Estimating causes of maternal death in data-sparse contexts

Michael Chong (joint work with Monica Alexander & Marija Pejcinovska)
University of Toronto Graduate Student Research Day | May 25, 2022

Context

- **Goal** is to estimate, for each country, the distribution of maternal deaths across 7 categories of causes of death
- **Data** consist of cause-specific counts by country-year
  - can come from different sources (e.g. vital registration, studies)
  - varies in coverage (proportion of deaths in the country captured by the data)
  - varies in quality
  - counts may be missing
  - some countries may have no data at all

Model for observed cause proportions

vector of cause-specific counts

$$\mathbf{y_i} \sim \text{Multinomial}(d_i, \mathbf{p_i})$$

vector of cause proportions

$$\mathbf{p_i} = (p_{i,1}, \ldots, p_{i,7})$$

$$\log\left(\frac{p_{i,j}}{p_{i,7}}\right) = \beta_{0,j} + \beta_{r(c(i)),j} + u_{c(i),j} + q_{i,j}$$

Model for observed cause proportions

$$\log\left(\frac{p_{i,j}}{p_{i,7}}\right) = \beta_{0,j} + \beta_{r(c(i)),j} + u_{c(i),j} + q_{i,j}$$

region effect

$$\beta_{r,j} \sim \text{Normal}(0, \sigma_\beta^2)$$

country effect

$$(u_{c,1}, \ldots, u_{c,6}) \sim \text{MVN}(\mathbf{0}, \Sigma)$$

$$\Sigma = V\Omega V$$

$$\Omega \sim \text{LKJ}(1)$$

data quality adjustment

For observations of type 1 (high quality)

$$q_{i,j} = 0$$

For types 2, 3, and 4 (lower quality)

$$q_{i,j} \sim \text{Normal}(0, \sigma^2_{\text{type}(i)})$$

Partial data

For data where all 7 causes are recorded,

$$\mathbf{y_i} \sim \text{Multinomial}(d_i, \mathbf{p_i})$$

$$\mathbf{p_i} = (p_{i,1}, \ldots, p_{i,7})$$

$$p_{i,j} = \frac{g_{i,j}}{\sum_{l=1}^{7} g_{i,l}}$$

$$g_{i,j} := \frac{p_{i,j}}{p_{i,7}} = \exp(\beta_{0,j} + \beta_{r(c(i)),j} + u_{c(i),j} + q_{i,j})$$

If, say, the first category is missing,

only non-missing categories

$$\mathbf{y'_i} \sim \text{Multinomial}(d_i, \mathbf{p'_i})$$

$$\mathbf{p'_i} = (p'_{i,2}, \ldots, p'_{i,7})$$

reduced category multinomial

$$p'_{i,j} = \frac{g_{i,j}}{\sum_{l=2}^{7} g_{i,l}}$$

Implementation via Poisson trick makes this easy for arbitrary combinations of missing causes

Estimation of true proportions

At the country level

weights based on data coverage

$$\log\left(\frac{p^*_{c,j}}{p^*_{c,7}}\right) = \beta_{0,j} + \beta_{r(c),j} + w_c \cdot u_{c,j} + (1 - w_c) \cdot \tilde{u}_{c,j}$$

estimated country effect for c

$\tilde{\mathbf{u}}_c \overset{\text{RNG}}{\sim} \text{MVN}(0, \Sigma)$ — generic realization of country effect

