Meta-Regression Case Studies in Infectious Disease Epidemiology

Joshua L. Warren
Department of Biostatistics
Yale University

04/14/2023

Infectious Disease Epidemiology

- Study of how and why infectious diseases emerge and spread among different populations[1]
- Investigating what strategies prevent or contain the spread of disease at the population level[1]
- Data gathered/collected from multiple sources to answer questions of interest
- Various modeling strategies used (e.g., statistical and mathematical/mechanistic/transmission modeling)

Role of the Statistician

- Studies can be time-consuming and resource intensive, while only focusing on a very specific setting (e.g., spatial location, disease outbreak, population)
- Meta-analyses/-regressions are important tools for generalizing study findings; results potentially useful for policy makers
- An opportunity for statisticians to get involved to ensure appropriate models are being used, and to develop new methodology for unique/emerging data sources

Statistical Considerations (My Experiences)

- Spatial and spatiotemporal correlation
- Multivariate outcomes
- Hierarchical modeling
- Survival data
- Causal inference
- Changepoint methods

Two Recent Examples for Today

- *Global, regional, and national estimates of tuberculosis incidence and case detection among incarcerated persons: A systematic analysis*
  - Collaborators: Leonardo Martinez, Ted Cohen, Jason Andrews, many others
- *An analytical framework to extrapolate results across multiple sites: Real-world examples from rotavirus clinical trials*
  - Collaborators: Virginia Pitzer, Ottavia Prunas, Elizabeth Sajewski, Ernest Asare, many others

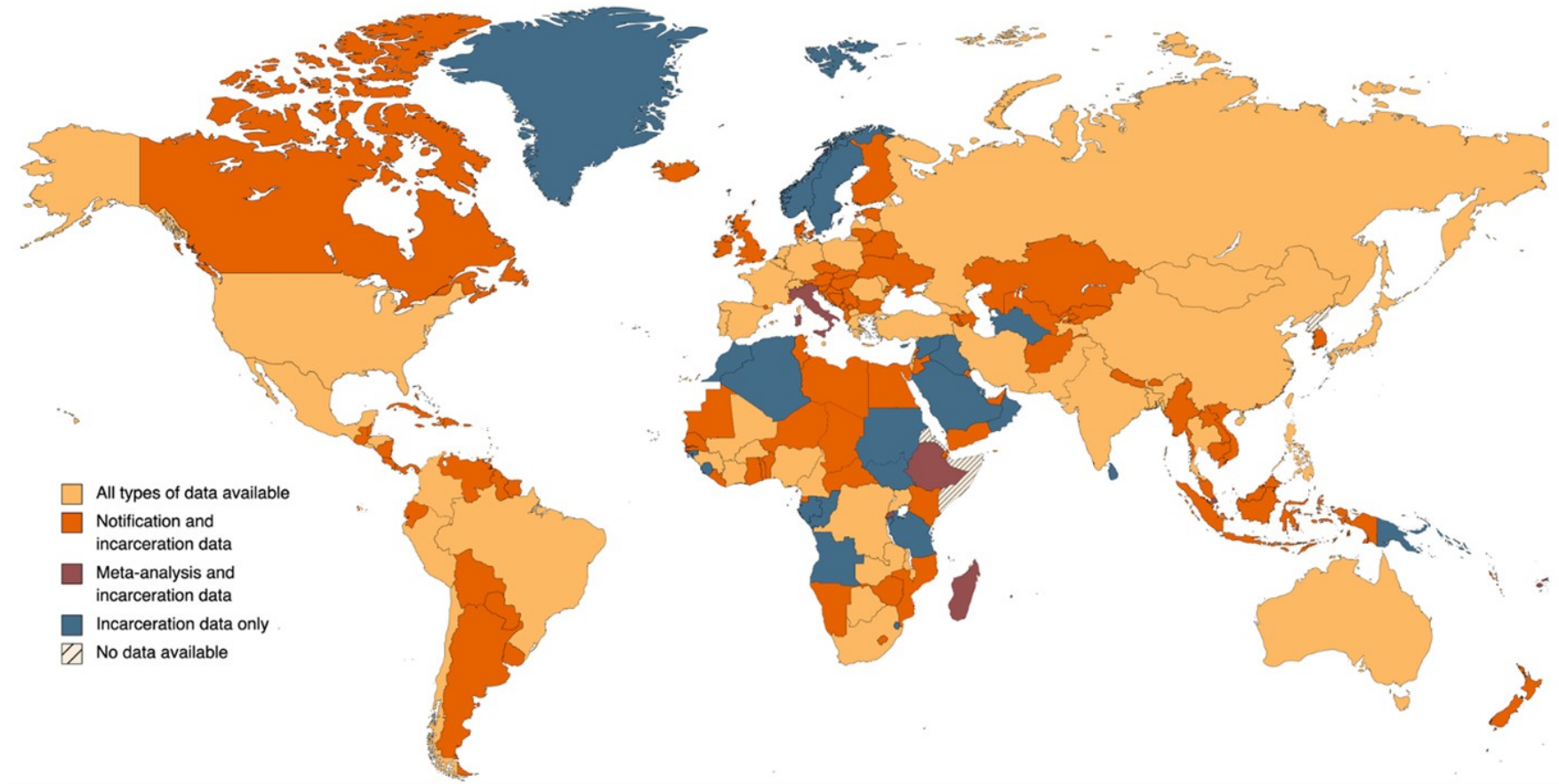
Tuberculosis (TB) in Prisons

- Studies of TB in prisons often report **incidence** and/or **prevalence** estimates (**and uncertainty**) for a given location/population and point in time
- Separately, many countries collect **TB notifications** across time, and also monitor the **prisoner population size**
- Study Goals:
  - Quantify the global TB burden among incarcerated persons across time
  - Develop an interpretable meta-regression model to jointly model all data sources, correctly characterize their uncertainty, and interpolate in countries/years with sparse/no data

