Car-bicycle crashes in Michigan counties

Biostats 682

Andrew Soncrant, Shikun Wang, Brook Luers

April 17, 2017

# 1 Introduction

This project investigates car crashes in Michigan involving cyclists. After five cyclists died in a June, 2016, crash in Kalamazoo, news outlets discussed rising numbers of car-bicycle crashes across the state (e.g. [1], [2]). Our analysis describes *rates* of car-bicycle crashes, per 100,000 residents, in each of Michigan's 83 counties between 2004 and 2015. We hope to describe trends in cyclist safety while accounting for differences in population and economic conditions among Michigan counties.

## 1.1 Data

We obtained crash reports for cyclist-involved crashes between 2004 and 2015 from Michigan Traffic Crash Facts [3]. These crash reports were used to compute the number of crashes in each county during each of these 12 years. Annual estimates of county population, per capita income, and unemployment rate were taken from the Michigan Department of Technology, Management and Budget [4], and we obtained the land area of each county from the U.S. Census Bureau [5]. To obtain more symmetric distributions and similar scales for these covariates, we transformed county land areas and per-capita income to a logarithmic scale.

Figure 1 displays the observed number of crashes per 100,000 people each year, highlighting the 10 most populous counties and the counties surrounding Washtenaw county (which includes Ann Arbor). Figure 2 displays the change in the crash rate between 2004 and 2015 versus the average crash rate for the 20 counties with the highest average crash rates.

## 2 Model

Denote the observed number of crashes in county $c$ in year $t$ by $y_{ct}$. We assumed that this observed number of crashes reflects the county's population and the underlying crash rate in year $t$. We sought to build a model so that

$$\mathrm{E}\left(y_{ct}\right) = n_{ct}\eta_{ct}$$

where $n_{ct}$ is the population of county $c$ in year $t$ (in units of 100,000 people) and $\eta_{ct}$ is the rate of crashes per 100,000 people. Inferences regarding $\eta_{ct}$ allow comparisons of county-level crash rates across time.

We formulated a hierarchical Poisson regression to directly model the county-specific crash rates $\eta_{ct}$ as a function of time, per-capita income, unemployment rate, and county land area. Specifically, we chose

$$\begin{aligned}
y_{ct} &\sim \text{Poisson}(\lambda_{ct}) \quad \text{county } c,\ \text{year } t = 0, 1, \ldots, 11 \\
\log(\lambda_{ct}) &= \log(n_{ct}) + \mu_{ct} + \epsilon_{ct} \\
\mu_{ct} &= \beta_{0c} + \beta_{1c}t \\
&\quad + \beta_{2c}(\text{log per-capita income})_{ct} \\
&\quad + \beta_{3c}(\text{unemployment rate})_{ct} \\
&\quad + \alpha_1(\text{log land area})_c \\
\epsilon_{ct} &\sim N(0, \sigma^2)
\end{aligned}$$

This parameterizes the rate $\eta_{ct} = e^{\mu_{ct}}$ in terms of $\mu_{ct}$, a linear combination of covariates. The coefficients $\beta_{jc}$ represent, for each county, a linear trend in the crash rate over time and a multiplicative relationship between the crash rate and per-capita income, unemployment rate, and land area. The term $\epsilon_{ct}$ accounts for overdispersion and is independent of the regression coefficients. The population counts $n_{ct}$ are considered known constants, so $\log(n_{ct})$ is the "offset" in this log-linear model. Since each county's land area does not vary over time, we modeled $\alpha_1$ independently of $\beta_{jc}$.

## 2.1 Prior elicitation

To choose priors for the regression coefficients $\beta_{1j}$ and $\alpha_1$ we can consider their interpretation as multiplicative effects on the rate of crashes in a given county-year period. For example, with fixed values of per-capita income, land area, and unemployment rate, our model implies that each year the rate $\eta_{ct}$ of crashes is multiplied by $e^{\beta_1}$. Values of $\beta_1$ with magnitude near 0.01 correspond to annual changes of about 1 percent in the crash rate; magnitudes of $\beta_1$ near 0.5 correspond to roughly a 60-percent change in the crash rate per year. To understand the coefficients corresponding to log-transformed covariates, consider $\beta_2$. If we compare one county with twice the per-capita income of another county and fix the values of $t$ and the other covariates, then the ratio of the crash rate for the high-income county to that of the low-income county is $2^{\beta_2}$. We expect the magnitude of $\beta_2$ to be less than 1, since it seems unlikely that counties differing by a factor of 2 in income should have rates that differ by a factor of 2. This type of reasoning suggests that most of the prior mass for the regression coefficients should be between $-1$ and 1.

