Bayesian hierarchical factor regression models to infer cause of death from verbal autopsy data

Kelly Moran, Amy Herring, David Dunson, Liz Turner

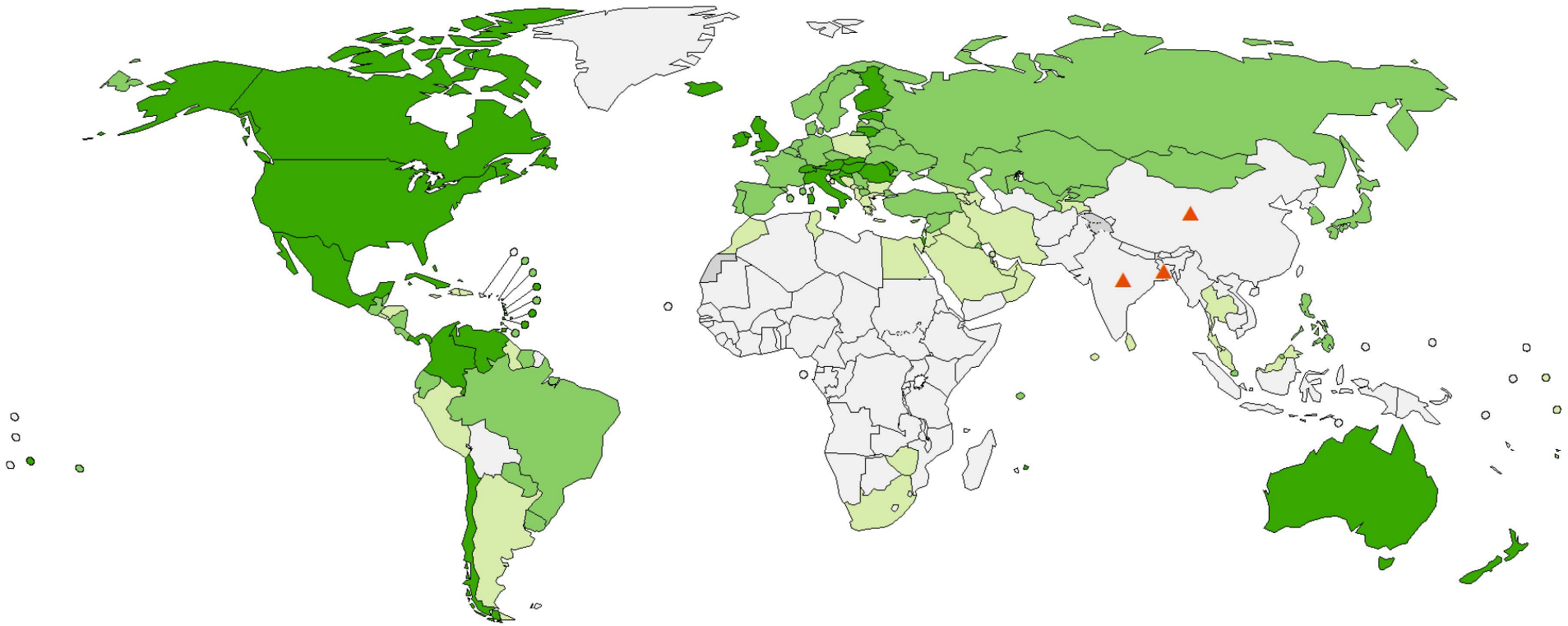
Duke University

kelly.r.moran@duke.edu

October 4, 2019

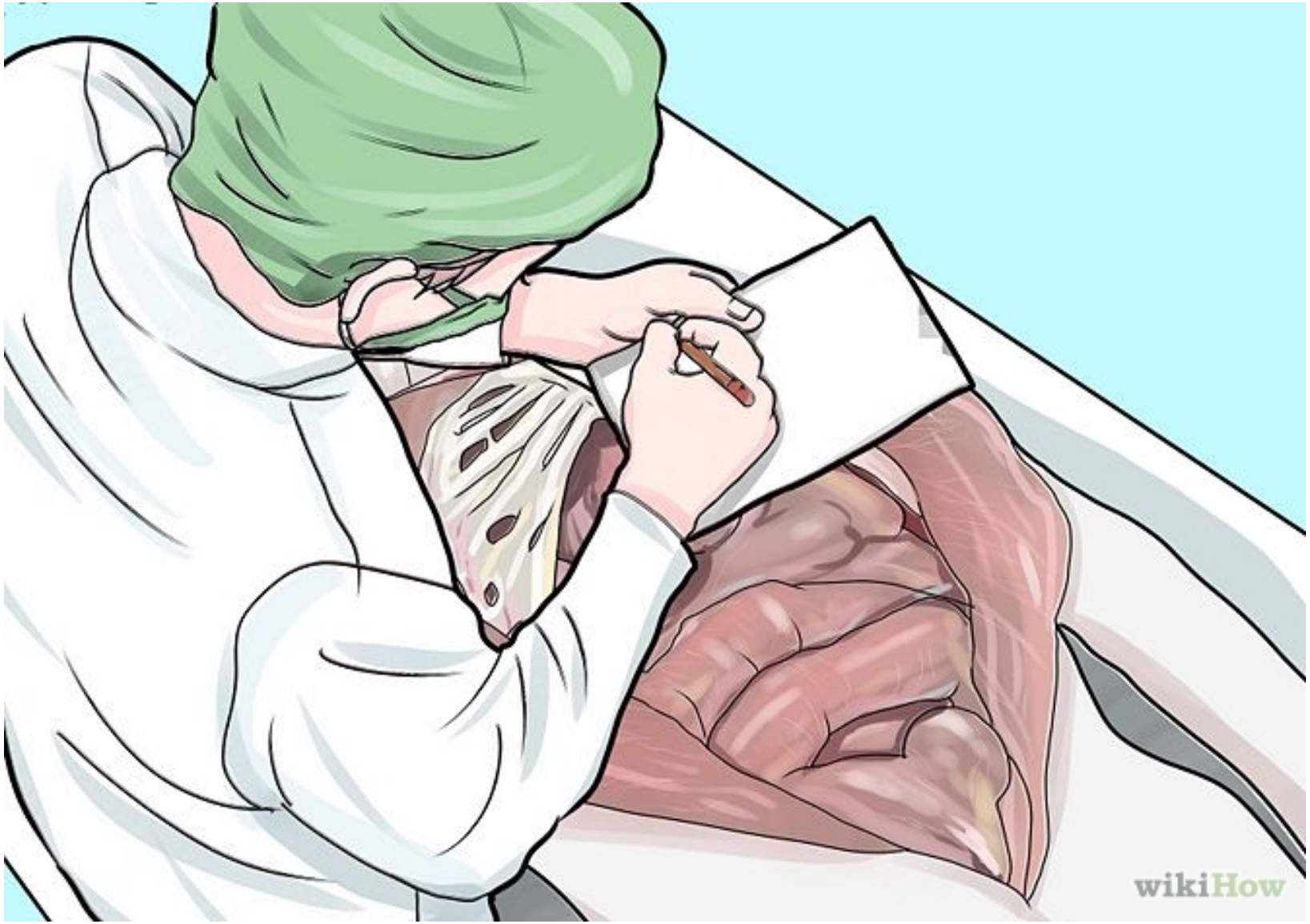
In the developing world, what are people dying from?

Cause-of-death information by country, 2014



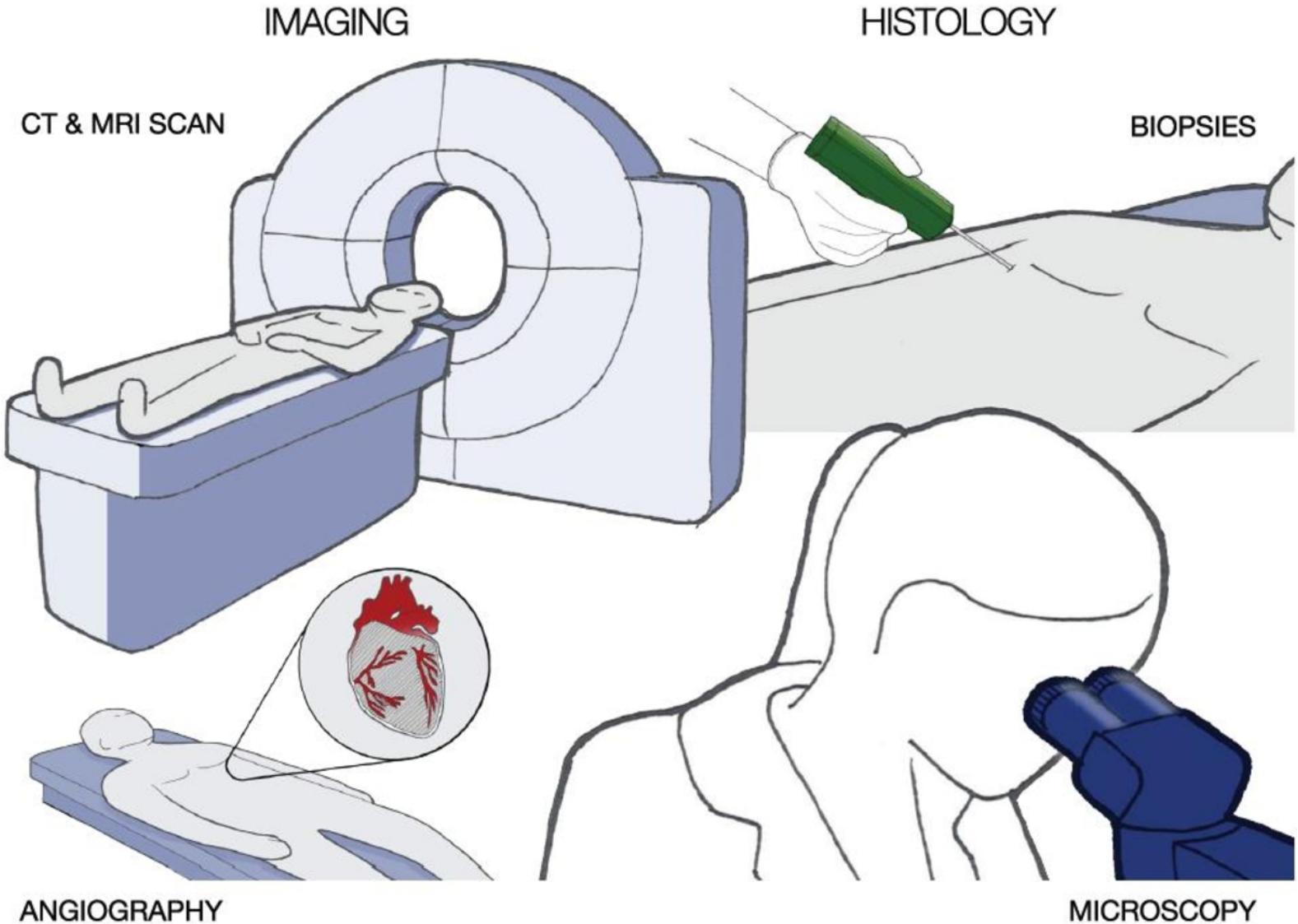
*From Nichols et al. (2018) "The WHO 2016 verbal autopsy instrument: An international standard suitable for automated analysis"

Methods: Full autopsy



*From https://www.wikihow.com/Perform-an-Autopsy-on-a-Human-Being.

Methods: Minimally invasive autopsy



*From "How to implement a Minimally Invasive Autopsy (MIA) procedure in a hospital setting; a practical guideline for radiologists".

Methods: Verbal autopsy

*From https://www.unfpa.org/fr/node/13319.

Verbal autopsy framework

The **verbal autopsy** (VA) is "a protocolised procedure that allows the classification of causes of death through analysis of data derived from structured interviews with family, friends, and caregivers."