Data Collection

- Systematic literature review:
  - January 1980 - August 2020
  - Studies must have estimated/reported TB incidence and/or prevalence among incarcerated populations
  - 48 countries with incidence estimates, 92 with prevalence estimates
- Received TB notification data from 152 of 199 countries
- Incarceration data from 193 countries

Data Availability



Proposed Statistical Model Summary

- Use all observed data/estimates (i.e., incidence, prevalence, notifications) within a single meta-regression framework (i.e., no multi-stage approximation)
- Leverage the completeness in space and time of the notification data, and its strong (hypothesized) association with incidence, to interpolate incidence estimates in unobserved countries/years
- Similarly, use the more complete prevalence estimates to provide additional predictive information for incidence
- Work in hierarchical Bayesian setting to allow for more natural uncertainty quantification and prediction

Incidence Rates (Observed Data)

$$\widehat{\theta}_{ijtkl} \overset{\text{ind}}{\sim} \text{N}\left(\theta_{ijtkl}, \widehat{\sigma}^2_{itjkl}\right);$$

$$i = 1, \ldots, r;\ j = 1, \ldots, c_i;\ t = 1, \ldots, s_{ij};$$

$$k = 1, \ldots, n_{ijt};\ l = 1, \ldots, m_{ijtk}$$

- $\widehat{\theta}_{ijtkl}$: Estimated log incidence rate from cohort l, within study k, within year t, within country j, within region i
- $\widehat{\sigma}^2_{itjkl}$: Standard error
- θ_{ijtkl}: True but unobserved log incidence rate

Prevalence Rates (Observed Data)

$$\widehat{\psi}_{ijtkl} \overset{\text{ind}}{\sim} \text{N}\left(\psi_{ijtkl}, \widehat{\tau}^2_{itjkl}\right)$$

- $\widehat{\psi}_{ijtkl}$: Estimated log prevalence rate
- $\widehat{\tau}^2_{itjkl}$: Standard error
- ψ_{ijtkl}: True but unobserved log prevalence rate

Notification Rates (Observed Data)

$$Z_{ijt} \overset{\text{ind}}{\sim} \text{Negative Binomial}\left(p_{ijt}, r\right),$$

$$p_{ijt} = \frac{r}{r + \lambda_{ijt}},$$

$$\ln\left(\lambda_{ijt}\right) = O_{ijt} + \mathbf{x}_{ijt}^{\mathsf{T}}\boldsymbol{\beta}_z + \alpha_z + \mu_{zi} + \kappa_{zij}$$

- Z_{ijt}: Number of notifications
- O_{ijt}: Log of the estimated prison population
- $\mathbf{x}_{ijt}$: Vector of covariates
- α_z: Global intercept
- μ_{zi}: Region-specific intercept
- κ_{zij}: Country-specific intercept

Incidence Rates (Latent Process)

$$\theta_{ijtkl} \overset{\text{ind}}{\sim} \text{N}\left(\theta_{ijt} + \pi_{\theta g(i,j,t,k)}, \sigma^2_\theta\right),$$

$$\theta_{ijt} = \max\left[\ln\left(\frac{Z_{ijt}}{\exp\{\text{O}_{ijt}\}} + c\right)\gamma_\theta + \mathbf{x}^{\text{T}}_{ijt}\boldsymbol{\beta}_\theta + \alpha_\theta + \mu_{\theta i} + \kappa_{\theta ij} + \eta_{\theta ijt}, \ln\left(\frac{Z_{ijt}}{\exp\{\text{O}_{ijt}\}} + c\right)\right]$$

- θ_{ijt}: Log incidence rate
- $\ln\left(\frac{Z_{ijt}}{\exp\{\text{O}_{ijt}\}} + c\right)$: Log notification rate, corrected for cases of 0

Incidence Rates (Latent Process)

- α_θ: Global intercept
- $\mu_{\theta i}$: Region-specific intercept
- $\kappa_{\theta ij}$: Country-specific intercept
- $\eta_{\theta ijt}$: Year-specific intercept
- $\pi_{\theta g(i,j,t,k)}$: Study-specific intercept
- σ^2_θ: Describes variability across cohorts within a study

Prevalence Rates (Latent Process)

$$\psi_{ijtkl} \overset{\text{ind}}{\sim} \text{N}\left(\ln\left(D\right) + \ln\left(\exp\left\{\theta_{ijt}\right\} - \frac{Z_{ijt}}{\exp\left\{\text{O}_{ijt}\right\}}\right) + \pi_{\psi g(i,j,t,k)}, \sigma^2_{\psi}\right)$$

- Mean structure comes from literature
- D: Duration of disease
- $\exp\left\{\theta_{ijt}\right\}$: Incidence rate
- $\frac{Z_{ijt}}{\exp\left\{\text{O}_{ijt}\right\}}$: Notification rate
- $\pi_{\psi g(i,j,t,k)}$: Study-specific intercept
- σ^2_{ψ}: Describes variability across cohorts within a study