We chose the following prior parameterizations:

$$\begin{pmatrix} \beta_{0c} \\ \beta_{1c} \\ \beta_{2c} \\ \beta_{3c} \end{pmatrix} \sim \text{Normal}\left( \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix}, W\Omega W \right) \qquad \text{where } W = \begin{pmatrix} \tau_0 & & & \\ & \tau_1 & & \\ & & \tau_2 & \\ & & & \tau_3 \end{pmatrix}$$

$$\Omega = \text{Correlation matrix for } \beta \qquad \Omega \sim \text{LKJcorr}(1)$$

$$\tau_i \sim \text{Half-Cauchy}(0,1) \qquad \sigma \sim \text{Half-Cauchy}(0,1)$$

$$\alpha_1 \sim N(0, 0.5^2) \qquad \theta_i \sim N(0, 0.5^2)$$

$$\tau_i, \theta_i \text{ independent across } i = 0, 1, 2, 3$$

The LKJ correlation distribution with parameter 1, described in [6] and defined by [7], provides a uniform distribution over the set of correlation matrices (postive-definite, symmetric matrices with diagonal elements equal to 1). The Half-Cauchy$(0, 1)$ distribution has density $f(\sigma) \propto (1 + \sigma^2)^{-1}$ restricted to the positive real line.

## 3 Data Analysis

We first consider the raw crash rates, summarized in Figures 1 and 2. In general, most counties had less than 30 crashes per 100,000 residents, and the crash rates seem to have decreased during this 11-year period. Ingham county (which contains Lansing) had a relatively high annual crash rate which remained roughly constant over time, while Shiawassee, St.Joseph and Wexford counties are safer with a decreasing trend of crash rate. Grand Traverse and Isabella counties had increasing trends of crash rate between 2004 and 2015, which indicates that they are potentially dangerous for bicyclists. Considering the right panel of Figure 1, since 2008 Washtenaw county has had a higher crash rate than most of its surrounding counties. We will now verify these observations with posterior inference for our model.

We fit the model described in Section 2 with Stan [6] and `rstan` [8], using 4 Markov chains with $10,000$ iterations per chain. The first $5,000$ iterations from each chain were discarded. Trace plots (not shown) and values of $\hat{R}$, which were less than 1.1 for all model parameters, indicate convergence of the Markov chains.

Tables 1 and 2 contain posterior summaries for the population-level regression coefficients and scale parameters, respectively. Posterior means and intervals for $\theta_1$ and $\theta_2$ suggest that crash rates are declining over time and that high-income years are associated with higher car-bicycle crash rates. County unemployment rate and land area do not seem to predict the crash rate.

From 2004 to 2015, the average car-bicycle crash rate for all Michigan counties was decreasing. The posterior mean of $e^{\theta_1}$ is 0.966, corresponding to roughly a 4-percent annual decrease in the crash rate across all counties. This is consistent with our observations in Figure 1 and 2. To understand $\theta_2$, consider a high-income county with twice the per-capita income of a low-income county (in the same year) with equal values of the other covariates. The posterior mean of $2^{\theta_2}$ is 1.24, indicating a 24 percent higher crash rate in the high-income county.

Figure 3 displays the posterior mean of the crash rate $e^{\mu_{ct}}$ and the time trend $e^{\beta_{0c}+\beta_{1c}t}$ for all county-years, with Washtenaw and its surrounding counties highlighted. The right panel of Figure 3, along with the posterior means and intervals for $e^{\beta_{1c}}$ in Figure 4, confirm that the rate of car-bicycle crashes has been decreasing slightly for most Michigan counties between 2004 and 2015. Again, Ingham county stands out as having the highest crash rate across all Michigan counties during this time period. After controlling for income, unemployment rate and land area, Washtenaw county had essentially no change in its car-bicycle crash rate between 2004 and 2015, while most other counties had rates decreasing by about 3 percent each year.

