The Report Committee for Brian Russell Knab certifies that this is the approved version of the following report:

Infectious Disease and the South Texas Colonias

APPROVED BY

SUPERVISING COMMITTEE:

Sahotra Sarkar, Supervisor

Peter Ward

Infectious Disease and the South Texas Colonias

by

Brian Russell Knab, B.A.

REPORT

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE IN STATISTICS

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2016

Infectious Disease and the South Texas Colonias

Brian Russell Knab, M.S. Stat.
The University of Texas at Austin, 2016

Supervisor: Sahotra Sarkar

In this study, I investigated infectious disease in Texas, with a focus on the impacts of poverty and lack of infrastructure in the South Texas colonias on rates of infectious disease. I used Bayesian statistical methods, and in particular, hierarchical conditional autoregressive Poisson regression to model county-level rates of hospitalization across the state. According to that model, and with at least 97.5% probability, the average risk of hospitalization is greater in counties containing colonias as compared to counties which do not for the following infectious disease categories tracked by the Texas Department of State Health Services:

> Amebiasis, Brucellosis, Candidiasis, Chickenpox, Coccidoidomycosis, Ill defined intestinal infections, Intestinal infections due to other organisms, Bacterial food poisoning, Rickettsioses, Salmonella infections, Typhus, Viral Exanthemata, Pulmonary tuberculosis, Septicemia, Shigellosis, Diseases due to Coxsackie virus, Typhoid and paratyphoid fevers, and Whooping cough

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction, Data, and Methods

1.1 Introduction

The State of Texas describes the colonias on its southern border as residential areas "that may lack some of the most basic living necessities, such as potable water and sewer systems, electricity, paved roads, and safe and sanitary housing." The substandard living conditions, the State notes, which include "dilapidated homes, a lack of potable water and sewer and drainage systems, and floodplain locations" render "many colonias an ideal place for the proliferation of disease." [6]

This latter claim – that many colonias are an ideal place for the proliferation of disease – has been borne out by research. Researchers have found elevated rates of Hepatitis A [15], Cryptosporidium parvum [16], and human herpesvirus [7] in the children living in colonias, as compared to children living in urban border communities and San Antonio. Doyle and Bryan (2000) [9] – studying infectious disease morbidity over the entire US Region bordering Mexico – found elevated rates of morbidity due to botulism, brucellosis, diphtheria, hepatitis A, mumps, rabies, rubella, salmonellosis, and shigellosis in the U.S.-Mexico border region, as compared to the country as a whole.

The study that follows is very much in the vein of (Doyle and Bryan 2000) – I am interested in rates of infectious disease in the border counties – but with some major differences. First, my study is focused on Texas counties, and whether there is a measurable effect of the colonias on rates of infectious disease therein, where Doyle and Bryan were concerned with the entire border region spanning Texas, New Mexico, Arizona, and California. Second, Doyle and Bryan rely on morbidity as reported by the National Notifiable Diseases Surveillance System, where I rely on hospitalization data. Third, Doyle and Bryan report *raw morbidity risk ratios*, but they do not quantify nor allow for chance or uncertainty in those reports. They report, for example, that diptheria morbidity risk is 1.85 times higher in the border counties as compared to near-border counties. But while it is true that – over the 9 year period they studied (1990 - 1998) – the observed rate of diptheria morbidity was 1.85 times higher in the border counties, this was because of only 2 reported cases of diptheria over 9 years. Similarly, Doyle and Bryan report that plague and rabies morbidity risk was nearly 4 times higher (3.43, and 3.86 respectively) in border counties as compared to non-border counties, on the basis of 3 reported cases of rabies, and 4 reported cases of plague, in the border counties. The worry that arises in the face of such small numbers is that chance fluctuations might lead to drastically different conclusions. In the following, I avoid this problem by using Bayesian statistical methods to quantify the uncertainty in the risk estimates.

1.2 Data

I drew the data mainly from the Texas Department of State Health Services. From the years 1999 through 2009, the Department collected – from all Texas hospitals – information on every hospitalization that occurred in the state. That data includes, for each hospitalization, the principal diagnosis and the patient's zip code, from which I was able to compile a total hospitalization incidence count for 111 infectious disease categories, per year, for every county in the state.[4]

Beyond that, I also used data from the 2010 U.S. Census, to get the population, median income, and the demographic makeup of each Texas county.[1][5] Finally, to get the total number of colonias in each county, I counted the total number listed on the Texas Secretary of State website.[2]

1.3 Methods

The study proceeded in two stages.

1.3.1 Stage 1

Hospitals used 111 general diagnosis codes when reporting infectious disease hospitalizations to the Texas Department of State Health Services. Each of those codes corresponds to an infectious disease or infectious disease category. (There is, for example, one code for human immunodeficiency virus, and one code for 'other parapoxvirus infections,' which subsumes bovine stomatitis, and sealpox.) Appendix C contains a break down of those disease

categories that I will be discussing in more detail below.

In the first stage of the study, I calculated the statewide rate of hospitalization for each disease category for each year; i.e.

$$\hat{R}_{\text{Statewide},1999} = \frac{\text{Incidence in 1999}}{\text{Texas Population}}$$

Using that rate I derived an expected hospitalization incidence for each county, for each year, assuming equal risk across the state; i.e., assuming each person in the state was equally likely to be hospitalized as a result of a given disease.[1] For Hidalgo county, for example, that expected incidence is given by

$$\hat{E}_{\text{Hidalgo},1999} = \hat{R}_{\text{Statewide, }1999} \times \text{Hidalgo Population}$$

I also calculated the observed risk ratio, which again for Hidalgo county is given by

$$\hat{RR}_{\text{Hidalgo},1999} = \frac{\text{Observed Incidence in 1999}}{\hat{E}_{\text{Hidalgo},1999}}$$

This ratio represents the observed departure in Hidalgo county from the expected incidence under the assumption of equal statewide risk.

Finally, I computed the average observed risk ratio for the counties Hidalgo, Cameron, Starr, El Paso, Webb and Maverick, which are the counties with the largest numbers of colonias, and then selected for further study those diseases where that average was greater than 1; i.e., where, on average, the observed incidence was greater than expected under the assumption of equal statewide risk.

[1] This method suggested by Gelfand et al.[10, chapter 13]

1.3.2 Stage 2

The first stage yielded 26 diseases or disease categories where average hospitalization incidence appeared higher than expected in Hidalgo, Cameron, Starr, Maverick, El Paso, and Webb counties. In the second stage, I investigated those diseases more closely by modeling hospitalization incidence using Bayesian hierarchical conditional autoregressive Poisson regression.[2]

For each disease category, I assumed that hospitalization incidence, in a given year, followed a separate Poisson distribution in each county, with offset $\hat{R}_{\text{County}}$ – the expected count under the assumption of equal statewide risk, as described above. Thus

$$\text{Incidence}_i|\lambda_{\text{County}(i)} \sim \text{Poisson}(\hat{R}_{\text{County}(i)}\lambda_{\text{County}(i)})$$

I then used the percentage of the county population which is Hispanic or Latino, the median income for the county, and the number of colonias in the county to estimate $\log(\lambda_{\text{County}})$.

$$\log(\lambda_{\text{County}(i)}) = X\beta + \phi_{\text{County}(i)}$$

To account for any spatial autocorrelation in the Poisson means, I let the error

[2]The hierarchical part of the model is similar to that described in chapter 15 of (Gelman and Hill 2007)[12]

ϕ_i of county i be drawn from a conditional auto-regressive prior:[3]

$$\phi_i|\phi_{j,j\neq i}, \alpha, \tau_\phi \sim N(\alpha\frac{1}{N_j}\sum_{i\sim j}\phi_j, \tau_\phi^{-1})$$

where $i \sim j$ if county j neighbors i and

N_j is the number of counties bordering county i.

Finally, I completed the model with uninformative priors on the remaining parameters, including the regression coefficients.

$$\beta \sim N(0, 5) \qquad\qquad \alpha \sim \text{Unif}(0, 1)$$

$$\tau_\phi \sim \text{Gamma}(2, 2)$$

I generated Markov Chain Monte Carlo simulations from the joint posterior distribution of all of the parameters with the language STAN. [4] Appendix D contains the STAN code.

My main concern, throughout the study, was the risk ratio of hospitalizations for a given disease. Let, for example, $\hat{R}_{\text{HIDALGO}}$ be the expected incidence in Hidalgo under the assumption of equal statewide risk, as described above. And let R_{HIDALGO} be the true rate in Hidalgo county. Then the *risk ratio* for Hidalgo county is given by

$$\lambda_{Hidalgo} = \frac{R_{\text{HIDALGO}}}{\hat{R}_{\text{HIDALGO}}}$$

[3]Conditional autoregressive models were first discussed in (Besag et al 1991). For advances see (Gelfand and Vounatsou 2003) and (Jin et al 2005).

[4]The speed of the MCMC simulation was greatly improved by adopting the exact sparse CAR specification of Max Joseph.[14]

The risk ratio for Hidalgo county, in other words, describes how much larger (or smaller) the rate of hospitalizations is in Hidalgo, as compared to the expected rate of hospitalizations in Hidalgo under the assumption of equal statewide risk. Notice that if the risk ratio for some disease in some county is 1, then the rate of hospitalizations due to that disease in that county is no larger than expected given the observed statewide risk of contracting that disease.

Chapter 2

Results

2.1 Stage 1 Results

There were 26 disease categories where hospitalization incidence was, on average, higher than expected under the assumption of equal statewide risk in Hidalgo, Cameron, Starr, Maverick, El Paso, and Webb counties. Those disease categories are:[1]

> Actinomycotic infections, Amebiasis, Brucellosis, Candidiasis, Chickenpox, Coccidiodomycosis, Herpes simplex, Herpes zoster, Ill defined intestinal infections, Intestinal Infections due to other organisms, Bacterial food poisoning, Ricekttsioses, Salmonella infections, Other enterovirus diseases of the central nervous system, Other non-arthropod-borne viral diseases of the central nervous system, Other protozoal intestinal diseases, Other typhus, Other viral exanthemata, Pulmonary tuberculosis, Septicemia, Shigellosis, Specific diseases due to Coxsackie virus, Streptococcal sore throat and scarlet fever, Typhoid and paratyphoid Fevers, Viral and chlamydical infections, and Whooping Cough

[1] See Appendix C for a description of which specific diseases these categories subsume.

Figure 2.1: Observed vs. Expected Incidence, Pulmonary Tuberculosis.

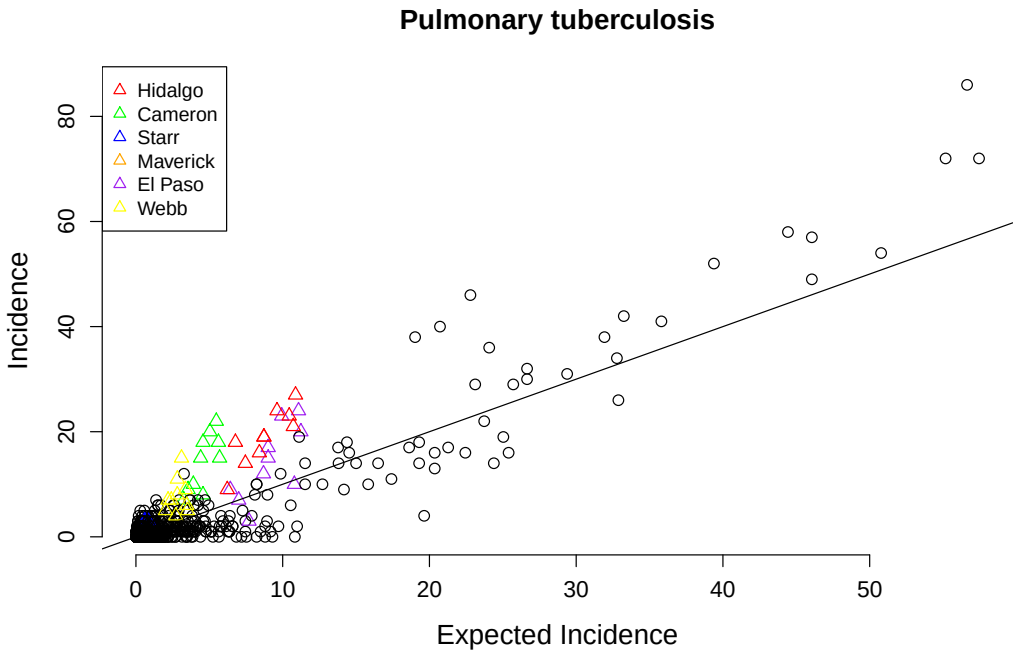


Figure 2.1, for example, is a plot of the observed hospitalization incidence against the expected incidence under equal statewide risk for pulmonary tuberculosis. The solid line has unit slope, and each point is an observed incidence in a particular county. Thus a point will fall on that line, just in case the observed incidence in the associated county equals the expected incidence. Notice that all of Hidalgo, Cameron, Starr, Maverick, El Paso and Webb counties appear to have higher average incidence of pulmonary tuberculosis than expected. Similar plots, for the other 25 disease categories described above can be found in appendix B.

2.2 Stage 2 Results

2.2.1 Summary Results

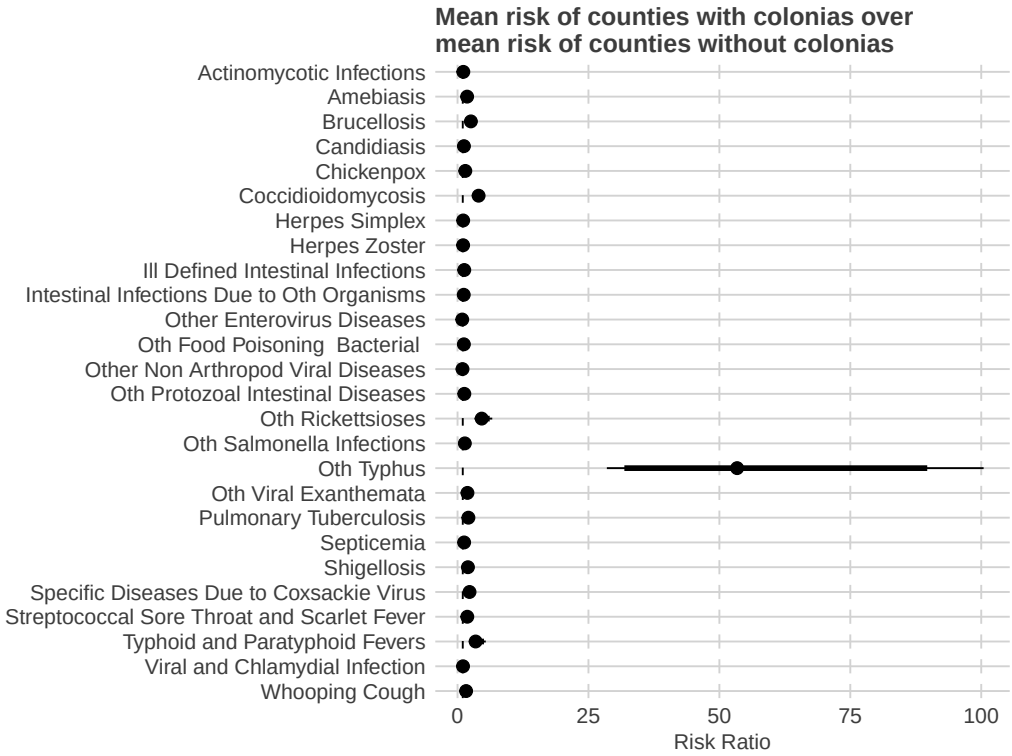
Turning to stage 2, I begin with some summary results. Figure 2.2 shows the 95% highest posterior density credible intervals for the mean of the risk ratio, for counties which contain colonias, divided by the mean risk ratio for counties which do not contain colonias; i.e.,

$$\Delta = \frac{\texttt{mean}(\Lambda_C)}{\texttt{mean}(\Lambda_N)} \quad \text{where} \begin{cases} \lambda_{\text{County}(i)} \in \Lambda_C & \text{if County(i) contains colonias,} \\ \lambda_{\text{County}(i)} \in \Lambda_N & \text{Otherwise.} \end{cases}$$

The disease category *Other Typhus* stands out in figure 2.2; the average *Other Typhus* risk ratio for counties with colonias is with 95% probability between 25 and 100 times larger than the average risk ratio in counties which do not contain colonias.

Figure 2.3 shows the same plot as Figure 2.2, but excluding *Other Typhus*, to increase fidelity for the other disease categories. Table 2.1 contains the estimates and 95% credible intervals depicted in figures 2.2 and 2.3. And what table 2.1 and figures 2.2 and 2.3 show is that, for 24 of the 26 disease categories under study, the estimated average risk ratio in counties which contain colonias is more than 100% of the average risk ratio in counties which do not contain colonias. Each, however, of the 95% credible intervals for Actinomycotic infections, Herpes simplex, Herpes zoster, and Viral and chlamydial infections, extends below 1. Thus, for each of those disease categories, the probability that – on average – counties which contain colonias have a lower risk of hospitalization than counties which do not is non-neglible; it is at least 2.5%.

Figure 2.2: Mean risk ratio of counties with colonias over mean risk of ratio counties without colonias.

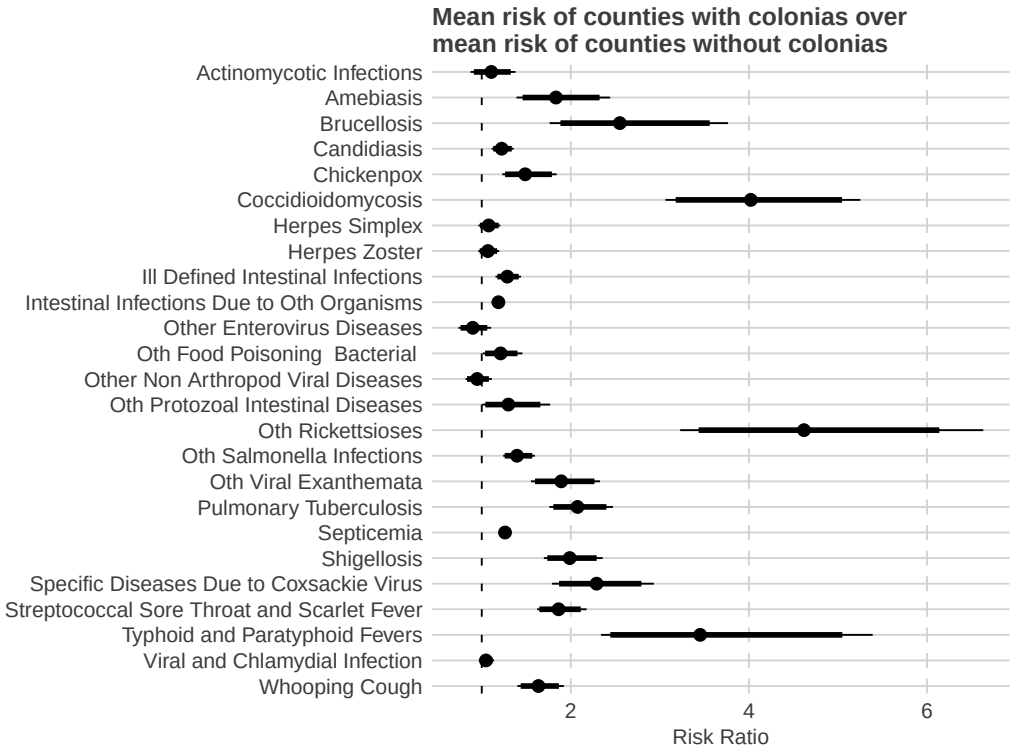


2.2.2 Specific Results

Turning to more specific results, I will describe the results for amebiasis in detail; the general lessons will then extend to the other 25 disease categories studied closely here.

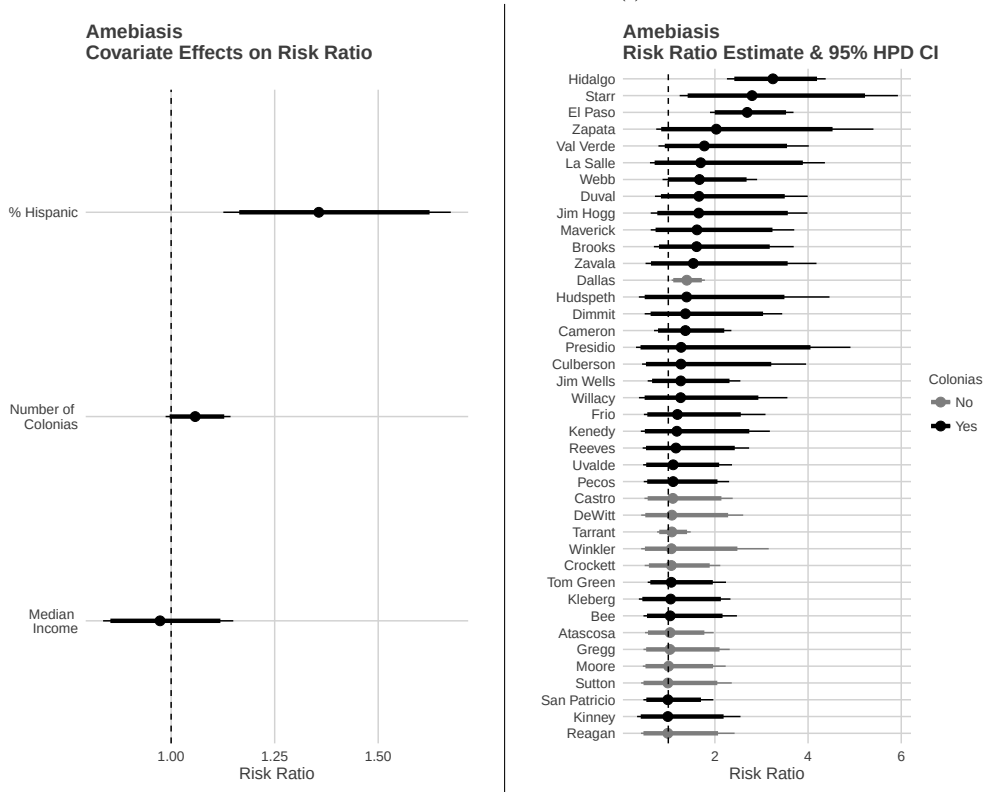
The two plots in figure 2.4 are similar; they both show posterior estimates, and 95% highest posterior density credible intervals. Figure 2.4a shows the credible intervals for the effect of the standardized covariates on the average risk ratio. It gives, in other words, estimates and 95% posterior credible

Figure 2.3: Mean risk ratio of counties with colonias over mean risk of ratio counties without colonias, excluding 'Other Typhus'.



intervals for the average effect of a single standard deviation change in each of the studied covariates on the risk ratio for any given county. According to Figure 2.4a, the model estimates that as the percentage of the county population and the number of colonias which is hispanic or latino increases, and as the median income decreases, the risk of amebiasis tends to increase. Figure 2.4b shows the 95% credible interval for the risk ratio, $\lambda_{\text{COUNTY}(i)}$, for the 40 counties with the highest median risk ratios. In Hidalgo county, for example, the rate of Amebiasis hospitalization is, with 95% probability, between 2 and 4

Figure 2.4: Amebiasis; Medians and 95% Highest Posterior Density Credible Intervals for County Level Risk Ratios, $\lambda_{\text{COUNTY}(i)}$, and Covariate Effects



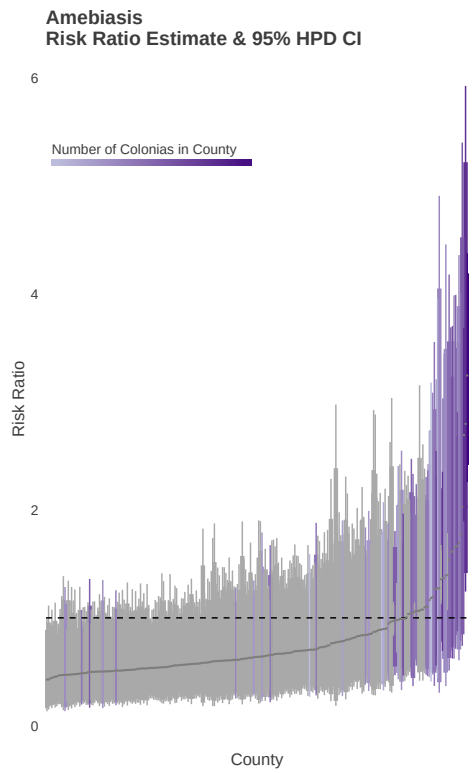
(a) Covariate Effects

(b) 40 Highest Median Counties

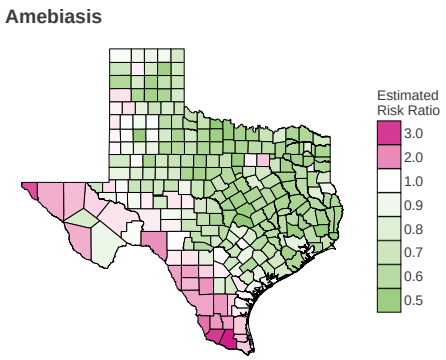
times higher than you'd expect under the assumption of equal statewide risk. Those counties in figure 2.4b which contain colonias are highlighted.

