Probabilistic Cause-of-death Assignment using Verbal Autopsies

Tyler H. McCormick 1, Zehang Li 1, Clara Calvert 2, Amelia C. Crampin 2,3, Kathleen Kahn 4, and Samuel J. Clark 1,3,4

1University of Washington, 2London School of Hygiene and Tropical Medicine, 3ALPHA Network, London, 4INDEPTH Network, Ghana

Introducing

- Fewer than one-third of deaths worldwide are assigned a cause [1].
- Verbal autopsy (VA) used to assess cause of death and estimate cause-specific mortality fraction (CSMF).
- Interview with caregivers/relatives → data describing the signs and symptoms leading up to the death.

Automated VA methods

Learn connections between symptoms and causes using:

- **Gold standard data**
  - Multiple methods proposed by The Institute for Health Metrics and Evaluation (IHME) such as Tariff [2].
  - Early work by King and Lu [3].
- **Expert inputs**
  - InterVA [4]: widely used and also supported by the WHO.
  - Information from physicians in the form of ranked lists of signs and symptoms associated with each cause of death.

**Problem: Uncertainties exist in**

- Population cause distribution (C)
- Individual symptoms (S)
- Physician provided relationships (P_{is})
- Physician coded causes (G)

**Ideas:**

- Quantify uncertainties at all levels.
- Goals of inference:
  - \( y_i \in \{1, ..., C\} \): cause for death \( i \);
  - \( \pi \in \{\pi_1, ..., \pi_C\} \): population CSMF.
- Data with noise:
  - \( \tilde{s}_i \): signs/symptoms for death \( i \).
- \( P_{\pi|s} \): ranking matrix of conditional probabilities, i.e., “A+”, “A”, ...

**Model specification**

- Population CSMFs:
  - \( \pi_c = \exp \theta_c / \sum \exp \theta_c \)
  - Individual symptoms given causes:
    - \( s_{ij}/y = c \sim \text{Bernoulli}(P(s_{ij}|y = c)) \)
  - Individual causes of death given CSMF:
    - \( y_i/\pi_1, ..., \pi_C \sim \text{Multinomial}(\pi_1, ..., \pi_C) \)
  - Truncated Beta prior for ranked \( P_{is} \):
    - \( P_{L(s|c)} \sim \text{Beta}(a_{s|c}, M - a_{s|c}) \)
    - \( P_{L(s|c)} \leq (\sum_{c'} P_{L(s|c')} - 1)P_{L(s|c)+1} \)
- Computation Posterior not available in closed form. Obtain samples using MCMC where most steps have conjugate priors; \( \pi \) is sampled with a Metropolis-Hastings step.

**Physician coding**

- Some surveys reviewed by physicians.
- Each death coded by multiple physicians, each assign a cause.
- Certain level of physician bias is inevitable.

**Two-stage model**

1. Debias physicians’ tendencies [6]

- Use the broad categories of debiased cause distribution: \( Z_i = \{z_{1i}, ..., z_{Ci}\} \)

2. Assignments:

\[
P(y_i|\pi_i, S_i, Z_i) = \sum_{g=1}^{C} P(y_i|\pi_i, \eta_i = g)P(\eta_i = g|Z_i)
\]

where \( \eta_i \) is the latent indicator for category assignments.

**References**