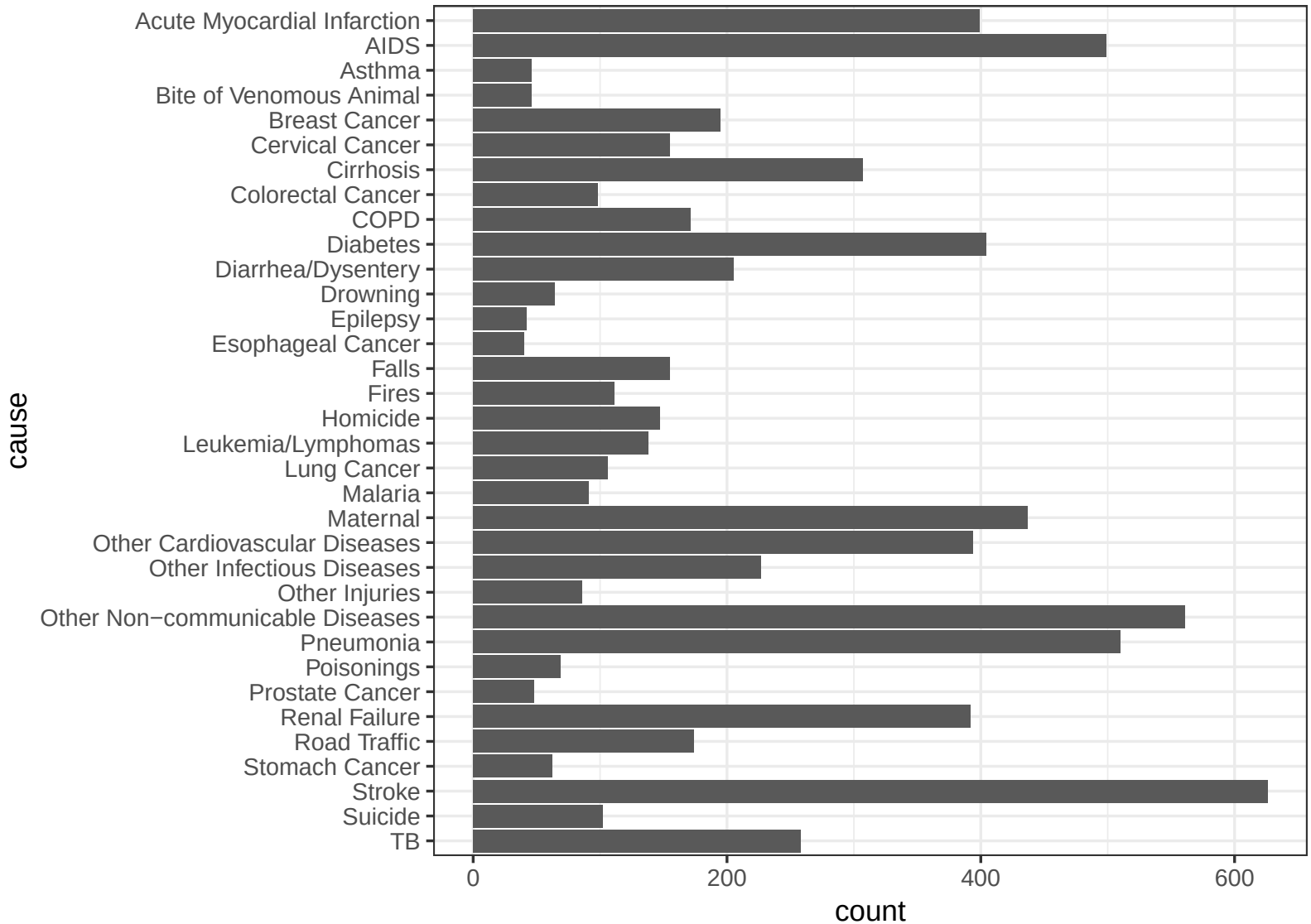
*From Bassat et al. (2013) "Development of a post-mortem procedure to reduce the uncertainty regarding causes of death in developing countries"

PHMRC data

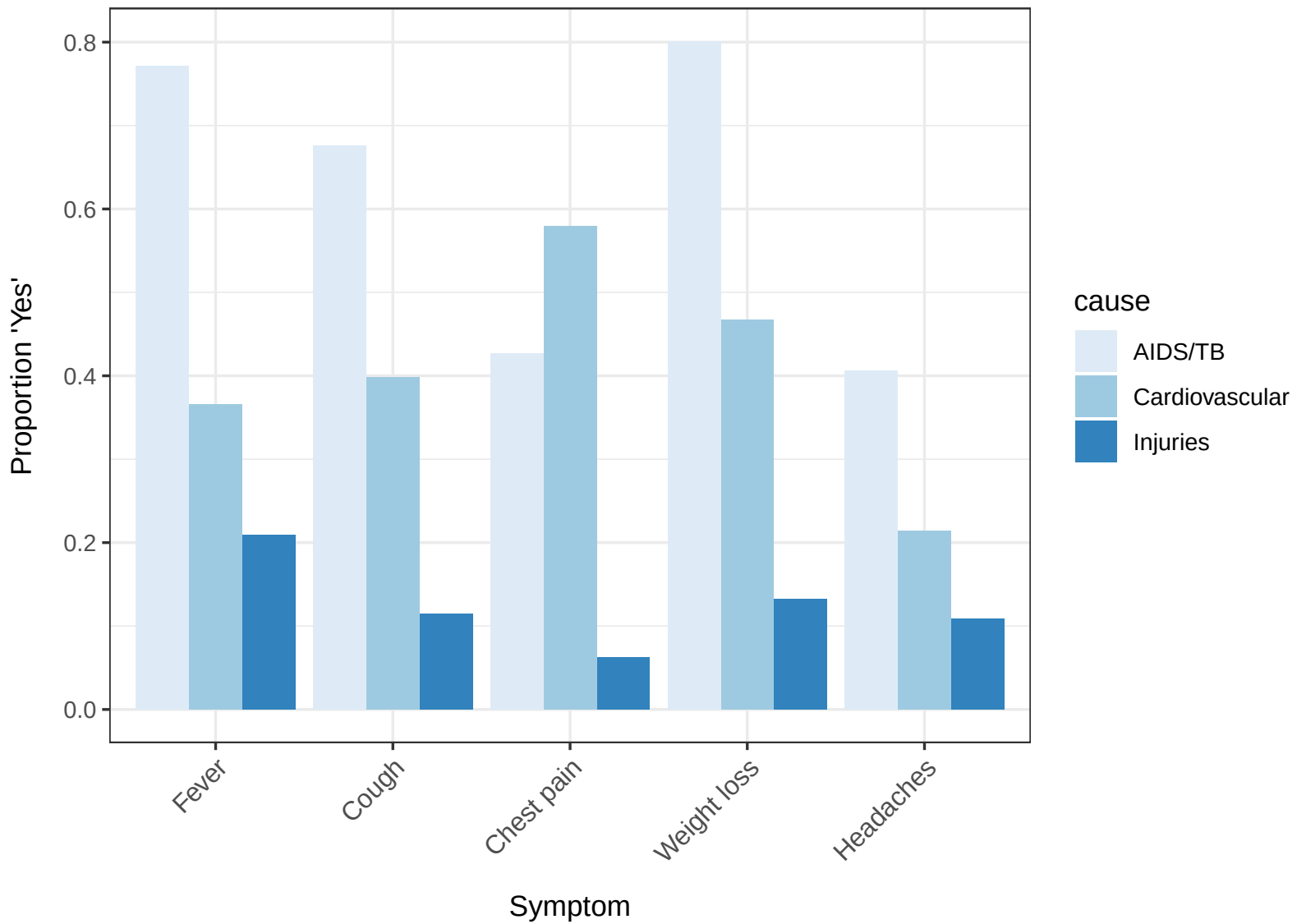
The Population Health Metrics Research Consortium (PHMRC) created a "Gold Standard" VA database for training/testing VA models.

- Includes 7,836 adults, for whom the broad list of causes for analysis number 34 and ≈ 200 symptoms are commonly included in analyses
- Data collected from 2007-2010 across six sites in four countries
- Questions include binary, numeric, categorical, and narrative; e.g.:
  - Did (s)he have breathlessness?
  - For how many days did (s)he have breathlessness?
  - During the illness that led to death did his/her breathing sound like any of the following: [stridor/grunting/wheezing]?

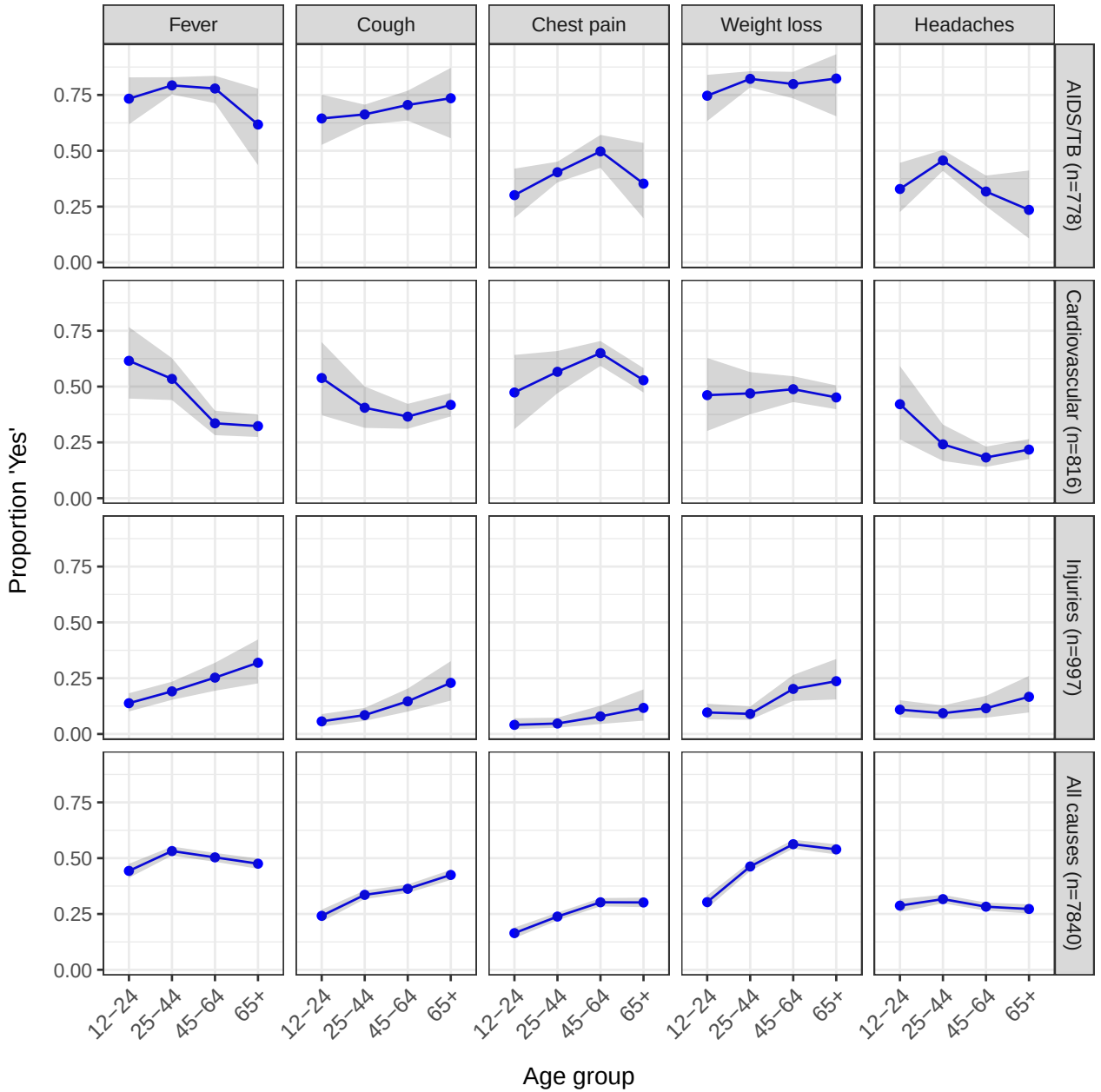
Data: Causes


cause
Acute Myocardial Infarction
AIDS
Asthma
Bite of Venomous Animal
Breast Cancer
Cervical Cancer
Cirrhosis
Colorectal Cancer
COPD
Diabetes
Diarrhea/Dysentery
Drowning
Epilepsy
Esophageal Cancer
Falls
Fires
Homicide
Leukemia/Lymphomas
Lung Cancer
Malaria
Maternal
Other Cardiovascular Diseases
Other Infectious Diseases
Other Injuries
Other Non−communicable Diseases
Pneumonia
Poisonings
Prostate Cancer
Renal Failure
Road Traffic
Stomach Cancer
Stroke
Suicide
TB
0
200
400
600
count

