Application of Hierarchical Bayes to Estimate Relative Risk of Suicide in Idaho Counties

A Thesis
Presented in Partial Fulfilment of the Requirements for the
Degree of Master of Science
with a
Major in Statistical Science
in the
College of Graduate Studies
University of Idaho
by
Ernestina Osei Boateng

Major Professor: Michelle Wiest, Ph.D.
Committee Members: Timothy Johnson, Ph.D.; Clarissa Richardson, Ph.D.
Department Administrator: Christopher Williams, Ph.D.

May 2020

Authorization to Submit Thesis

This thesis of Ernestina Osei Boateng, submitted for the degree of Master of Science with a major in Statistical Science and titled "Application of Hierarchical Bayes to Estimate Relative Risk of Suicide in Idaho Counties," has been reviewed in final form. Permission, as indicated by the signatures and dates given below, is now granted to submit final copies to the College of Graduate Studies for approval.

Major Professor: ______________________________ Date____________
Michelle Wiest, Ph.D.

Committee
Members: ______________________________ Date____________
Timothy Johnson, Ph.D.

______________________________ Date____________
Clarissa Richardson, Ph.D.

Department
Administrator: ______________________________ Date____________
Christopher Williams, Ph.D.

Abstract

Suicide is the intention of harming oneself to cause death. Idaho is amongst the states with growing rates of suicide deaths and suicide is a major public health issue. Most studies conducted on suicide have investigated suicide rates at state levels but not small areas such as counties or demographic groups. This study was conducted to develop a reliable model to characterize suicide at the county level.

Relative risk of suicide for all counties in Idaho were obtained using Poisson hierarchical Bayes model and the performance was compared to standardized mortality ratio estimates. Relative risk estimates obtained using Poisson hierarchical Bayes had low standard errors compared to the relative risk estimates obtained using standardized mortality ratio. These estimates allowed us to observe that age adjusted relative risk of suicide was consistently above that of the state of Idaho in Custer, Lemhi, Shoshone and Nez Perce. Also, Madison was the only county with age adjusted relative risk of suicide consistently below the state of Idaho over the time period.

County level suicide risk factors such as unemployment, education and social connectivity could be incorporated into the model to predict relative risk of suicide. Local government and public health experts could use predictions from the model to target suicide preventive measures.

Acknowledgements

My sincere gratitude to the Almighty God for making my thesis a success. I would like to thank my major professor Dr. Michelle Wiest for her countless sacrifices and support throughout the academic year. Her guidance, support and concerns about this project always motivated me to work harder. She was instrumental by providing insights and expertise in directing me not to only succeed in this research but future works as well. I am eternally grateful to her.

Also my sincere appreciation to my committee members Dr. Timothy Johnson and Dr. Clarissa Richardson for their immense contribution, time and support towards this study. Special thanks to Dr. Erkan Buzbas for his willingness to help at anytime.

My sincere thanks to Samantha Morgan for her efforts in making the data used in this study available.

Dedication

To my parents Mr and Mrs Anane Boateng, and siblings; Frank Owusu Boateng, Mavis Animwaa Boateng, Richard Anka Boateng. I am so grateful for all the support and financial sacrifices you made throughout my schooling. I do not know how life would be without your love, prayers and encouragement.