# 4 Discussion

The goal of this project is to investigate claims made by the media that bicycle crashes are on the rise in Michigan (e.g. [1], [2]). We take our analysis one step further and perform Bayesian analysis at the county level. In our analysis, we account for changes in demographic factors such as income, unemployment rate, and population over time. We conclude that bicycle crash rates have indeed changed over time, but not in the way hypothesized by the media; bicycle riding in Michigan has actually become more safe over time. Our analysis, however, is not without its potential flaws. The features included in our model were chosen by convenience, in that demographic data on the state and county level is readily available from the census and other government agencies. We were unable to include weather data in our analysis, which one could reasonably assume has an impact on rider safety—both because of how weather affects traffic patterns and road conditions, and because of how climate and weather have been changing over time. Furthermore, from a Bayesian standpoint, we don't have the expertise in this area to assume anything other than rather uninformative priors in our model. Our goal is that the information gleaned in this study can be used to help set prior distributions in future analyses.

## 5 Tables and Figures

| | $\theta_0$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\alpha_1$ |
|---|---|---|---|---|---|
| Posterior mean | -0.0097 | -0.0349 | 0.3067 | -0.0018 | -0.0687 |
| Lower 80% CI | -0.6303 | -0.0418 | 0.1681 | -0.0093 | -0.2744 |
| Upper 80% CI | 0.6189 | -0.0281 | 0.4464 | 0.0052 | 0.1373 |
| $\hat{R}$ | 0.9999 | 1.0000 | 1.0000 | 1.0007 | 1.0001 |

Table 1: Posterior means, intervals and $\hat{R}$ for population-level regression coefficients.

| | $\tau_0$ | $\tau_1$ | $\tau_2$ | $\tau_3$ | $\sigma$ |
|---|---|---|---|---|---|
| Posterior mean | 0.3643 | 0.0183 | 0.0341 | 0.0133 | 0.0647 |
| Lower 80% CI | 0.0867 | 0.0133 | 0.0073 | 0.0043 | 0.0508 |
| Upper 80% CI | 0.5984 | 0.0236 | 0.0569 | 0.0219 | 0.0791 |
| $\hat{R}$ | 1.0037 | 1.0000 | 1.0134 | 1.0022 | 1.0196 |

Table 2: Posterior means, intervals and $\hat{R}$ for scale parameters.

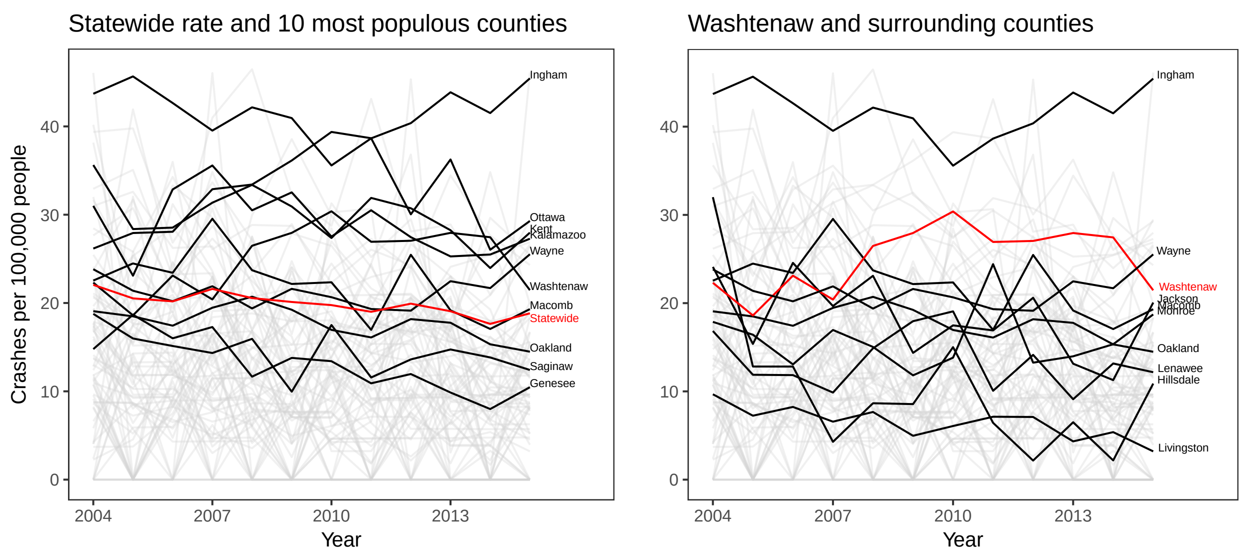


Figure 1: Observed number of crashes per 100,000 people, $y_{ct}/n_{ct}$. The left panel highlights the statewide rate (in red) and the 10 most populous counties. The right panel highlights Washtenaw county (in red) and its surrounding counties in southeast Michigan.