Turning to Figure 2.5, Figure 2.5a shows the 95% credible intervals for the county level risk ratios for all of the 254 counties in Texas. The purple gradient shows the number of colonias in a county, as reported by the Texas Secretary of State website. (Counties without any reported colonias are left grey.) Figure 2.5b is a map of the posterior estimate of the risk ratio in each

Figure 2.5: Amebiasis Risk Ratio 95% HPD CI by county, and Estimated Risk Ratio Map



(a) 95% HPD CI by County



(b) Risk Ratio Estimate Map

Texas county.

The plots in Figure 2.5 display clear patterns; generally speaking, counties which contain colonias, and which are along the border, have higher amebiasis risk ratios than those which do not contain colonias and are not along the border. Note also that among the counties whose estimated risk ratio *is* greater than 1, counties containing colonias are overrepresented. This is consistent with the results shown in Figure 2.3, and in Table 2.1.

Similar plots and maps for the 25 other disease categories I studied can be found in Appendix A.

Table 2.1: Median and 95% Highest Posterior Density Credible Interval for $\Delta = \frac{\texttt{mean}(\Lambda_C)}{\texttt{mean}(\Lambda_N)}$

Disease/Category	Median Δ	Δ 95% HPD CI
Actinomycotic Infections	1.11	$(0.89, 1.38)$
Amebiasis	1.84	$(1.38, 2.52)$
Candidiasis	1.22	$(1.11, 1.36)$
Chickenpox	1.48	$(1.22, 1.84)$
Brucellosis	2.56	$(1.75, 3.92)$
Coccidioidomycosis	4.05	$(3.06, 5.34)$
Herpes Simplex	1.08	$(0.96, 1.23)$
Herpes Zoster	1.07	$(0.96, 1.95)$
Ill Defined Intestinal Infections	1.28	$(1.16, 1.43)$
Other Enterovirus Diseases	0.90	$(0.74, 1.10)$
Intestinal Infections Due to Other Organisms	1.18	$(1.13, 1.25)$
Other Food Poisoning - Bacterial	1.22	$(1.01, 1.46)$
Other Non-Arthropod-Borne Viral Diseases	0.95	$(0.81, 1.12)$
Other Protozoal Intestinal Diseases	1.31	$(1.00, 1.76)$
Other Rickettsioses	4.60	$(3.23, 6.54)$
Other Salmonella Infections	1.40	$(1.22, 1.61)$
Other Typhus	53.86	$(28.33, 100.20)$
Other Viral Exanthemata	1.89	$(1.53, 2.33)$
Pulmonary Tuberculosis	1.76	$(1.42, 2.17)$
Septicemia	1.26	$(1.22, 1.30)$
Shigellosis	1.98	$(1.67, 2.35)$
Diseases Due to Coxsackie Virus	2.28	$(1.79, 2.94)$
Strep Throat and Scarlet Fever	1.85	$(1.62, 2.18)$
Typhoid and Paratyphoid Fevers	3.41	$(2.25, 5.40)$
Viral and Chlamydial Infection	1.04	$(0.96, 1.14)$
Whooping Cough	1.64	$(1.39, 1.94)$

Chapter 3

Discussion

For 24 of the 26 diseases investigated closely in this study, the estimated average risk ratio of hospitalization as a result of that disease, in counties containing colonias, is greater than the estimated average risk ratio in counties which do not contain colonias (See Table 2.1). Those diseases where the average risk ratio, in counties containing colonias, is – with at least 97.5% probability – higher than in counties without them, include

> Amebiasis, Brucellosis, Candidiasis, Chickenpox, Coccidoidomycosis, Ill defined intestinal infections, Intestinal infections due to other organisms, Protozoal intestinal diseases, Bacterial food poisoning, Rickettsioses, Salmonella infections, Typhus, Viral exanthemata, Pulmonary tuberculosis, Septicemia, Shigellosis, Diseases due to Coxsackie virus, Typhoid and paratyphoid fevers, and Whooping cough

The estimated difference in risk across counties containing colonias and those which do not, however, varies widely depending on the disease in question. The average risk of viral and chlamydial infection is estimated to be only 4% higher in counties which contain colonias, as compared to counties which do

not. The average risk of *Other Typhus* – a category which subsumes murine typhus, Brill's disease, and scrub typhus – is estimated at over 5000% higher in counties containing colonias as compared to those which do not. This latter is a consequence of the fact that almost all of the typhus hospitalizations which occurred between 1999 and 2009 were from patients living in the southernmost Texas counties (see the map in Figure A.14). Most of the hospitalizations – 834 of the total 1475 – were from patients living in one county alone, Hidalgo, which is the county with the largest number of colonias (933 are recognized on the Texas Secretary of State websie).

Hidalgo county is, in general, something of an outlier with respect to the diseases here studied. It is among the 15 counties with the highest estimated risk ratios for Amebiasis, Brucellosis, Herpes simplex, Ill defined intestinal infections, Bacterial food poisoning, Other rickettsioses, Other typhus, Other viral exanthemata, Pulmonary tuberculosis, Shigellosis, Specific diseases due to coxsackie virus, Streptococcal sore throat, Typhoid and paratyphoid fevers, and Whooping cough. (See the figures in Appendix A.)

That said, to the extent that it is the colonias that are effecting the risk of infectious disease in Hidalgo, that effect is not uniform across the state. There were zero reported cases, for example, of *Other Typhus* hospitalization in El Paso county over the 10 year study period, but there are 330 State recognized colonias in El Paso county. A similar thing is true for the category *Other Rickettsioses* – which subsumes Q Fever, Trench Fever, and Rickettsialpox – and the category *Typhoid and paratyphoid fevers*. The maps in appendix A

show that it is only southern Texas counties at an elevated risk of hospitalization due to these diseases. Many things could explain this, and further study is needed. Perhaps the difference between southern and western counties which contain colonias is due to differences in the conditions in the southern and western colonias. Or perhaps the difference is attributable to differences in environmental suitability for the diseases.

Turning to our covariates, Table 3.1 shows that for 24 of the 26 diseases under study, the 95% highest posterior density credible intervals for the effect of the number of colonias contains 1. Thus there is at least a 2.5% chance that – holding a county's demographics and median income fixed – an increase in the number of colonias in that county would have no effect, or even a negative effect, on the risk ratio for 24 of the 26 diseases. The only diseases this is not true of are Shigellosis, and Herpes Simplex, for which the 95% highest posterior density credible intervals for the effect of the number of colonias are $(1.04, 1.27)$ and $(1.02, 1.15)$ respectively.

But these results are deceptive, as the covariates here studied are correlated (see Table 3.2). As the number of colonias increases, the percentage of the county population which is Hispanic or Latino tends to increase, and median income tends to decrease. These correlations induce collinearity among the covariates, which inflates the variance of the covariate parameters, and leads to instability in those parameter estimates. (The pairwise correlations are not that strong, but see (Vatcheva et al (2016))[17] which shows how even weak pairwise correlations can have significant effects.) This is an unavoid-

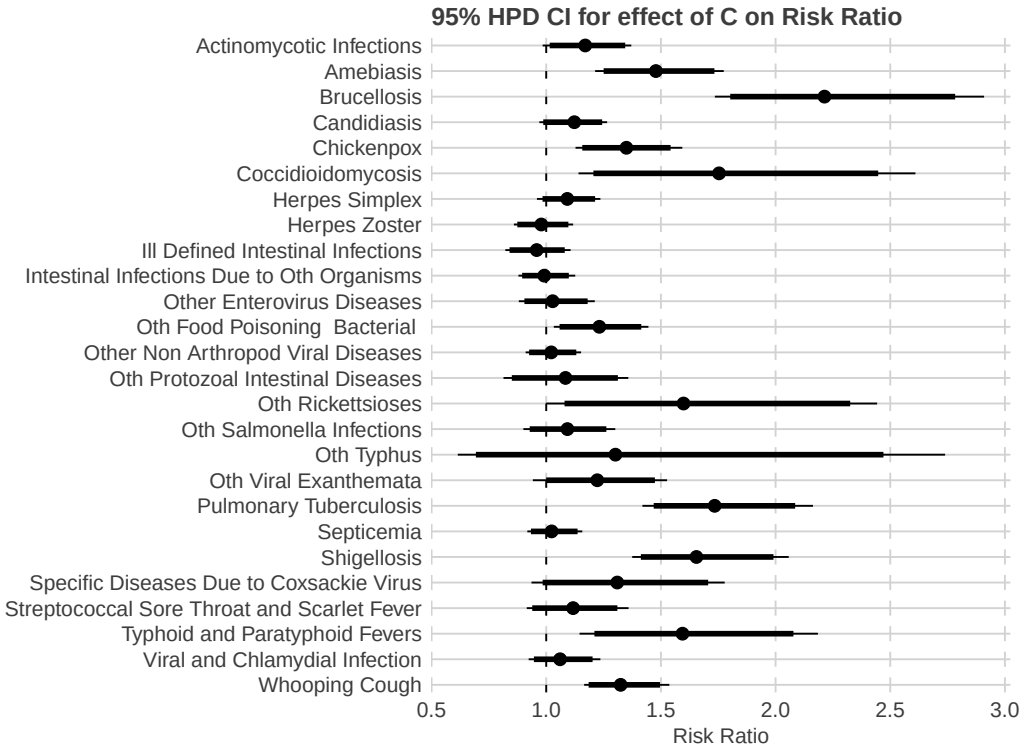
able consequence of the county-level data here considered; those counties which contain colonias are also, on average, the counties with the lowest median incomes, and with the highest percentages of Hispanics or Latinos. It is therefore difficult to discern any effect of the colonias *themselves*, apart from the effects of concomitant differences of demographics and income.

However, given the colonias tend to be majority (or entirely) Hispanic or Latino, and residents tend to have low incomes, we can consider the following quantity:

$$\mathrm{C} = \beta_{\%\ \textsc{Hisp}} + \beta_{\textsc{Num Colonias}} - \beta_{\textsc{Income}}$$

C describes the combined effect of increases in the percentage of hispanics or latinos in the county population, increases in the number of colonias in the county, and decreases in the median income. Figure 3.1 shows 90% and 95% highest posterior density credible intervals for C. And what Figure 3.1 shows is that the combined effect of a unit increase in the percentage of the population which is Hispanic or Latino, a unit increase in the number of colonias, and a unit decrease in the median income, is – with probability greater than 97.5% – an amplifying effect for Amebiasis, Brucellosis, Chickenpox, Coccidiomycosis, Bacterial Food Poisoning, Rickettsioses, Pulmonary Tuberculosis, Shigellosis, Typhoid and Paratyhpoid fevers, and Whooping Cough. The effect is largest for Amebiasis, Brucellosis, Pulmonary Tuberculosis, and Shigellosis. Thus an increase in the number of colonias in a county – along with a (typically concomitant) increase in the percentage of Hispanics or Latinos, and a (typically concomitant) decrease in the median income – with probability greater than

Figure 3.1: Effect of C = $\beta_{\%\ \text{Hisp}} + \beta_{\text{Num Colonias}} - \beta_{\text{Income}}$ on Risk Ratio



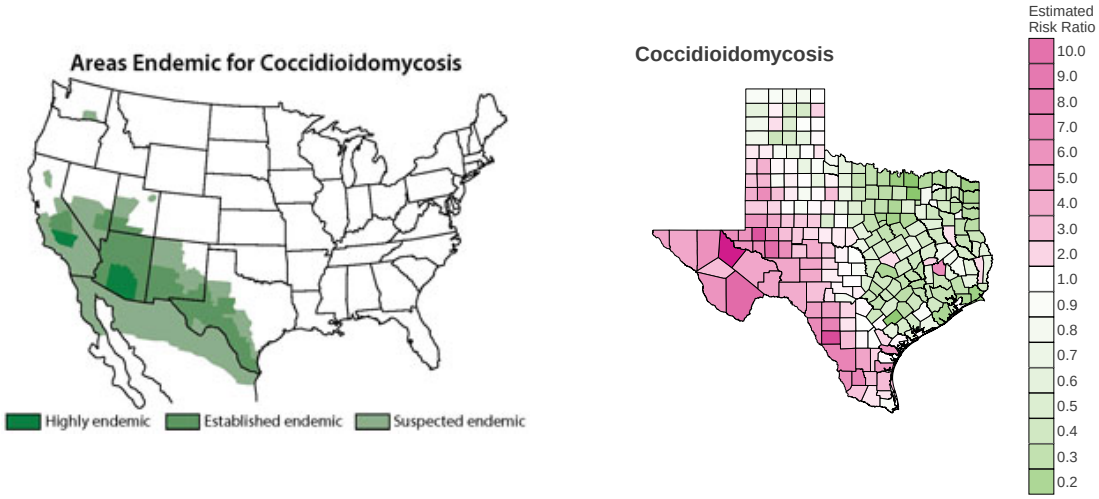
97.5%, will amplify the risk of those infectious diseases. Table 3.3 displays the estimates and 95% highest posterior density credible intervals for C, as computed for Amebiasis, Brucellosis, Chickenpox, Coccidiomycosis, Bacterial Food Poisoning, Rickettsioses, Pulmonary Tuberculosis, Shigellosis, Typhoid and Paratyphoid Fevers, and Whooping Cough.

The diseases for which increases in C – with at least 95% probability – lead to increased risk of hospitalization can be split into three categories:

Coccidioidomycosis is in a category on its own; it is a pulmonary infection which results from the inhalation of the fungus *Coccidioides*. That fungus is only present, in Texas, in the soil of West Texas. Figure 3.3 strongly suggests that it is simply those counties in which the fungus *Coccidioides* "is known to live or suspected to live" (according to the CDC)[3] that are at an elevated risk of Coccidioidomycosis. The elevated risk in counties containing colonias, in other words, potentially has little to do with the particular conditions of the colonias, and more to do with the coincidence of counties containing colonias, and those where the fungus *Coccidioides* lives. In fact, similar environmental coincidents could explain the elevated risks for all the of the diseases described in Table 3.2, but if so those environmental coincidents are less obvious than in the case of Coccidioidomycosis. Increased *hospitalization* from Coccidiomycosis could however be the result of a lack of access to non-emergency medical care in the colonias; again, further study is needed.

A second category – of diseases for which increases in C has an amplifying effect – includes Amebiasis, Brucellosis, Bacterial Food Poisoning, Shigellosis, Other Rickettsioses, and Typhoid and Paratyphoid fevers. *Other Rickettsioses* describes *non-tick-borne* rickettsioses (see Appendix C), including Q fever, Trench fever, and Rickettsialpox, each of which is transmitted by lice, mites and close contact with animals. Amebiasis is "most common in tropical areas with poor santitation."[3] Brucellosis is "an infection spread from animals to people, mostly by unpasteurized dairy products."[3] Bacterial Food Poisoning is caused by "Campylobacter, Salmonella, Shigella, E. Coli,

Figure 3.2: Approximate areas where Coccidiodes is known to live vs. Estimated Risk Ratio map



Listeria, botulism, [and] norovirus"[3] present in food. Shigellosis is a highly contagious diahreaal disease, which is spread through contact with an infected persons feces.[3] Typhoid fever is a bacterial disease "spread by contaminated food and water."[3] Each of these, in other words, are diseases which are exacerbated by poor sanitation. Recalling the State of Texas' claim that the colonias "lack some of the most basic living necessities ... [including] safe and sanitary housing," and that many "lack potable water and sewage and

drainage systems,"[6] it is plausible that the conditions in the colonias are directly responsible for an increased rate of hospitalization due to these diseases over the study period.

The last category – of diseases for which increases in C lead to increases in risk – includes Chickenpox, Pulmonary tuberculosis, and Whooping cough. The latter two are both respiratory diseases which are spread through contact with an infected person. Chickenpox and Whooping cough are both highly contagious and vaccine preventable. Pulmonary tuberculosis is less contagious, and partly vaccine preventable. The State of Texas notes that, in the colonias, "a lack of medical services compounds health problems,"[6] and here again it is plausible that this lack is directly contributing to increased hospitalizations as a result of Chickenpox, Pulmonary tuberculosis and Whooping cough.

3.1 Conclusions

For many of the infectious diseases studied here, the risks of hospitalization due to those diseases was higher in counties containing colonias as compared to counties which do not contain colonias. But importantly – if the increased risk of those diseases is due to the colonias – the effect is not uniform across the state, and the colonias are not monolithic. Counties in the southernmost part of the state, and in particular Hidalgo county, are the counties where the risks for many of the infectious diseases studied here appears to be greatest.

Table 3.1: Estimated effect, and 95% HPD CI for the effect of a single standard deviation increase in the number of colonias in the county, holding income and demographics fixed.

Disease	# Colonias Effect Estimate	Effect 95% HPD CI
Actinomycotic infections	0.98	$(0.91, 1.07)$
Amebiasis	1.05	$(0.98, 1.14)$
Brucellosis	1.06	$(0.98, 1.17)$
Candidiasis	0.99	$(0.93, 1.05)$
Chickenpox	1.00	$(0.94, 1.08)$
Coccidioidomycosis	0.99	$(0.86, 1.15)$
Herpes Simplex	1.08	$(1.02, 1.15)$
Herpes Zoster	1.00	$(0.95, 1.06)$
Ill Defined Intestinal Infections	1.04	$(0.98, 1.10)$
Intestinal Infections due to other Organisms	1.03	$(0.97, 1.10)$
Other enterovirus diseases of cent nerv sys	1.04	$(0.97, 1.12)$
Other Food Poisoning Bacterial	1.05	$(0.98, 1.13)$
Other non-arthropod-borne viral diseases	0.99	$(0.93, 1.06)$
Other protozoal intestinal diseases	1.03	$(0.94, 1.15)$
Other Rickettsioses	1.04	$(0.92, 1.16)$
Other Salmonella Infections	1.02	$(0.96, 1.10)$
Other Typhus	1.07	$(0.90, 1.28)$
Other Viral Exanthemata	1.03	$(0.96, 1.11)$
Pulmonary Tuberculosis	1.00	$(0.93, 1.07)$
Septicemia	1.00	$(0.94, 1.06)$
Shigellosis	1.15	$(1.04, 1.27)$
Specific Diseases due to Coxsackie Virus	1.07	$(0.97, 1.19)$
Strep Throat and Scarlet Fever	1.00	$(0.93, 1.08)$
Typhoid and paratyphoid Fevers	1.03	$(0.93, 1.15)$
Viral and Chlamydial Infection	1.06	$(0.98, 1.13)$
Whooping Cough	1.04	$(.98, 1.107)$

Table 3.2: Estimated pairwise correlations, and p-values under the assumption of no correlation, for studied covariates

	Med Income	# of Colonias	% Hispanic
Median Income	-	$\hat{r} = -0.14$, $p = .017$	$\hat{r} = -0.24$, $p < .001$
# of Colonias	$\hat{r} = -0.14$, $p = .017$	-	$\hat{r} = .33$, $p < .001$
% Hispanic	$\hat{r} = -0.24$, $p < .001$	$\hat{r} = .33$, $p < .001$	-

Table 3.3: Estimated effect, and 95% HPD CI for the effect of a single standard deviation increase in the number of colonias in the county, holding income and demographics fixed.

Disease	# Colonias Effect Estimate	Effect 95% HPD CI
Amebiasis	1.05	$(0.98, 1.14)$
Brucellosis	1.06	$(0.98, 1.17)$
Coccidioidomycosis	0.99	$(0.86, 1.15)$
Herpes Simplex	1.08	$(1.02, 1.15)$
Ill Defined Intestinal Infections	1.04	$(0.98, 1.10)$
Intestinal Infections due to other Organisms	1.03	$(0.97, 1.10)$
Other Food Poisoning Bacterial	1.05	$(0.98, 1.13)$
Other Rickettsioses	1.04	$(0.92, 1.16)$
Other Salmonella Infections	1.02	$(0.96, 1.10)$
Other Typhus	1.07	$(0.90, 1.28)$
Other Viral Exanthemata	1.03	$(0.96, 1.11)$
Pulmonary Tuberculosis	1.00	$(0.93, 1.07)$
Septicemia	1.00	$(0.94, 1.06)$
Shigellosis	1.15	$(1.04, 1.27)$
Specific Diseases due to Coxsackie Virus	1.07	$(0.97, 1.19)$
Strep Throat and Scarlet Fever	1.00	$(0.93, 1.08)$
Viral and Chlamydial Infection	1.06	$(0.98, 1.13)$
Whooping Cough	1.04	$(.98, 1.107)$

Table 3.4: Estimated effect, and 95% HPD CI for $C = \beta_{\%\ \text{Hisp}} + \beta_{\text{Num Colonias}} - \beta_{\text{Income}}$.

Disease	C Effect Estimate	Effect 95% HPD CI
Amebiasis	1.48	$(1.23, 1.79)$
Brucellosis	2.58	$(1.73, 3.01)$
Chickenpox	1.34	$(1.14, 1.59)$
Coccidioidomycosis	1.75	$(1.16, 2.57)$
Other Food Poisoning Bacterial	1.20	$(1.02, 1.45)$
Other Rickettsioses	1.60	$(1.06, 2.53)$
Pulmonary Tuberculosis	1.76	$(1.42, 2.17)$
Shigellosis	1.64	$(1.36, 1.98)$
Typhoid and Paratyphoid Fevers	1.61	$(1.17, 2.22)$
Whooping Cough	1.32	$(1.17, 1.52)$