Table of Contents

List of Figures

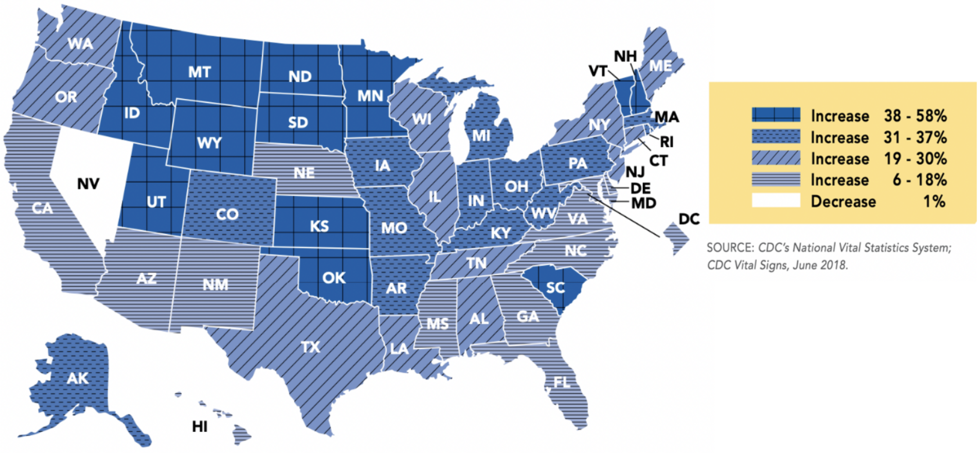
CHAPTER 1

Introduction

1.1 Background of Study

Suicide is the second leading cause of death in the world. Sadly, every 40 seconds someone commits suicide somewhere in the world [1]. In the United States (US), suicide is the tenth leading cause of death [2] and therefore is an important public health issue [1].

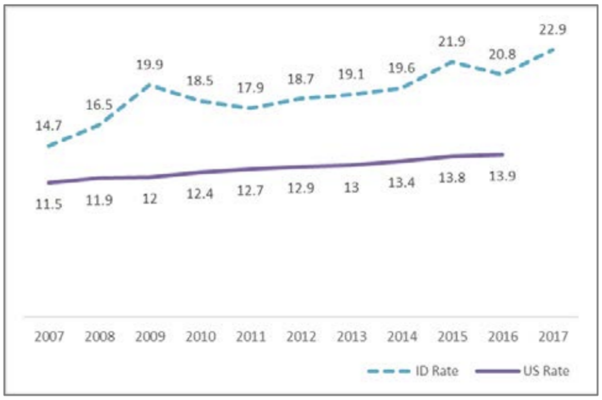
In the US alone, about 10.6m people have serious thoughts of committing suicide [3]. Commonly, mental condition is seen as the cause of suicide, but suicide is rarely caused by one condition and 54 percent of suicide issues reported did not have a known mental issue [4]. In 2018, 48,344 Americans died by suicide. In 2017, there were 1.4m suicide attempts with a total of 47,173 completed suicides in the US with an average of one death in every eleven minutes [3]. In 2016, there were a total of 44,965 deaths by suicide in the US. Thus on average every twelve minutes an individual dies by suicide [5]. The rate of suicide keeps increasing with no assurance of decreasing. In 2018, the race with highest age adjusted suicide rate was American Indians and Alaska natives with 22.1 per 100,000 followed by whites with 18 per 100,000. Black or African Americans together with Asians Pacific Islander had the lowest rates, 7.2 and 7.0 per 100,000 respectively [6]. Idaho was among the states with the highest increase in suicide rates from 1999 to 2016 [7]. Figure 1.1 shows the percentage increase in suicide rates in the US.



SOURCE: *CDC's National Vital Statistics System; CDC Vital Signs, June 2018.*

Figure 1.1: Suicide rates
Source: Center for Disease Control National Vital Statistics System

Though the population size of Idaho is relatively smaller than most states, Idaho ranked fifth among the top ten states with leading causes of death by suicide in 2017 [8]. Moreover, suicide rates in Idaho was consistently higher than the overall rates in the US from 2007-2017 [5], making suicide the eighth leading cause of death in Idaho from 2010 to 2014 [10]. Though the rates of suicides decreased from 2015 to 2016 in Idaho, suicide rate was about 50 percent above the national average [5]. On the average, every 24 hours someone dies by suicide in Idaho [10]. Figure 1.2 shows the rates of suicide in Idaho and overall rates of suicide in the United States from 2007 to 2017.