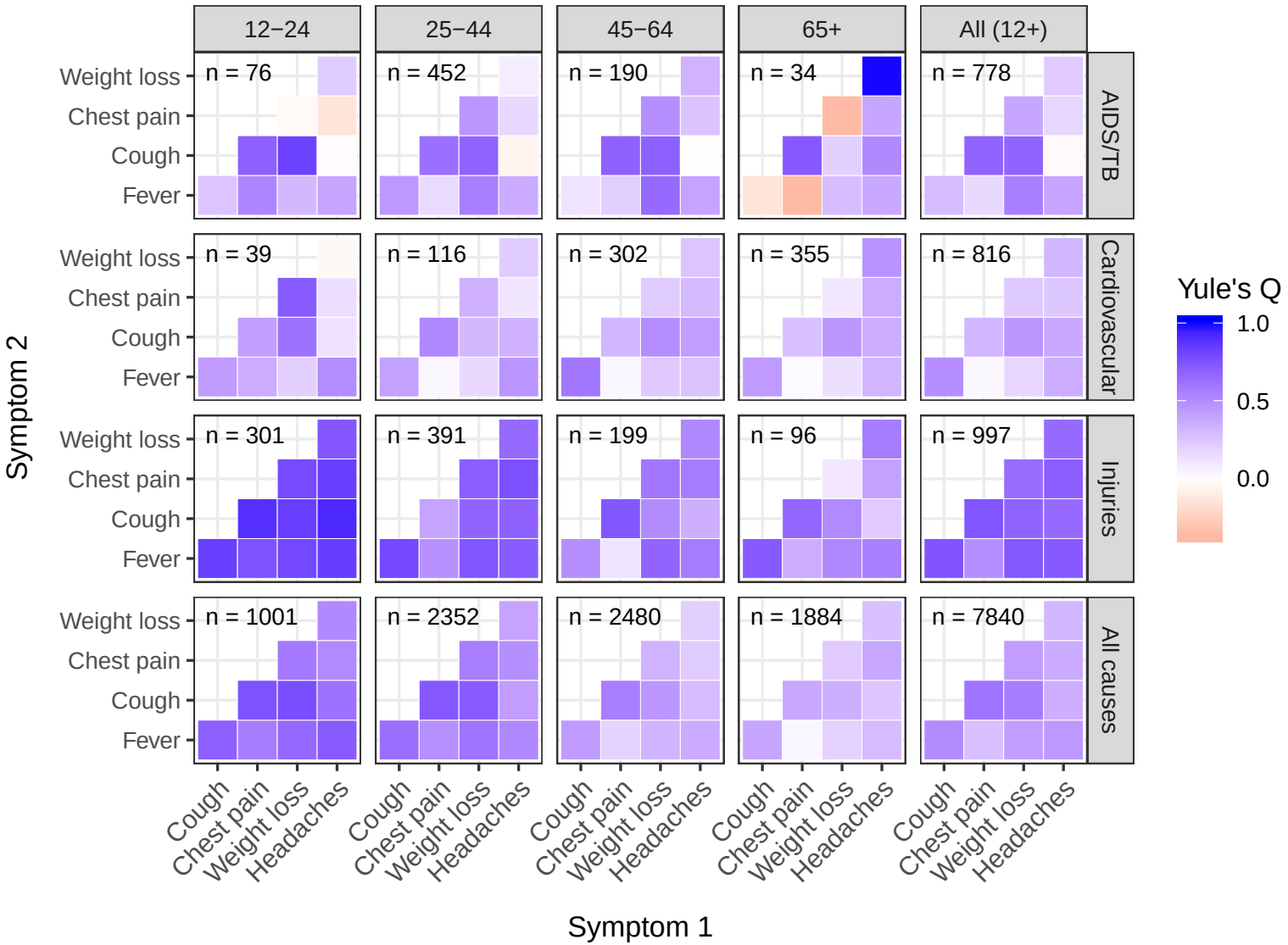
Data: Symptoms by cause



Data: Covariate dependence in symptom prevalence

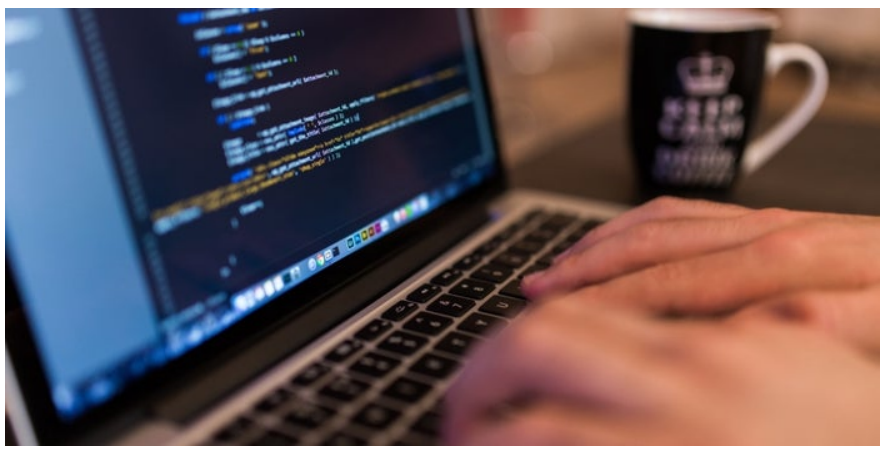


Data: Covariate dependence in symptom association



Analyzing verbal autopsy data

- Physician coding
  - Expensive
  - Not reproducible
  - Relies on expert judgment
- Computer coding
  - Inexpensive
  - (Can be) reproducible
  - Relies on algorithms, training data, and/or expert judgment



Modeling goals

- Model symptom mean and association by cause of death
- Share information across causes (via hierarchical modeling)
- Allow both the conditional (on cause) mean and the conditional association between symptoms to vary with covariates
- Probabilistically predict cause of death for an individual given their symptoms
- Improve on cause of death (COD) and cause-specific mortality fraction (CSMF) estimation relative to state-of-the-art VA algorithms

Model structure

The goal is to learn the cause of death y_i given symptoms $\boldsymbol{s}_i$.

$$\pi(y_i = c|\boldsymbol{s}_i) = \frac{\pi(\boldsymbol{s}_i|y_i = c)\pi(y_i = c)}{\Sigma_{h=1}^{C}\pi(\boldsymbol{s}_i|y_i = h)\pi(y_i = h)}, i = 1 \dots N.$$

Model structure

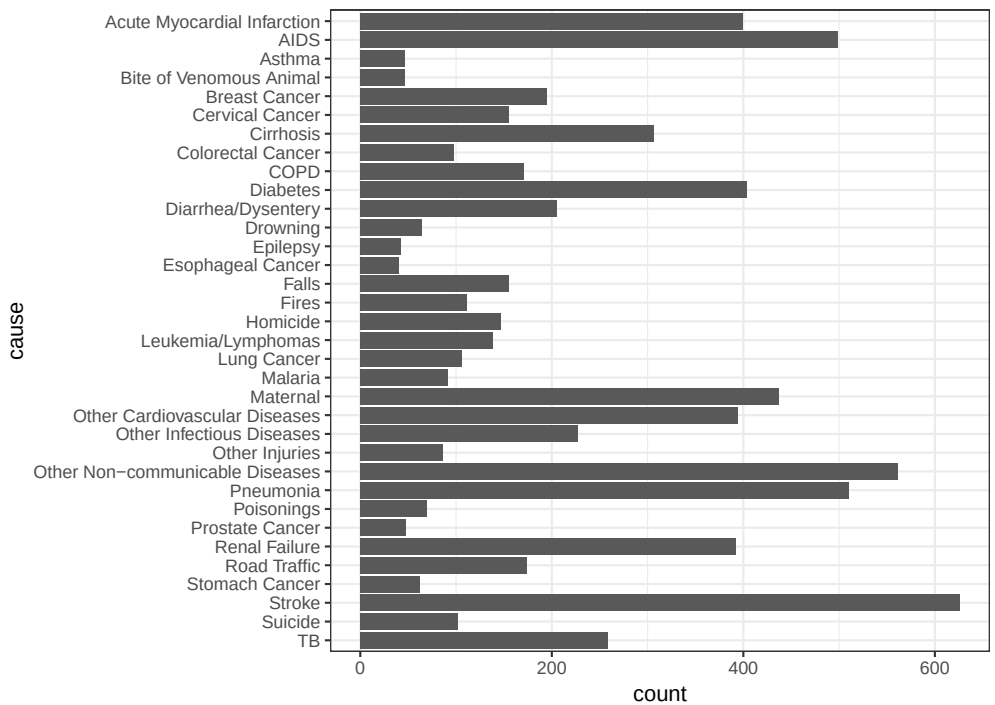
The goal is to learn the cause of death y_i given symptoms $\boldsymbol{s}_i$.

$$\pi(y_i = c|\boldsymbol{s}_i) = \frac{\pi(\boldsymbol{s}_i|y_i = c)\pi(y_i = c)}{\sum_{h=1}^{C}\pi(\boldsymbol{s}_i|y_i = h)\pi(y_i = h)}, i = 1 \dots N.$$

Prior over causes

$$\{\Pr(y_i = 1), \ldots, \Pr(y_i = C)\} \sim \text{Dirichlet}(a_1, ..., a_C)$$

Under the assumption that little is known about the CSMF in the region of interest, but that the distribution of deaths across causes is non-uniform, set $a_1 = \ldots = a_C < 1$.



Model structure

Recall the goal is to learn the cause of death y_i given symptoms $\boldsymbol{s}_i$.

$$\pi(y_i = c|\boldsymbol{s}_i) = \frac{\pi(\boldsymbol{s}_i|y_i = c)\pi(y_i = c)}{\sum_{h=1}^{C}\pi(\boldsymbol{s}_i|y_i = h)\pi(y_i = h)}, i = 1 \ldots N.$$

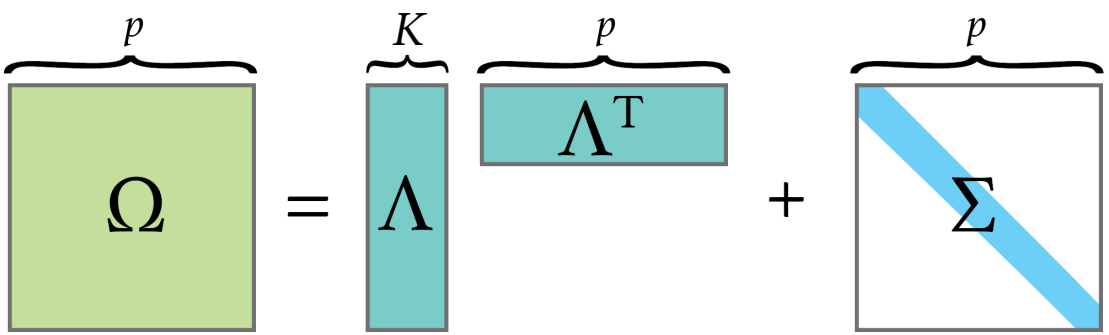
Likelihood of symptoms given cause

Introduce a factor model to account for correlation between symptoms.

- The traditional factor model is:

$$\begin{aligned} \boldsymbol{z}_i = \Lambda \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad & \boldsymbol{\eta}_i \sim \mathsf{N}(0, I_K), \\ & \boldsymbol{\epsilon}_i \sim \mathsf{N}(\mathbf{0}_p, \Sigma_0), \quad \Sigma_0 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2) \\ & i = 1, \ldots, N. \end{aligned}$$

- The prior induced on the latent z_i by integrating out the unknown $\boldsymbol{\eta}_i$ is then $\boldsymbol{z}_i | y_i \sim \mathsf{N}(\mathbf{0}_p, \Lambda\Lambda' + \Sigma_0)$.



Symptoms aren't continuous

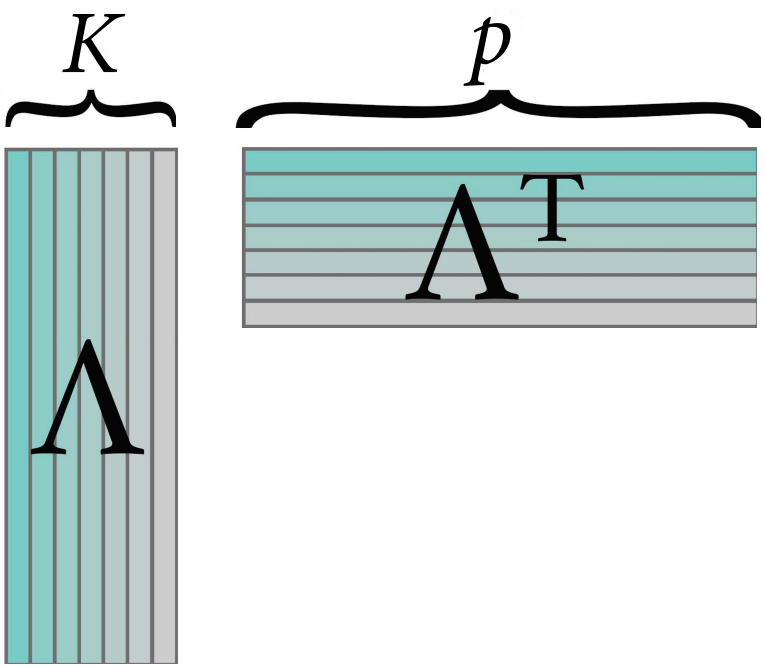
In order to allow this framework to encompass symptoms of mixed type, define z_{ij} to be continuous 'latent symptoms.'

Then assume $s_{ij} = f_j(z_{ij})$, $j = 1, \ldots, p$, where $f_j()$ depends on the symptom.

- E.g., for binary s_{ij} such as "Did the decedent have a fever?", $f_j(z_{ij}) = 1(z_{ij} > 0)$.
- E.g., for continuous s_{ij} such as "What was the decedent's highest temperature?", $f_j(z_{ij}) = z_{ij}$.
- E.g., for count type symptoms such as "For how many days did the decedent have a fever?", $f_j(z_{ij})$ can be defined using nonparametric Bayes count process models [Canale and Dunson (2013)].

Number of factors isn't known

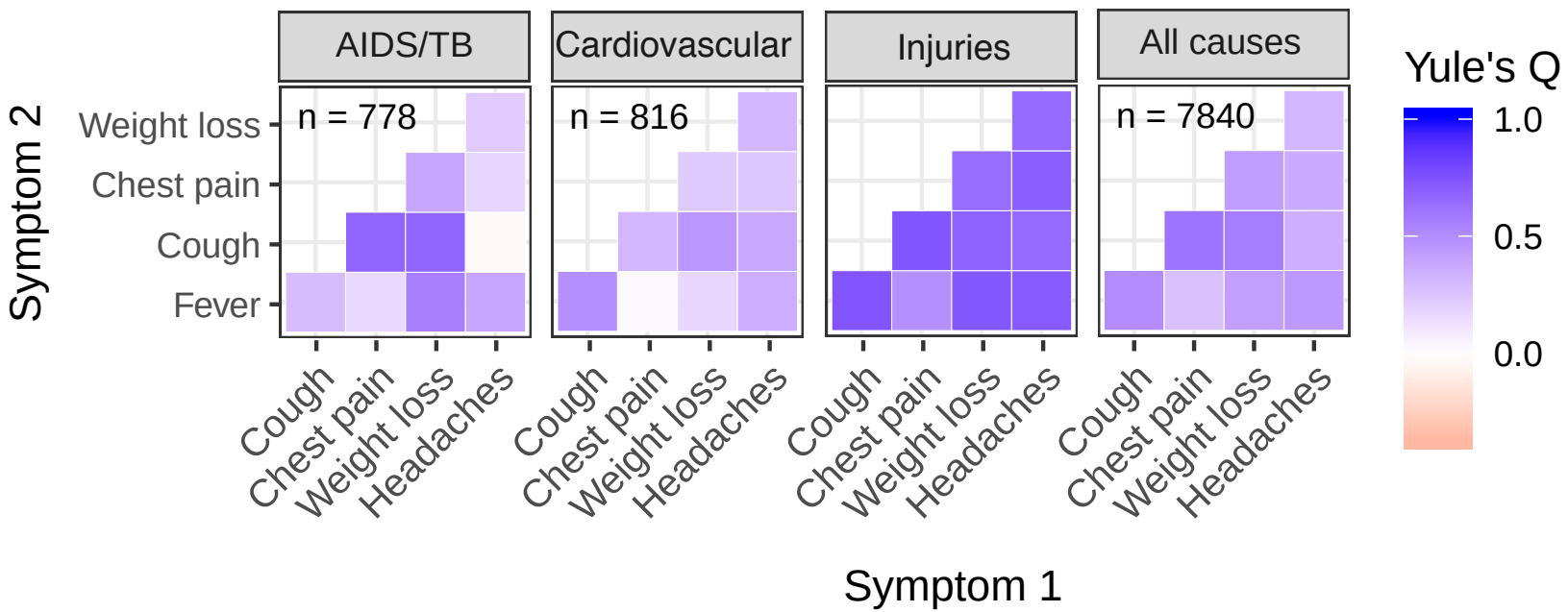
Implement stochastic shrinkage on columns of the $p \times K$ loadings matrix Λ [Bhattacharya and Dunson (2011)] so number of factors K can be learned.