Aggregation to the regional level

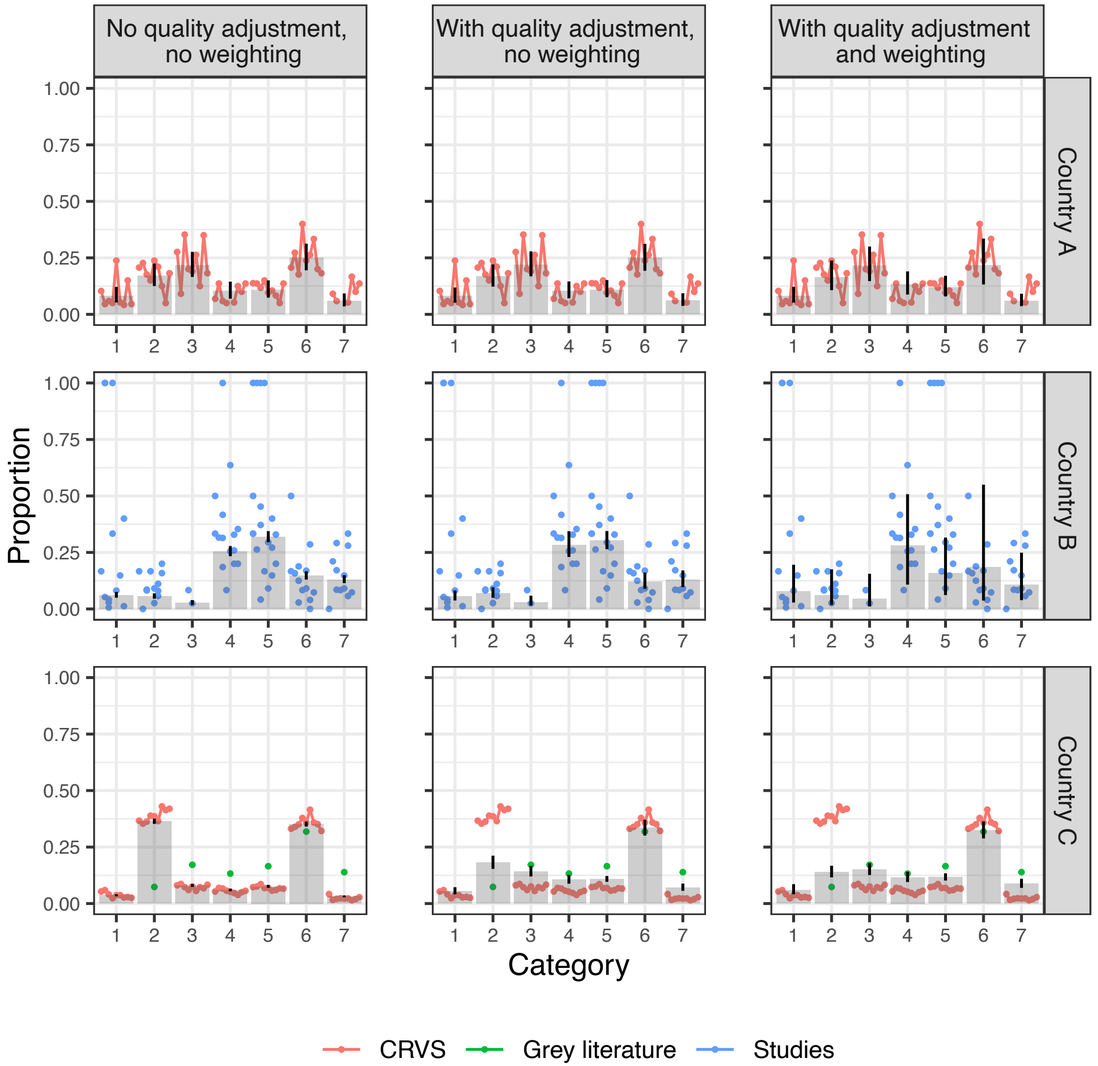
country-year maternal death counts (obtained externally)