Appendices

Appendix A

Plots: Particular Results, All Diseases

Figure A.1: Actinomycotic Infections

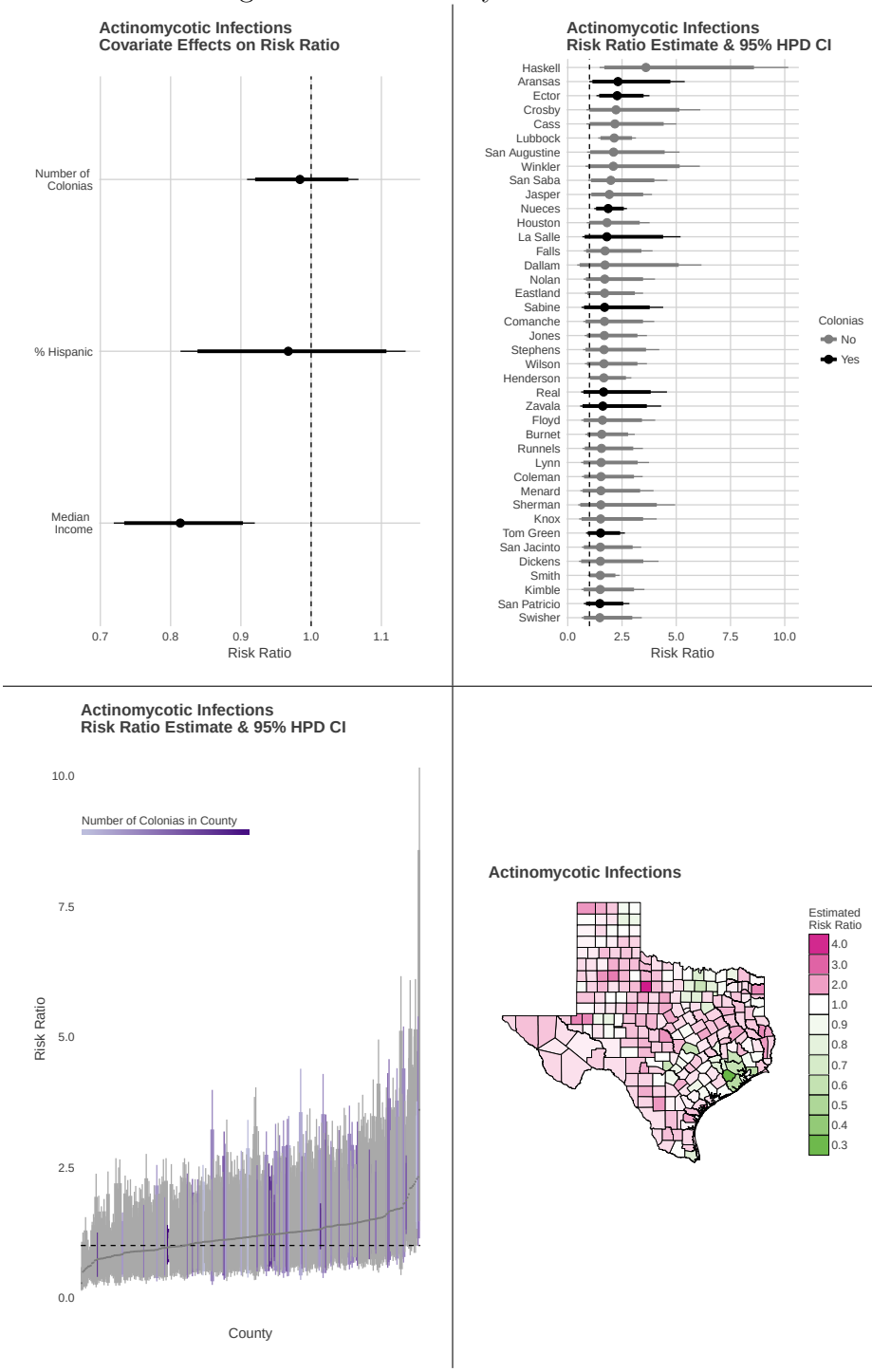


Figure A.2: Amebiasis

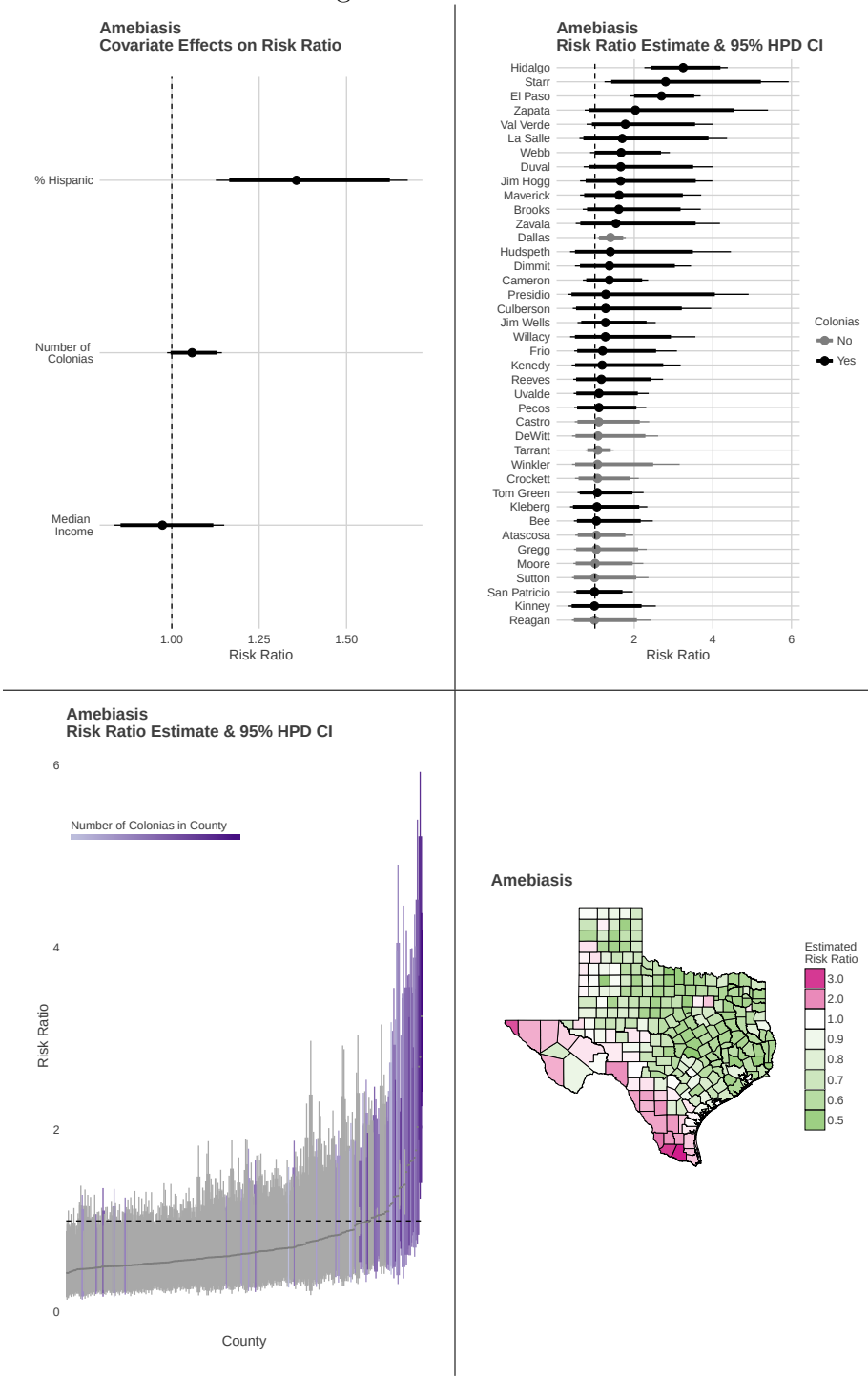


Figure A.3: Brucellosis

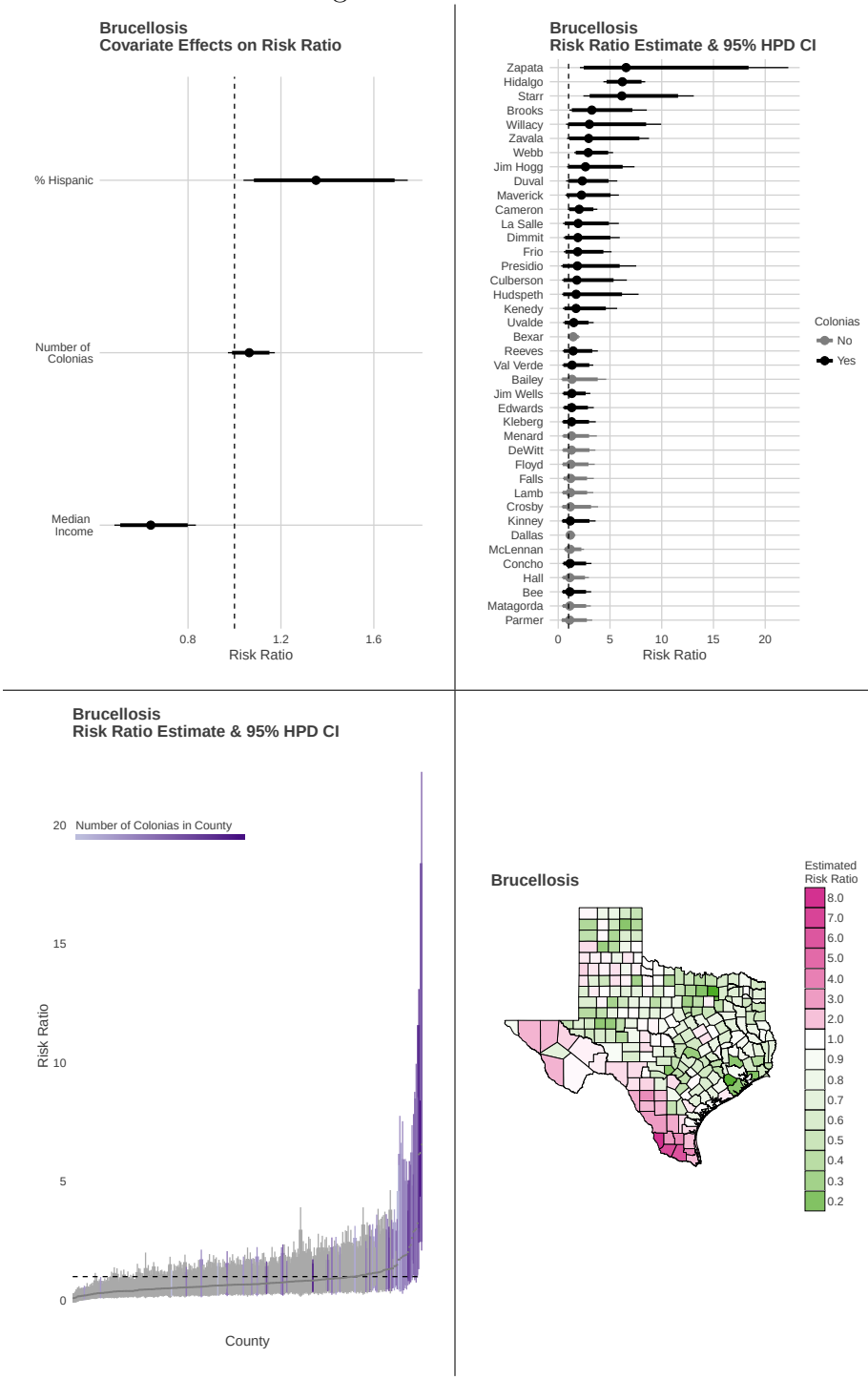


Figure A.4: Candidiasis

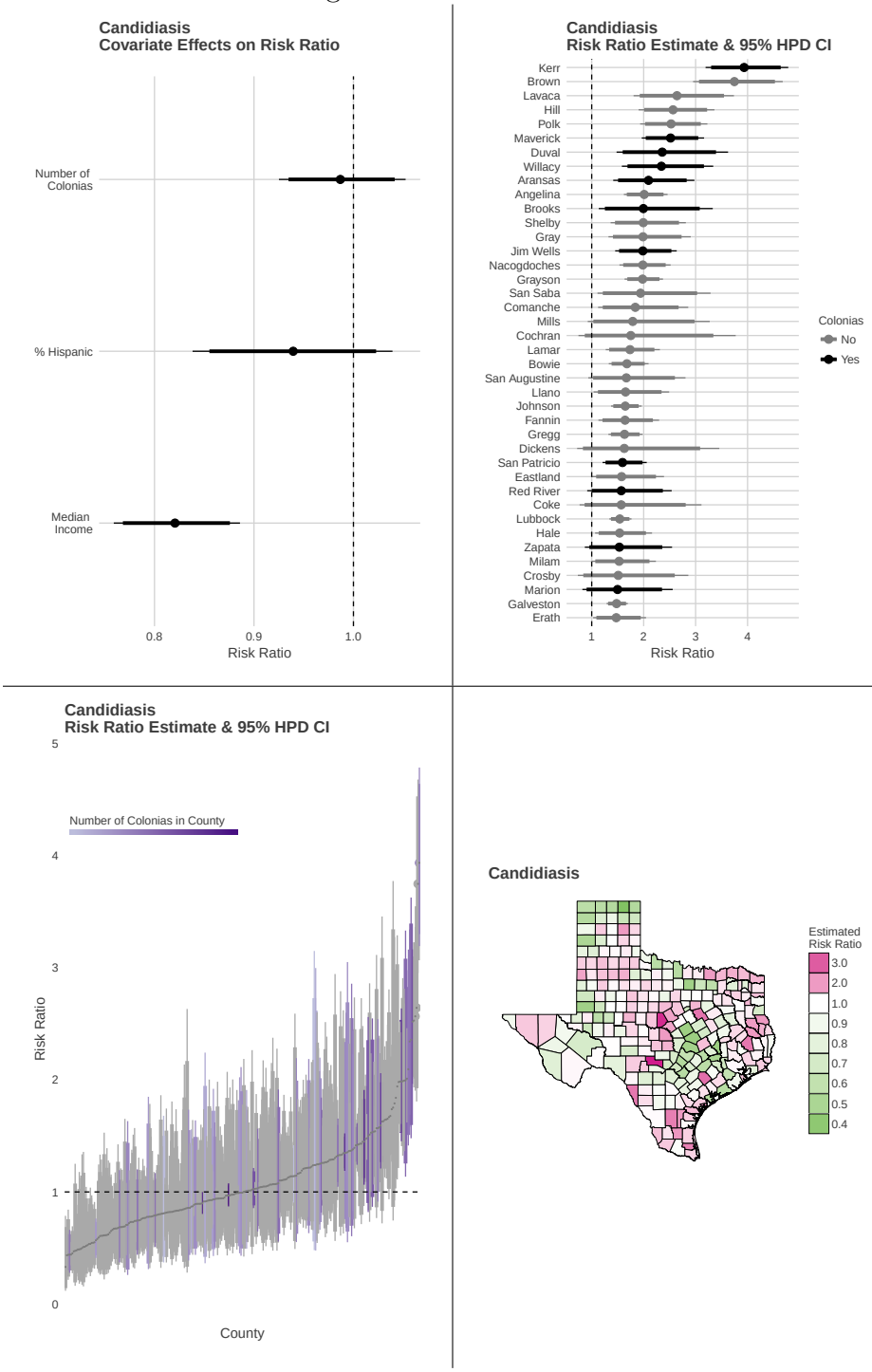


Figure A.5: Chickenpox

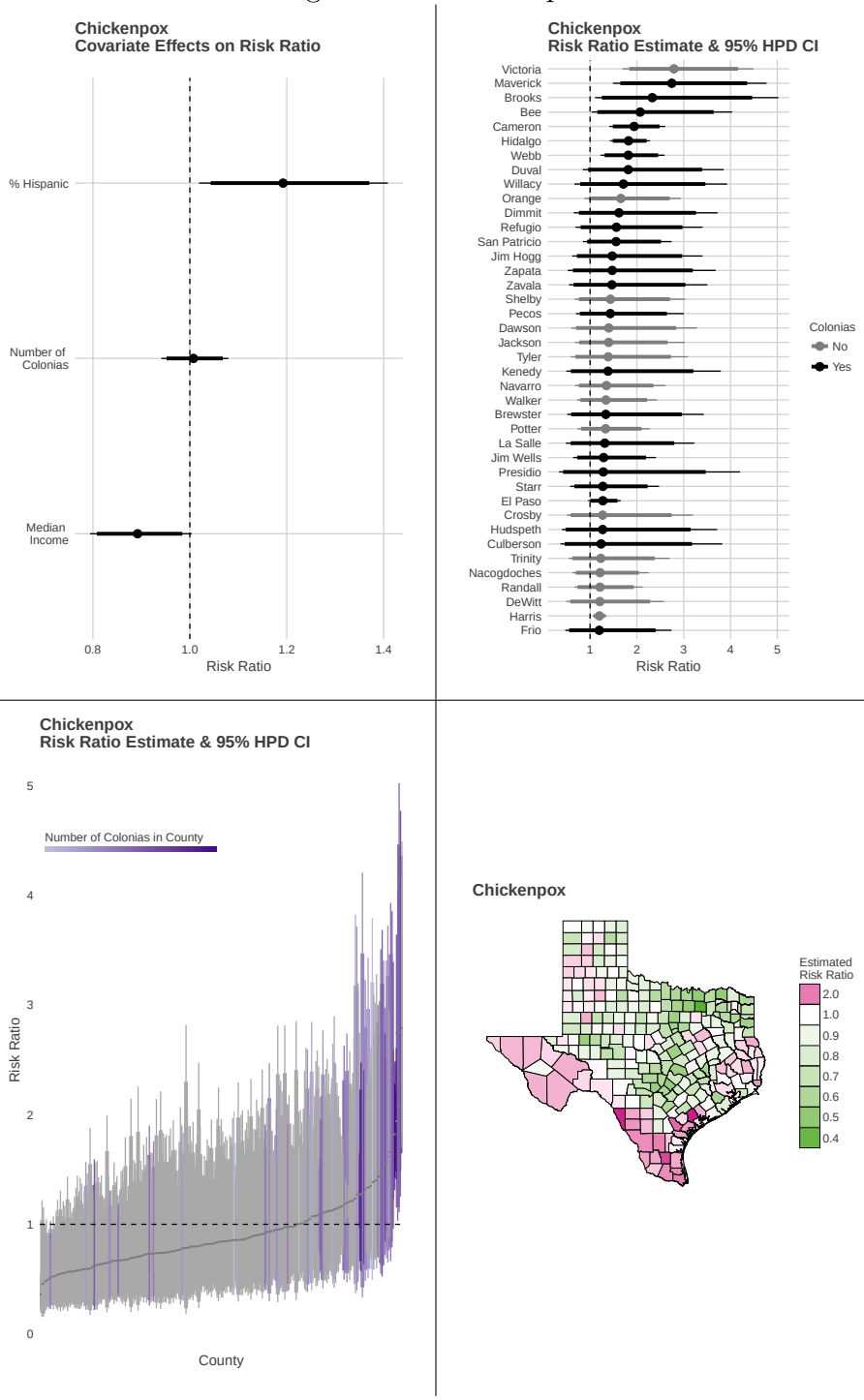


Figure A.6: Coccidioidomycosis

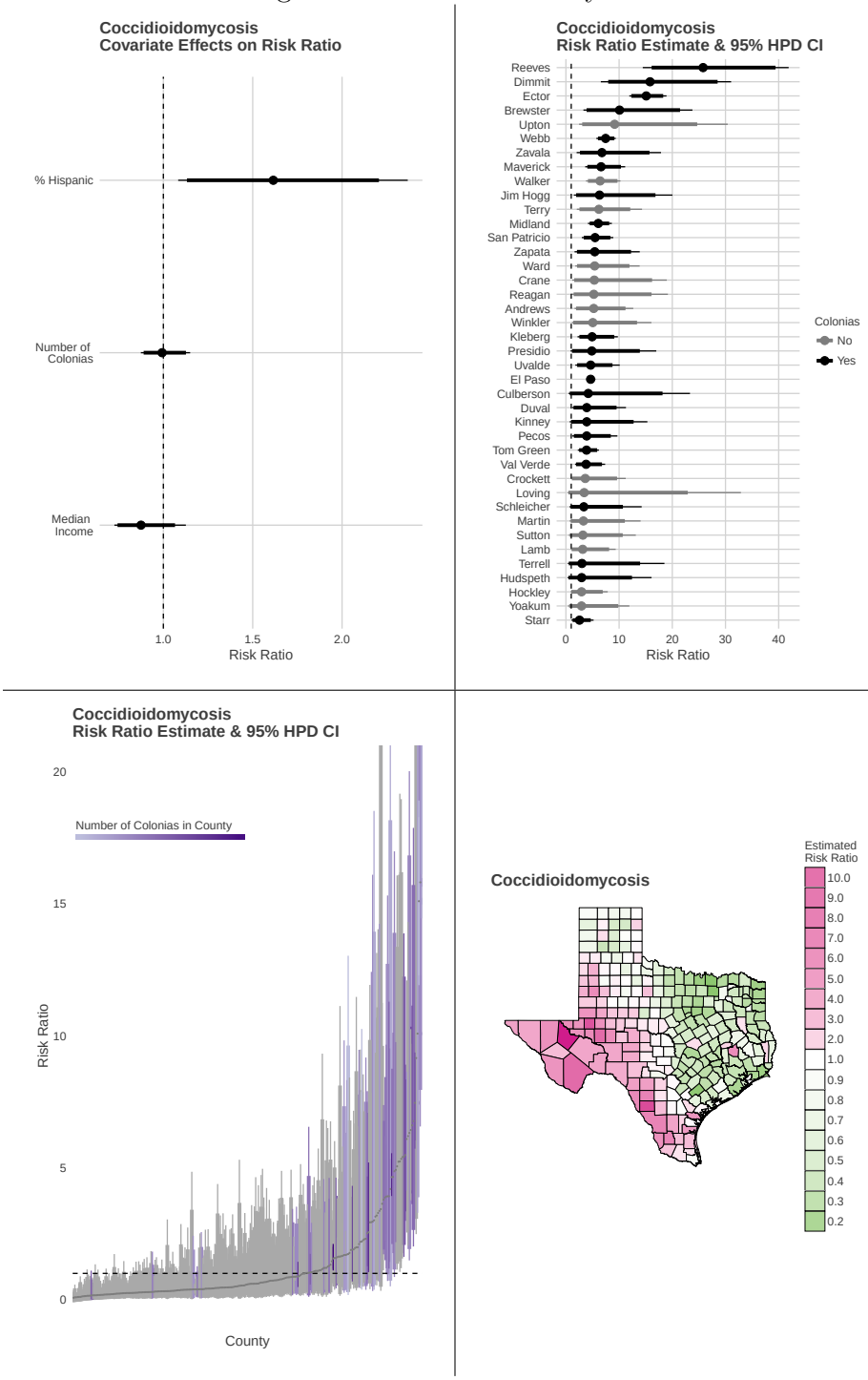


Figure A.7: Herpes Simplex

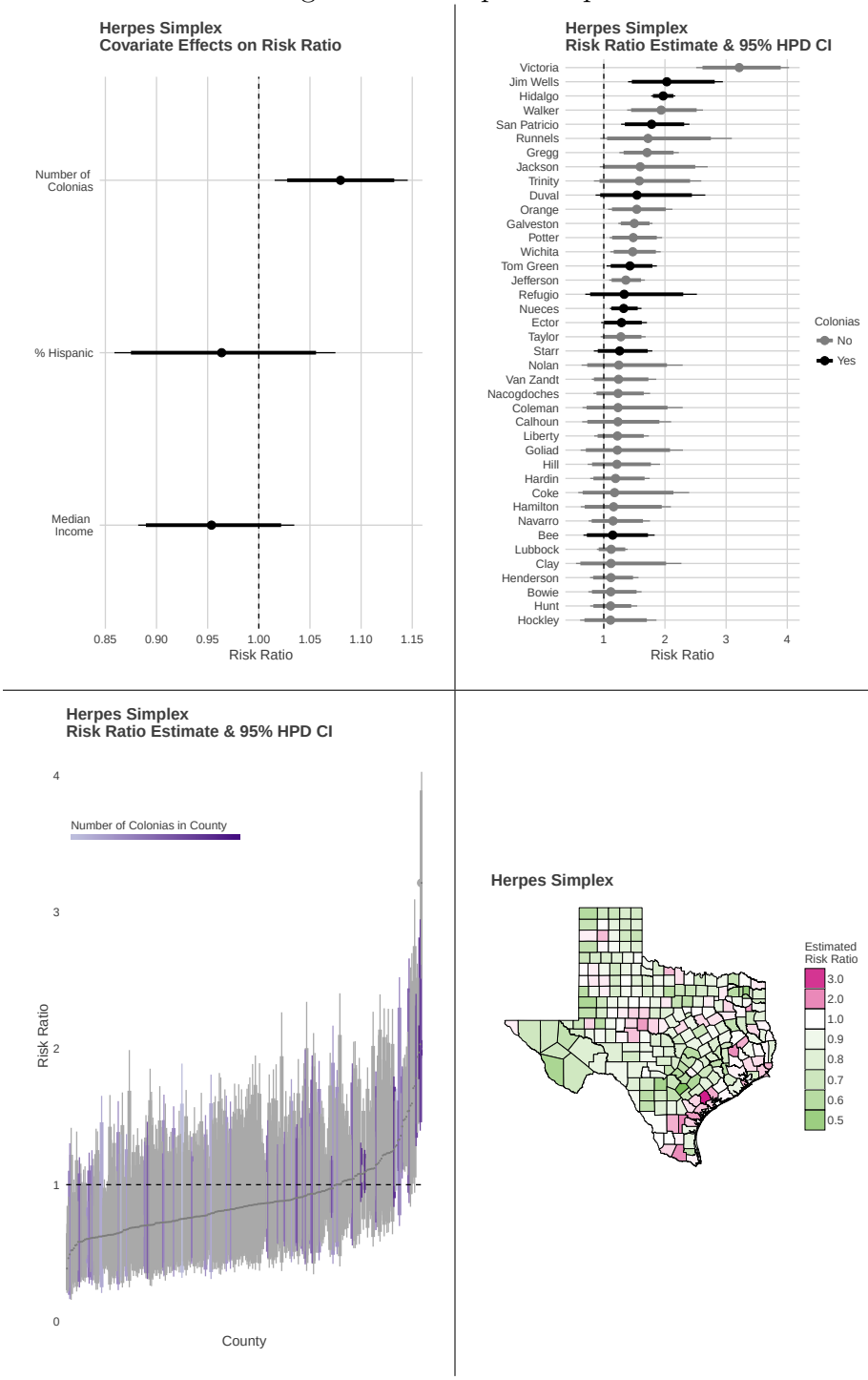


