Umeå Centre for
Global Health Research

WHO Collaborating Centre
for Verbal Autopsy

Verbal Autopsy Technical Bulletin – November 2015

# Performance of various automated coding methods for Verbal Autopsy

Various automated coding methods for verbal autopsy have been developed in recent years. This Bulletin sets out the major published evaluations of current models as a guide for users wishing to make a choice between available methods, or compare them with physician coding.

The main currently available automated methods are Tariff (*Population Health Metrics* 2011; 9:31), InsilicoVA (http://arxiv.org/abs/1411.3042) and InterVA (*Global Health Action* 2012; 5:19281). All methods comprise three essential components: encapsulated knowledge (which may be derived from data and/or human expert knowledge); algorithmic logic; and software implementations. Current versions of all the models make use of subsets of the WHO 2014 VA standard (http://www.who.int/healthinfo/statistics/verbalautopsystandards) for inputs and outputs. It is expected that forthcoming versions of the various models will offer compatibility with the full range of inputs and outputs defined in the WHO 2014 VA standard.

Evaluations of automated VA methods include a range of test data sets and outcome parameters. Test datasets may include physician-coded verbal autopsy material, hospital assigned causes of death, or deliberately designed reference datasets such as the Population Health Metrics Research Consortium (PHMRC) dataset (*Population Health Metrics* 2011; 9:27). One of the most critical characteristics of an evaluation is whether it involves a model implementation in which the model's encapsulated knowledge is entirely independent of the test dataset, or whether the same dataset is used for model building and testing.

**Evaluations where model knowledge is independent of the test dataset**

A multi-methods comparison (*BMC Medicine* 2014; 12:20) reported similar performance between Tariff and InterVA. The newer InsilicoVA model (http://arxiv.org/abs/1411.3042) reported greater accuracy than Tariff and InterVA, indicating its potential advantages after further development. InterVA has been positively evaluated against the PHMRC reference dataset (*BMC Medicine* 12:23) and against physician coding (*Journal of Global Health* 2015; 5:010402). InterVA has also been specifically evaluated in relation to biomedical outcomes: HIV infection (*Global Health Action* 2013; 6:22448) and sickle cell disease (*BMC Medicine* 2014; 12:65).

**Evaluations where model knowledge is based on the test dataset**

An initial evaluation of the Tariff method, using the PHMRC reference dataset to build and test the model, reported good performance (*Population Health Metrics* 2011; 9:31). A similar comparative study (*BMC Medicine* 2014; 12:5) reported superior performance for Tariff over other methods.