Cause-specific symptom covariance

Allow covariance between latent symptoms to depend on cause of death, as in the Bayesian factor model of Kunihama et al. (2018).

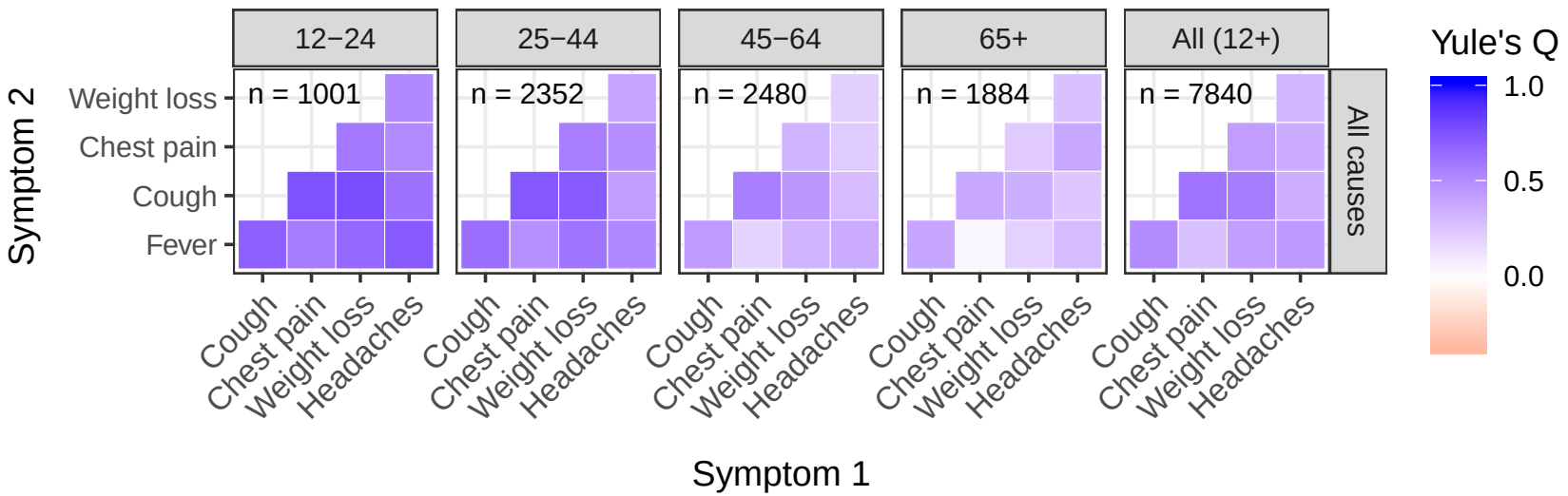
$$\boldsymbol{z}_i = \Lambda_{y_i}\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim \mathrm{N}(\mathbf{0}_K, I_K), \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}_p, \Sigma_0),$$
$$\boldsymbol{z}_i | y_i \sim \mathrm{N}(\mathbf{0}_p, \Lambda_{y_i}\Lambda_{y_i}{}' + \Sigma_0).$$



Covariate-dependent hierarchical symptom covariance

Allow covariance between latent symptoms to vary with covariates $\boldsymbol{x}_i$ as in the covariance regression of Fox and Dunson (2015). Model hierarchically because many causes have few observed deaths.

$$\boldsymbol{z}_i = \Lambda_{y_i}(\boldsymbol{x}_i)\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim \mathrm{N}(\mathbf{0}_K, I_K), \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}_p, \Sigma_0),$$
$$\boldsymbol{z}_i | y_i \sim \mathrm{N}(\mathbf{0}_p, \Lambda_{y_i}(\boldsymbol{x}_i)\Lambda_{y_i}(\boldsymbol{x}_i)' + \Sigma_0).$$



Covariate-dependent hierarchical symptom covariance

Decompose $p \times K$ loadings matrix $\Lambda_{y_i}(\boldsymbol{x}_i)$ as in Fox and Dunson (2015):

$$\begin{aligned}\Lambda_{y_i}(\boldsymbol{x}_i) &= \Theta_{y_i}\boldsymbol{\xi}_{y_i}(\boldsymbol{x}_i),\\ \Theta_{y_i} &\in \mathbb{R}^{p\times L},\\ \boldsymbol{\xi}_{y_i}(\boldsymbol{x}_i) &= \{\xi_{i,lk}(\boldsymbol{x}_i),\ l = 1,\ldots,L,\ k = 1,\ldots,K\}.\end{aligned}$$

Column-specific shrinkage via Battacharya and Dunson (2011), info shared across causes via common mean Δ_{jl}. For $j = 1,\ldots,p, l = 1,\ldots,L$:

$$\theta_{y_i,jl} \sim \mathrm{N}(\Delta_{jl}, \phi_{\Theta,jl}^{-1}\tau_{\Theta,l}^{-1}), \quad \phi_{\Theta,jl} \sim \mathrm{Ga}(\gamma_\Theta/2, \gamma_\Theta/2), \quad \tau_{\Theta,l} = \prod_{h=1}^{l}\delta_{\Theta,h}.$$

Set $\boldsymbol{\xi}_{y_i,lk}(\boldsymbol{x}_i)$ with a hierarchical linear model, $l = 1,\ldots,L, k = 1,\ldots,K$:

$$\xi_{y_i,lk}(\boldsymbol{x}_i) = \boldsymbol{\beta}_{y_i,lk}^T\boldsymbol{x}_i, \quad \boldsymbol{\beta}_{y_i,lk} \sim \mathrm{N}_B(\boldsymbol{\mu}_{\beta_{lk}}, \Sigma_{\beta_{lk}}).$$

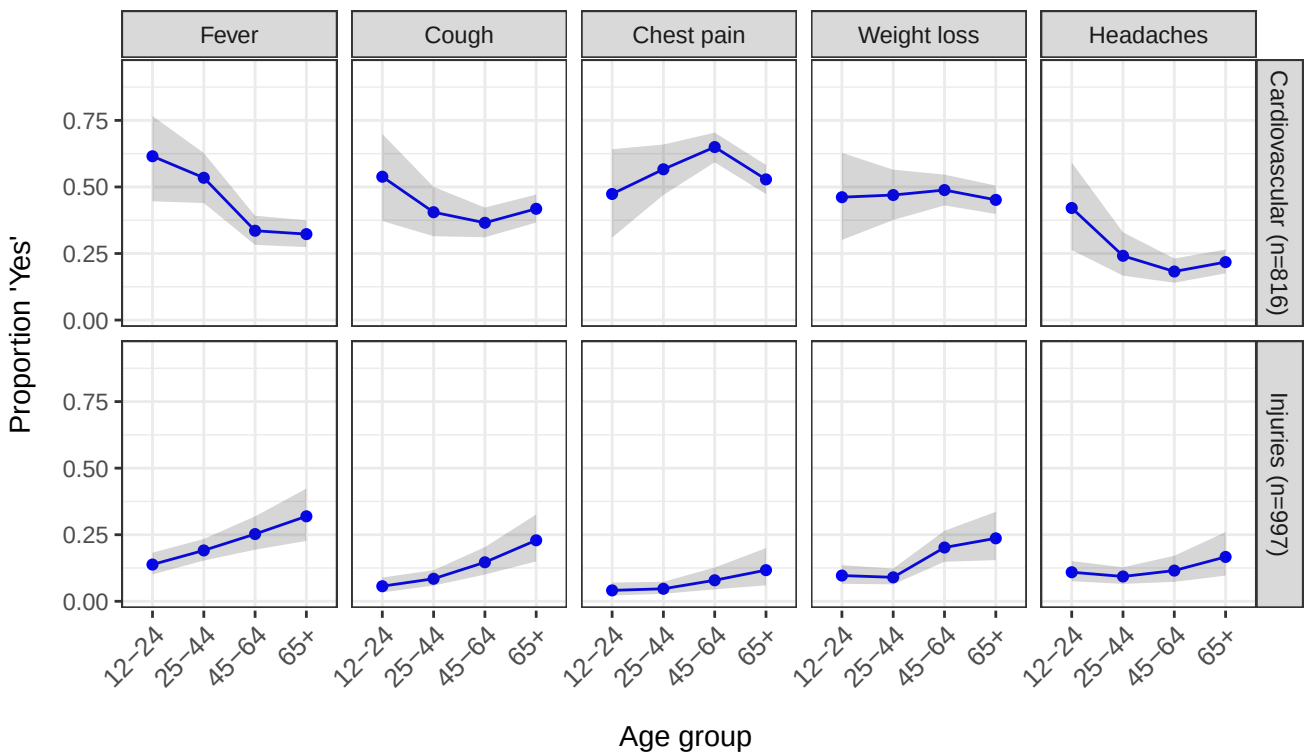
Cause-specific covariate-dependent symptom mean

Allow cause-specific and covariate-dependent latent symptom mean.
Model hierarchically because many causes have few observed deaths.

$$\boldsymbol{z}_i = \Lambda_{y_i}(\boldsymbol{x}_i)\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim \mathrm{N}(\boldsymbol{\psi}_{y_i}(\boldsymbol{x}_i), I_K), \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}_p, \Sigma_0),$$

$$\boldsymbol{z}_i | y_i \sim \mathrm{N}(\Lambda_{y_i}(\boldsymbol{x}_i)\boldsymbol{\psi}_{y_i}(\boldsymbol{x}_i), \Lambda_{y_i}(\boldsymbol{x}_i)\Lambda_{y_i}(\boldsymbol{x}_i)' + \Sigma_0),$$

$$\psi_{y_i,k}(\boldsymbol{x}_i) = \boldsymbol{\alpha}_{y_i,k}^T \boldsymbol{x}_i, \quad \boldsymbol{\alpha}_{y_i,k} \sim \mathrm{N}_B(\boldsymbol{\mu}_{\alpha_k}, \Sigma_{\alpha_k}), \quad k = 1, \ldots, K.$$



Determining COD for new observations

For person $i^* \in U^*$, where U^* denotes the group of individuals having unknown COD, calculate

$$\pi(y_{i^*} = c|\boldsymbol{s}_{i^*}) = \frac{\pi(\boldsymbol{s}_{i^*}|y_{i^*} = c)\pi(y_{i^*} = c)}{\sum_{c'=1}^{C}\pi(\boldsymbol{s}_{i^*}|y_{i^*} = c')\pi(y_{i^*} = c')}$$

for each potential cause c, and sample from the resulting discrete distribution.

Then compute the population distribution of causes for individuals in U^*:

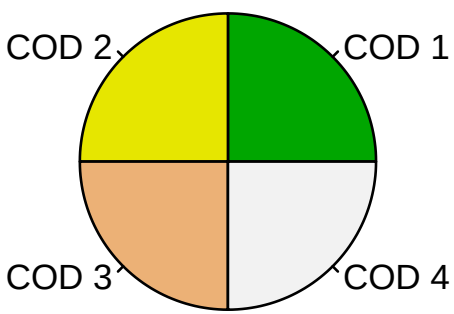
$$\text{CSMF}_{U^*} = \left(\frac{1}{|U^*|}\sum_{i^* \in U^*} 1(y_{i^*} = 1), \ldots, \frac{1}{|U^*|}\sum_{i^* \in U^*} 1(y_{i^*} = C)\right).$$

Simulation setup

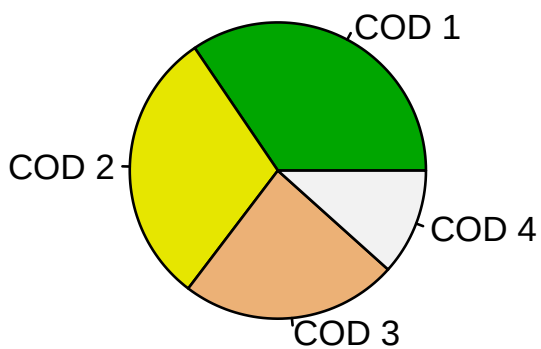
The goal is to mimic a scaled down version of PHMRC data and PHMRC-based simulation studies.