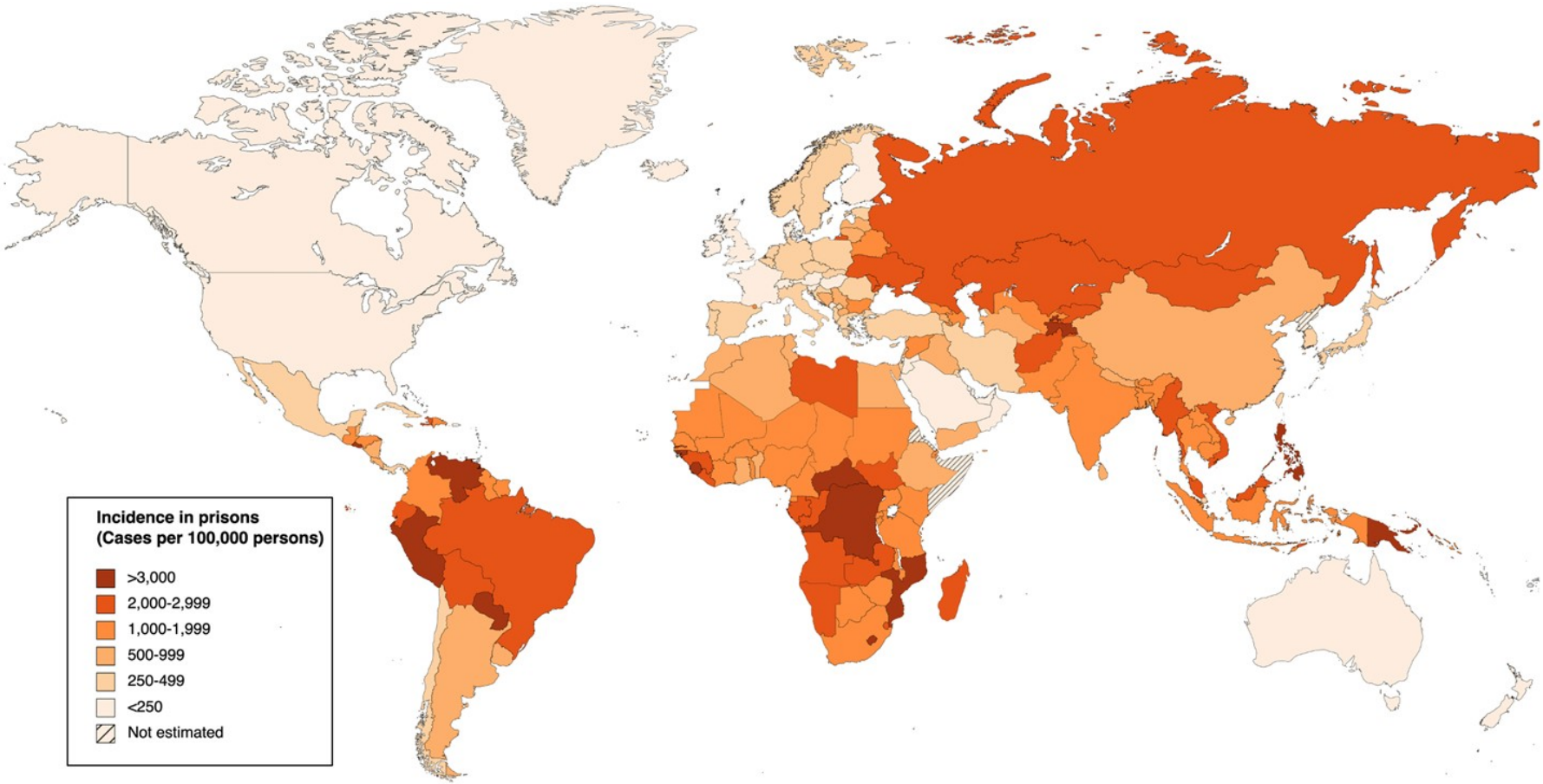
Random Effect Distributions

- $\mu_{zi} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\mu z}\right)$, $\mu_{\theta i} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\mu \theta}\right)$
- $\kappa_{zij} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\kappa z}\right)$, $\kappa_{\theta ij} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\kappa \theta}\right)$
- $\eta_{\theta ijt} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\eta \theta}\right)$
- $\pi_{\theta g(i,j,t,k)} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\pi \theta}\right)$, $\pi_{\psi g(i,j,t,k)} \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_{\pi \psi}\right)$