Figure A.8: Herpes Zoster

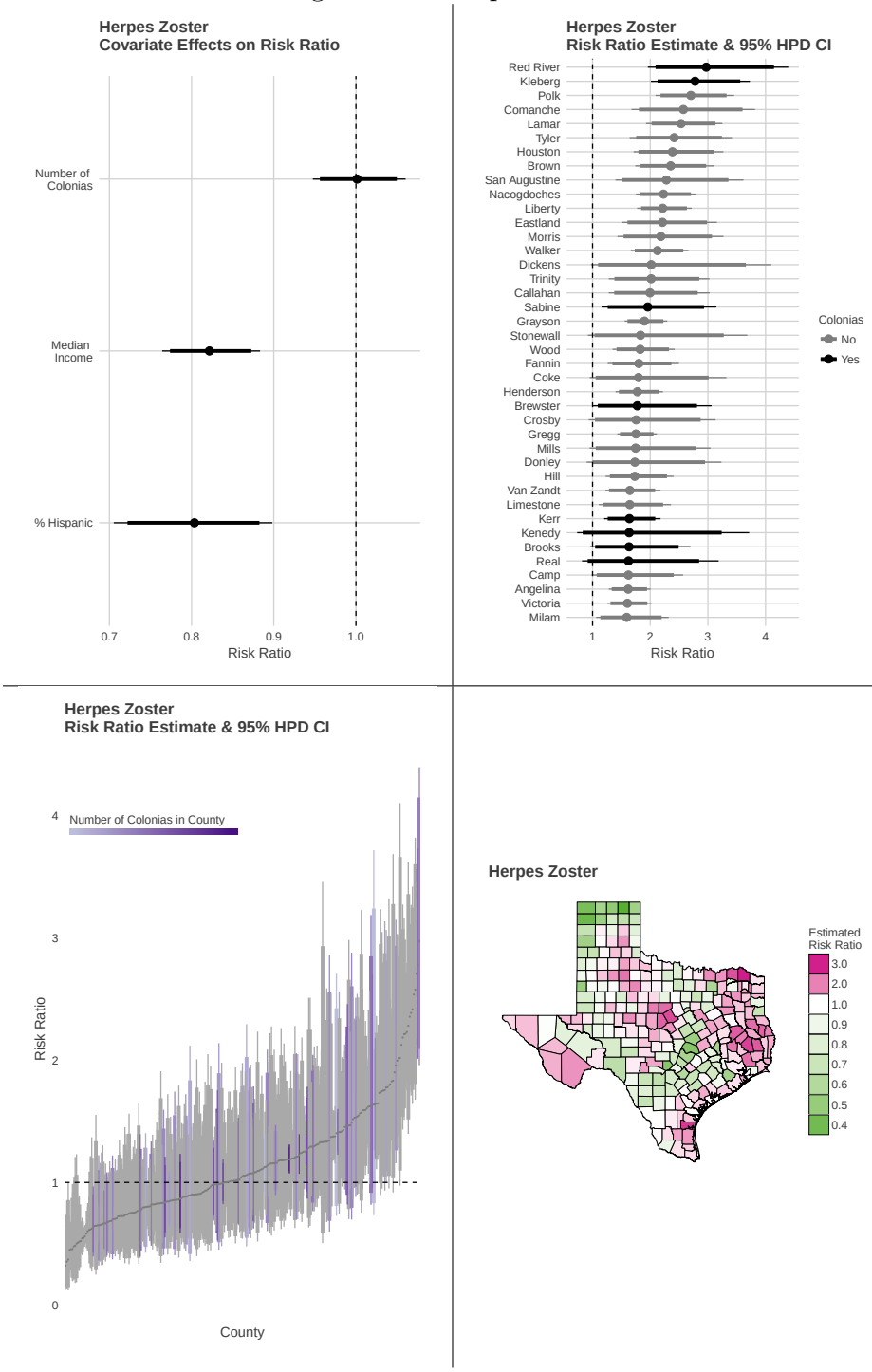


Figure A.9: Ill Defined Intestinal Infections

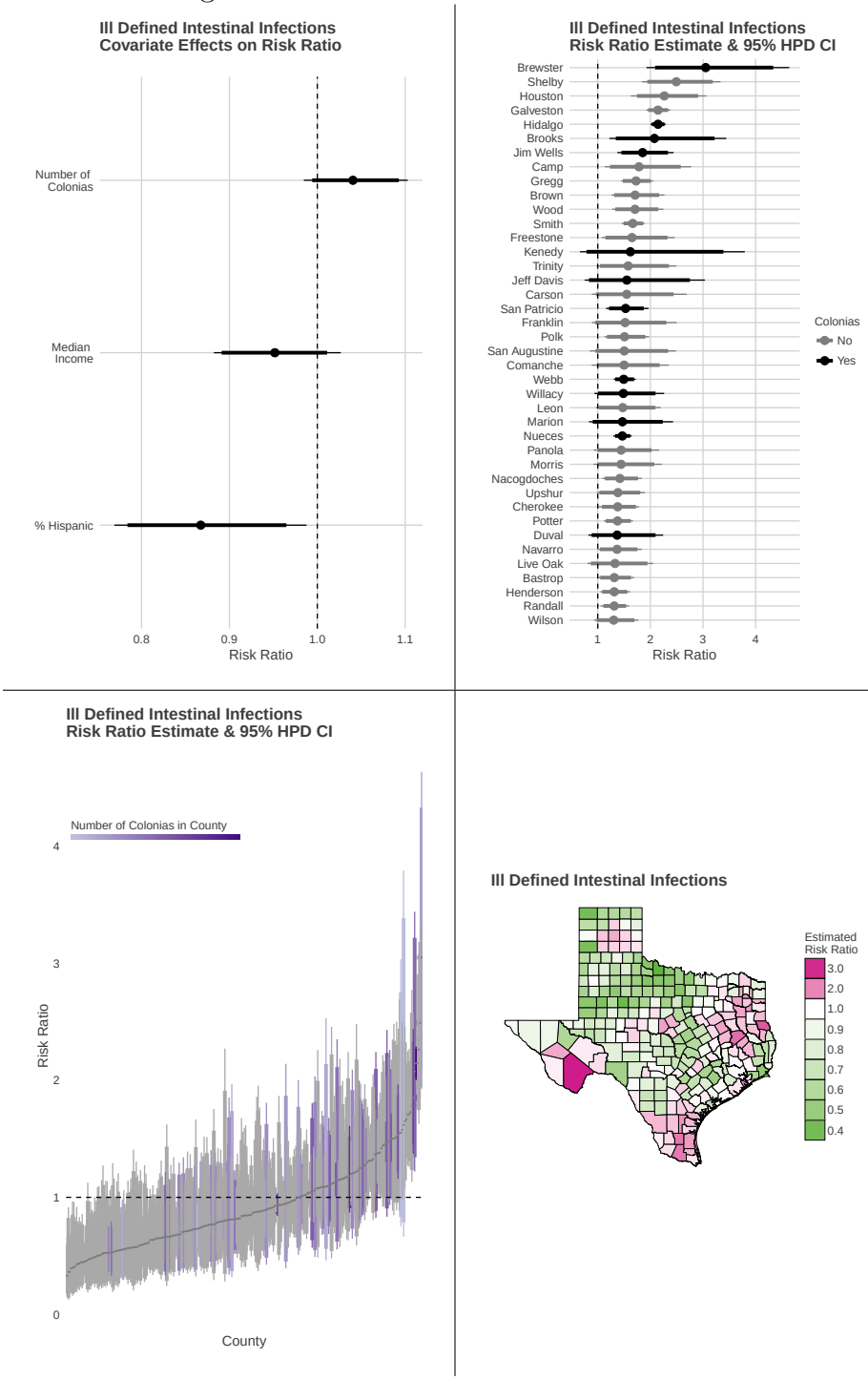


Figure A.10: Intestinal Infections due to Other Organisms

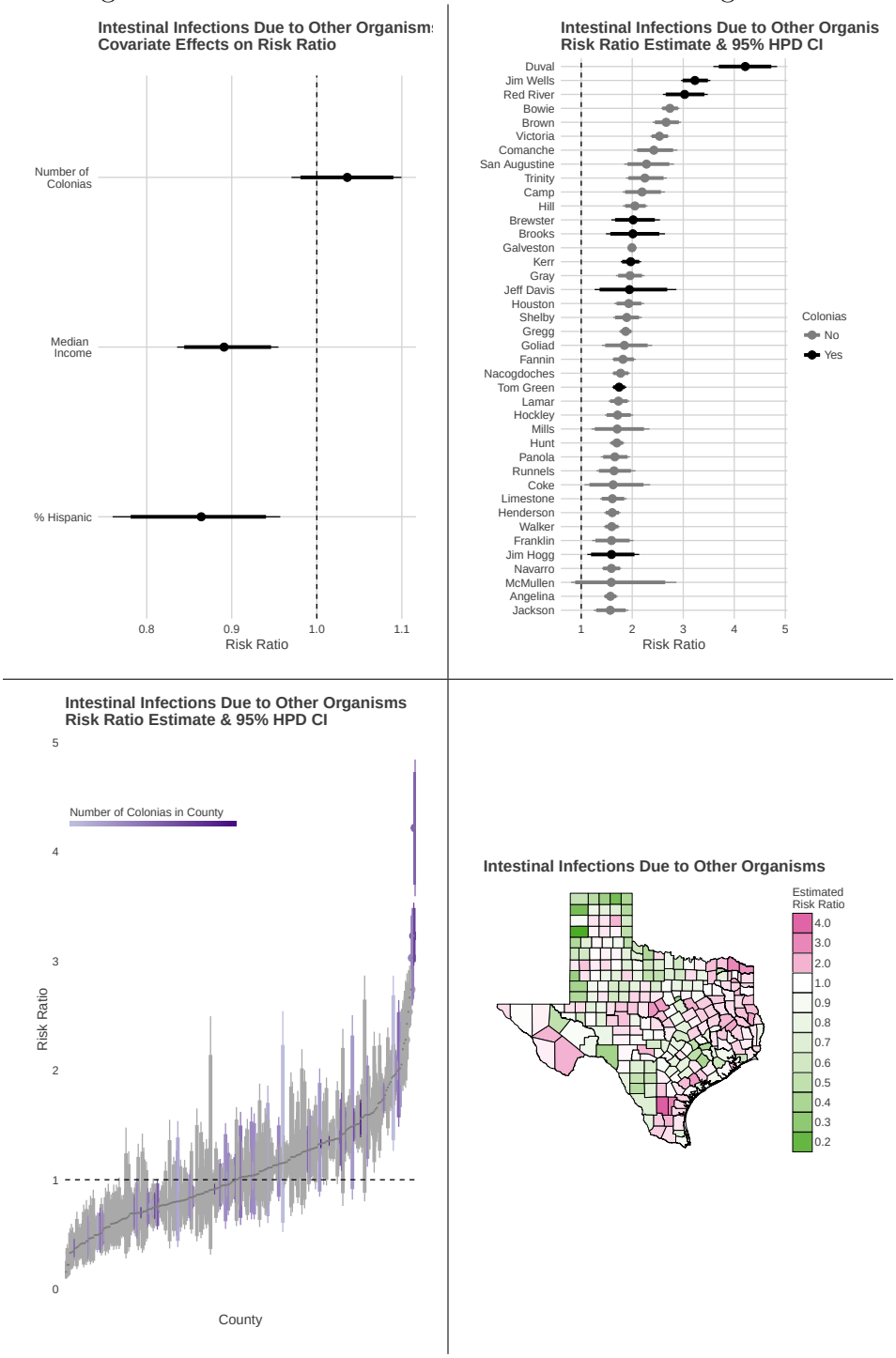


Figure A.11: Other enterovirus diseases of the central nervous system

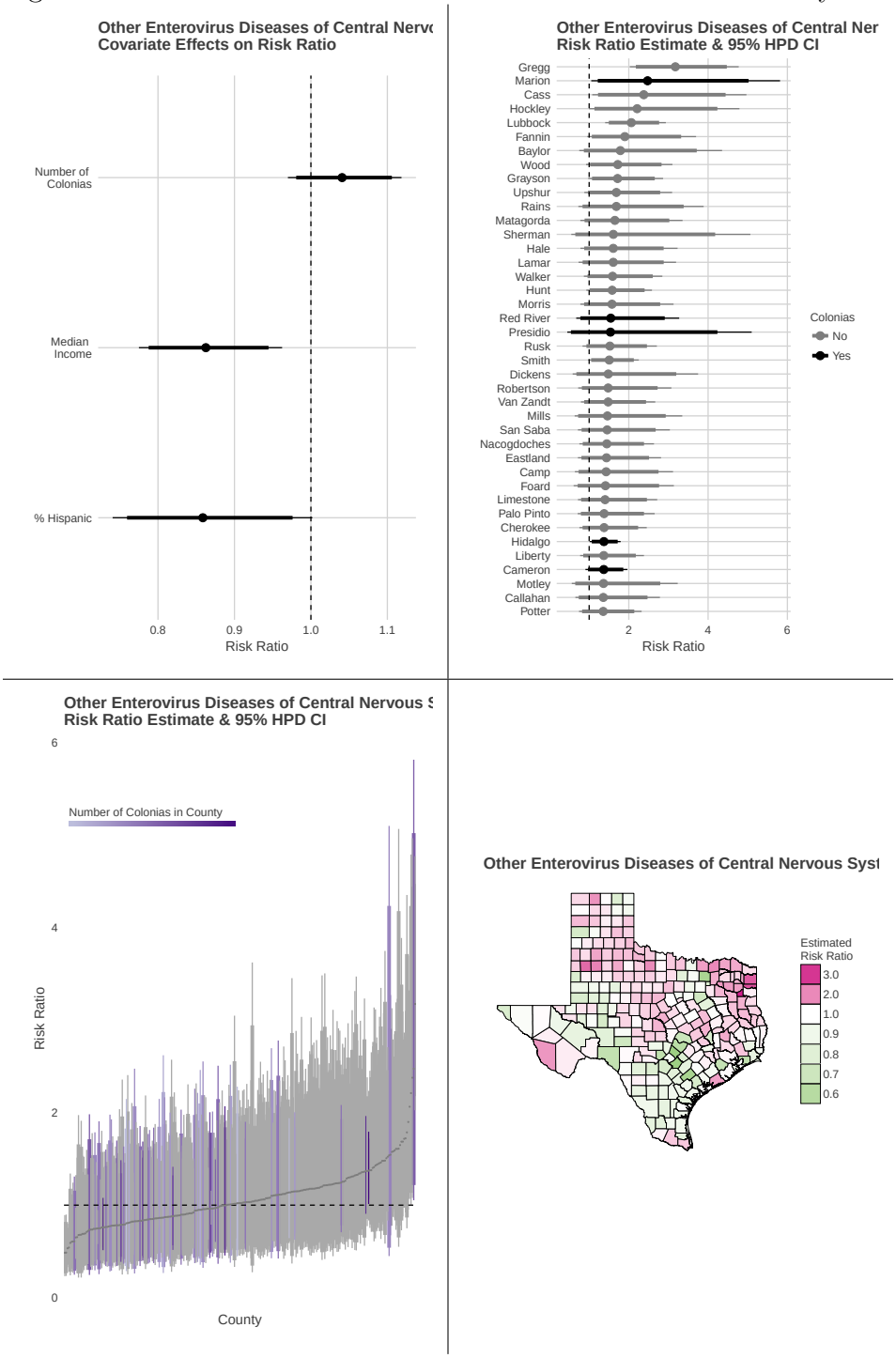


Figure A.12: Other Food Poisoning Bacterial

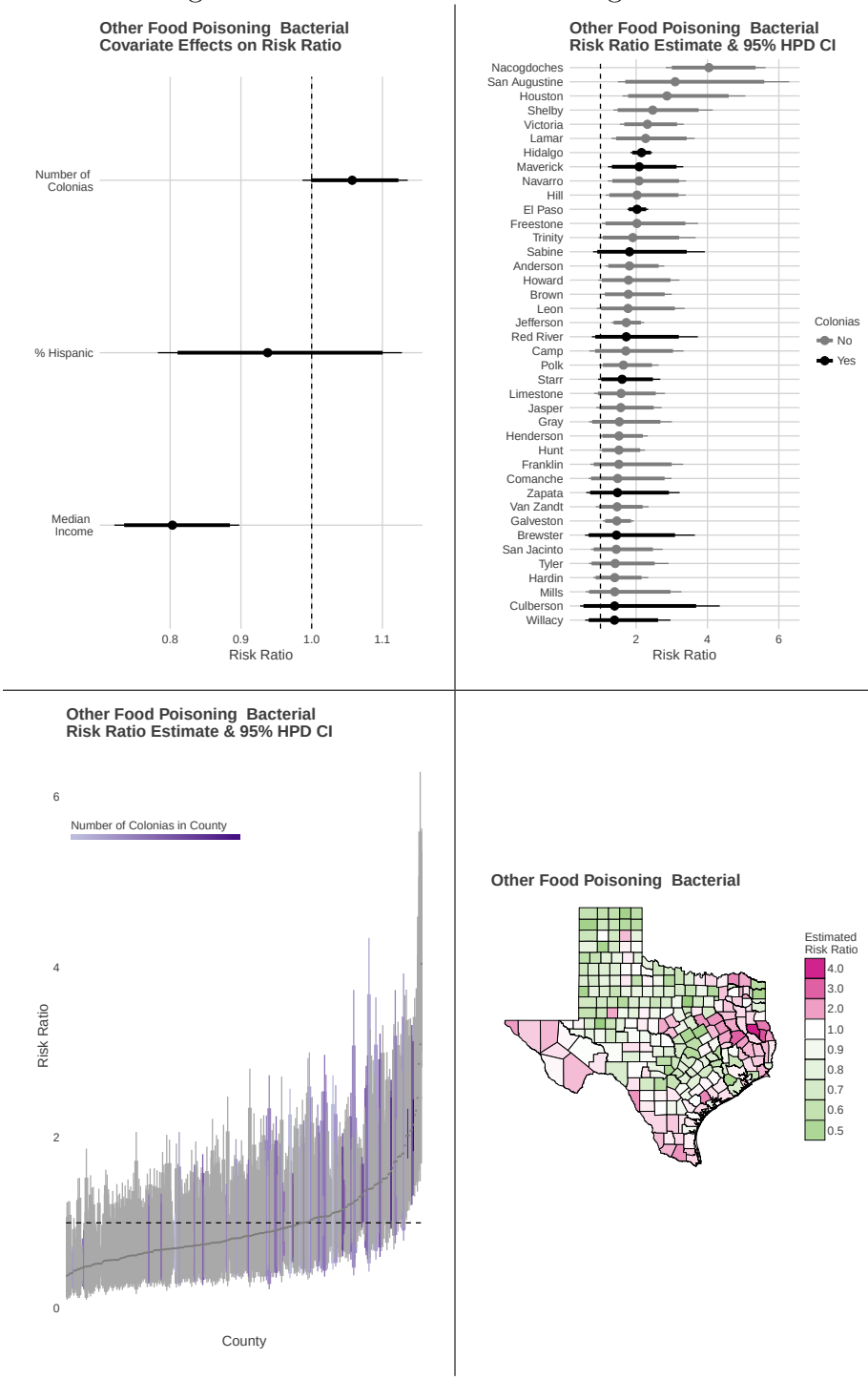


Figure A.13: Other non-arthropod-borne viral diseases of the central nervous system