- $C = 4$ "causes"
- $N = 928$ "deaths" (75% train, 25% test)
- $P = 21$ "symptoms"
- 50 simulated data sets per setting for this talk, 1000 in paper

Train CSMF

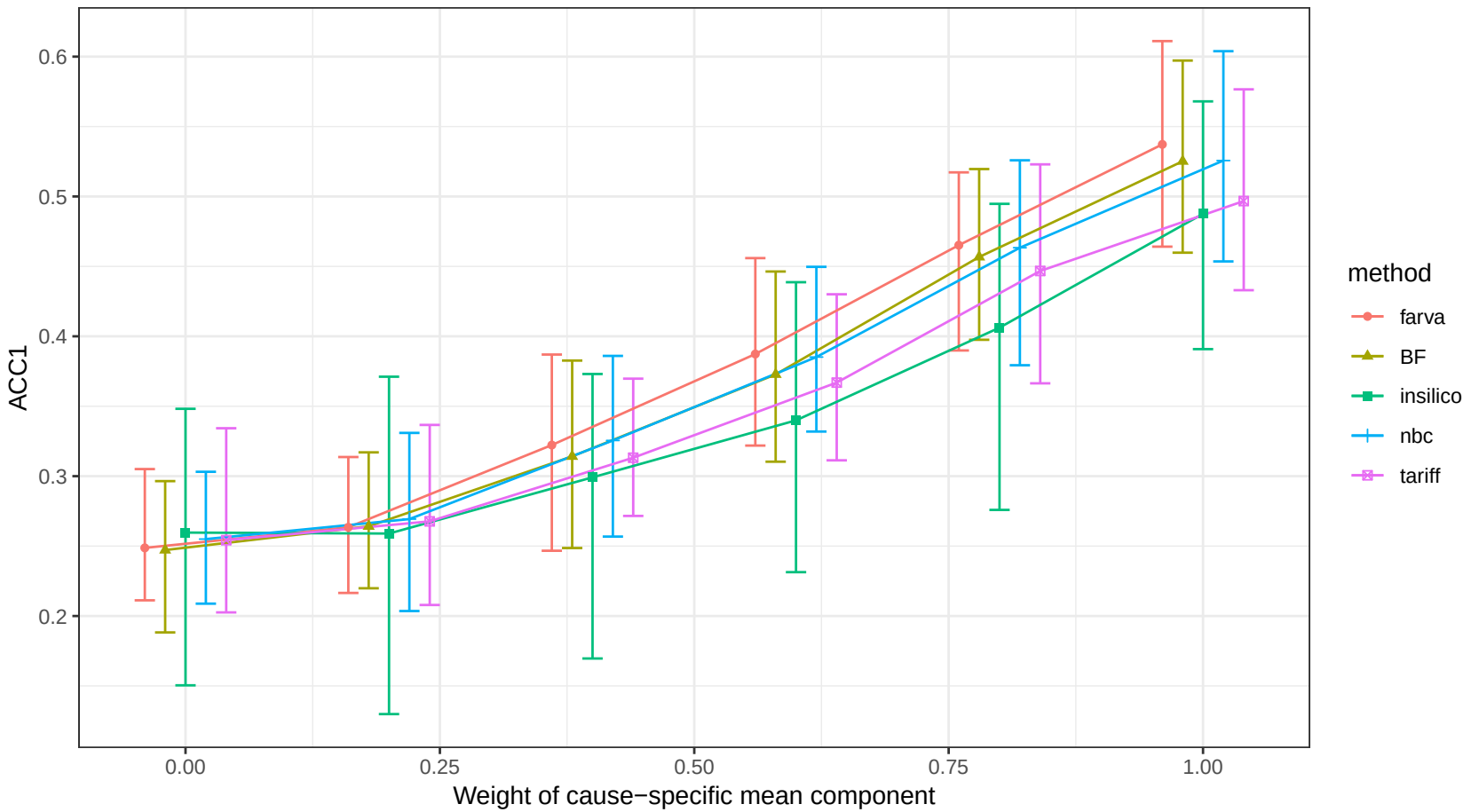


Test CSMF



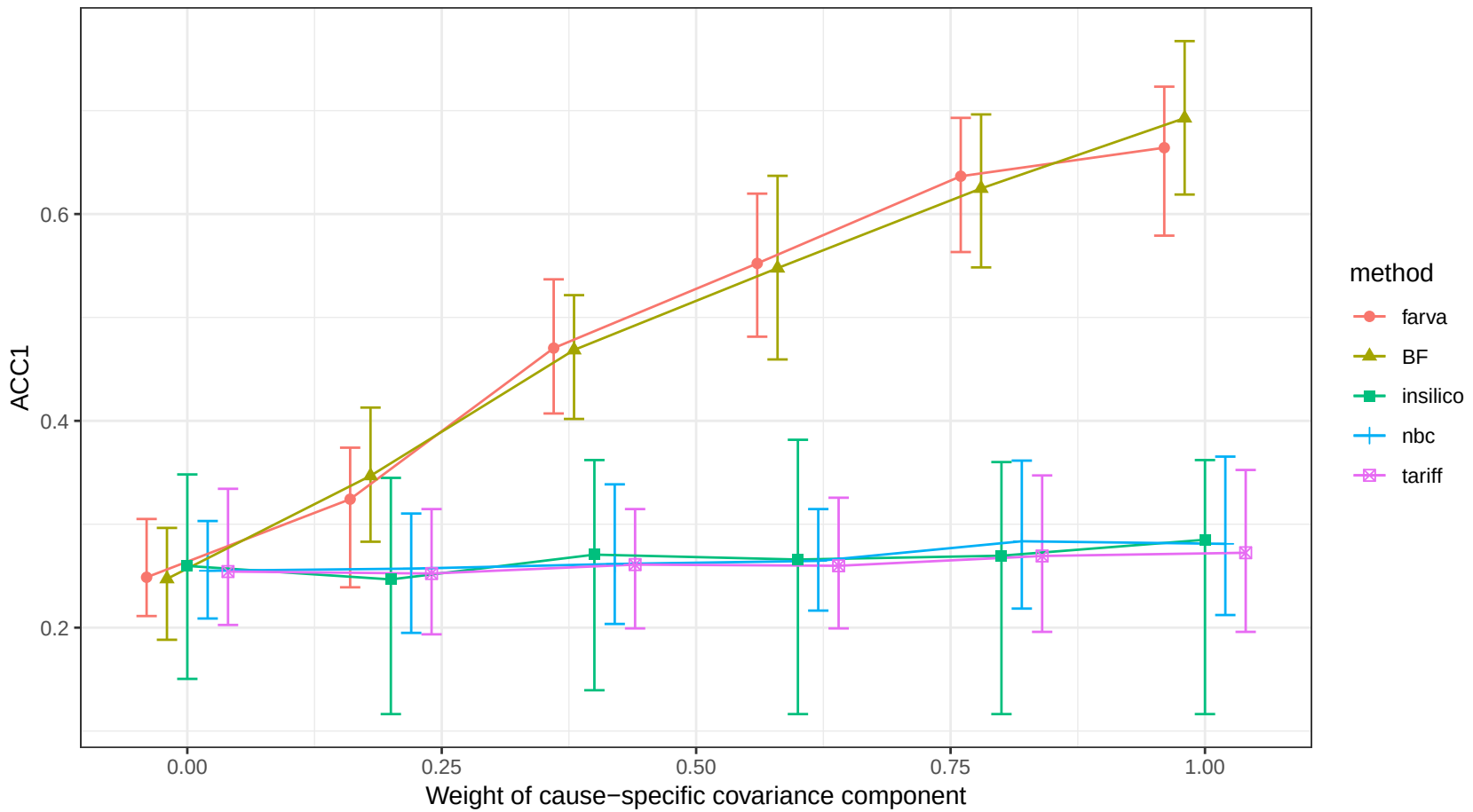
Simulation settings (cause specificity of mean)

- Mean structure comprised of some common component and some cause-specific component, with
  $\mu_{\text{cause } c} = (1 - w) \cdot m_{\text{common}} + w \cdot m_{\text{cause } c}$.
- Common covariance structure across causes.



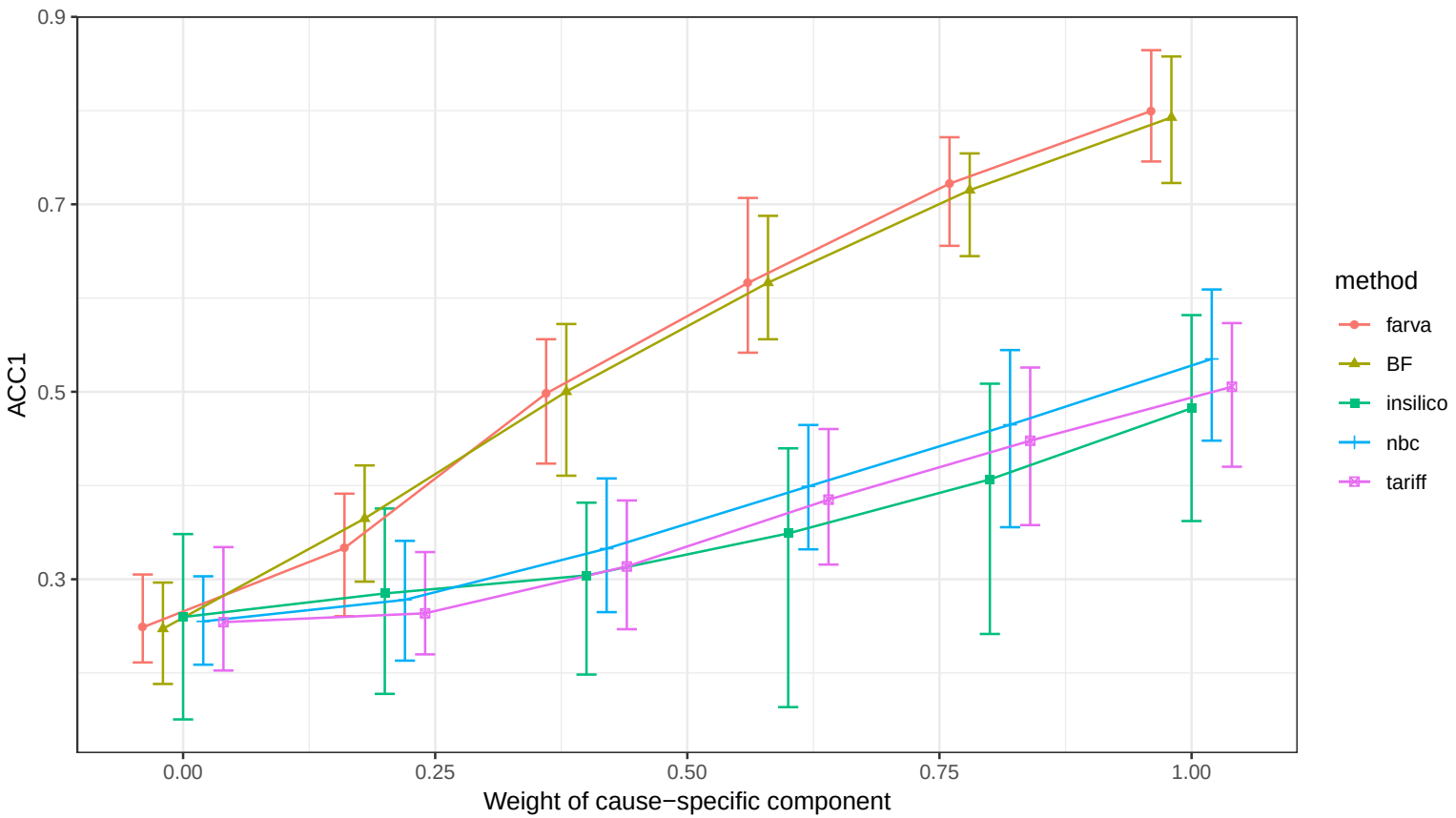
Simulation settings (cause specificity of covariance)

- Covariance structure comprised of some common component and some cause-specific component, with
  $\Sigma_{\text{cause } c} = (1 - w) \cdot s_{\text{common}} + w \cdot s_{\text{cause } c}$.
- Common mean structure across causes.