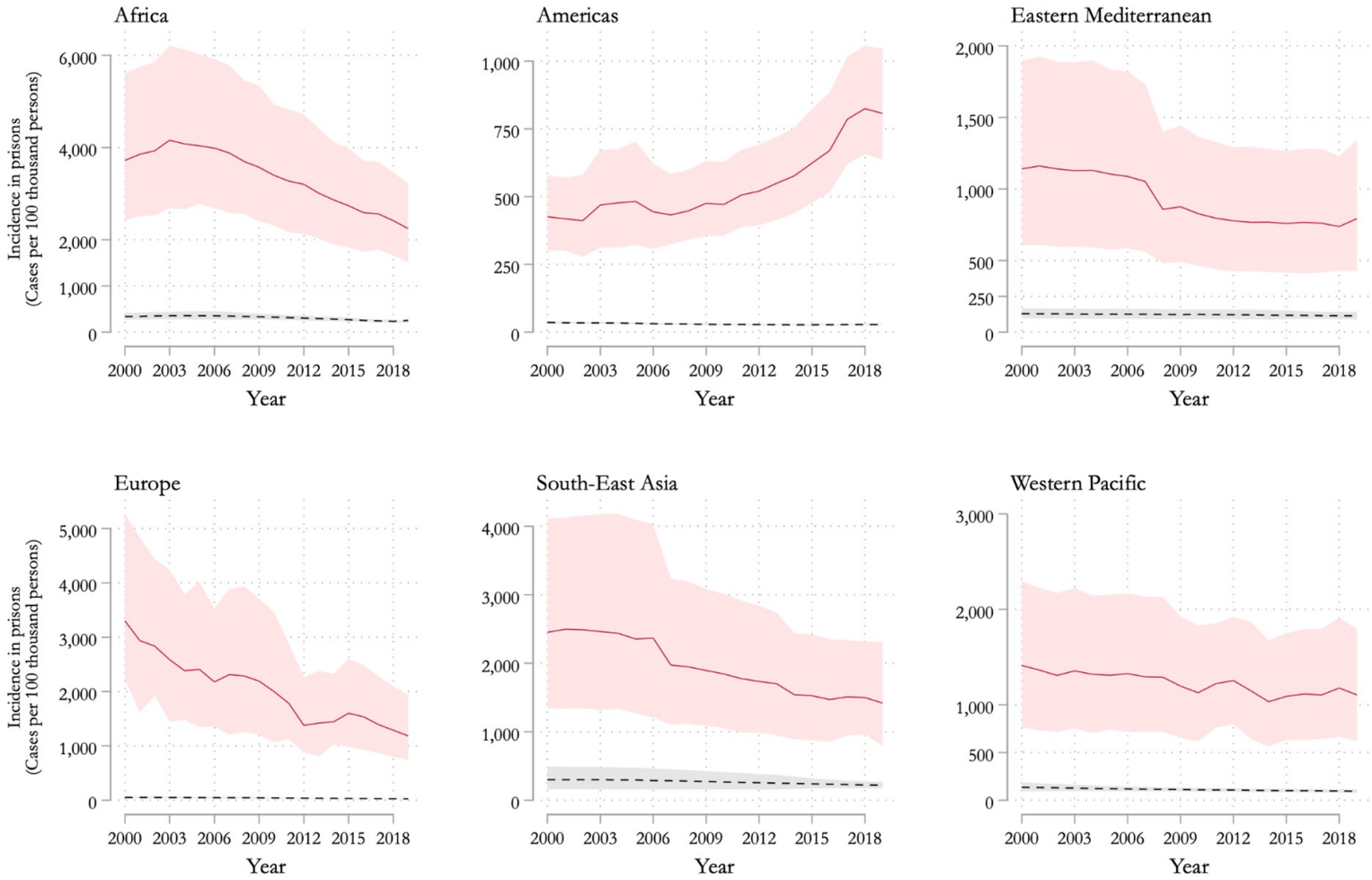
Model Fitting and Inference

- Weakly informative prior distributions used throughout
- Markov chain Monte Carlo (MCMC) posterior sampling algorithm used for making posterior inference
- Main quantities of posterior interest: $\exp\{\theta_{ijt}\}$ for all i, j, t
- Bayesian framework makes posterior inference and interpolation straightforward

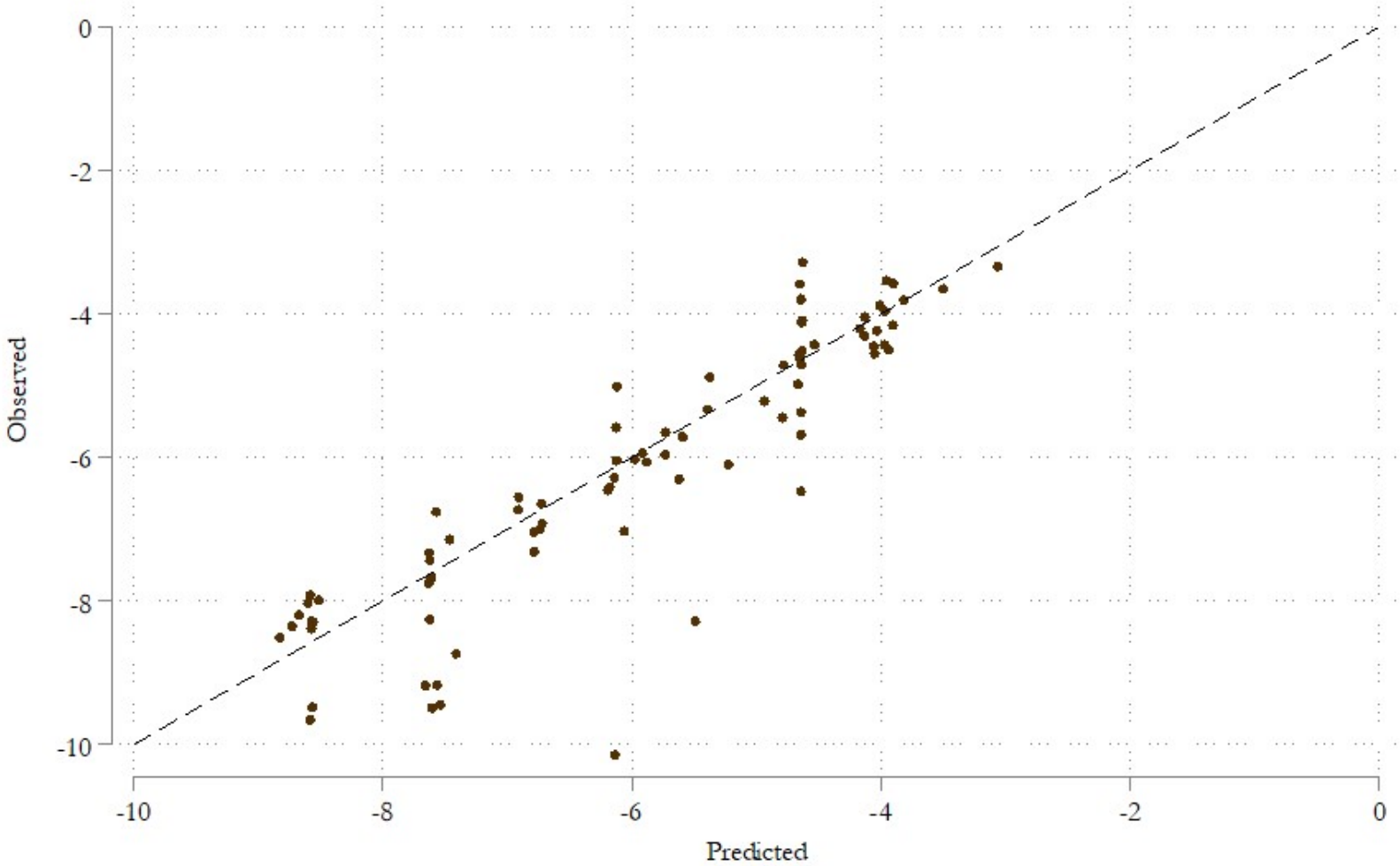
Estimated TB Incidence, 2019



Estimated TB Incidence in WHO Regions



10-Fold Cross Validation Results



95% prediction intervals included the true values 94.1% of the time

Conclusions

- Extraordinarily high TB incidence rate among incarcerated persons globally
- Incarcerated populations must be addressed with interventions specifically tailored to improve diagnoses and prevent transmission as a part of the broader global TB control effort
- Need for more comprehensive and standardized country-level data collection/reporting effort to improve estimates moving forward