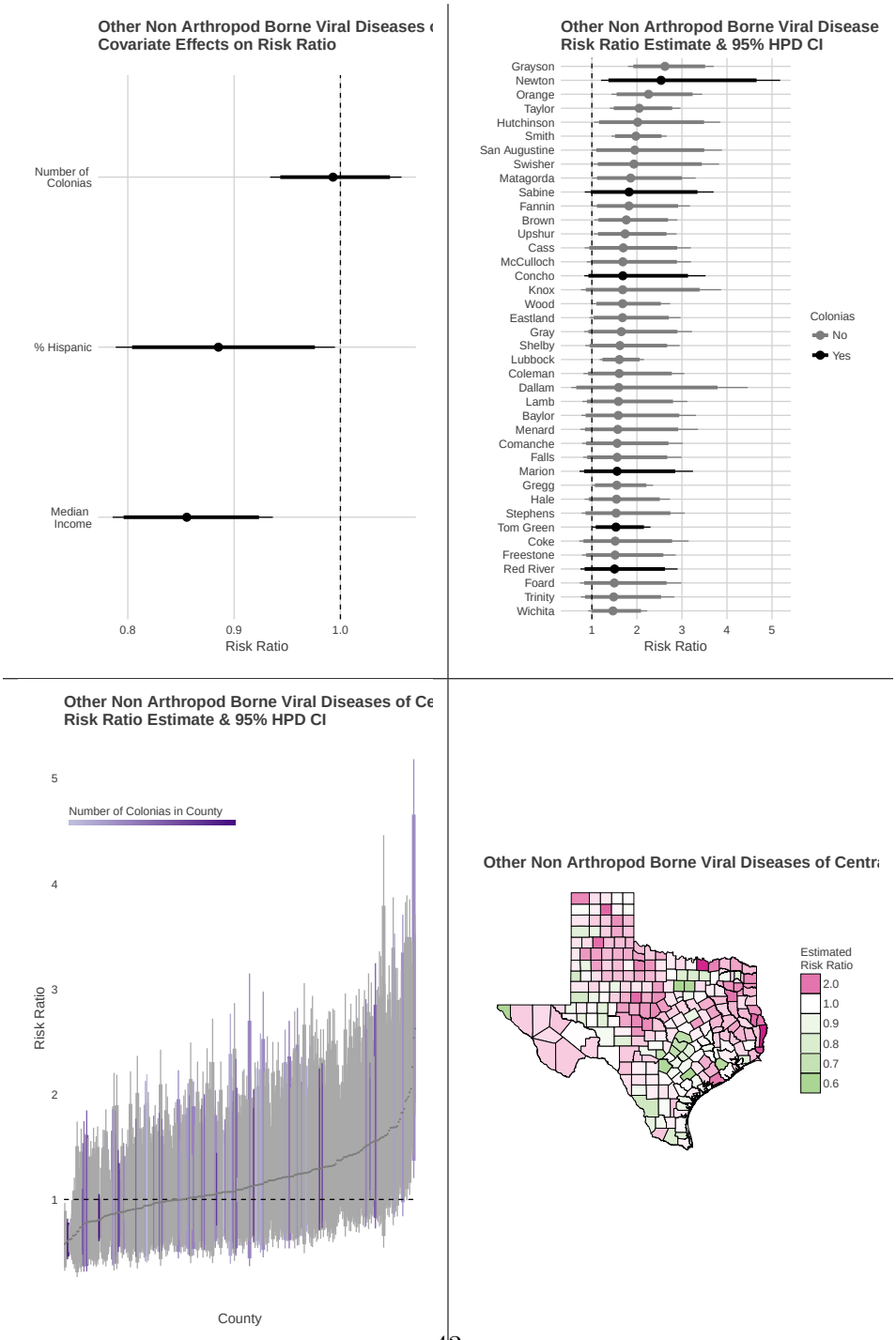


Figure A.14: Other protozoal intestinal diseases

Figure A.15: Other Rickettsioses

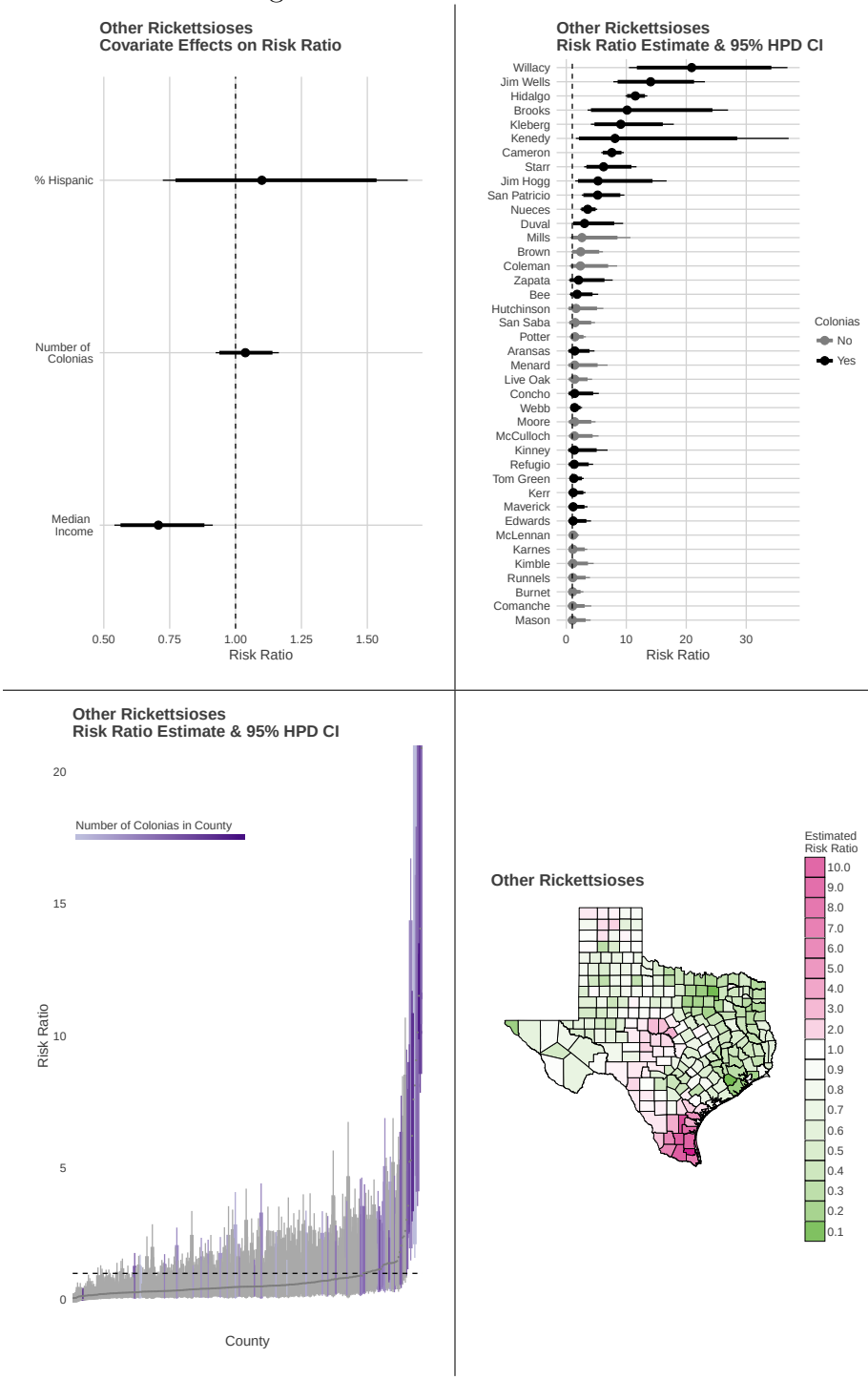


Figure A.16: Other Salmonella Infections

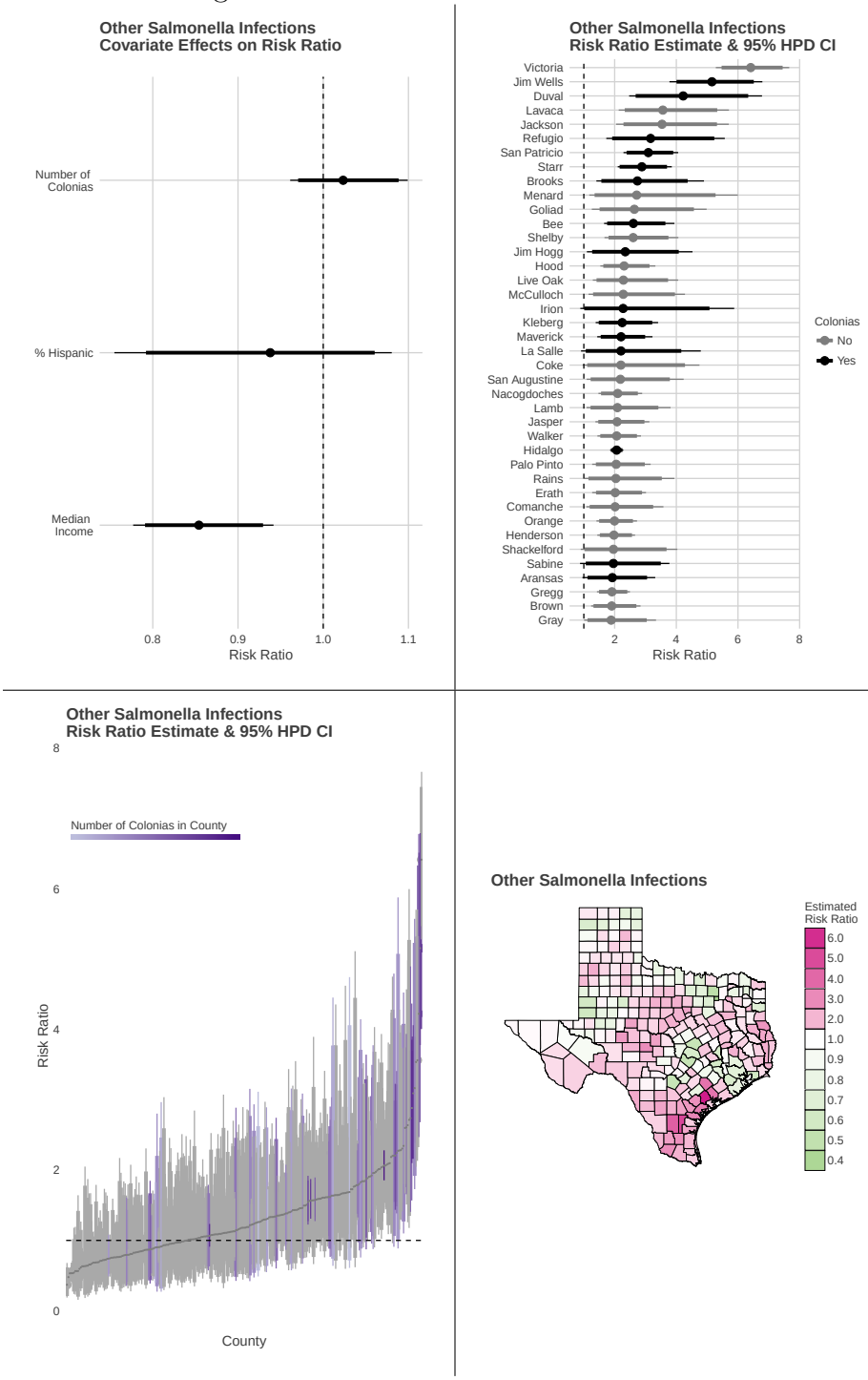


Figure A.17: Other Typhus

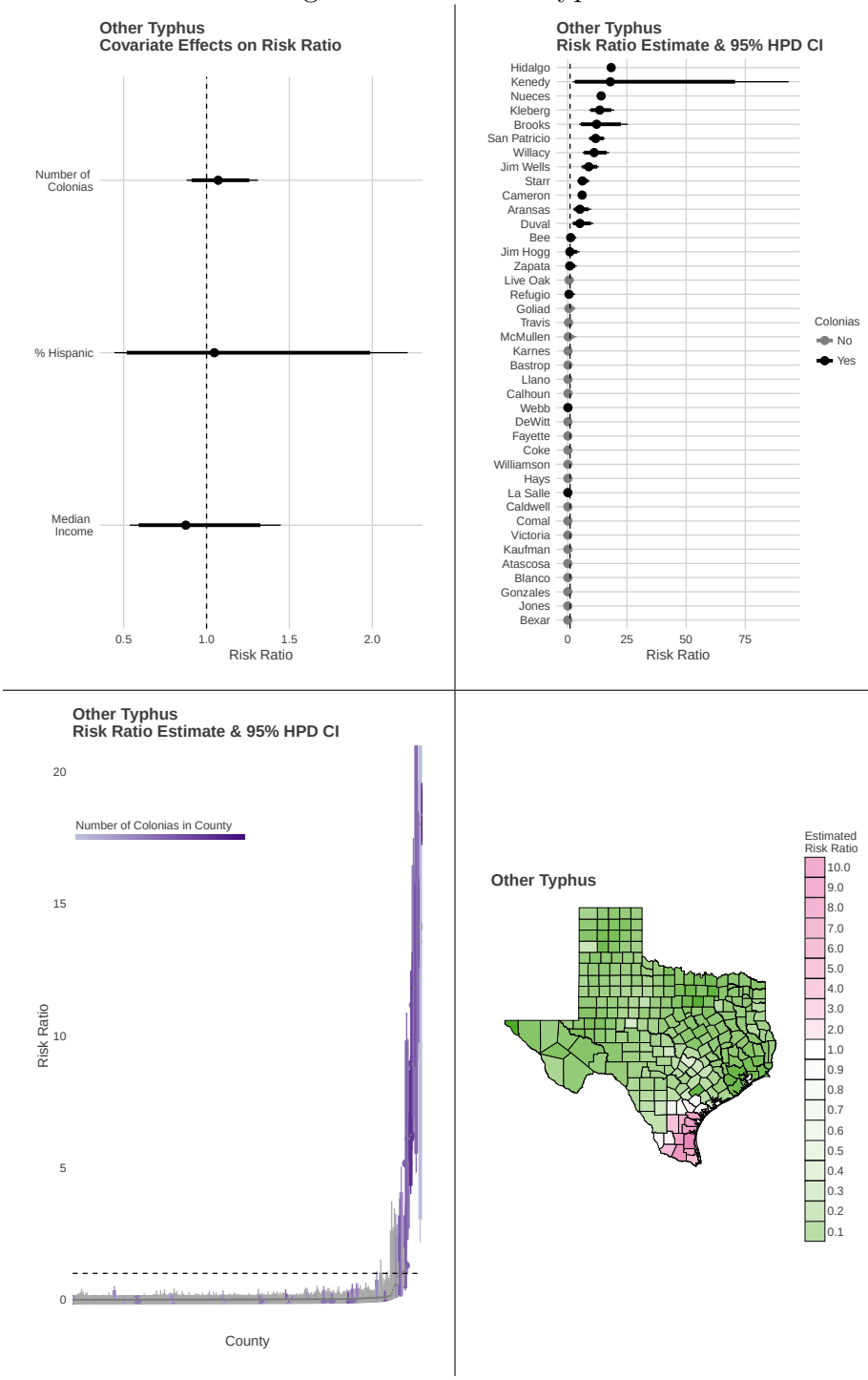


Figure A.18: Other Viral Exanthemata

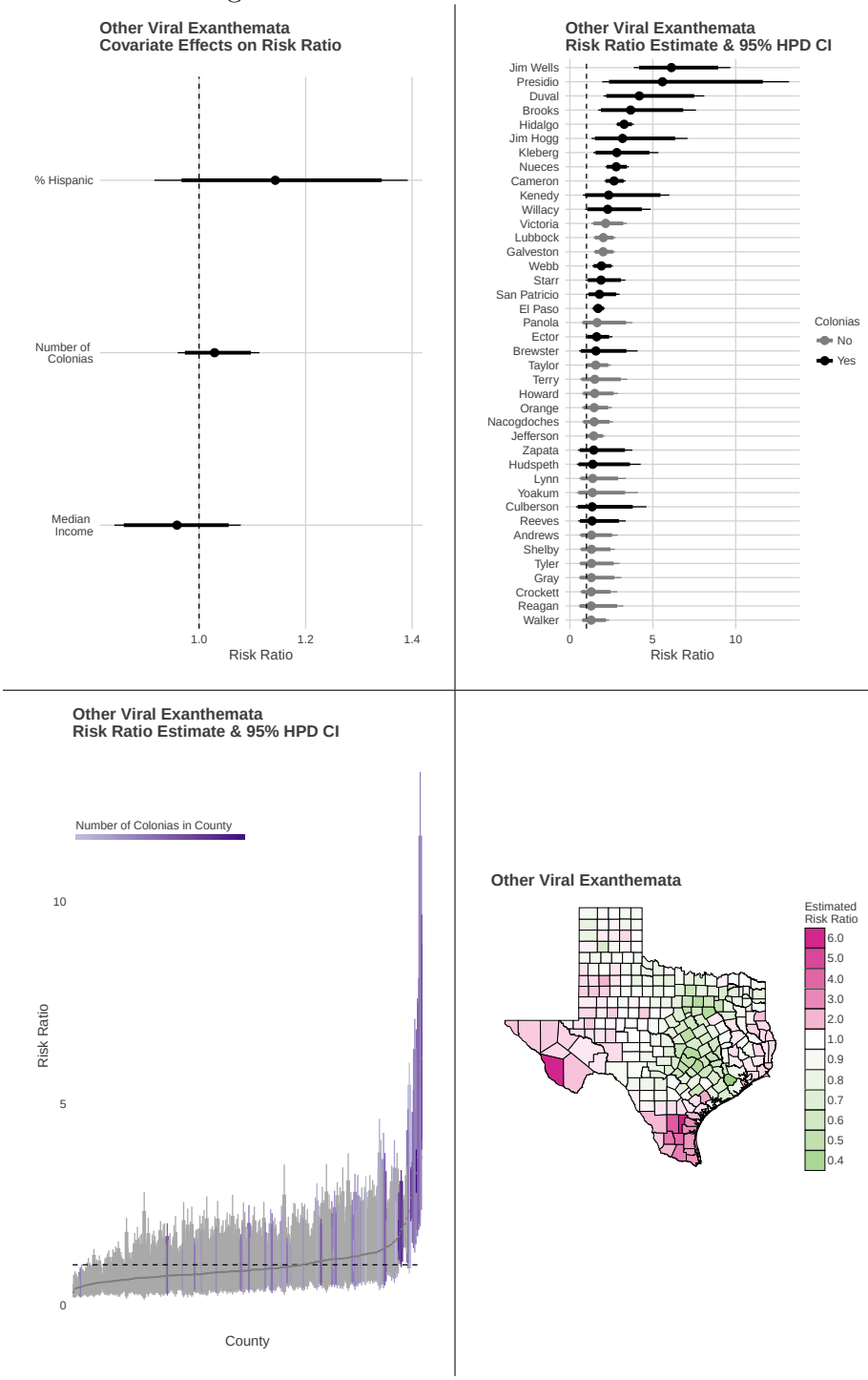


Figure A.19: Pulmonary Tuberculosis

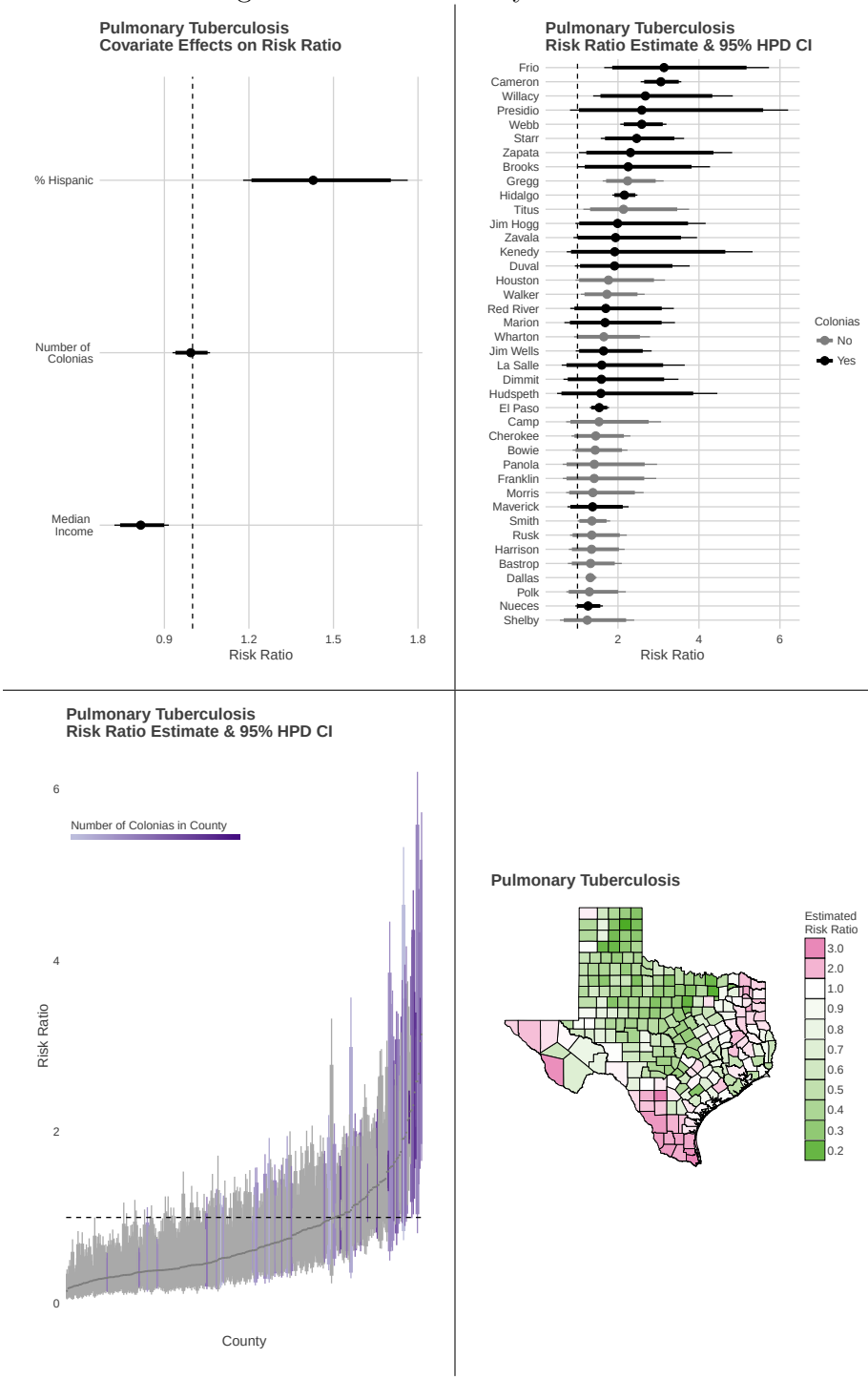


Figure A.20: Septicemia

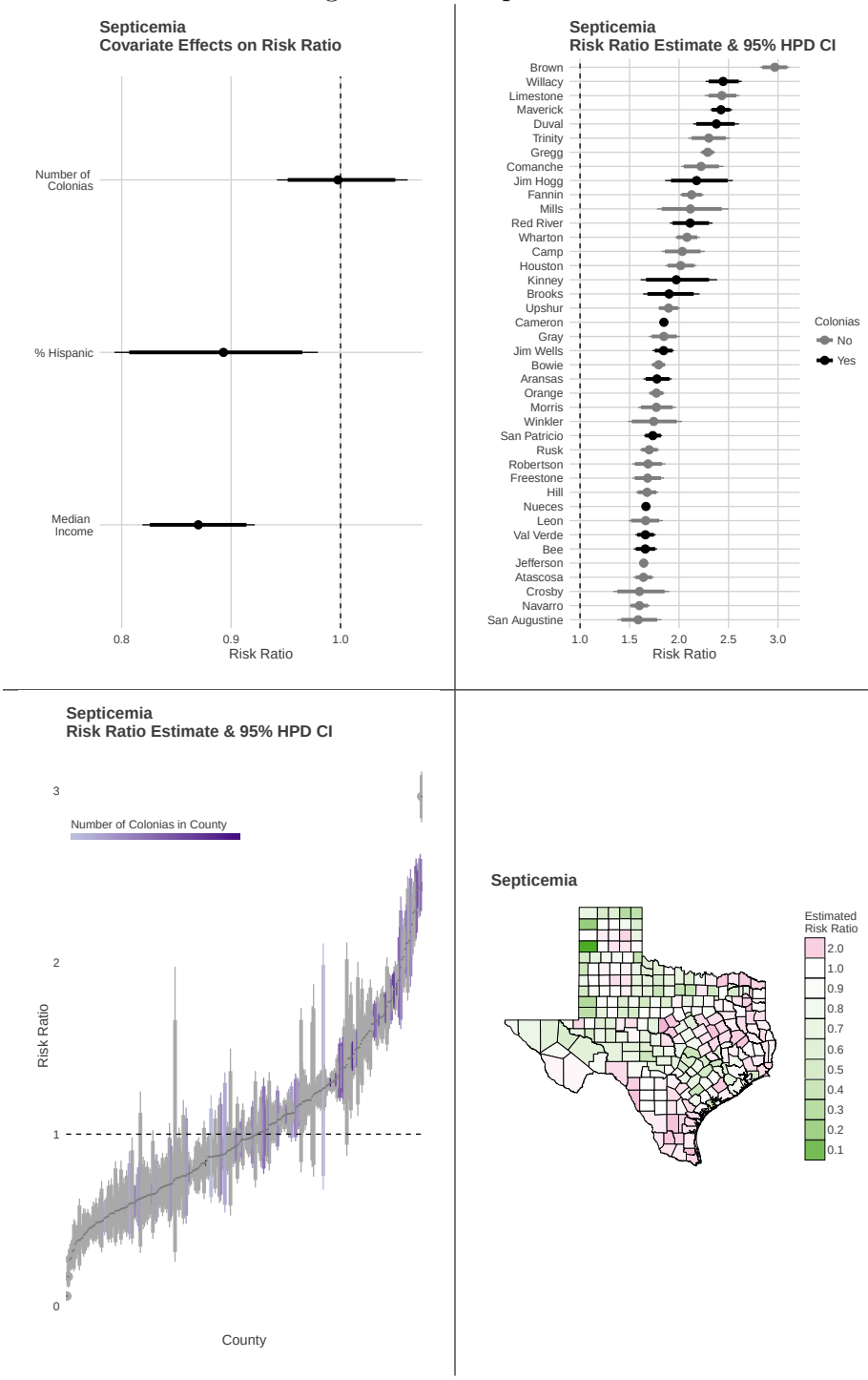