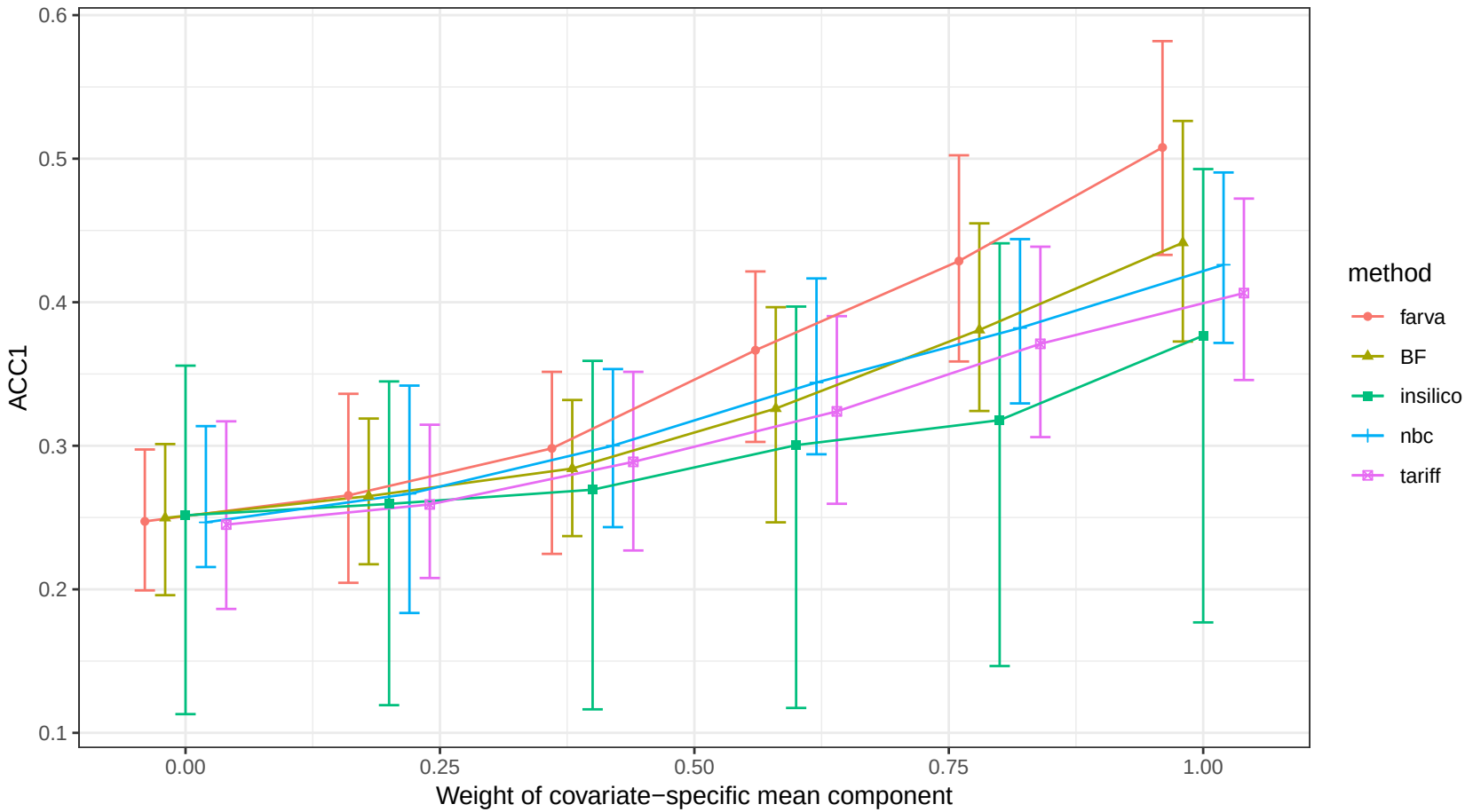
Simulation settings (cause specificity of mean and covariance)

- Mean and covariance structure comprised of some common component and some cause-specific component:
  $\mu_{\text{cause } c} = (1 - w) \cdot m_{\text{common}} + w \cdot m_{\text{cause } c}$ and
  $\Sigma_{\text{cause } c} = (1 - w) \cdot s_{\text{common}} + w \cdot s_{\text{cause } c}$



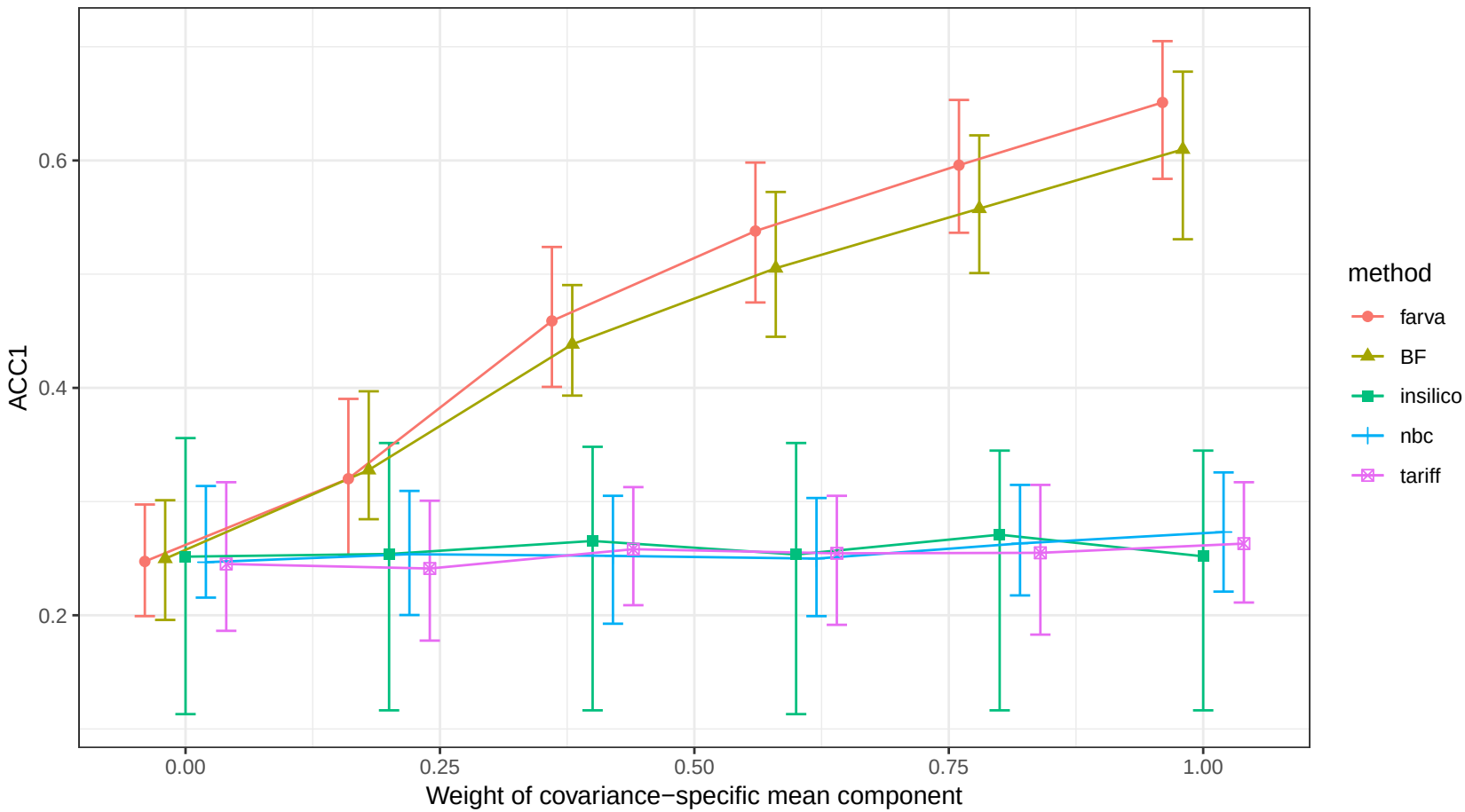
Simulation settings (group specificity of mean)

- Mean structure comprised of some common (cause-specific) component and some group-specific component, with $\mu_{\text{group-specific}} = (1 - w) \cdot m_{\text{common}} + w \cdot m_{\text{group}}$.
- Independent covariance structure shared across causes/groups.



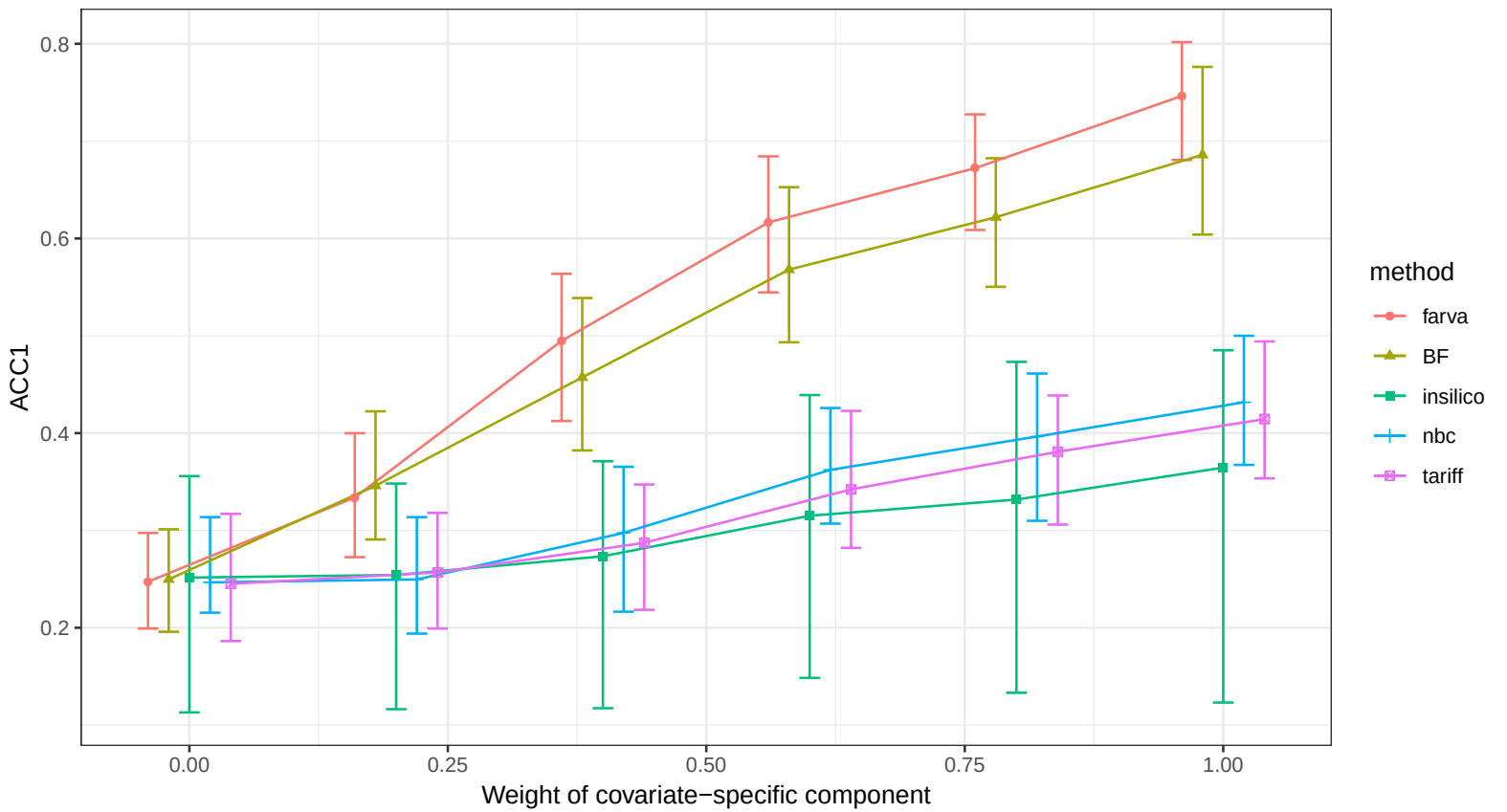
Simulation settings (group specificity of covariance)

- Covariance structure comprised of some common (cause-specific) component and some group-specific component, with $\Sigma_{\text{group-specific}} = (1 - w) \cdot s_{\text{common}} + w \cdot s_{\text{group}}$.
- Mean structure shared across causes/groups.



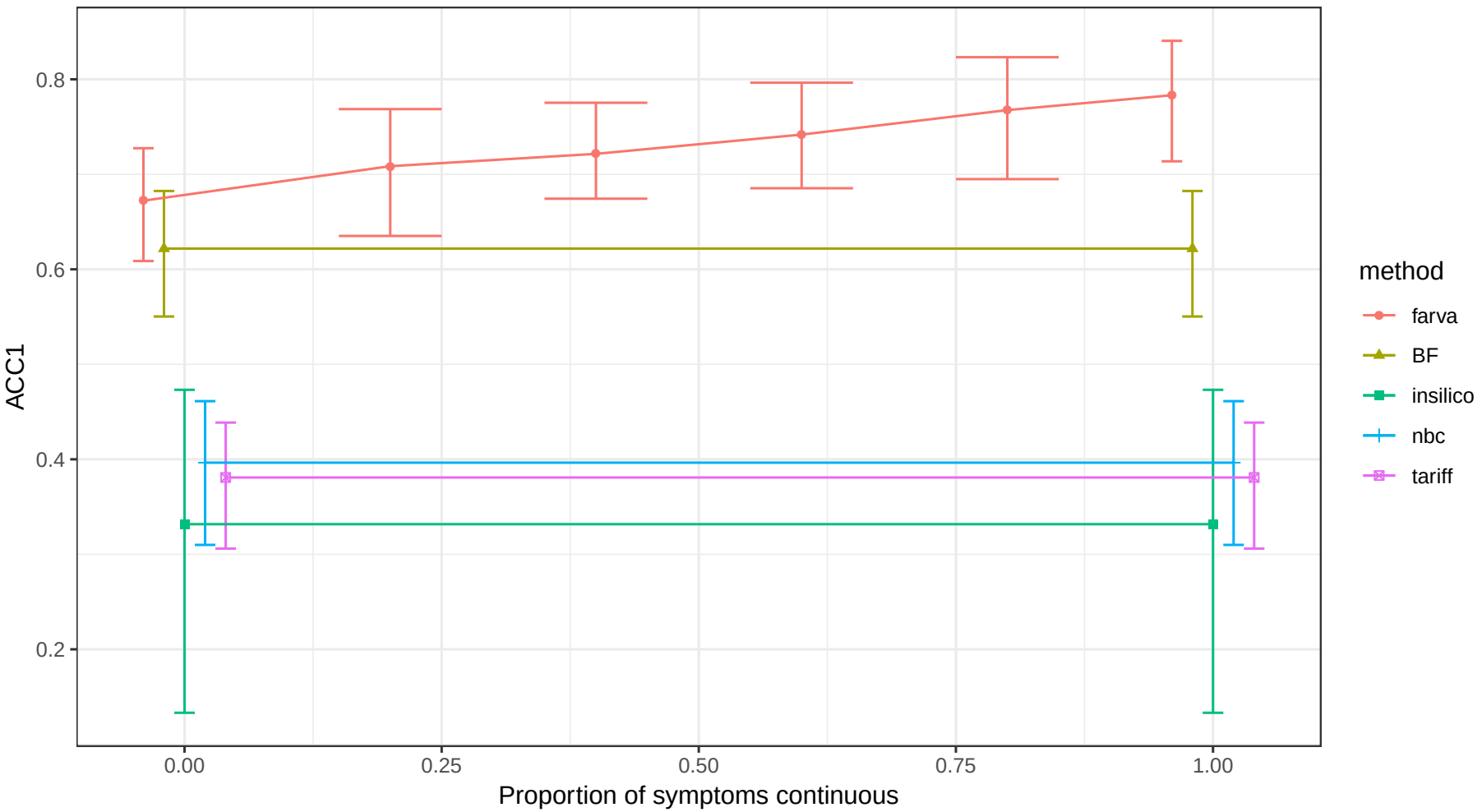
Simulation settings (group specificity of mean and covariance)

- Mean and covariance structure comprised of some common (cause-specific) component and some group-specific component:
  $\mu_{\text{group-specific}} = (1 - w) \cdot m_{\text{common}} + w \cdot m_{\text{group}}$ and
  $\Sigma_{\text{group-specific}} = (1 - w) \cdot s_{\text{common}} + w \cdot s_{\text{group}}$.



