

Understanding the eradication of malaria in the United States 1920-1950

Adam Howes (ath19@ic.ac.uk)

1 Background

- Malaria eradicated in the US during the years 1920-1950
- Interest in learning why this happened, or which factors were associated with faster eradication
- Have county level malaria data from various sources, with some limitations
- Have county level covariate data from various sources, with some limitations

2 Data

- Malaria data as follows...
- Covariate data in the following categories
 - County
 - Drainage
 - Farmland
 - Mortality
 - People
 - Socioeconomic
 - Weather
 - Zooprophylaxis

3 Model

- As implemented in `fit_sae-Y20-50`
- Uses `malariadata.csv` and `southern13_areas.geojson`
- Calculates mortality rate per 100,000
- For categorical data, takes some single fixed numeric value.
 - This could be improved using imputation via fitting a model, which I have started in `impute_numeric-Y33-37` but not finished
- Model defined for all years 1920 to 1950, and the counties from `southern13_areas.geojson`
- For covariates not available in a particular year and county, I imputed using `missForest()` and normalised the output (see `impute_covariates`)
 - I don't think the imputed covariates are good. Just looking at the time series output you get it's not what you'd want
- Model formula uses a fixed effect for each covariate, a Besag model on space at the country level, and an AR1 model on time at the year level
- Use a Poisson likelihood adapted to account for non-integer counts
- Fit the model with the R-INLA implementation of INLA

4 Results

- Generate plausible seeming posterior means at a state level

- Don't particularly like the way that the data for 1933-1937 is handled
- Strange upswing at the end of the time period where there is no data
- Generate estimates of association with malaria rate of the covariates
 - Only association, no causation
 - Covariate imputation algorithm I don't think is good
 - Perhaps they're of interest anyway, I'm not sure

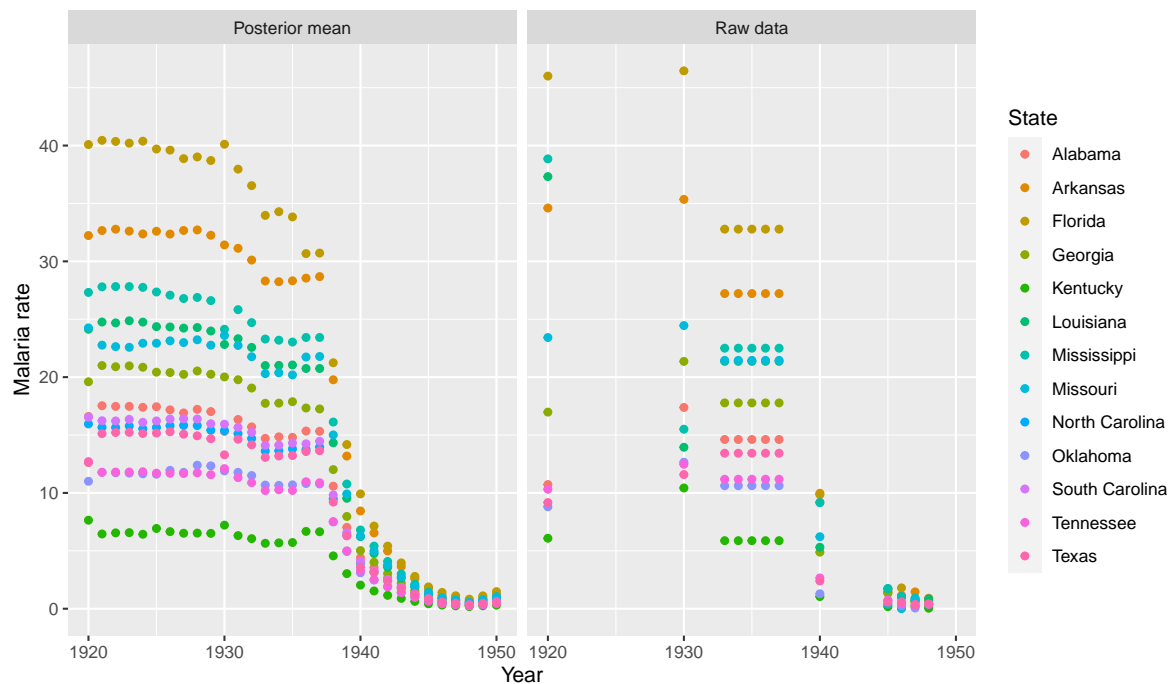


Figure 1: Posterior mean of malaria rate at a state level as compared with the raw data.