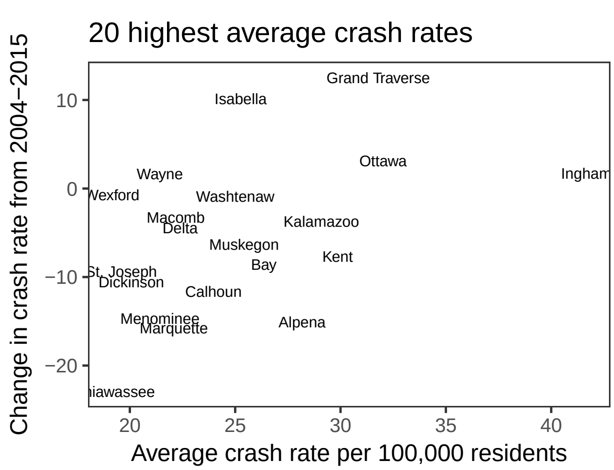


Figure 2: Change in crash rate versus average crash rate for the 20 counties with the highest average crash rates.

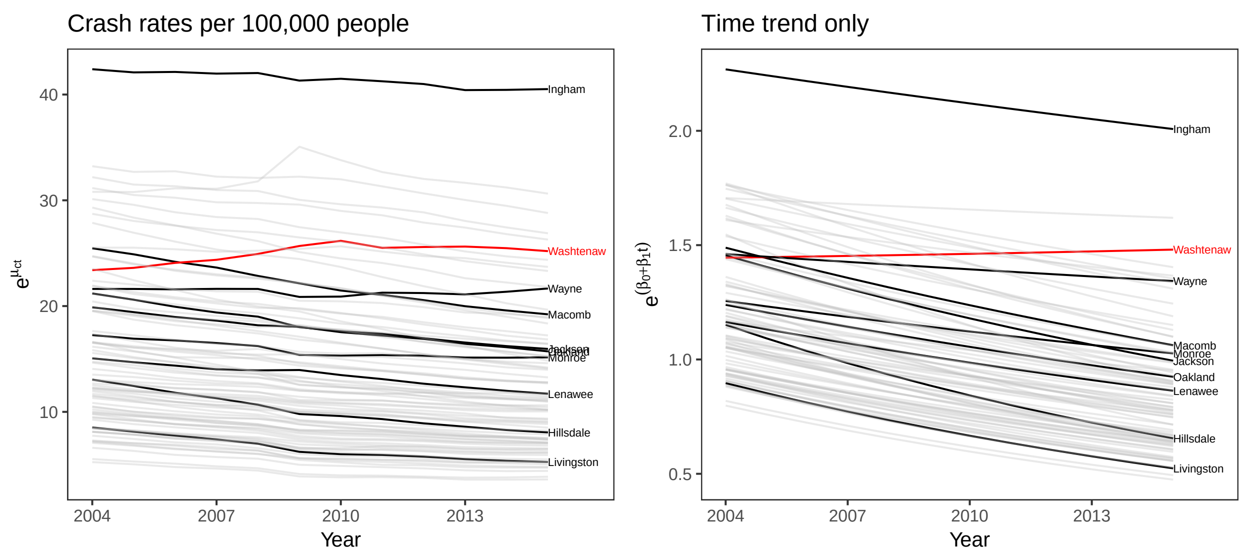


Figure 3: Posterior means for the crash rate time trend ($e^{\beta_0+\beta_1 t}$) and the overall crash rate in each county. Washtenaw (in red) and surrounding counties are highlighted.

