Full Dataset for Adult Module

Random sampling without replacement by cause

25% test dataset

Resampled with replacement based on uninformative Dirichlet draw

Resampled test dataset

For each death in resampled test dataset, scores for each cause are calculated by summing tariffs (from train dataset tariff matrix) for endorsed items

Each cause score for each death is ranked relative to the corresponding cause score distribution in the train data

Cause assignment decided by cause with highest rank on each death

Training dataset is resampled with replacement to have a uniform cause distribution

75% train dataset

Tariff computed for each cause, for each sign/symptom

Tariff matrix

For each death in train dataset, scores for each cause are calculated by summing tariffs for endorsed items

For each death in resampled test dataset, scores for each cause are calculated by summing tariffs (from train dataset tariff matrix) for endorsed items

Cause-specific mortality fraction accuracy and chance-corrected concordance computed for all deaths in resampled test dataset

Repeated 500 times