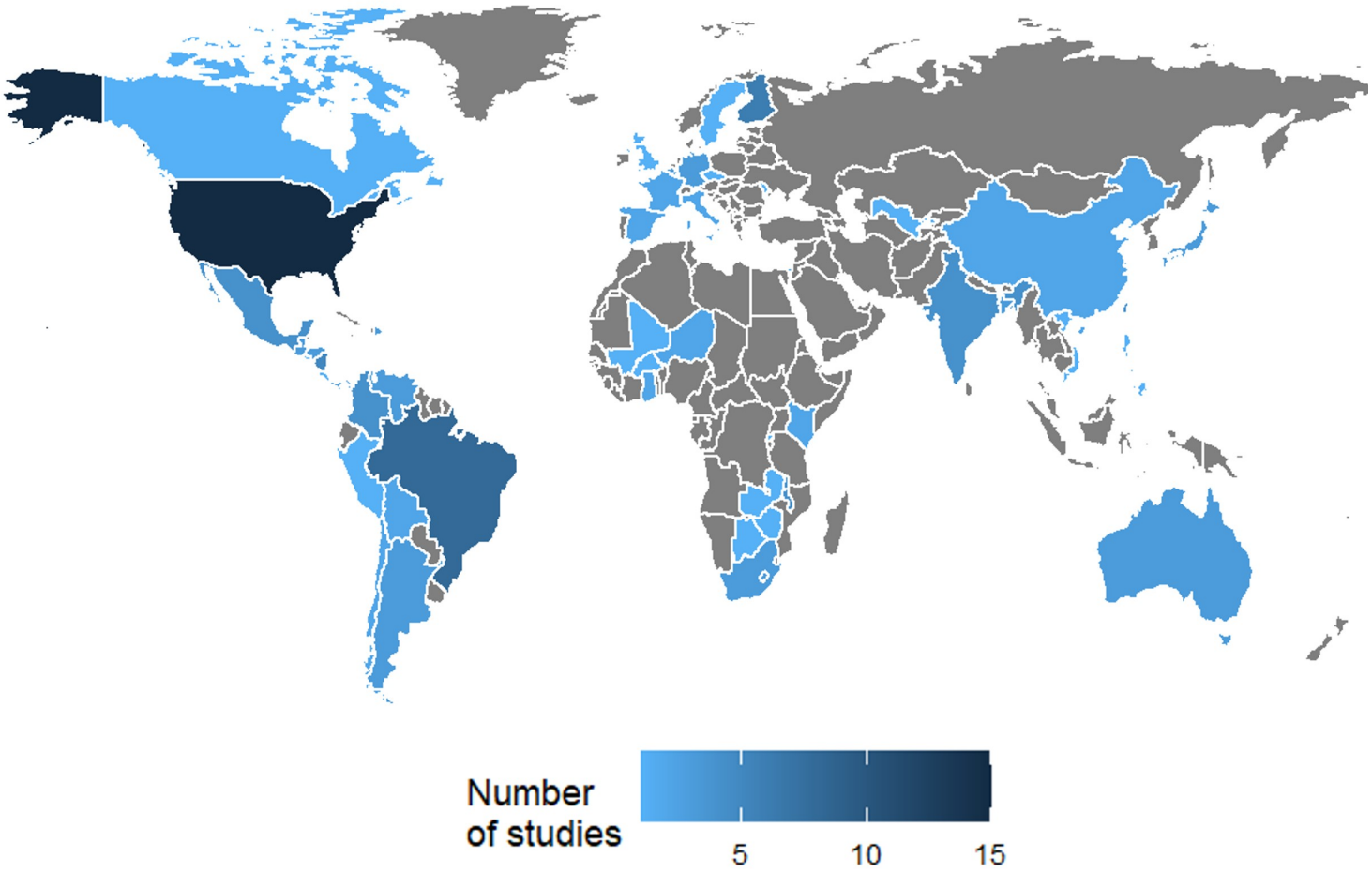
Vaccine Efficacy and Effectiveness

- Efficacy: Impact under ideal conditions (randomized controlled trial)
- Effectiveness: Impact under real world conditions (case-control study)
- Can differ by setting and depend on many underlying factors
- Study Goal:
  - Extrapolate results of studies conducted in multiple countries/settings for use by decision-makers in low-and middle-income countries