Figure A.21: Shigellosis

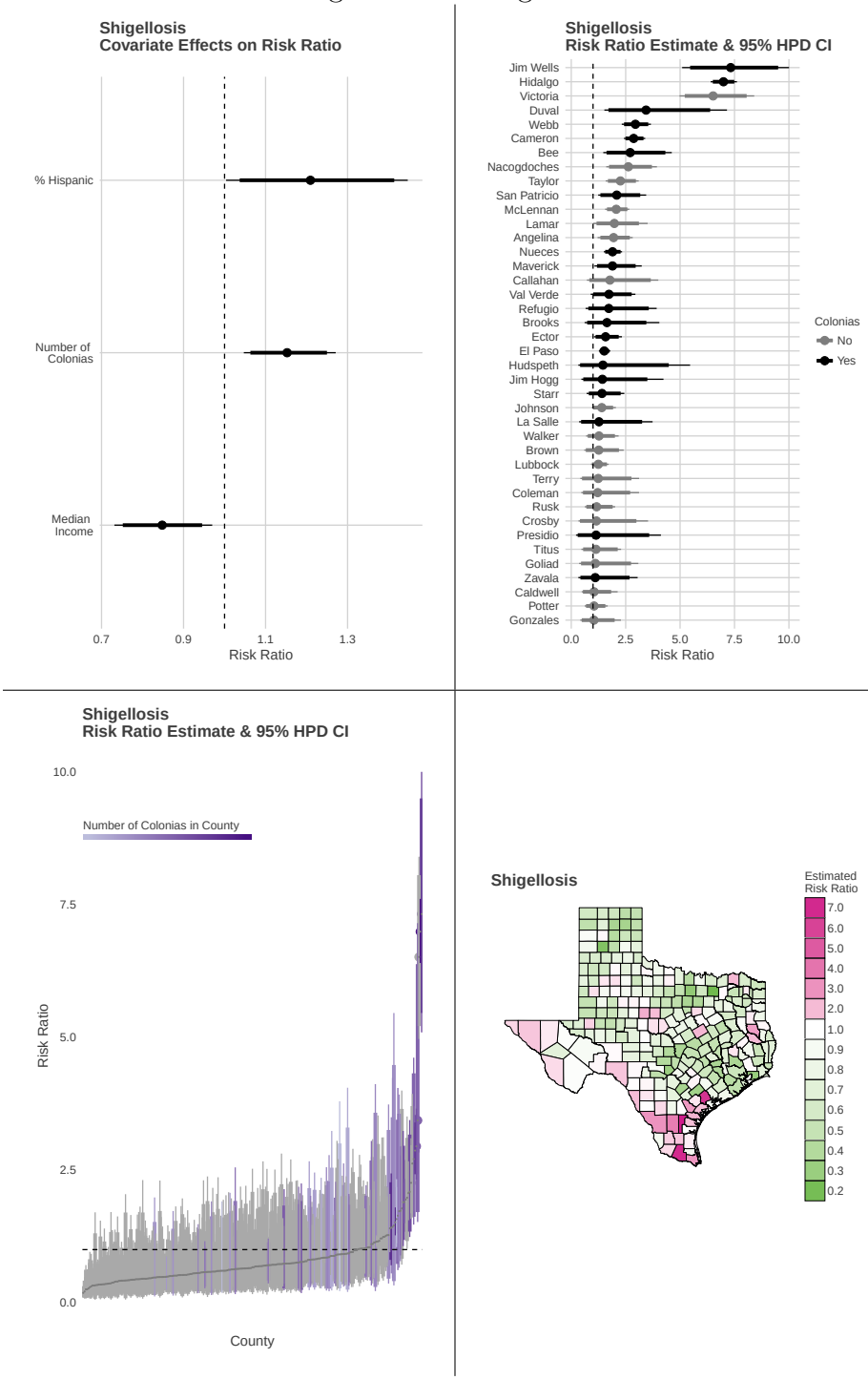


Figure A.22: Specific Diseases due to Coxsackie Virus

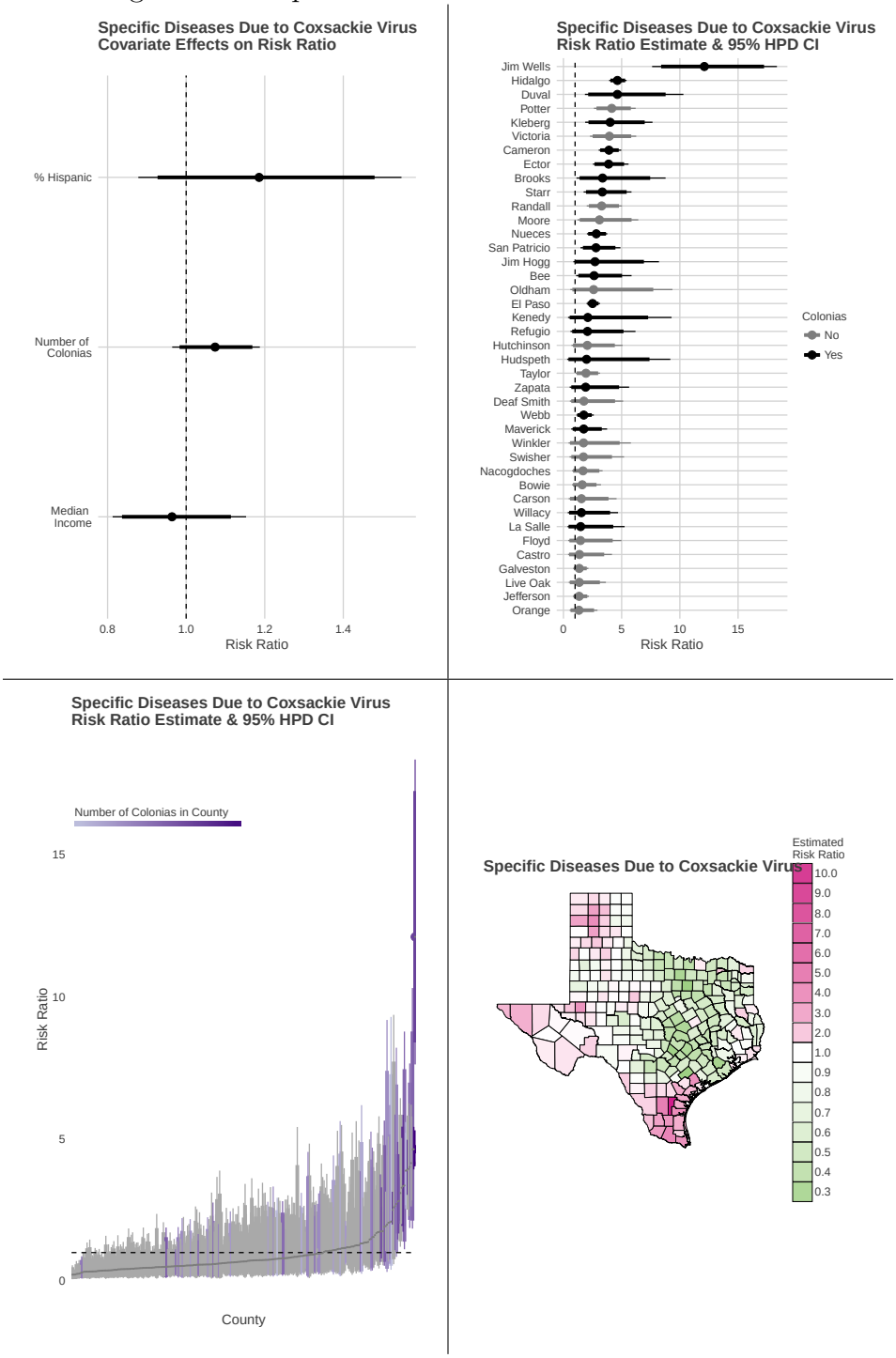


Figure A.23: Streptococcal Sore Throat and Scarlet Fever

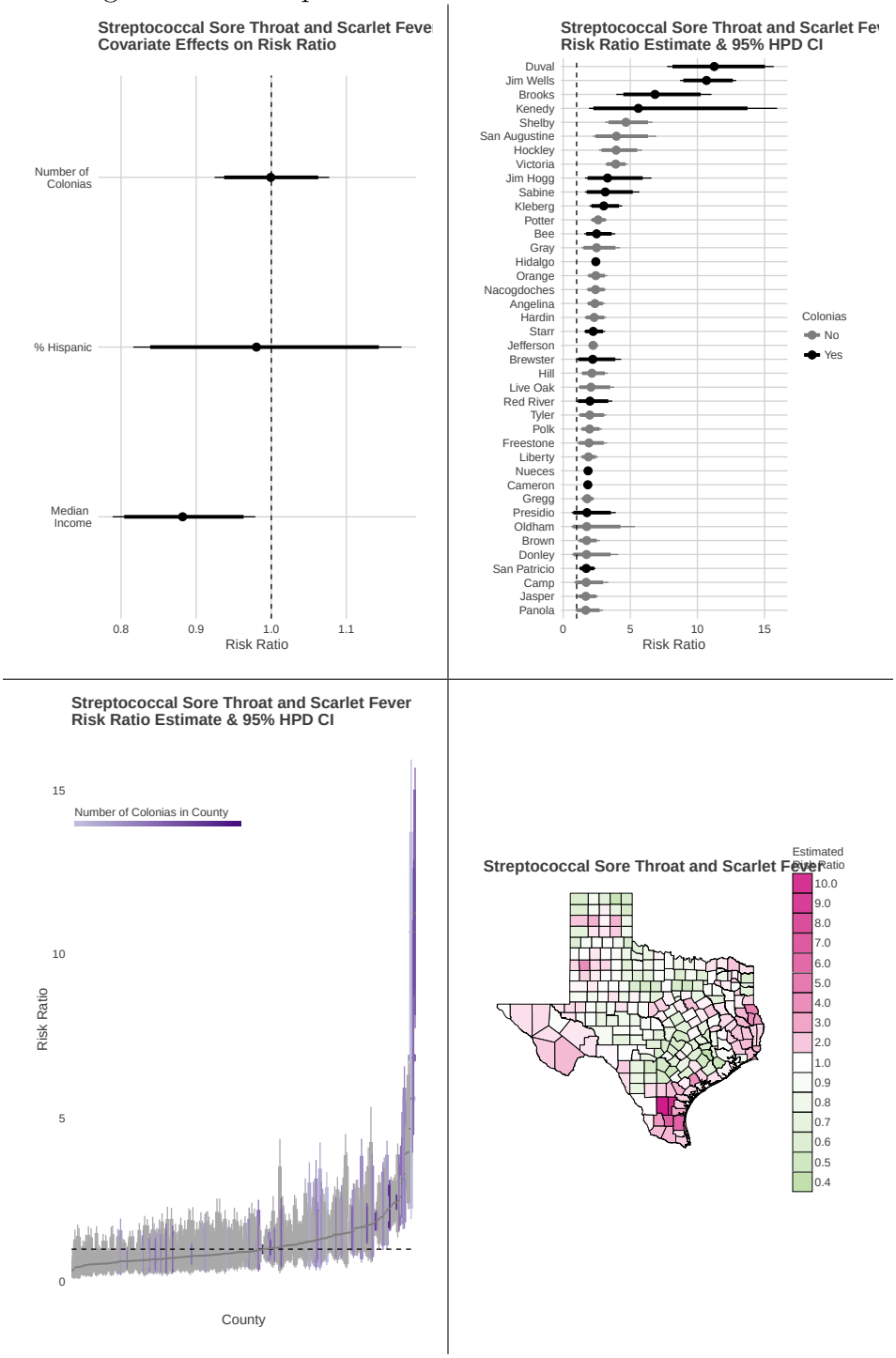


Figure A.24: Typhoid and Paratyphoid Fevers

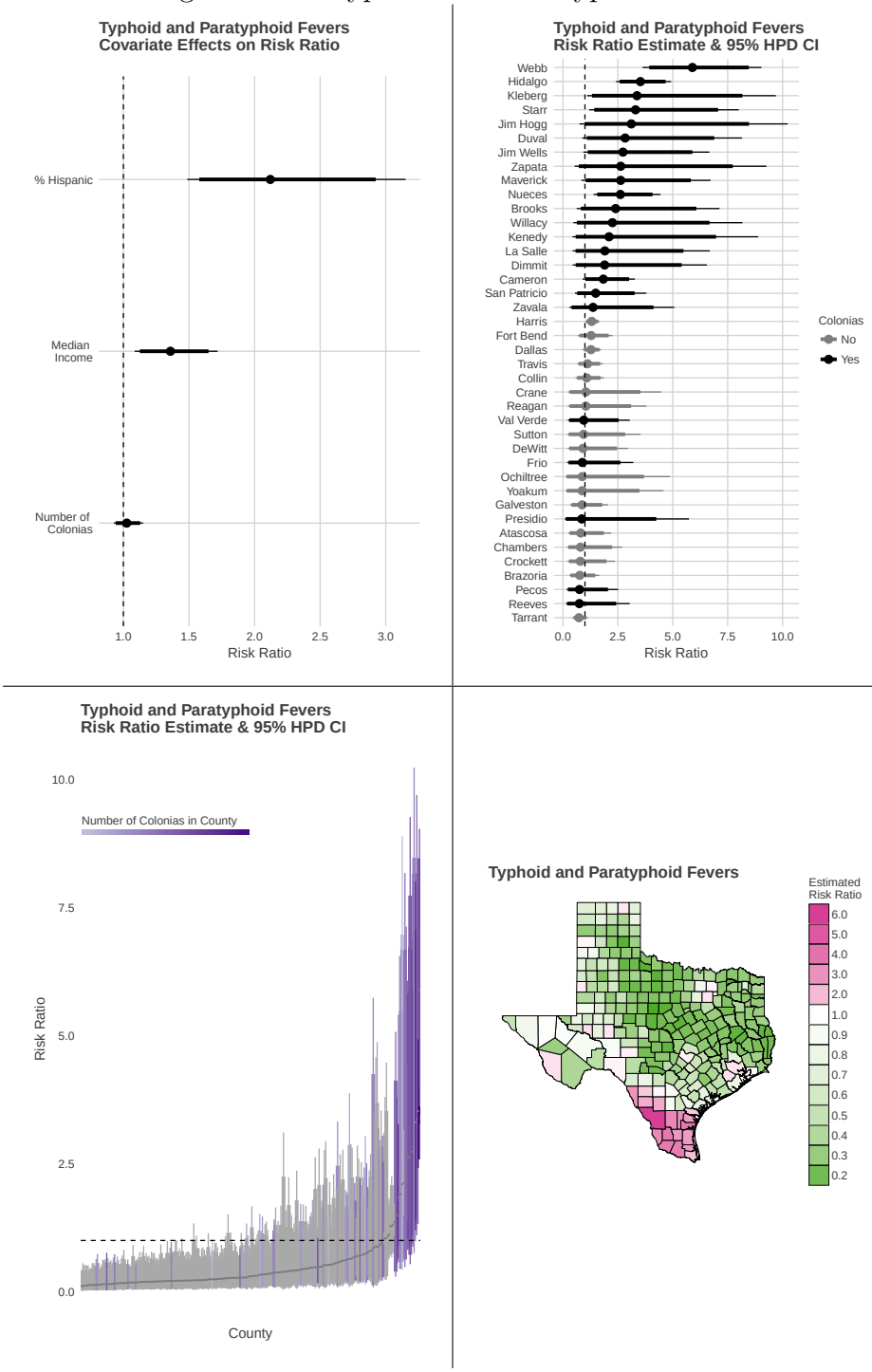


Figure A.25: Viral and Chlamydial Infection in Conditions Classified Elsewhere and of Unspecified Site

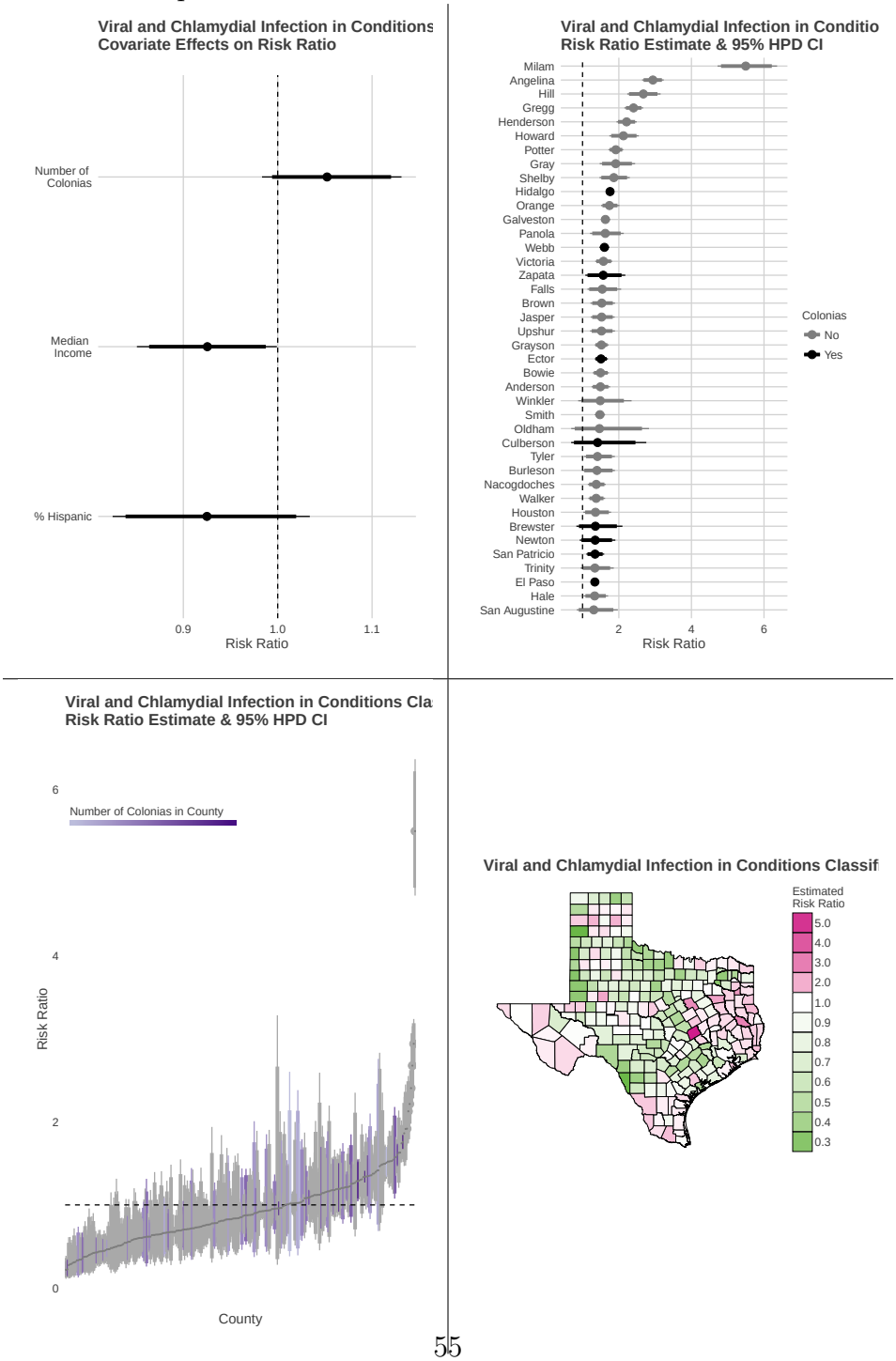
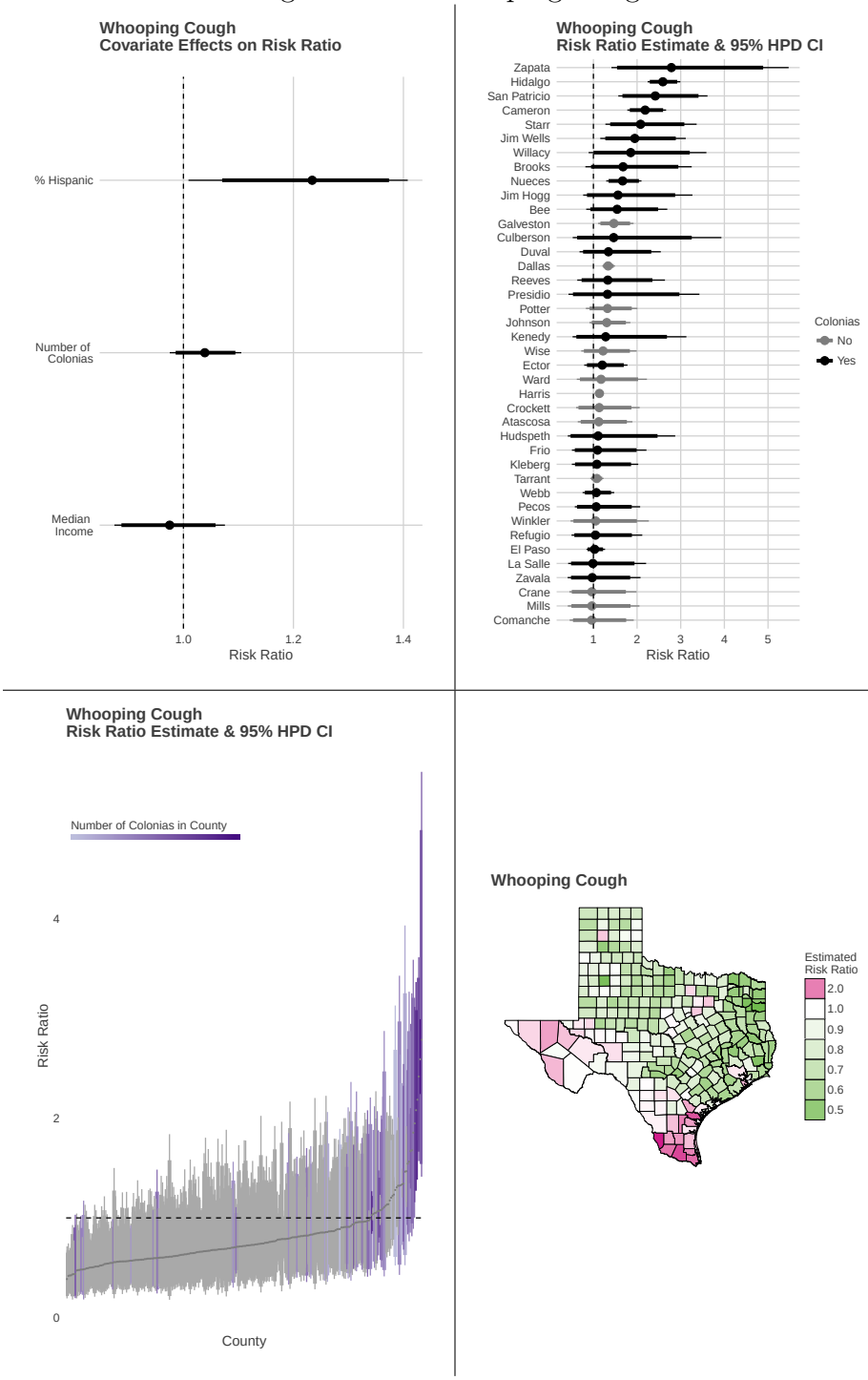