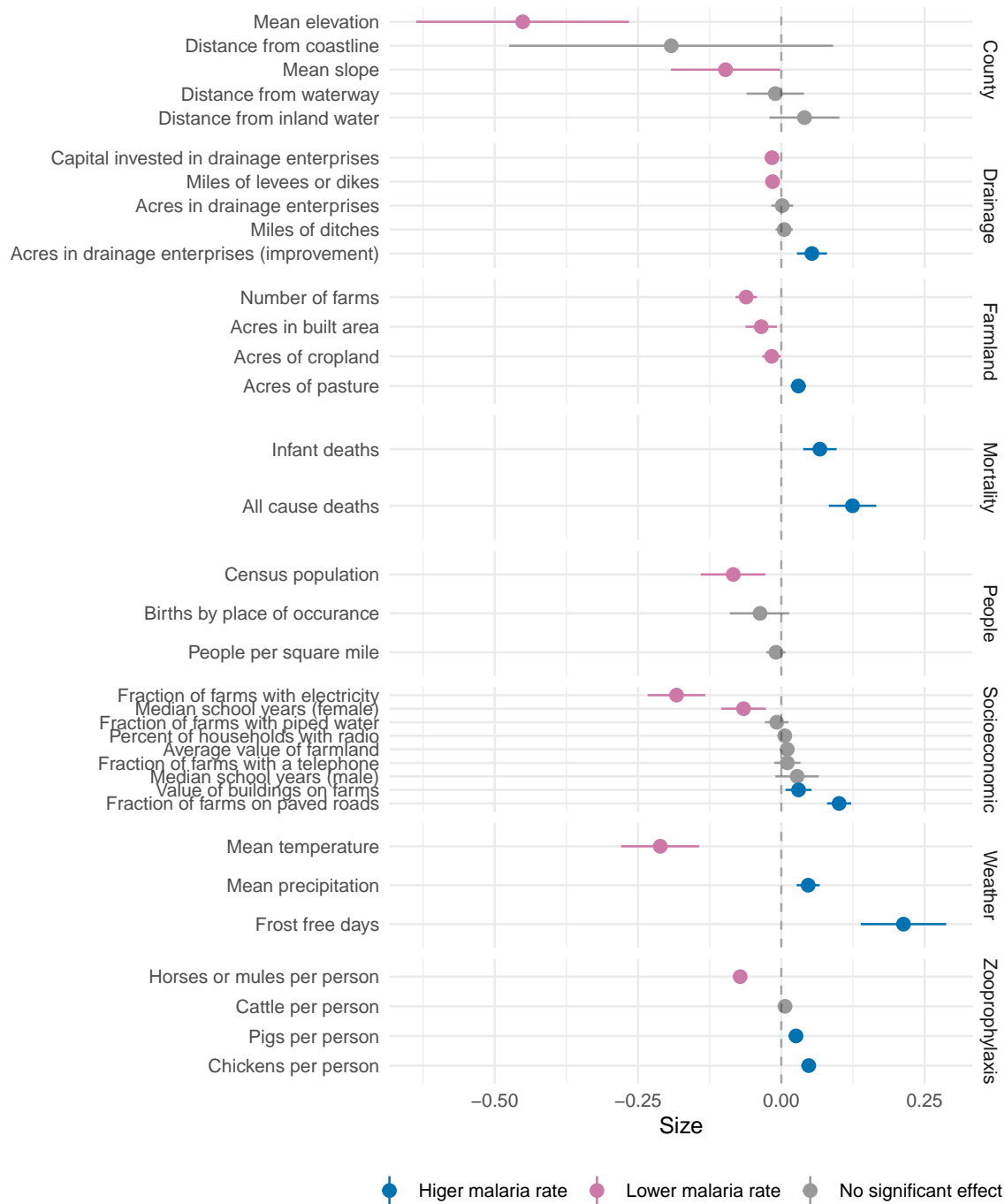


Figure 2: Posterior mean and credible intervals for regression coefficient parameters for each covariate.