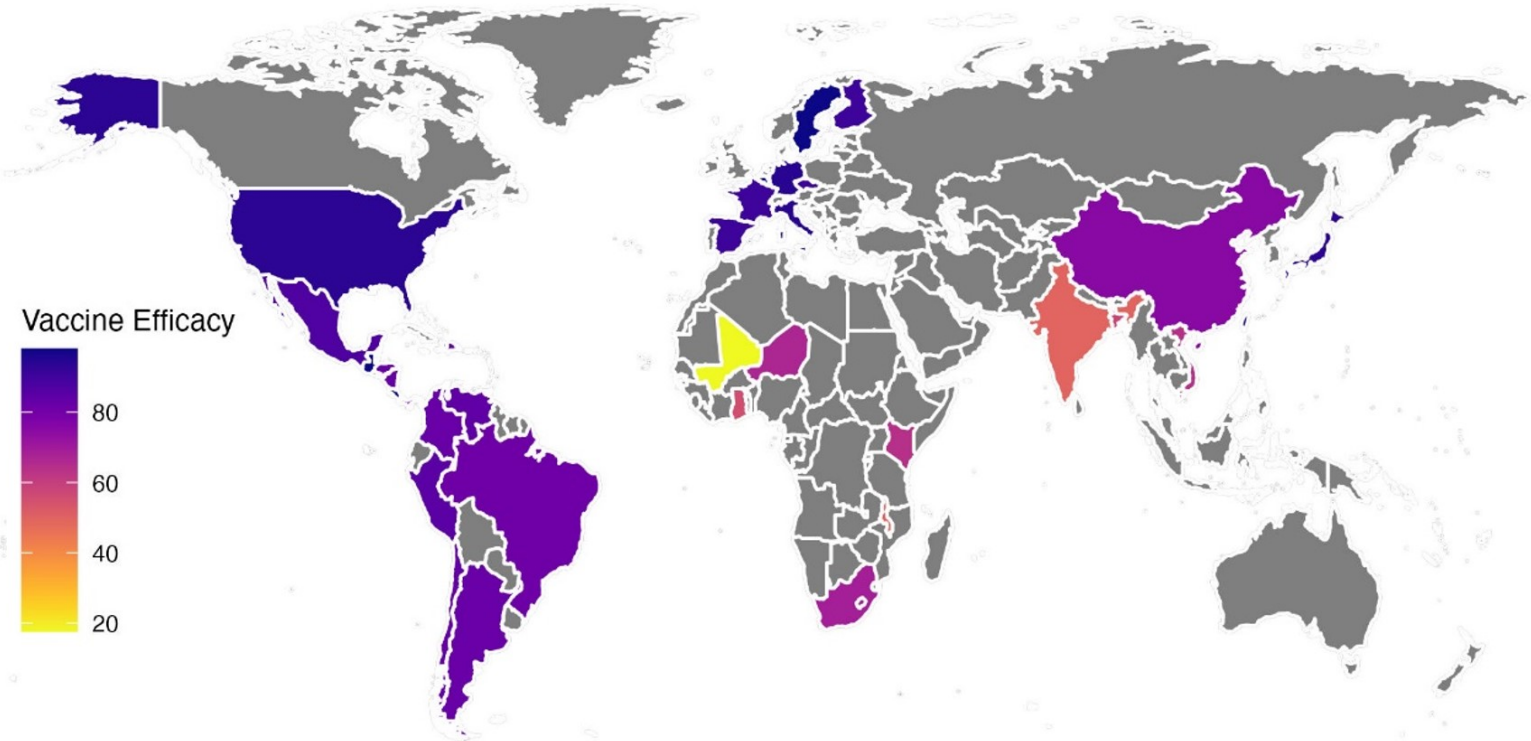
Data Collection

- Literature review for rotavirus vaccine efficacy and effectiveness estimates
  - Identified 74 studies including 33 vaccine efficacy and 54 vaccine effectiveness estimates
- Collected country-specific predictors that may explain variability in efficacy and effectiveness
  - Diarrhea prevalence among children
  - Population density
  - GDP per capita
  - Safe sanitation %
  - Safe drinking water %
  - Poverty %

Data Availability



Rotavirus Vaccine Efficacy



Proposed Statistical Model Summary

- Joint meta-regression modeling framework for vaccine efficacy and effectiveness
- Accounts for the fact that study-specific estimates may include individuals from different countries (i.e., mixture effect)
- Country-specific predictors used to explain variability in estimates, and predict unobserved countries
- Efficacy estimates used to predict effectiveness

Observed Data Model

$$z_i;\ \widehat{\theta}_i \overset{\text{ind}}{\sim} \mathrm{N}\left(\theta_i, \widehat{\sigma}_i^2\right),\ i = 1, \ldots, n;$$

- $\widehat{\theta}_i, \widehat{\sigma}_i^2$: Log estimate and variance, respectively,
  - $z_i = 0$: Vaccine efficacy
  - $z_i = 1$: Vaccine effectiveness
- θ_i: True value (unobserved latent process)
- n: Total number of studies

Latent Process Model

$$\theta_i \overset{\text{ind}}{\sim} \text{N}\left(\sum_{j=1}^{n_c} w_{ij}\eta_{z_i j}, \sigma^2_{\theta z_i}\right)$$

- $\eta_{z_i j}$: True log estimate of vaccine efficacy (η_{0j}) or vaccine effectiveness (η_{1j}) from country j
  - n_c total countries
- w_{ij}: Proportion of total people in study i that are from country j
- $\sum_{j=1}^{n_c} w_{ij}\eta_{z_i j}$: Weighted average of country-specific log estimates
- $\sigma^2_{\theta z_i}$: Describes study-level variability

Latent Process Model: Vaccine Efficacy

$$\eta_{0j} \overset{\text{ind}}{\sim} \text{N}\left(\mu + \mathbf{x}_j^{\mathsf{T}}\boldsymbol{\beta} + \phi_{s(j)}, \sigma^2_{\eta_0}\right), \; j = 1, \ldots, n_c$$

- $\mathbf{x}_j$: Vector of country j-specific predictors (excludes intercept)
- $\phi_{s(j)}$: Super region random effect
- $\phi_k \overset{\text{iid}}{\sim} \text{N}\left(0, \sigma^2_\phi\right), \; k = 1, \ldots, n_s$
- $\sigma^2_{\eta_0}$: Describes the country-level variability