$$d^*_{ct,j} = p^*_{c,j} \cdot d_{ct}$$

$$d^*_{R,j} = \sum_{R(c)=R} d^*_{ct,j}$$

$$p^*_{R,j} = \frac{d^*_{R,j}}{\sum_{l=1}^{7} d^*_{R,l}}$$

Case studies: data scenarios



high quality data, high coverage

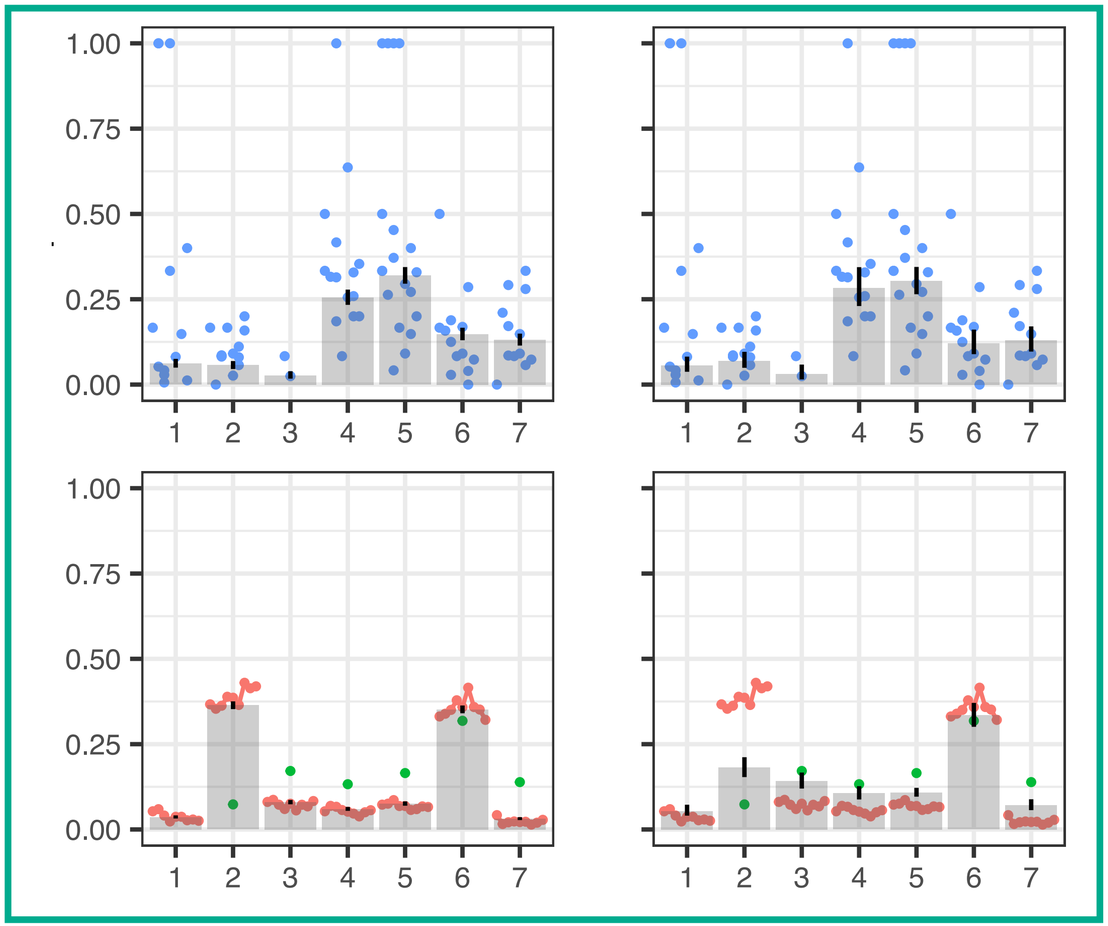
low coverage

low quality data

Case studies: low quality data

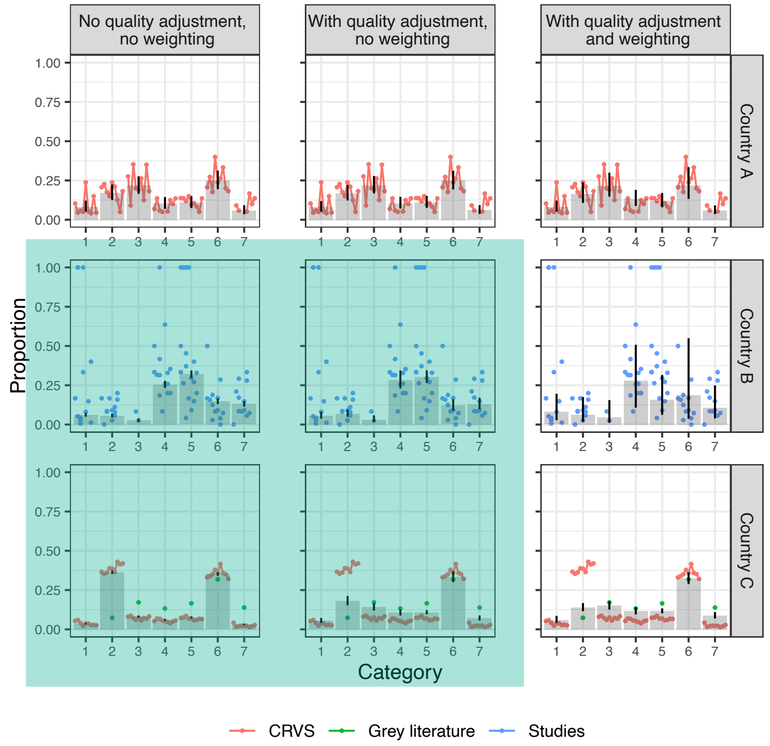
without **quality adjustment**