Simulation settings (proportion continuous data)

- Mean and covariance structure comprised of 20% common (cause-specific) component and 80% group-specific components.
- Proportion of data that is continuous is varied.

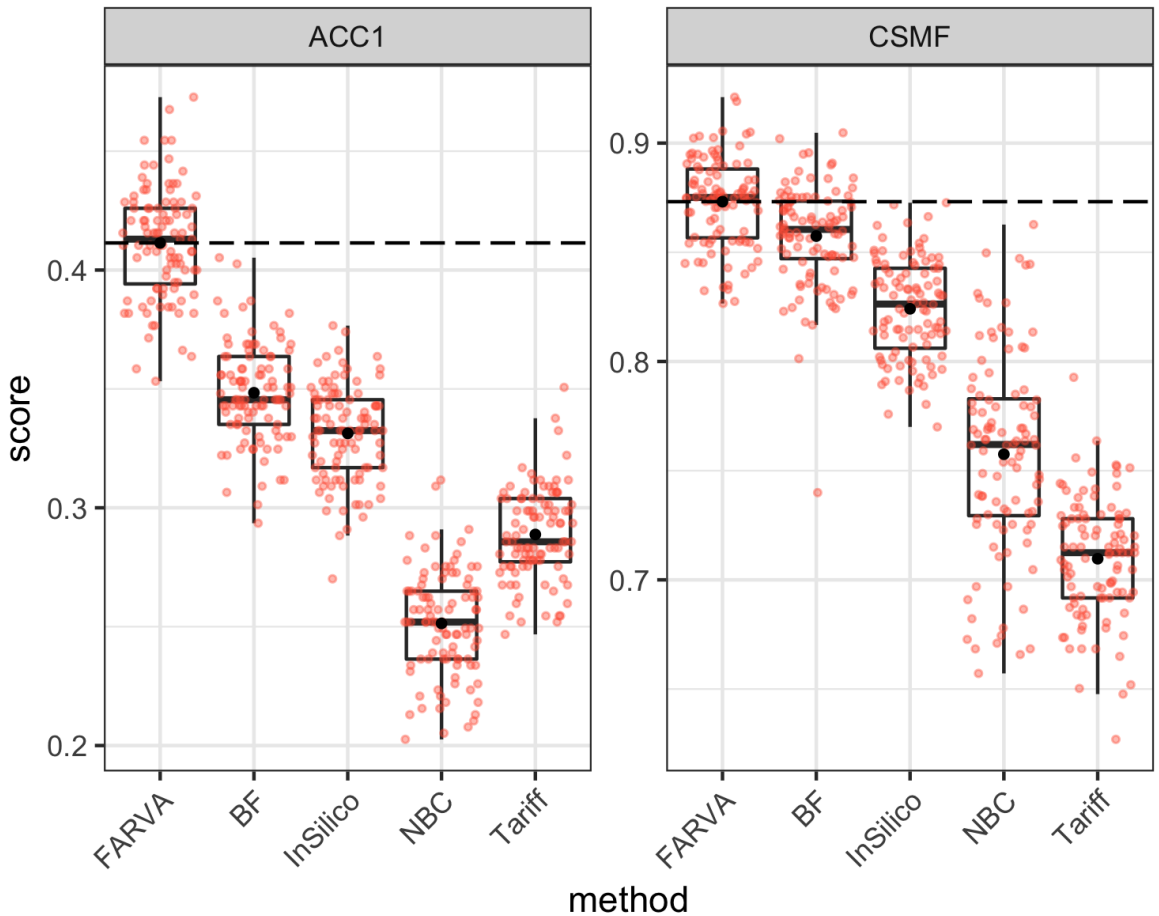


PHMRC data runs

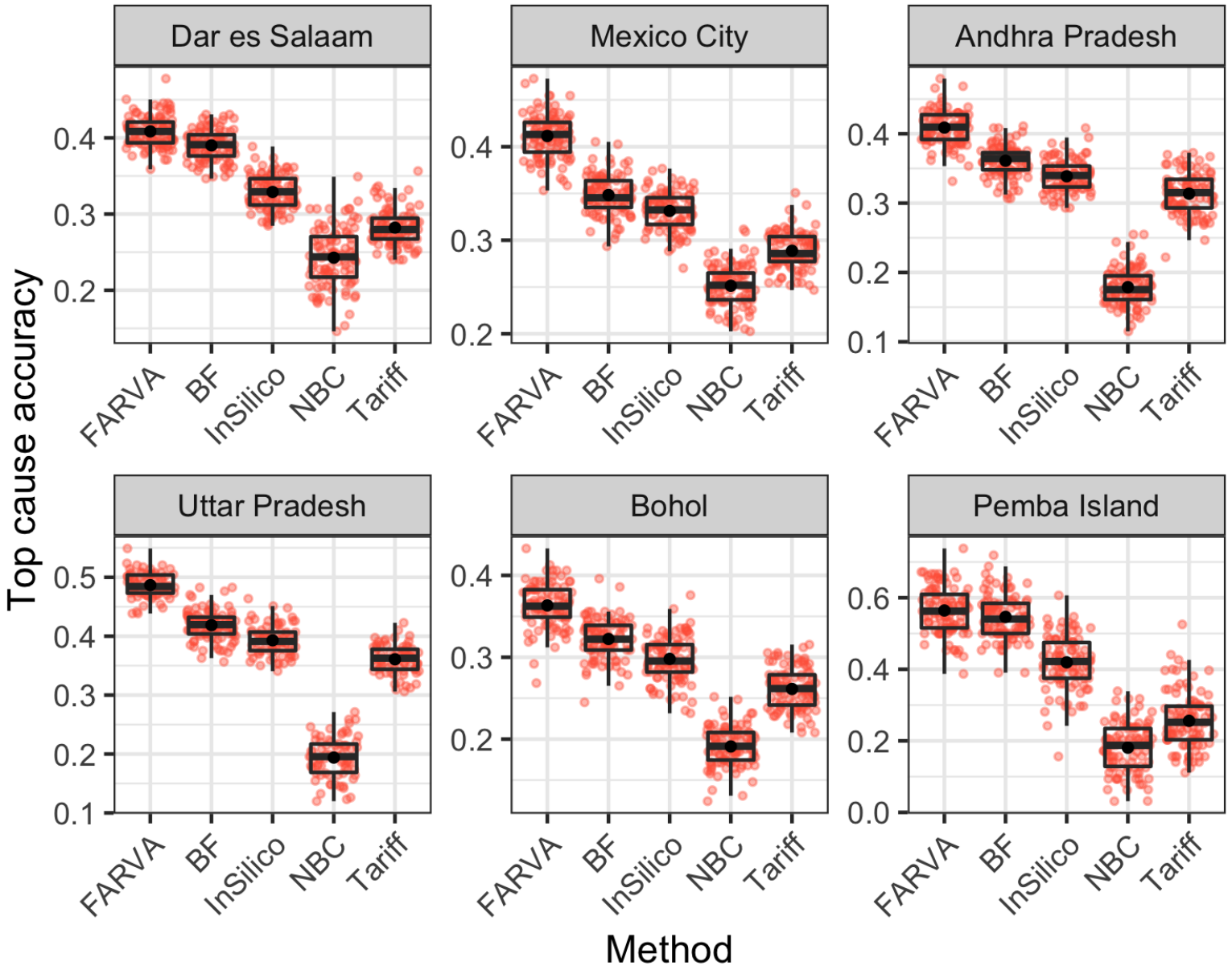
For each location assessed:

- Data split into 75% training, 25% test.
- Data cleaning steps used in **OpenVA** software performed, i.e. all variables converted to dichotomous symptoms matching those used in InterVA algorithm.
- Each model run, with FARVA including whether or not each decedent was an elder (≥ 65) as a covariate.
- Running: repeat the above 100 times in all locations.