Figure A.26: Whooping Cough



Appendix B

Plots: Observed vs. Expected Incidence Under Equal Statewide Risk

Figure B.1: Observed vs. Expected Incidence, Actinomycotic Infections

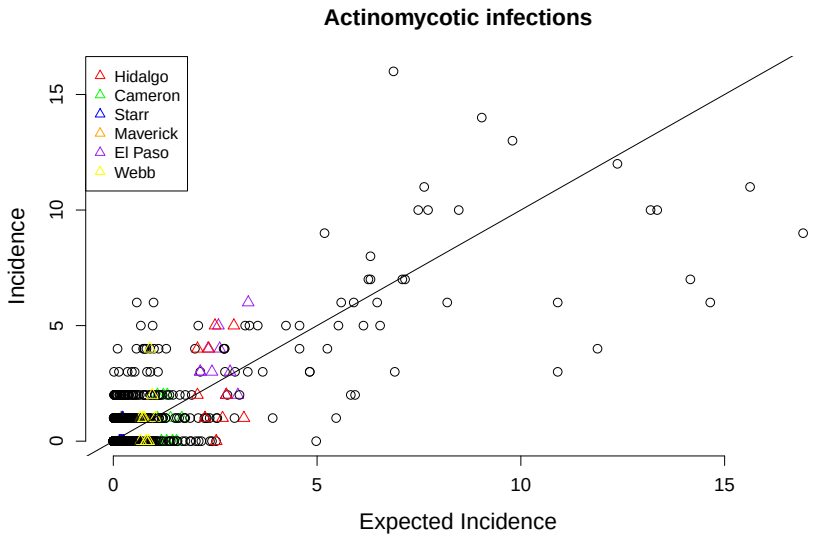


Figure B.2: Observed vs. Expected Incidence, Amebiasis

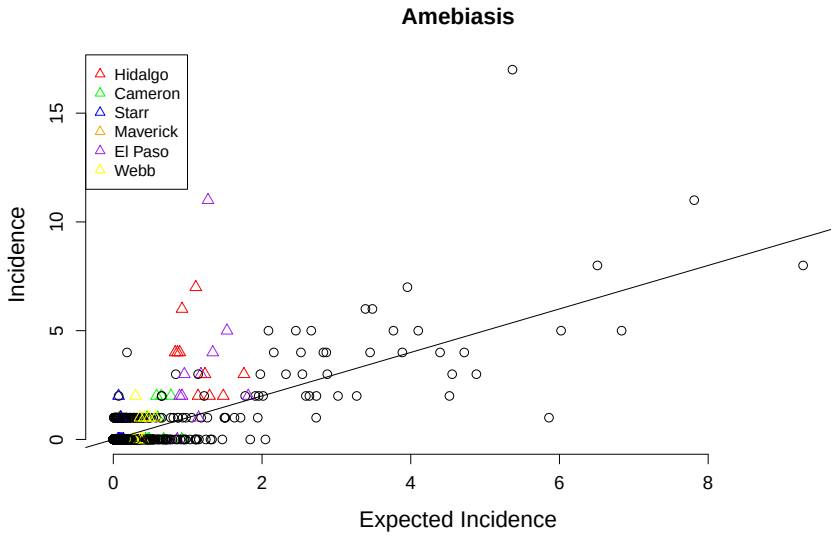


Figure B.3: Observed vs. Expected Incidence, Brucellosis

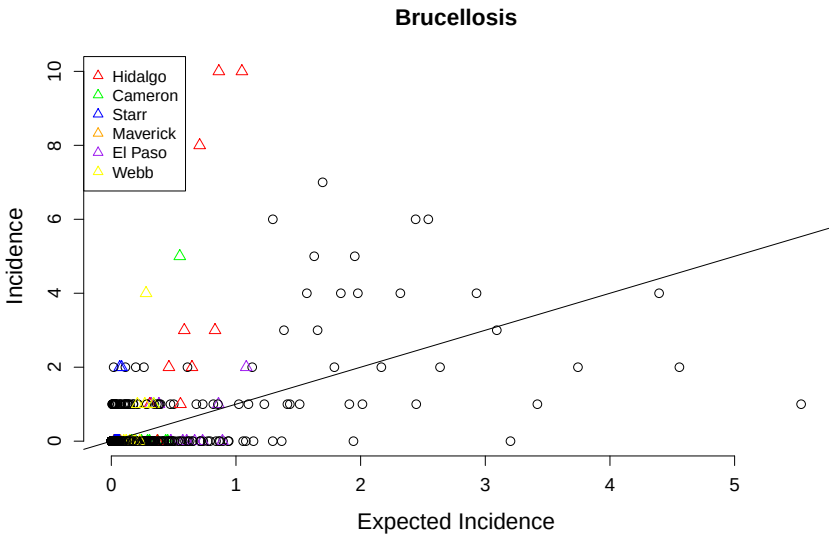


Figure B.4: Observed vs. Expected Incidence, Candidiasis

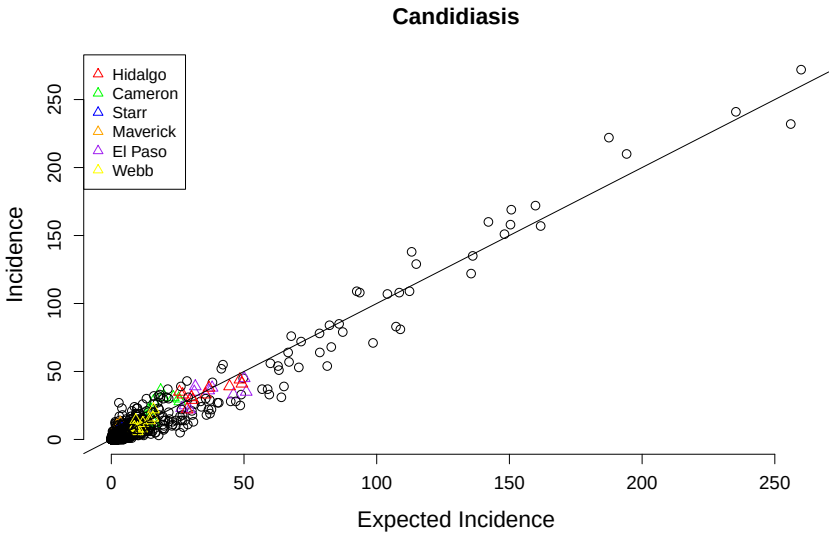


Figure B.5: Observed vs. Expected Incidence, Chickenpox

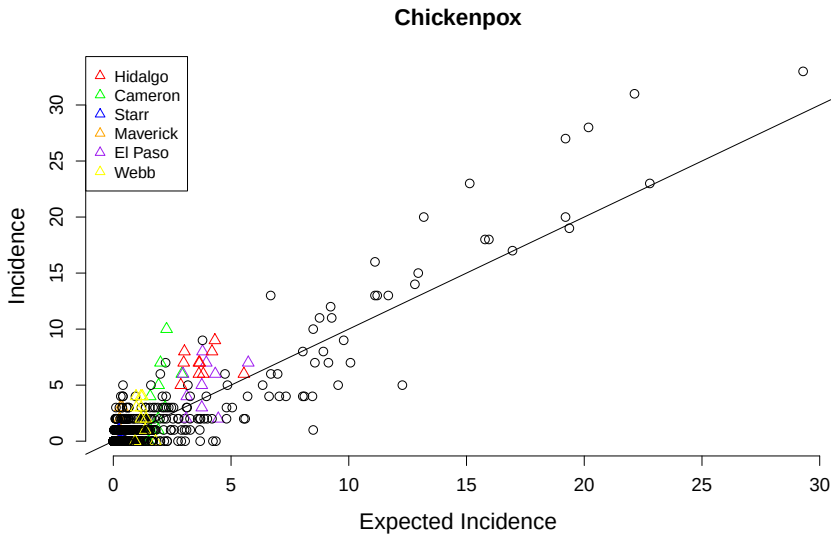


Figure B.6: Observed vs. Expected Incidence, Coccidioidomycosis

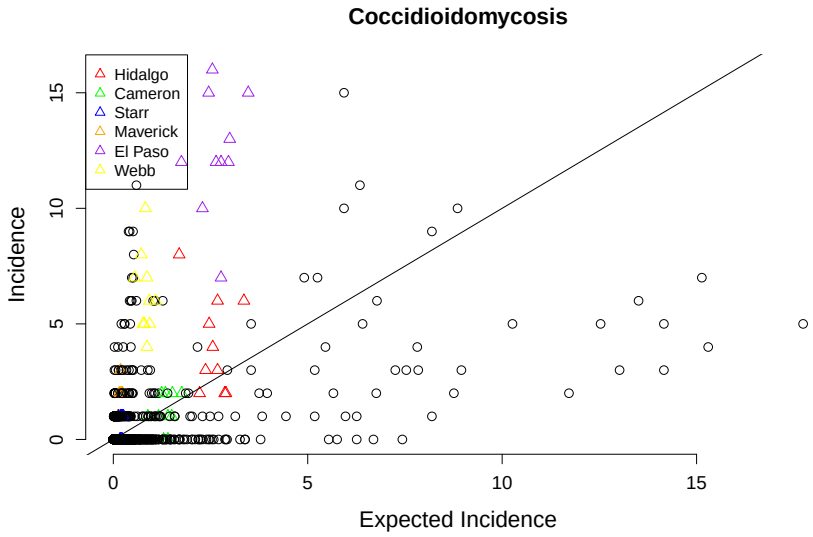


Figure B.7: Observed vs. Expected Incidence, Herpes Simplex

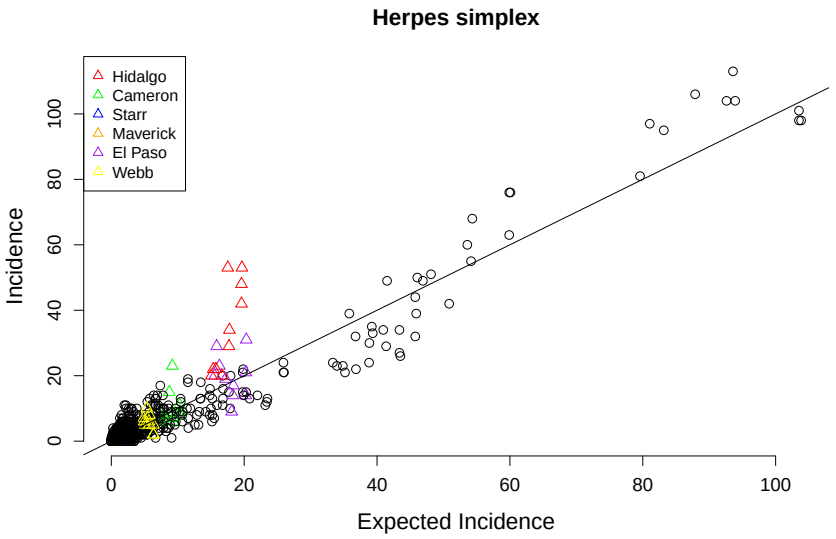


Figure B.8: Observed vs. Expected Incidence, Herpes Zoster

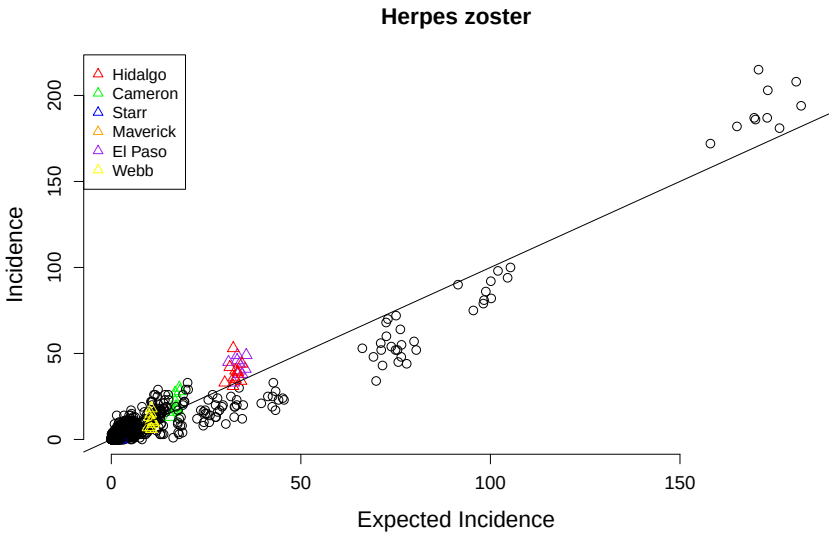


Figure B.9: Observed vs. Expected Incidence, Ill Defined Intestinal Infections

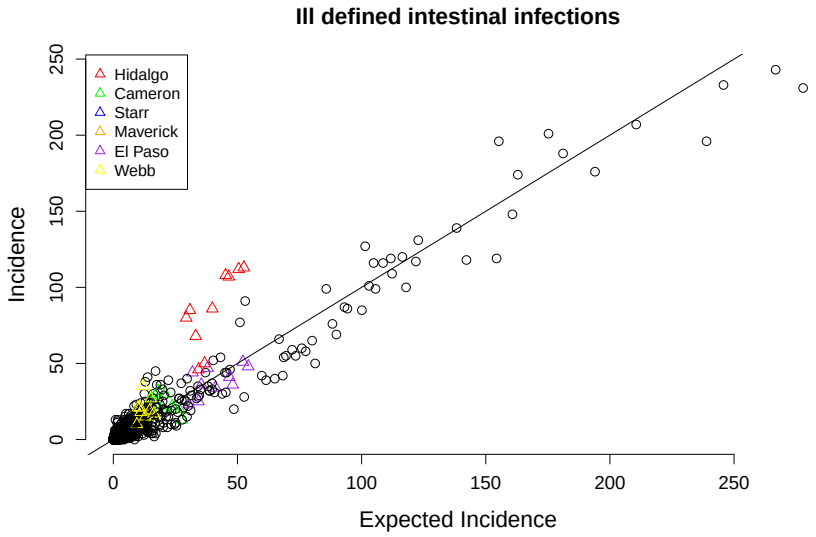


Figure B.10: Observed vs. Expected Incidence, Intestinal Infections due to Other Organisms

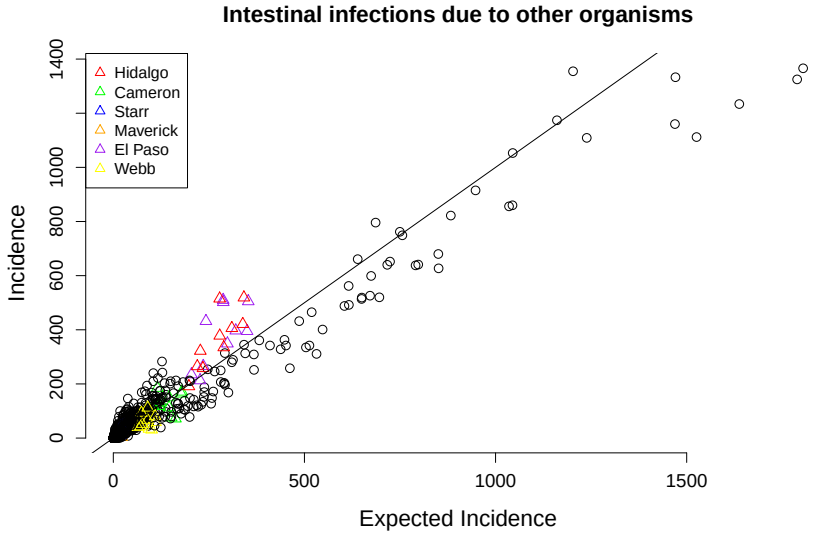


Figure B.11: Observed vs. Expected Incidence, Other enterovirus diseases of the central nervous system

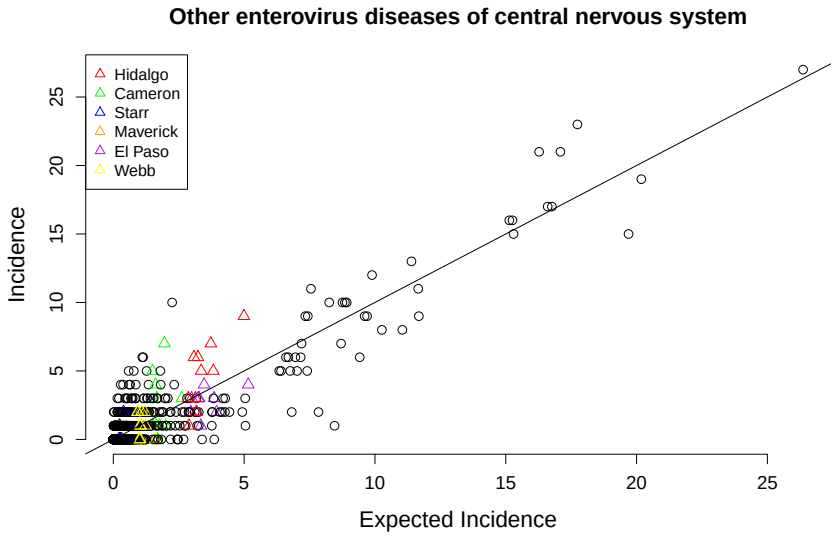


Figure B.12: Observed vs. Expected Incidence, Other non-arthropod borne viral diseases of the central nervous system

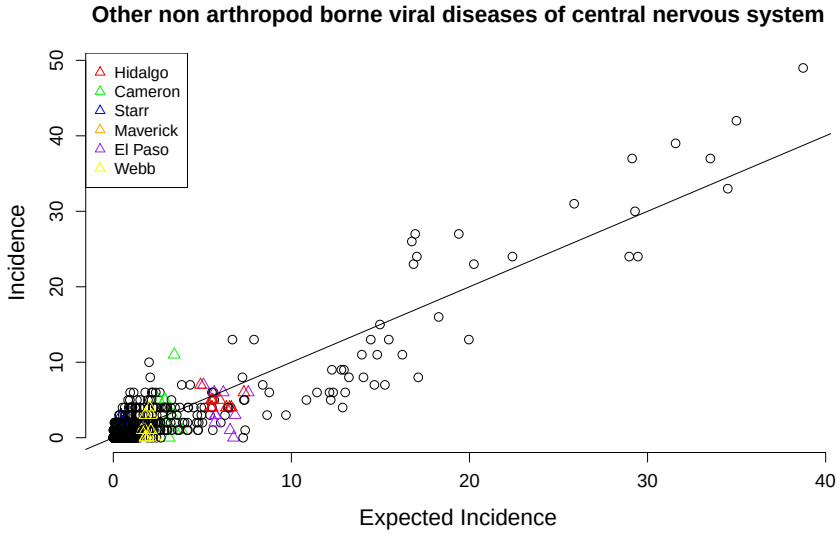


Figure B.13: Observed vs. Expected Incidence, Other protozoal intestinal diseases

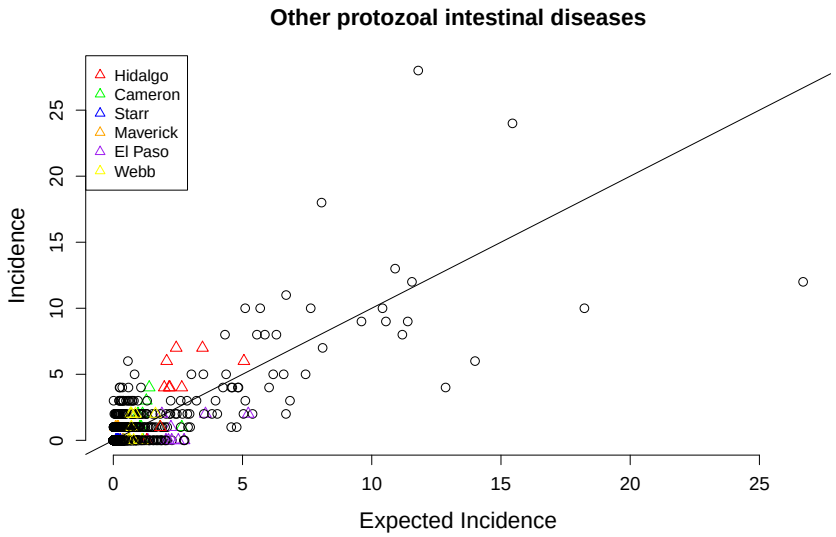


Figure B.14: Observed vs. Expected Incidence, Other Food Poisoning Bacterial

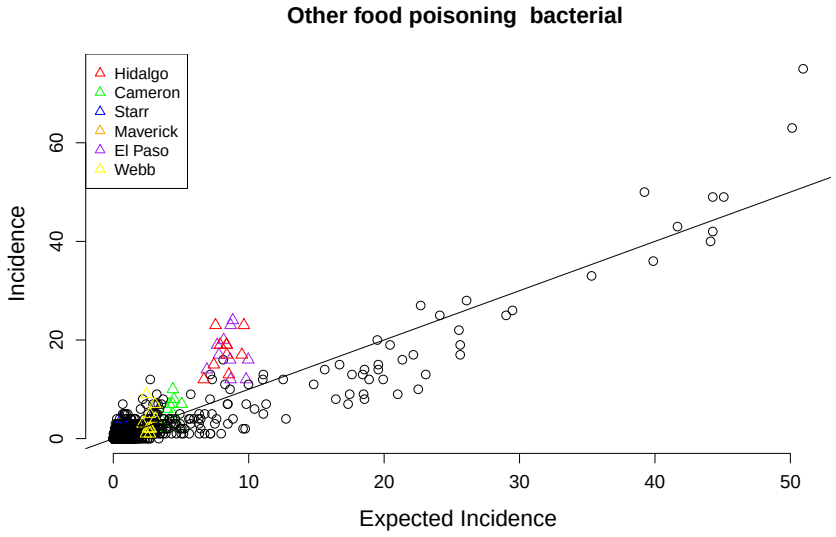


Figure B.15: Observed vs. Expected Incidence, Other Rickettsioses

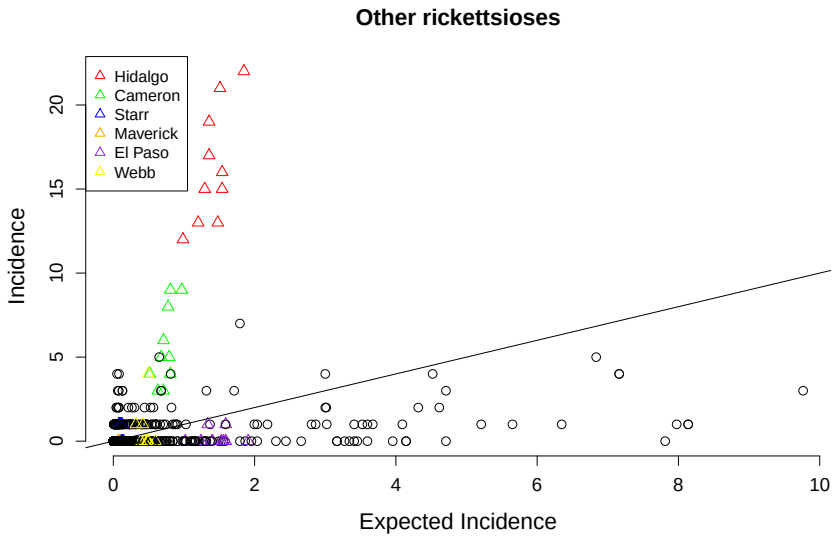


Figure B.16: Observed vs. Expected Incidence, Other Salmonella Infections

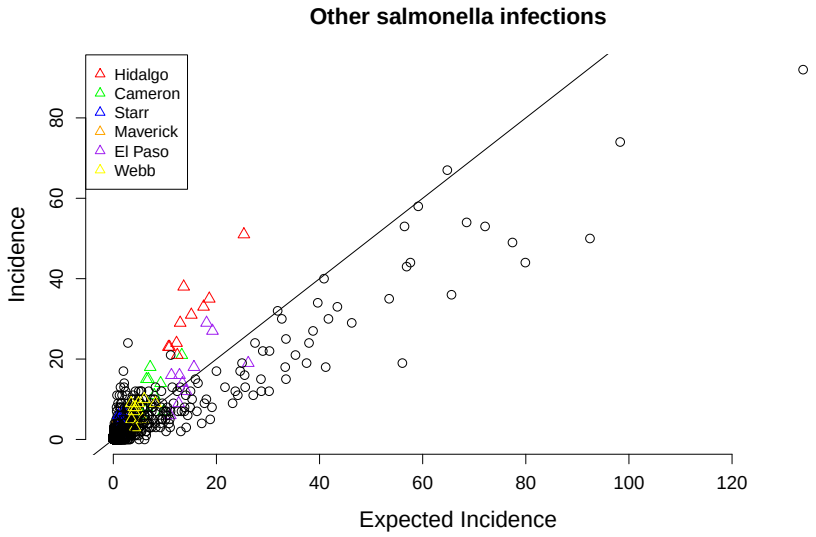


Figure B.17: Observed vs. Expected Incidence, Other Typhus

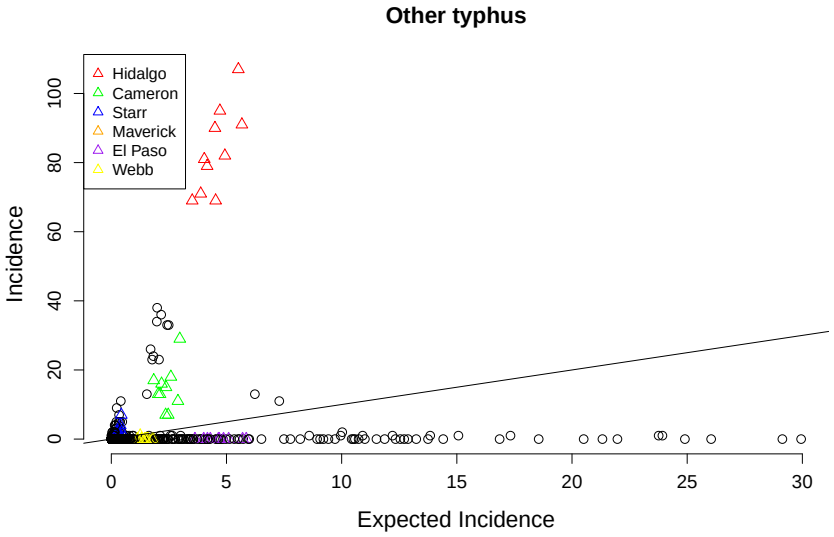


Figure B.18: Observed vs. Expected Incidence, Typhoid and paratyphoid fevers

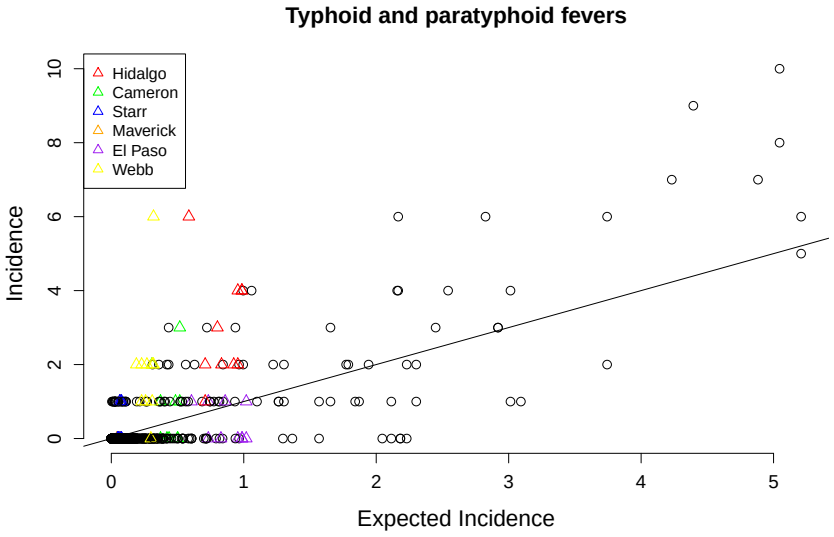


Figure B.19: Observed vs. Expected Incidence, Other Viral Exanthemata

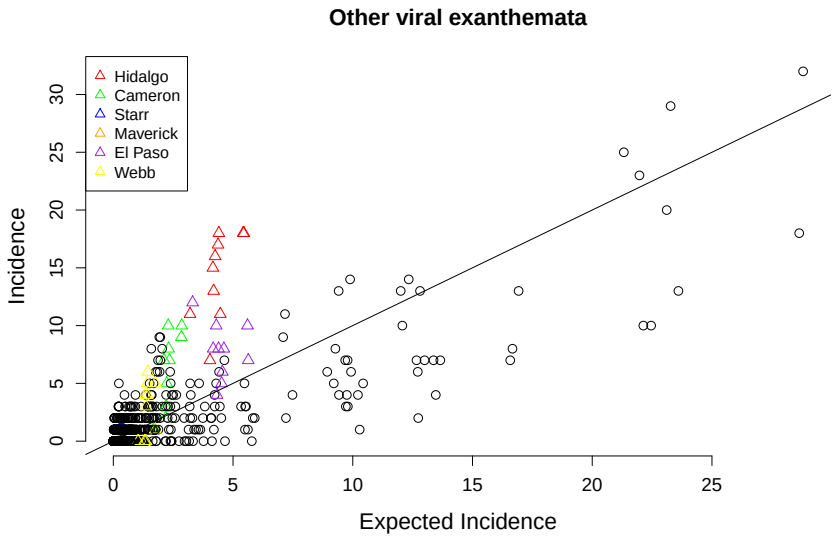


Figure B.20: Observed vs. Expected Incidence, Pulmonary Tuberculosis

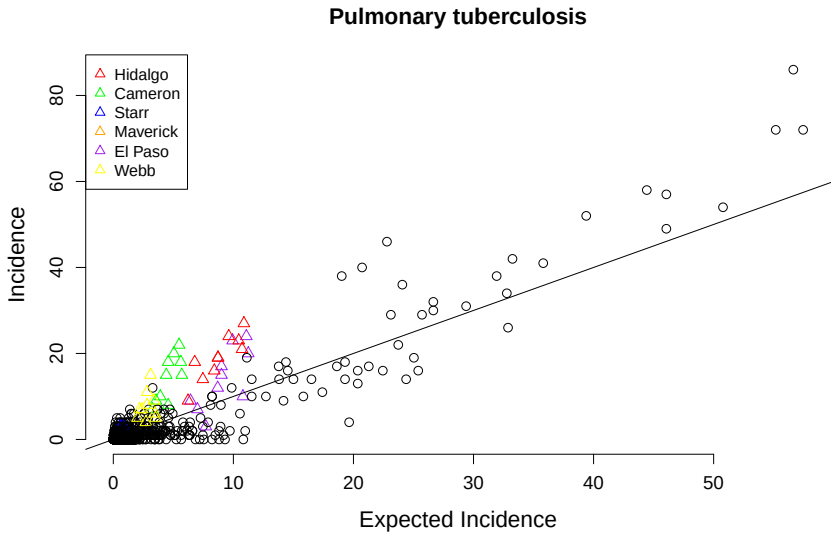


Figure B.21: Observed vs. Expected Incidence, Septicemia

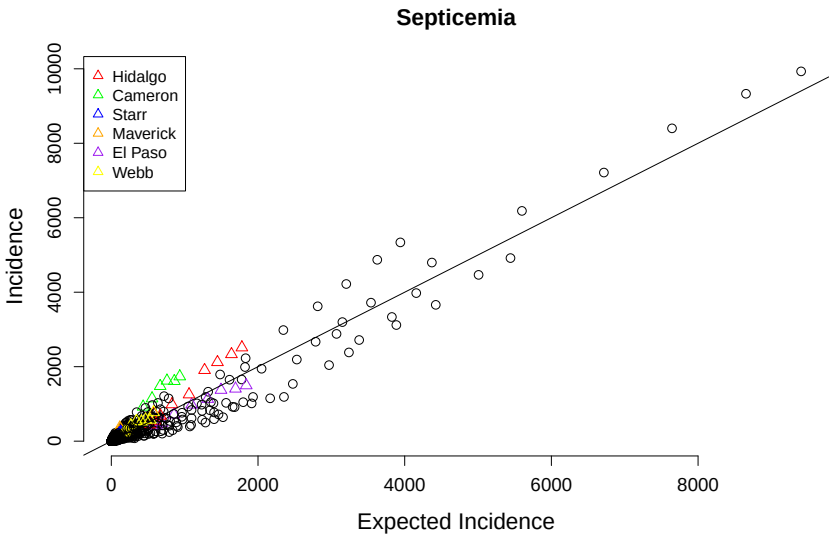


Figure B.22: Observed vs. Expected Incidence, Shigellosis

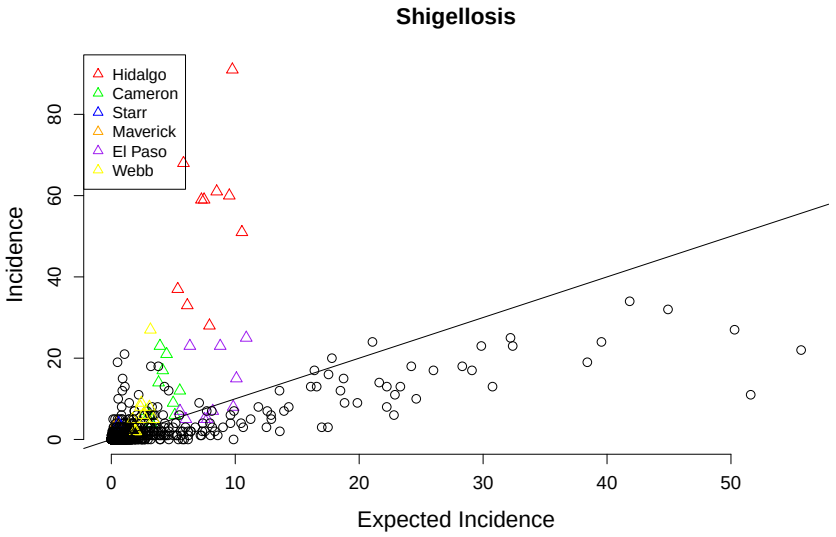


Figure B.23: Observed vs. Expected Incidence, Specific Diseases due to Coxsackie Virus