include **quality adjustment**



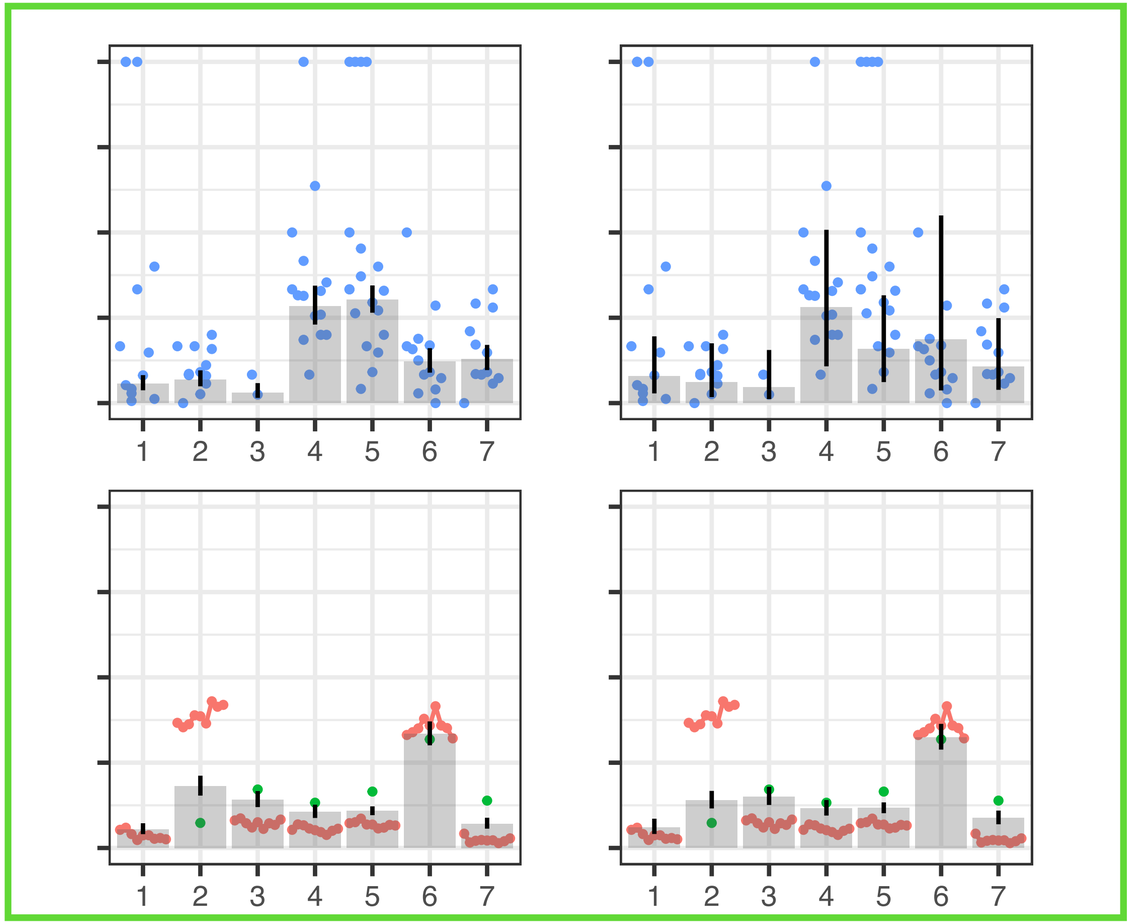
similar quality

low quality



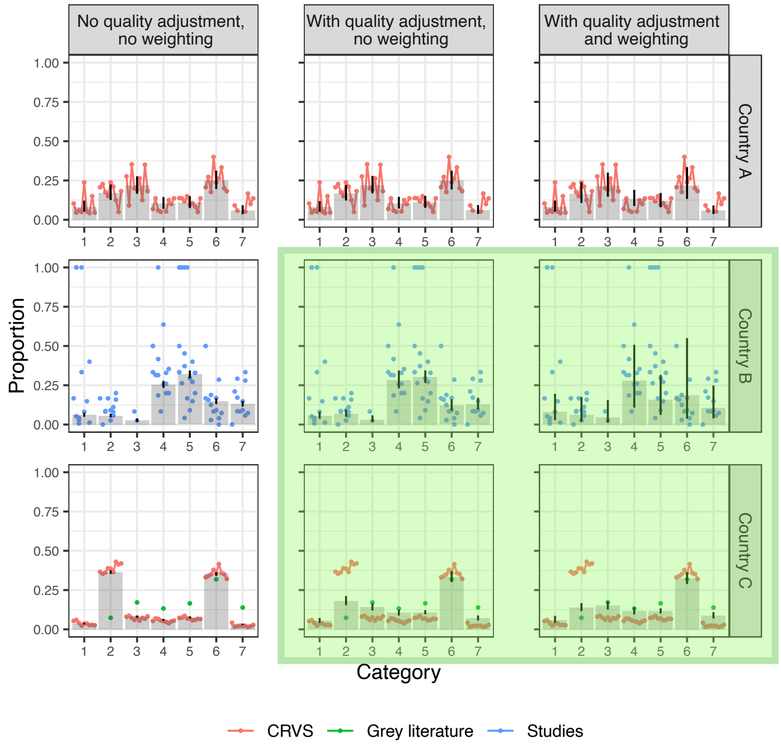
Case studies: low coverage

without **weighting** after **weighting**



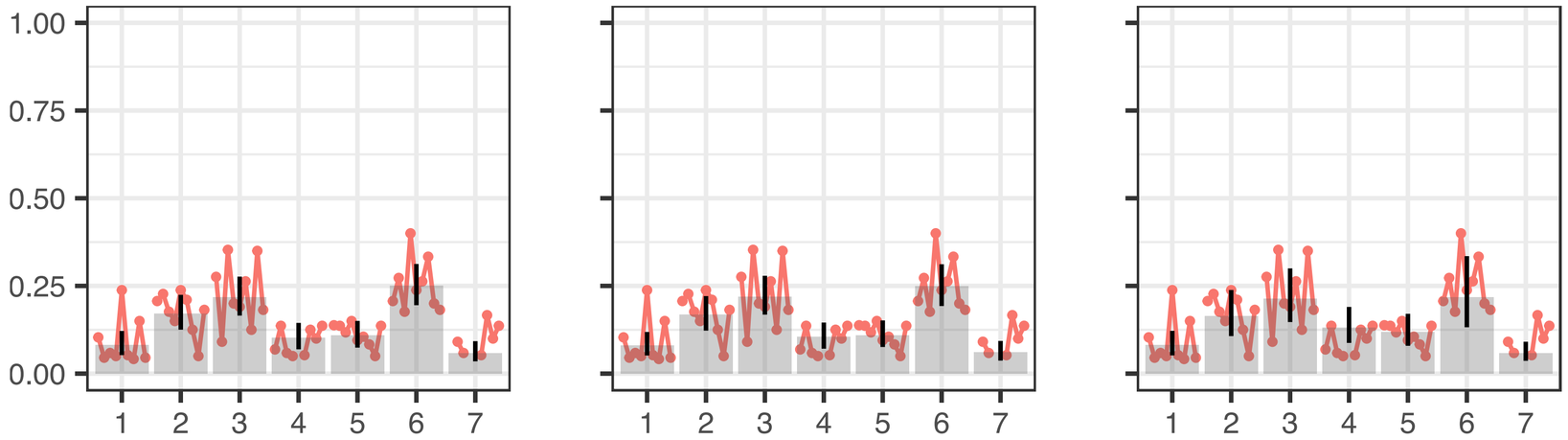
low coverage

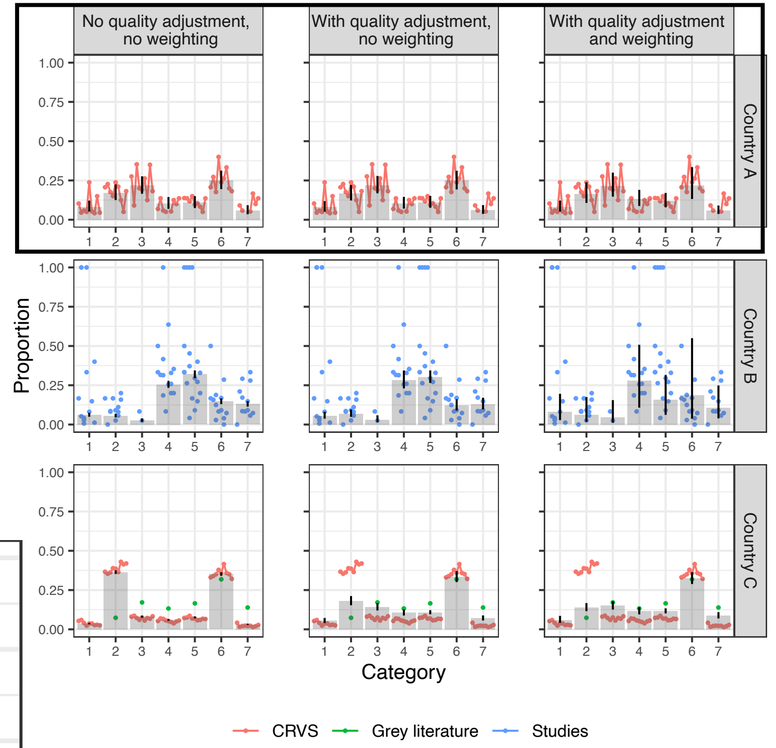
high coverage



Case studies: reliable data

high quality data, high coverage scenario





Summary

Presented a model framework to estimate cause of death distributions which:

- adjusts for variability from unreliable data
- reflects increased uncertainty in countries with low data coverage
- allows for arbitrary combinations of missing categories in observations

Acknowledgements

Thanks to Marija & Monica, and to Emma for organizing!