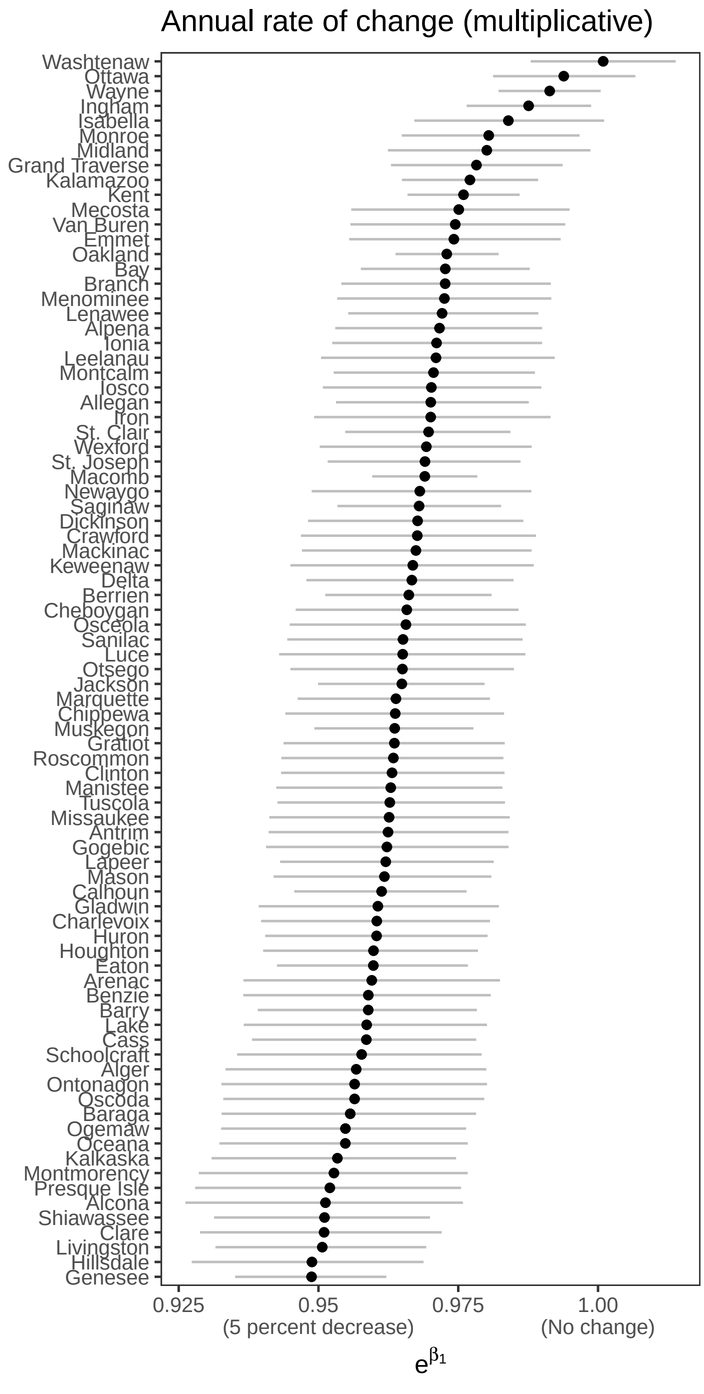


Figure 4: Posterior means for the annual percent change in crash rate ($e^{\beta_{1c}}$), removing the effect of unemployment rate, income and land area. Grey bars are 80-percent credible intervals.

# References


[1] Lindsay Knake. "Car-bike crashes drop in Ann Arbor, but fatalities increase". In: *The Ann Arbor News* (July 1, 2016). URL: http://www.mlive.com/news/ann-arbor/index.ssf/2016/07/ann_arbor_sees_decline_in_car-.html.

[2] Katrease Stafford. "Fatal bicyclist crashes surged 57% in Michigan last year". In: *Detroit Free Press* (July 13, 2016). URL: http://www.freep.com/story/news/local/michigan/2016/07/13/fatal-bicyclist-crashes-surged-57-michigan-last-year/86809398/.

[3] Michigan Office of Highway Safety Planning. *Michigan Traffic Crash Facts*. 2017. URL: https://www.michigantrafficcrashfacts.org/.

[4] Management Michigan Department of Technology and Budget. *Michigan Bureau of Labor Market Information and Strategic Initiatives*. Data Search. 2017. URL: http://milmi.org/datasearch.

[5] United States Census Bureau. *2016 U.S. Gazetteer Files*. 2016. URL: https://www.census.gov/geo/maps-data/data/gazetteer2016.html.

[6] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.14.0*. 2016. URL: http://mc-stan.org.

[7] Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. "Generating random correlation matrices based on vines and extended onion method". In: *Journal of Multivariate Analysis* 100.9 (2009), pp. 1989–2001.

[8] Stan Development Team. *RStan: the R interface to Stan*. R package version 2.14.1. 2016. URL: http://mc-stan.org/.