

Full Dataset
for Adult
Module

Repeated 500 times

Random sampling
without
replacement by
cause

25% test
dataset

75% train
dataset

Resampled with replacement
based on uninformative Dirichlet
draw

Tariff computed for each
cause, for each sign/symptom

Resampled
test dataset

Tariff
matrix

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

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

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

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

Cause assignment
decided by cause with
highest rank
on each death

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