Latent Process Model: Vaccine Effectiveness

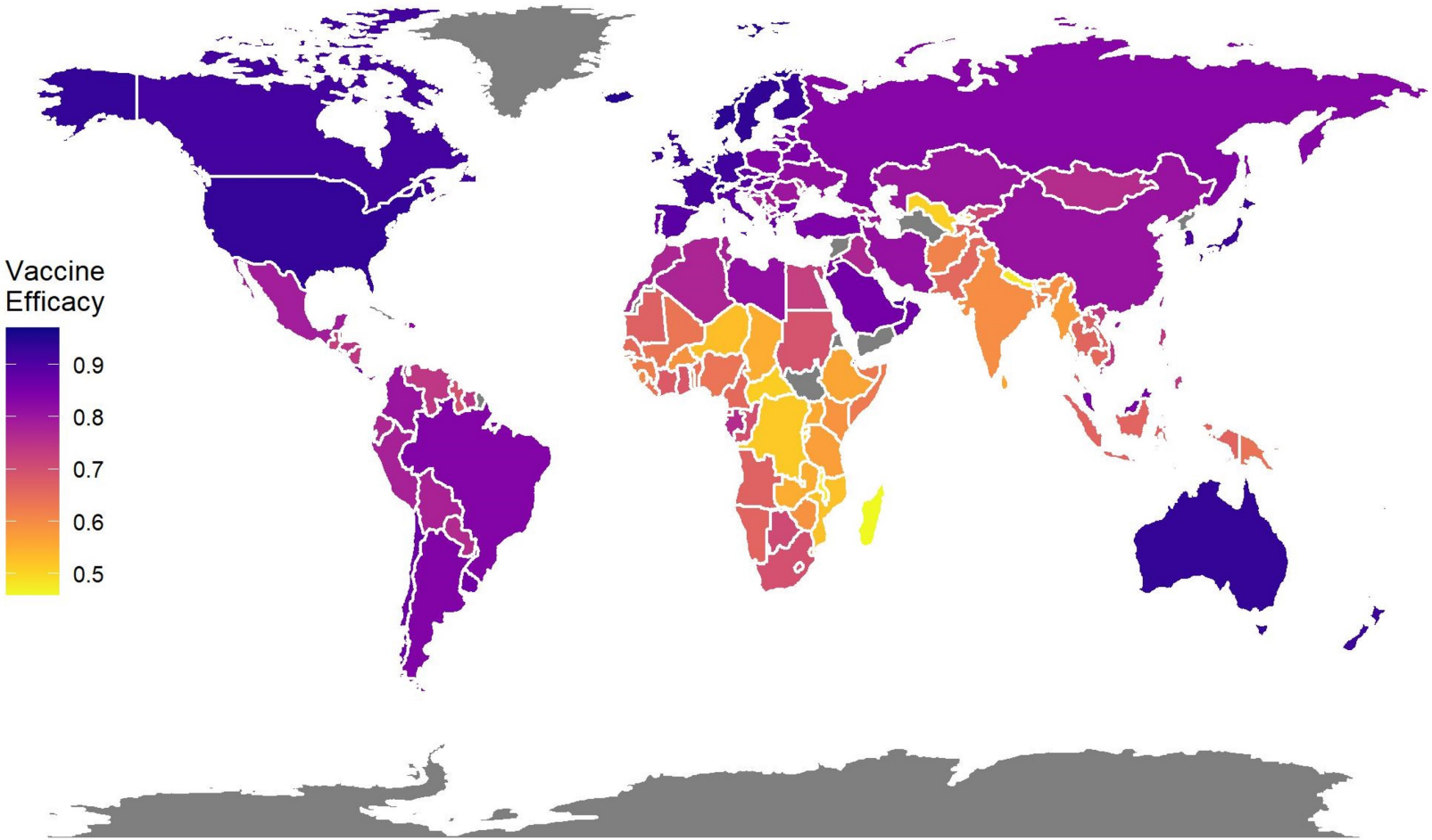
$$\eta_{1j} \overset{\text{ind}}{\sim} \mathsf{N}\left(\gamma_0 + \gamma_1 \eta_{0j}, \sigma^2_{\eta_1}\right)$$

- Vaccine efficacy from the same country used to predict the corresponding vaccine effectiveness
- γ_0, γ_1: Intercept and slope terms that connect vaccine effectiveness with vaccine efficacy
- $\sigma^2_{\eta_1}$: Describes country-level variability in this relationship