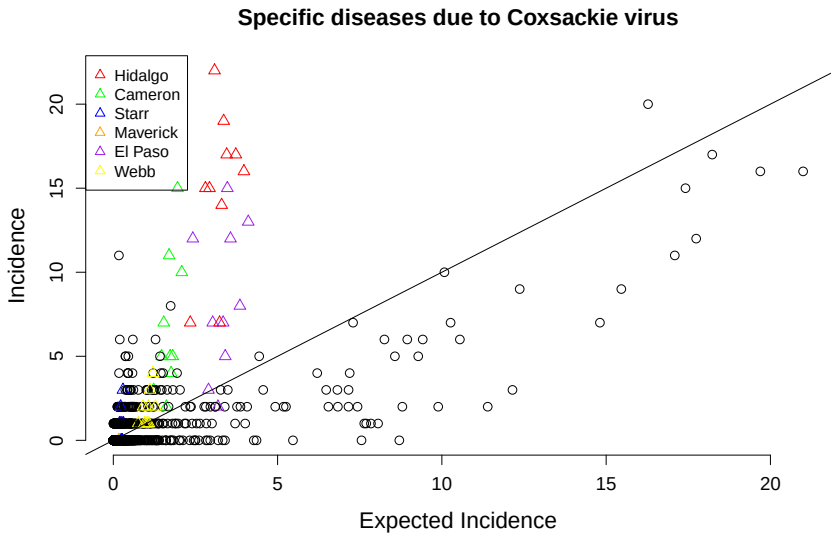


Figure B.24: Observed vs. Expected Incidence, Streptococcal Sore Throat and Scarlet Fever

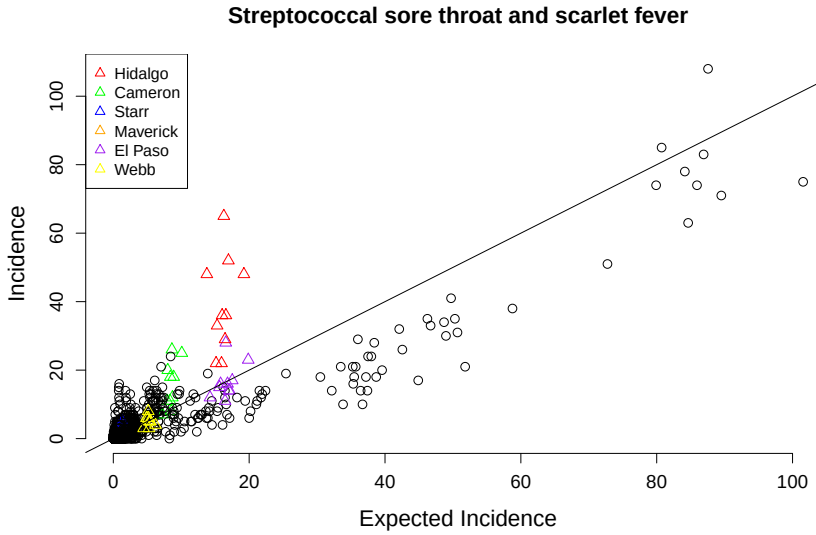


Figure B.25: Observed vs. Expected Incidence, Viral and Chlamyidal Infections

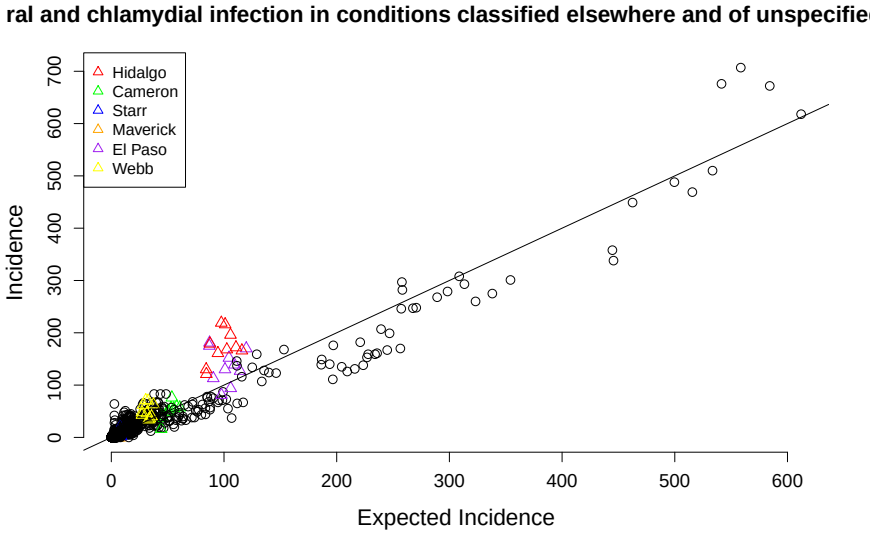
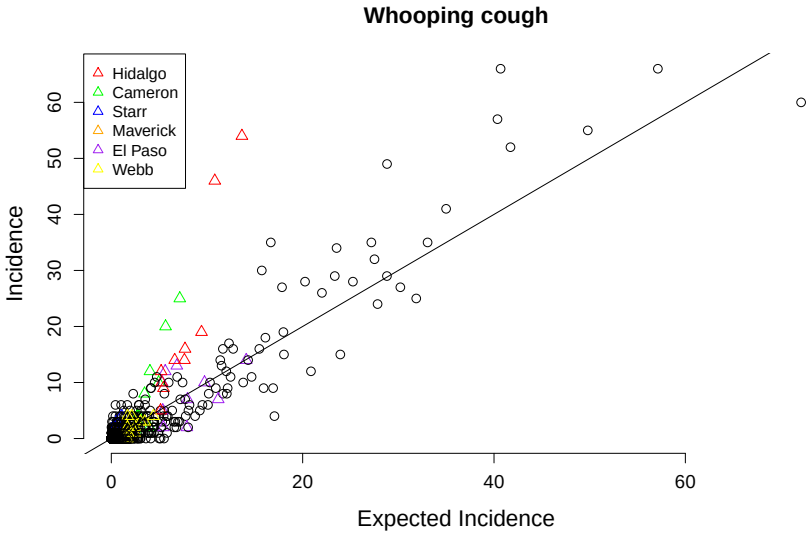


Figure B.26: Observed vs. Expected Incidence, Whooping Cough



Appendix C

Disease Category Lookup Tables

Table C.1: Actinomycotic Infections through Herpes Simplex

Disease Category	Subsumes
Actinomycotic Infections	Actinomycotic infections
Amebiasis	Acute amebic dysentery; Chronic intestinal amebiasis; Amebic nondysenteric colitis; Amebic liver abscess; Amebic lung abscess; Amebic brain abscess; Amebic skin ulceration; Amebic infection of other sites; Amebiasis unspecified.
Brucellosis	Brucella melitensis; Brucella abortus; Brucella suis; Brucella canis; Other brucellosis; Brucellosis unspecified.
Candidiasis	Candidiasis
Coccidioidomycosis	Primary coccidioidomycosis (pulmonary); Primary extrapulmonary coccidioidomycosis; Coiccidioidal meningitis; Other forms of progressive coccidioiomycosis; Chronic pulmonary coccidioidomycosis; Pulmonary coccidioidomycosis unspecified; Coccidioicomycosis unspecified
Herpes Simplex	Eczema herpeticum; Genital herpes; Genital herpes, unspecified; Herpetic vulvovaginitis; Herpetic ulceration of vulva; Herpetic infection of penis; Other; Herpetic gingivostomatitis; Herpetic meningoencephalitis; With ophthalmic complications; With unspecified ophthalmic complication; Herpes simplex dermatitis of eyelid; Dendritic keratitis; Herpes simplex disciform keratitis; Herpes simplex iridocyclitis; Other; Herpetic septicemia; Herpetic whitlow; With other specified complications; Visceral herpes simplex; Herpes simplex meningitis; Herpes simplex otitis externa; Herpes simplex myelitis; Other; With unspecified complication; Herpes simplex without mention of complication

Table C.2: Herpes Zoster through Intestinal Infections due to Other Organisms

Disease Category	Subsumes
Herpes Zoster	With meningitis, With other nervous system complications, With unspecified nervous system complication, Geniculate herpes zoster, Postherpetic trigeminal neuralgia, Postherpetic polyneuropathy, Herpes zoster myelitis, Other, With ophthalmic complications, Herpes zoster dermatitis of eyelid, Herpes zoster keratoconjunctivitis, Herpes zoster iridocyclitis, Other, With other specified complications, Otitis externa due to herpes zoster, Other, With unspecified complication, Herpes zoster without mention of complication
Ill Defined Intestinal Infections	Infectious colitis, enteritis, and gastroenteritis; Colitis, enteritis, and gastroenteritis of presumed infectious origin; Infectious diarrhea; Diarrhea of presumed infectious origin
Intestinal Infections due to Other Organisms	Escherichia coli [E coli]; E coli, unspecified; Enteropathogenic E coli; Enterotoxigenic E coli; Enteroinvasive E coli; Enterohemorrhagic E coli; Other intestinal E coli infections; Arizona group of paracolon bacilli; Aerobacter aerogenes; Proteus (mirabilis) (morganii); Other specified bacteria; Staphylococcus; Pseudomonas; Campylobacter; Yersinia enterocolitica; Clostridium difficile; Other anaerobes; Other gram-negative bacteria; Other; Bacterial enteritis, unspecified; Enteritis due to specified virus; Rotavirus; Adenovirus; Norwalk virus; Other small round viruses [SRV's]; Calicivirus; Astrovirus; Enterovirus NEC; Other viral enteritis; Other organism, not elsewhere classified

Table C.3: Other food poisoning - Bacterial through Other typhus

Disease Category	Subsumes
Other Food Poisoning - Bacterial	Staphylococcal food poisoning; Botulism food poisoning; Food poisoning due to Clostridium perfringens [C welchii]; Food poisoning due to other Clostridia; Food poisoning due to Vibrio parahaemolyticus; Other bacterial food poisoning; Food poisoning due to Vibrio vulnificus; Other bacterial food poisoning; Food poisoning, unspecified
Other Rickettsioses	Q fever; Trench fever; Rickettsialpox; Other specified rickettsioses; Rickettsiosis, unspecified
Other Salmonella Infections	Salmonella gastroenteritis; Salmonella septicemia; Localized salmonella infections; Localized salmonella infection, unspecified; Salmonella meningitis; Salmonella pneumonia; Salmonella arthritis; Salmonella osteomyelitis; Other; Other specified salmonella infections; Salmonella infection, unspecified
Other enterovirus diseases of the central nervous system	Other enterovirus diseases of the central nervous
Other non-arthropod-borne viral diseases of cent nerv system	Lymphocytic choriomeningitis, Meningitis due to adenovirus, Other specified non-arthropod-borne viral diseases of central nervous system, Unspecified non-arthropod-borne viral diseases of central nervous system
Other protozoal intestinal diseases	Balantidiasis, Giardiasis, Coccidiosis, Intestinal trichomoniasis, Cryptosporidiosis, Cyclosporiasis, Other specified protozoal intestinal diseases, Unspecified protozoal intestinal diseas
Other Typhus	Murine [endemic] typhus; Brill's disease; Scrub typhus; Typhus, unspecified

Table C.4: Other Viral Exanthemata through Shigellosis

Disease Category	Subsumes
Other Viral Exanthemata	Erythema infectiosum [fifth disease]; Other specified viral exanthemata; Viral exanthem, unspecified
Pulmonary Tuberculosis	Tuberculosis of lung, infiltrative; Tuberculosis of lung, nodular; Tuberculosis of lung with cavitation; Tuberculosis of bronchus; Tuberculous fibrosis of lung; Tuberculous bronchiectasis; Tuberculous pneumonia [any form]; Tuberculous pneumothorax; Other specified pulmonary tuberculosis; Pulmonary tuberculosis, unspecified
Septicemia	Streptococcal septicemia; Staphylococcal septicemia; Staphylococcal septicemia, unspecified; Methicillin susceptible Staphylococcus aureus septicemia; Methicillin resistant Staphylococcus aureus septicemia; Other staphylococcal septicemia; Pneumococcal septicemia [Streptococcus pneumoniae septicemia]; Septicemia due to anaerobes; Septicemia due to other gram-negative organisms; Gram-negative organism, unspecified; Hemophilus influenzae [H influenzae]; Escherichia coli [E coli]; Pseudomonas; Serratia; Other; Other specified septicemias; Unspecified septicemia
Shigellosis	Shigella dysenteriae; Shigella flexneri; Shigella boydii; Shigella sonnei; Other specified shigella infections; Shigellosis, unspecified

Table C.5: Specific Diseases due to Coxsackie Virus through Whooping Cough

Disease Category	Subsumes
Specific Diseases due to Coxsackie Virus	Herpangina; Epidemic pleurodynia; Coxsackie carditis; Coxsackie carditis, unspecified; Coxsackie pericarditis; Coxsackie endocarditis; Coxsackie myocarditis; Hand, foot, and mouth disease; Other specified diseases due to Coxsackie virus
Streptococcal Sore Throat and Scarlet Fever	Streptococcal sore throat; Scarlet fever
Typhoid and paratyphoid fevers	Typhoid fever, Paratyphoid fever A, Paratyphoid fever B, Paratyphoid fever C, Paratyphoid fever, unspecified
Viral and Chlamydial Infections	Adenovirus; ECHO virus; Coxsackie virus; Rhinovirus; Human papillomavirus; Retrovirus; Retrovirus, unspecified; Human T-cell lymphotrophic virus, type I [HTLV-I]; Human T-cell lymphotrophic virus, type II [HTLV-II]; Human immunodeficiency virus, type 2 [HIV-2]; Other specified retrovirus; Respiratory syncytial virus (RSV); Other specified viral and chlamydial infections; Hantavirus; SARS-associated coronavirus; Parvovirus B19; Other specified chlamydial infection; Other specified viral infection; Unspecified viral and chlamydial infections; Unspecified chlamydial infection; Unspecified viral infection
Whooping Cough	Bordetella pertussis [B pertussis]; Bordetella parapertussis [B parapertussis]; Whooping cough due to other specified organism; Whooping cough, unspecified organism

Appendix D

Code

D.1 Stan Code

```
/** NOTE: This code is only a slight modification
of Max Joseph's; see references. **/

functions {
  /**
  * Return the log probability of a proper conditional autoregressive (CAR) prior
  * with a sparse representation for the adjacency matrix
  *
  * @param phi Vector containing the parameters with a CAR prior
  * @param tau Precision parameter for the CAR prior (real)
  * @param alpha Dependence (usually spatial) parameter for the CAR prior (real)
  * @param W_sparse Sparse representation of adjacency matrix (int array)
  * @param n Length of phi (int)
  * @param W_n Number of adjacent pairs (int)
  * @param D_sparse Number of neighbors for each location (vector)
  * @param lambda Eigenvalues of D^{-1/2}*W*D^{-1/2} (vector)
  *
  * @return Log probability density of CAR prior up to additive constant
  */
  real sparse_car_lpdf(vector phi, real tau, real alpha,
    int[,] W_sparse, vector D_sparse, vector lambda, int n, int W_n) {
      row_vector[n] phit_D; // phi' * D
      row_vector[n] phit_W; // phi' * W
      vector[n] ldet_terms;

      phit_D = (phi .* D_sparse)';
      phit_W = rep_row_vector(0, n);
      for (i in 1:W_n) {
        phit_W[W_sparse[i, 1]] = phit_W[W_sparse[i, 1]] + phi[W_sparse[i, 2]];
        phit_W[W_sparse[i, 2]] = phit_W[W_sparse[i, 2]] + phi[W_sparse[i, 1]];
      }

      for (i in 1:n) ldet_terms[i] = log1m(alpha * lambda[i]);
```

```
      return 0.5 * (n * log(tau)
                    + sum(ldet_terms)
                    - tau * (phit_D * phi - alpha * (phit_W * phi)));
  }
}
data {
  int<lower = 1> n;   // total observations
  int<lower = 1> p;   // total spatial regions
  int<lower = 1> k;   // total number of covariates
  int<lower = 1> num_col_cnts;
  int<lower = 1> num_nocol_cnts;
  matrix[p, k] Z;     // covariates
  int cnty_idx[n];
  int cols_ids[num_col_cnts];      // ids of counties with colonias
  int nocols_ids[num_nocol_cnts];  // ids of counties without colonias
  int<lower = 0> y[n];
  vector[n] log_offset;
  matrix<lower = 0, upper = 1>[p,p] W; // adjacency matrix
  int W_p;                // number of adjacent region pairs
}
transformed data {
  int W_sparse[W_p, 2];   // adjacency pairs
  vector[p] D_sparse;     // diagonal of D (number of neigbors for each site)
  vector[p] lambda;       // eigenvalues of invsqrtD * W * invsqrtD

  { // generate sparse representation for W
  int counter;
  counter = 1;
  // loop over upper triangular part of W to identify neighbor pairs
    for (i in 1:(p - 1)) {
      for (j in (i + 1):p) {
        if (W[i, j] == 1) {
          W_sparse[counter, 1] = i;
          W_sparse[counter, 2] = j;
          counter = counter + 1;
        }
      }
    }
  }
  for (i in 1:p) D_sparse[i] = sum(W[i]);
  {
    vector[p] invsqrtD;
    for (i in 1:p) {
      invsqrtD[i] = 1 / sqrt(D_sparse[i]);
    }
```

```
    lambda = eigenvalues_sym(quad_form(W, diag_matrix(invsqrtD)));
  }
}
parameters {
  vector[p] phi; // Spatially Varying Random Effect
  real beta_0;
  vector[k] beta;
  real<lower = 0> tau;
  real<lower = 0, upper = 1> alpha;
}

transformed parameters{
  vector[p] mu;
  vector[p] rr;
  vector[k] rr_cov;

  mu = beta_0 + Z*beta + phi;
  rr = exp(mu);
  rr_cov = exp(beta);

}

model {
  phi ~ sparse_car(tau, alpha, W_sparse, D_sparse, lambda, p, W_p);

  tau ~ gamma(2,2);
  beta ~ normal(0,5);
  beta_0 ~ normal(0,5);

  for (i in 1:n){
    y[i] ~ poisson_log(log_offset[i] + mu[cnty_idx[i]]);
  }

}

generated quantities{
  real col_diff;
  real cont;
  col_diff = mean(exp(mu[cols_ids])) / mean(exp(mu[nocols_ids]));
  cont = exp(beta[1] + beta[2] - beta[3]);
}
```

Bibliography

[1] County characteristics datasets: Annual county resident population estimates by age, sex, race, and hispanic origin: April 1, 2010 to july 1, 2014. http://www.census.gov/popest/data/counties/asrh/2014/CC-EST2014-ALLDATA.html.

[2] Directory of colonias located in texas. http://www.sos.state.tx.us/border/colonias/reg-colonias/index.shtml. Accessed: 2016-11-20.

[3] Diseases and conditions. http://www.cdc.gov/DiseasesConditions/. Accessed: 2016-11-20.

[4] Hospital discharge data public use data file. https://web.archive.org/web/20160319012612/http://www.dshs.state.tx.us/THCIC/Hospitals/Download.shtm. Accessed: 2016-11-20.

[5] Small area income and poverty estimates: State and county estimates for 2010. https://www.census.gov/did/www/saipe/data/statecounty/data/2010.html. accessed: 2016-11-20.

[6] Texas secretary of state colonias faqs. http://www.sos.state.tx.us/border/colonias/faqs.shtml. Accessed: 2016-11-20.

[7] J. Baillargeon, C.T. Leach, J.H. Deng, S.J Gao, and H.B. Jenson. High prevalence of human herpesvirus 8 infection in south texas children. *The Journal of Medical Virology*, 67(4):542, August 2002.

[8] J. Besag, J. York, and A. Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, (1):1, 1991.

[9] T. Doyle and R. Bryan. Infectious disease morbidity in the us region bordering mexico, 1990-1998. *The Journal of Infectious Diseases*, 182(5):1503, 2000.

[10] A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp. *Handbook of Spatial Statistics*. CRC Press, Boca Raton, FL, 2010.

[11] A. Gelfand and P. Vounatsou. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11, 2003.

[12] A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2007.

[13] X. Jin, B. Carlin, and S. Banerjee. Generalized hierarchical multivariate car models for areal data. *Biometrics*, 61(4):950, 2005.

[14] Max Joseph. Exact sparse car models in stan. `http://mc-stan.org/documentation/case-studies/mbjoseph-CARStan.html`, 2016. Accessed: 2016-11-20.

[15] C.T. Leach, F.C. Koo, S.G. Hilsenbeck, and H.B. Jenson. The epidemiology of viral hepatitis in children in south texas: increased prevalence of hepatitis a along the texas-mexcio border. *The Journal of Infectious Disease*, 180(2):509, August 1999.

[16] C.T. Leach, F.C. Koo, T.L. Kuhls, S.G. Hilsenbeck, and H.B. Jenson. Prevalence of cryptosporidium parvum infection in children along the texas-mexico border and associated risk factors. *The American Journal of Tropical Medicine and Hygiene*, 62(5):656, May 2000.

[17] K.P. Vatcheva, M. Lee, J. McCormick, and M. Rabhar. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology*, 6(2):227, April 2016.

[18] Peter Ward. *Colonias and Public Policy in Texas and Mexico: Urbanization by stealth*. University of Texas Press, Austin, TX, 1999.