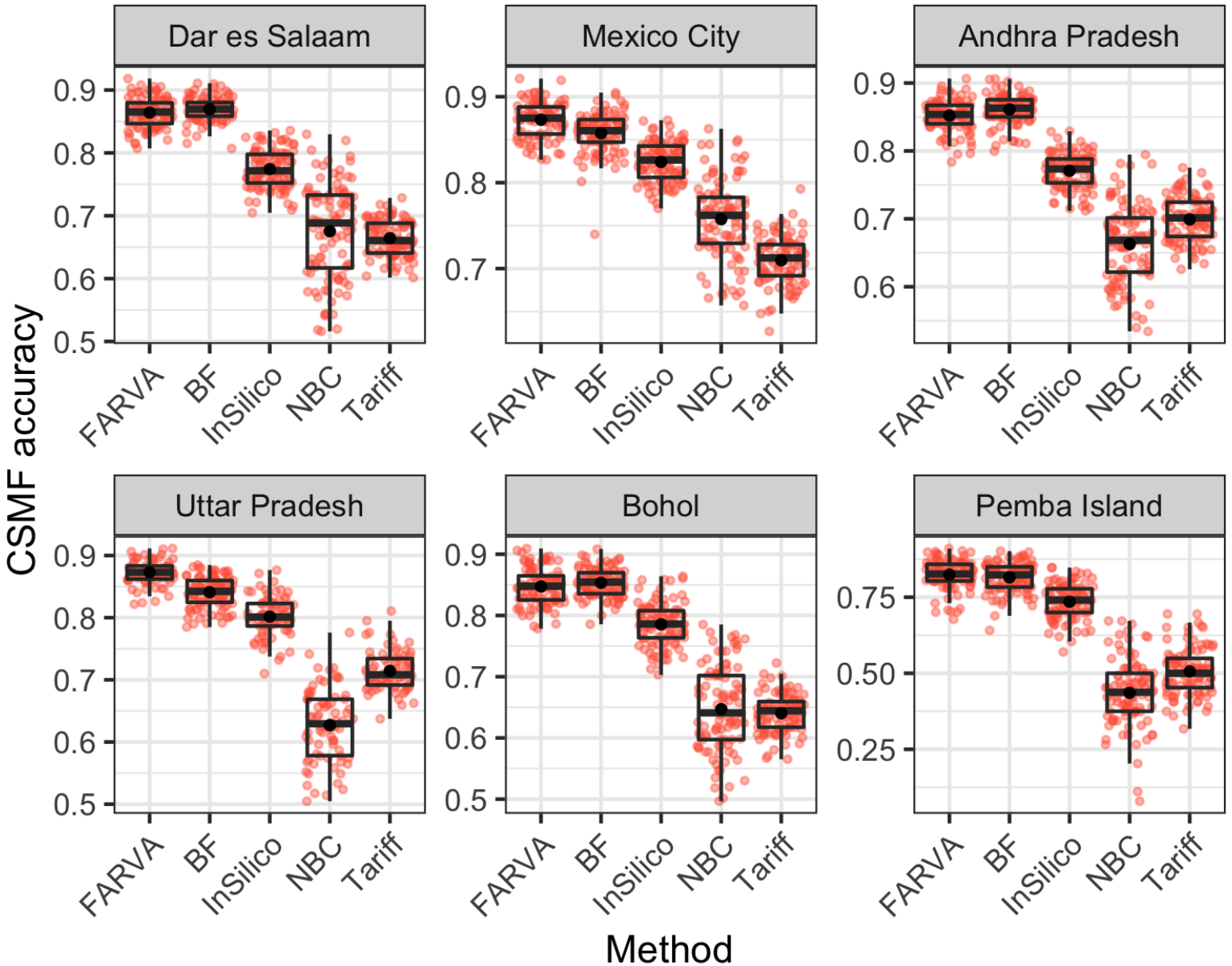
Mexico City performance



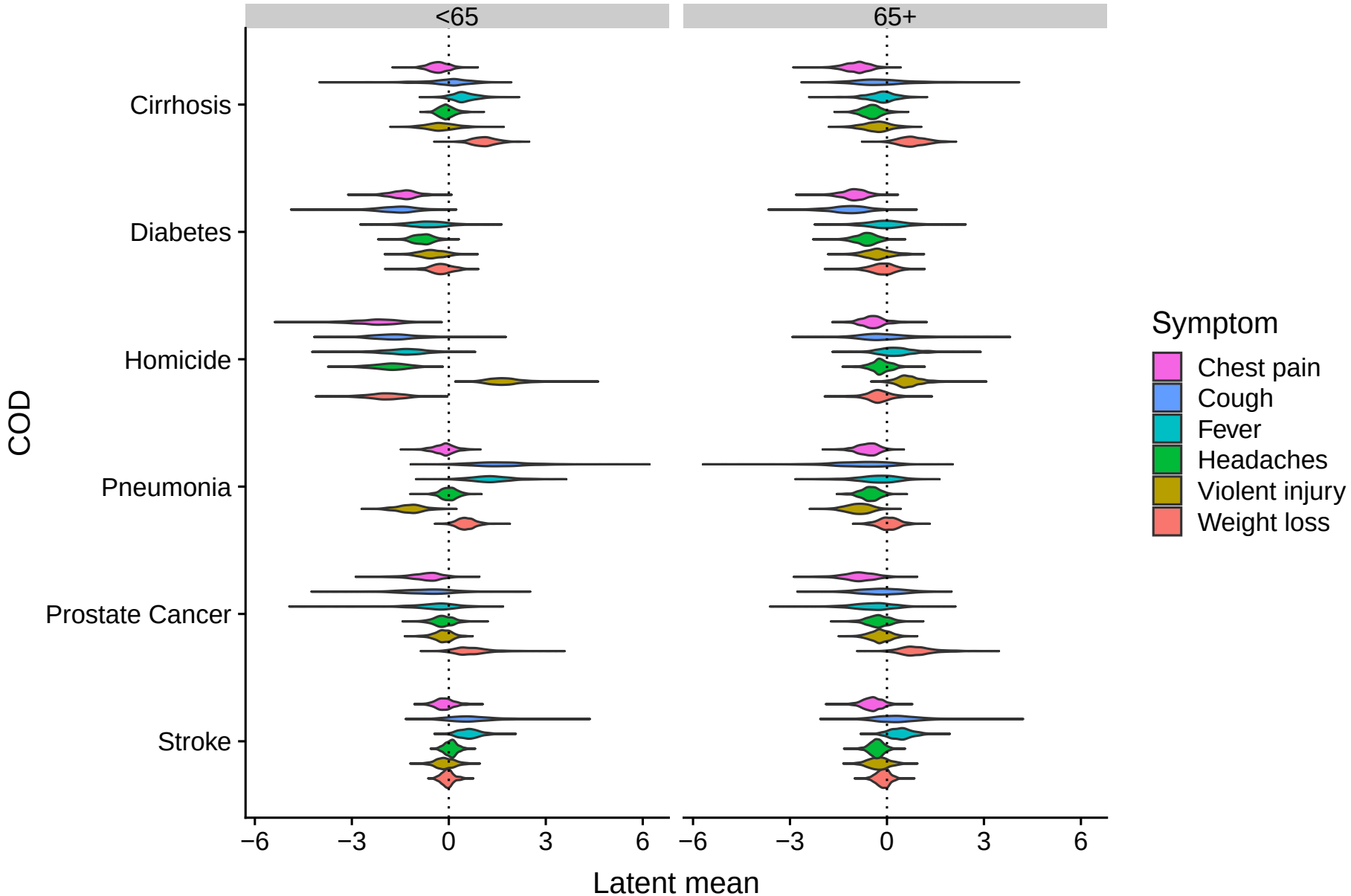
All locations (top cause accuracy)



All locations (CSMF accuracy)



Inference example



Future directions

- Discussed in paper (`https://arxiv.org/a/moran_k_1.html`):
  - Simulation study.
  - Inference on conditional symptom mean and associations.
  - Linking clinical, post mortem, and VA data.
- Package with example code available (`https://github.com/kelrenmor`)
- Open area of research:
  - Explicit modeling of missingness under MNAR assumption.
  - Selection of symptoms for analysis.
  - VA form modification (shortening) for unhelpful symptoms.
  - Utilizing free-text portion.
  - Sharing information between various questionnaires.

References

Anirban Bhattacharya and David B Dunson.

Sparse Bayesian infinite factor models.

Biometrika, pages 291–306, 2011.

Peter Byass, Daniel Chandramohan, Samuel J Clark, Lucia D'ambruoso, Edward Fottrell, Wendy J Graham, Abraham J Herbst, Abraham Hodgson, Sennen Hounton, Kathleen Kahn, et al.

Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool.

Global health action, 5(1):19281, 2012.

Daniele Durante.

A note on the multiplicative gamma process.

Statistics & Probability Letters, 122:198–204, 2017.

Emily B Fox and David B Dunson.

Bayesian nonparametric covariance regression.

The Journal of Machine Learning Research, 16(1):2501–2542, 2015.

References (continued)

Spencer L James, Abraham D Flaxman, and Christopher JL Murray.

Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies.

Population Health Metrics, 9(1):31, 2011.

Gary King, Ying Lu, et al.

Verbal autopsy methods with multiple causes of death.

Statistical Science, 23(1):78–91, 2008.

Tsuyoshi Kunihama, Zehang Richard Li, Samuel J Clark, and Tyler H McCormick.

Bayesian factor models for probabilistic cause of death assessment with verbal autopsies.

arXiv preprint arXiv:1803.01327, 2018.

Zehang Richard Li, Tyler H McCormick, and Samuel J Clark.

Using Bayesian latent gaussian graphical models to infer symptom associations in verbal autopsies.

arXiv preprint arXiv:1711.00877, 2018.

References (continued)

Tyler H McCormick, Zehang Richard Li, Clara Calvert, Amelia C Crampin, Kathleen Kahn, and Samuel J Clark.

Probabilistic cause-of-death assignment using verbal autopsies.

Journal of the American Statistical Association, 111(515):1036–1049, 2016.

Pierre Miasnikof, Vasily Giannakeas, Mireille Gomes, Lukasz Aleksandrowicz, Alexander Y Shestopaloff, Dewan Alam, Stephen Tollman, Akram Samarikhalaj, and Prabhat Jha.

Naive Bayes classifiers for verbal autopsies: comparison to physician-based classification for 21,000 child and adult deaths.

BMC medicine, 13(1):286, 2015.

Christopher JL Murray, Spencer L James, Jeanette K Birnbaum, Michael K Freeman, Rafael Lozano, and Alan D Lopez.

Simplified Symptom Pattern Method for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards.

Population health metrics, 9(1):30, 2011a.

References (continued)

Christopher JL Murray, Alan D Lopez, Robert Black, Ramesh Ahuja, Said Mohd Ali, Abdullah Baqui, Lalit Dandona, Emily Dantzer, Vinita Das, Usha Dhingra, et al.

Population health metrics research consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets.

Population health metrics, 9(1):27, 2011b.

Christopher JL Murray, Rafael Lozano, Abraham D Flaxman, Alireza Vahdatpour, and Alan D Lopez.

Robust metrics for assessing the performance of different verbal autopsy cause assignment methods in validation studies.

Population health metrics, 9(1):28, 2011c.

Christopher JL Murray, Rafael Lozano, Abraham D Flaxman, Peter Serina, David Phillips, Andrea Stewart, Spencer L James, Alireza Vahdatpour, Charles Atkinson, Michael K Freeman, et al.

Using verbal autopsy to measure causes of death: the comparative performance of existing methods.

BMC medicine, 12(1):5, 2014.

FARVA model (continued)

Recall $\boldsymbol{z}_i = \Lambda_{y_i}(\boldsymbol{x}_i)\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim \mathrm{N}(\boldsymbol{\psi}_{y_i}(\boldsymbol{x}_i), I_K), \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}_p, \Sigma_0)$

Define the entries of the latent mean vector hierarchically:

$$
\begin{aligned}
\psi_{y_i,k}(\boldsymbol{x}_i) =& \boldsymbol{\alpha}_{y_i,k}^T \boldsymbol{x}_i \\
& \boldsymbol{\alpha}_{y_i,k} \sim \mathrm{N}_B(\boldsymbol{\mu}_{\alpha_k}, \Sigma_{\alpha_k}), \\
& \boldsymbol{\mu}_{\alpha_k} \sim \mathrm{N}_B(A_0, L_0), \quad \Sigma_{\alpha_k} \sim \mathrm{IW}(v_0, D_0), \\
& k = 1, \ldots, K.
\end{aligned}
$$

Features:

- Latent symptom **mean** structure captured parsimoniously
- Latent symptom means are shrunk across causes

FARVA model (continued)

Recall $\boldsymbol{z}_i = \Lambda_{y_i}(\boldsymbol{x}_i)\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\eta}_i \sim \mathrm{N}(\boldsymbol{\psi}_{y_i}(\boldsymbol{x}_i), I_K), \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}_p, \Sigma_0)$.

Decompose loadings matrix as in Fox and Dunson (2015):

$$\begin{aligned}\Lambda_{y_i}(\boldsymbol{x}_i) &= \Theta_{y_i}\boldsymbol{\xi}_{y_i}(\boldsymbol{x}_i),\\ \Theta_{y_i} &\in \mathbb{R}^{p\times L},\\ \boldsymbol{\xi}_{y_i}(\boldsymbol{x}_i) &= \{\xi_{i,lk}(\boldsymbol{x}_i),\ l = 1,\ldots,L,\ k = 1,\ldots,K\}.\end{aligned}$$

Features:

- Symptom **covariance** is covariate-dependent, cause-specific, and modeled parsimoniously
- Symptom covariance shrunk across causes
- Number of factors K need only be an upper guess [Bhattacharya and Dunson (2011)]

FARVA model (continued)

Recall the factor loading matrix $\Lambda_{y_i}(\boldsymbol{x}_i)$ is decomposed as:

$$\Lambda_{y_i}(\boldsymbol{x}_i) = \Theta_{y_i}\xi_{y_i}(\boldsymbol{x}_i).$$

To share information across causes define the entries of each coefficient matrix $\Theta_c, c = 1, \ldots, C$, to share a common population level mean across causes Δ. Sparsity is induced on the population mean parameter for each entry in the coefficient matrix via the adaptive shrinkage prior of Bhattacharya and Dunson (2011).

$$\theta_{y_i,jl} \sim \mathsf{N}(\Delta_{jl}, \phi_{\Theta,jl}^{-1}\tau_{\Theta,l}^{-1}), \quad \phi_{\Theta,jl} \sim \mathsf{Ga}(\gamma_\Theta/2, \gamma_\Theta/2), \quad \tau_{\Theta,l} = \prod_{h=1}^{l} \delta_{\Theta,h},$$

$$\Delta_{jl} \sim \mathsf{N}(0, \phi_{\Delta,jl}^{-1}\tau_{\Delta,l}^{-1}), \quad \phi_{\Delta,jl} \sim \mathsf{Ga}(\gamma_\Delta/2, \gamma_\Delta/2), \quad \tau_{\Delta,l} = \prod_{h=1}^{l} \delta_{\Delta,h},$$

$$j = 1, \ldots, p, l = 1, \ldots, L,$$

FARVA model (continued)

Recall the factor loading matrix $\Lambda_{y_i}(\boldsymbol{x}_i)$ is decomposed as:

$$\Lambda_{y_i}(\boldsymbol{x}_i) = \Theta_{y_i}\boldsymbol{\xi}_{y_i}(\boldsymbol{x}_i).$$

To share information across causes define the entries of each of the predictor-dependent basis functions $\boldsymbol{\xi}_{y_i}(\boldsymbol{x}_i)$ using a hierarchical linear model:

$$\begin{aligned}
\xi_{y_i,lk}(\boldsymbol{x}_i) =& \boldsymbol{\beta}_{y_i,lk}^T \boldsymbol{x}_i \\
& \boldsymbol{\beta}_{y_i,lk} \sim \mathrm{N}_B(\boldsymbol{\mu}_{\beta_{lk}}, \Sigma_{\beta_{lk}}), \\
& \boldsymbol{\mu}_{\beta_{lk}} \sim \mathrm{N}_B(\mu_0, \Lambda_0), \quad \Sigma_{\beta_{lk}} \sim \mathrm{IW}(\nu_0, S_0), \\
& l = 1, \ldots, L, k = 1, \ldots, K.
\end{aligned}$$

Practical considerations

If you have "impossible" causes (e.g., prostate cancer for female decedents, maternal causes for male decedents):

- Fix $\pi(y_{i^*} = c|\boldsymbol{s}_{i^*}) = 0$ for all causes c s.t. person i^* could not have died of that cause.

If you have missing data, add a new step to the Gibbs sampler:

- Sample $\{z_{ij}\}$ for all i, j s.t. s_{ij} is missing. For i, j s.t. s_{ij} is missing, sample z_{ij} from $\mathrm{N}(\Lambda_{c[i],j\cdot}\boldsymbol{\eta}_i, \sigma_j^2)$, where $\Lambda_{c[i],j\cdot}$ denotes the j-th row of $\Lambda_{c[i]}$ and all parameters come from the most recent iteration of the Gibbs sampler.