 (Exact)	84%	40%	40%	85%
Level 4	85%	42%	43%	86%
Level 3	86%	45%	45%	87%
Level 2	86%	54%	58%	90%
Level 1	89%	76%	80%	93%