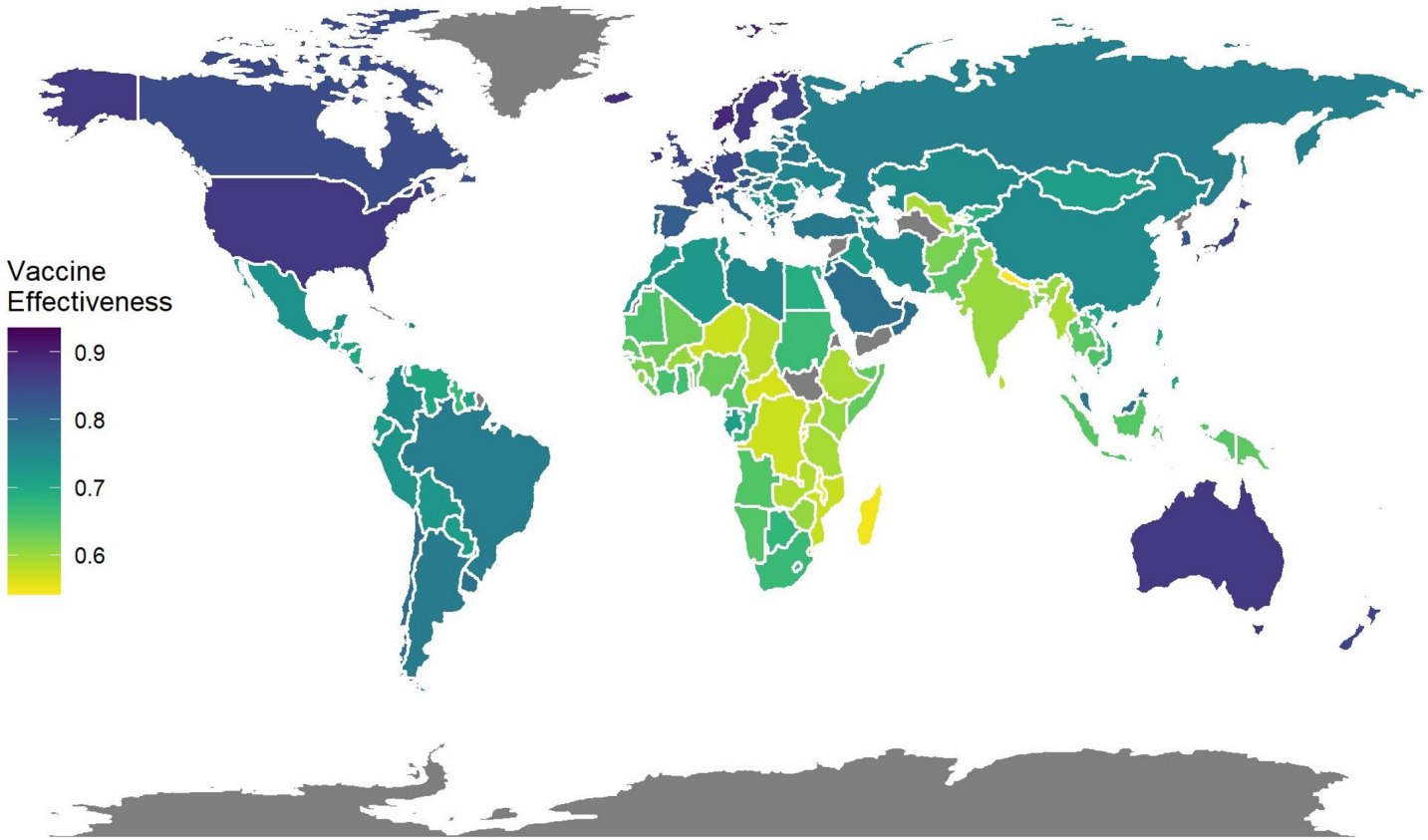
Model Fitting and Inference

- Weakly informative prior distributions used throughout
- MCMC posterior sampling algorithm used for making posterior inference
- Main quantities of posterior interest: η_{0j} and η_{1j} for all j
- Bayesian framework makes posterior inference and interpolation straightforward

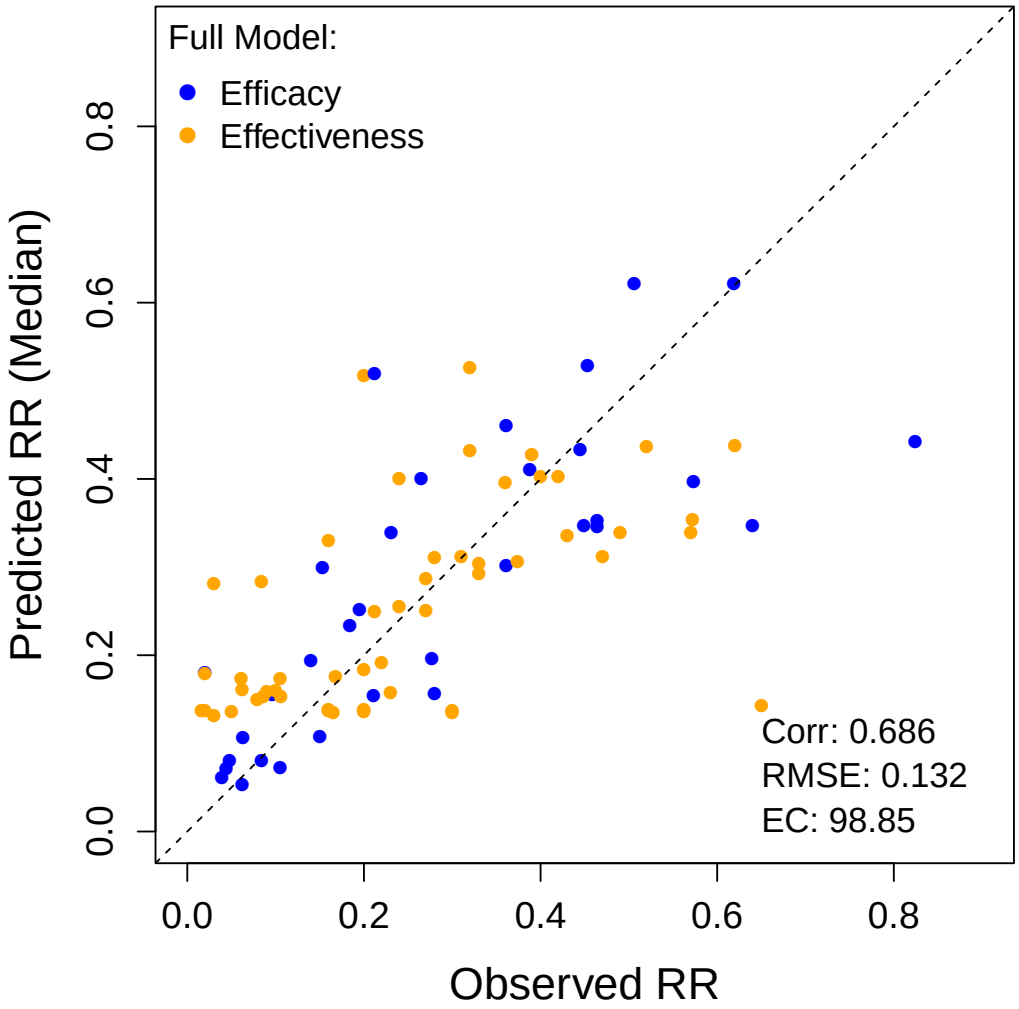
Vaccine Efficacy Estimation



Vaccine Effectiveness Estimation



Leave-One-Country-Out Cross Validation Results



Conclusions

- A limited number of widely available predictors can successfully predict vaccine efficacy in countries without study-based estimates
- Vaccine efficacy can be used to predict vaccine effectiveness
- In both projects, close communication with subject-matter experts helps to guide model development and results in methods that are effective and interpretable for a more general audience

References

1 Cornell University Public Health Infectious Disease Epidemiology, Website