***Rate per 100,000 population**

Figure 1.2: Suicide Rates
Source: Idaho Department of Health and Welfare

For a period of ten years (2005-2014), the suicide rate of males was significantly higher than that of females. Within the ten years period, suicide rate for females significantly increased from 4.8 to 10.1 but there was no significant increase in suicide rate for males(28.0-30.4) [11]. Also, suicide death by both male and females in Idaho exceeds the national average of both males and females in the US. Suicide is the second leading cause of death between the ages of 15-34 in both males and females and up to the age of 44 in males. Surprisingly, 105 children between the age group 6-18years died by suicide in Idaho, 27 of whom were below 14years [12]. Figure 1.3 shows the age adjusted suicide rates for all counties in Idaho for a five year period (2009-2013).

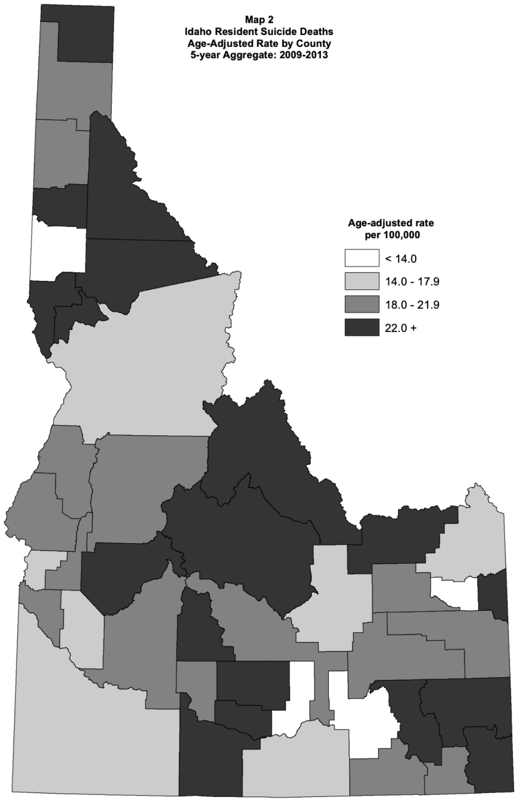


Figure 1.3: Age Adjusted Suicide Rates
Source: Idaho Department of Health and Welfare

Aside from the emotional stress caused by suicide, there is a heavy financial burden on the government. In 2010, suicide and self- reported injury cost Idaho $329,244,000 with an average of about $1,135,327 per suicide death [10]. In 2013, suicide cost the US government

$51billion for medical and work loss cost [9]. In 2015, the US government spent $69 billon on suicide and self-reported injuries [2]. Furthermore, Idaho ranked ninth in suicide deaths per capita with a rate of 19.6 per 100,000 in 2014. By 2016, suicide deaths per capita increased to 20.8 per 100,000 [5].

1.2 General Overview of Idaho Population

Idaho is the 14th largest state in the US with about 83,500sqm land. Majority of this land is managed by the federal government. While the state has a large land mass per square mile, the population is smaller than other states with fewer square miles. Even so, the population of Idaho increased from 1,570,773 to 1,754,208 people according to the US Center of Bureau population estimates from 2010 to 2018 [13]. Idaho is comprised of 44 counties with seven health districts [11]. The seven health districts are Panhandle, North Central, Southwest, Central, South Central, Southeastern and Eastern Idaho [14]. At least 90.6 percent of Idaho population graduated from high school and about 26.9 percent graduated from college. 54.6 percent of the population aged above sixteen were employed. The estimated median income was $53,084. 3.6 percent had income above $200k and six percent had income below $10k. An estimated number of 13.6 percent of Idaho population between 18-64 were below the poverty level [15]. Counties that experienced the greatest population change in Idaho from 2010-2015 included Ada, Canyon, Kootenai, Latah, Boundary, Bonneville, Valley, Teton, Jefferson and Twin Falls. ELmore, Custer, Camas, Butte, Shoshone, Fremont, Lemhi Clearwater and Clark were hit with the greatest loss [16]. Figure 1.4 shows the percentage increase and decrease for all counties in Idaho.

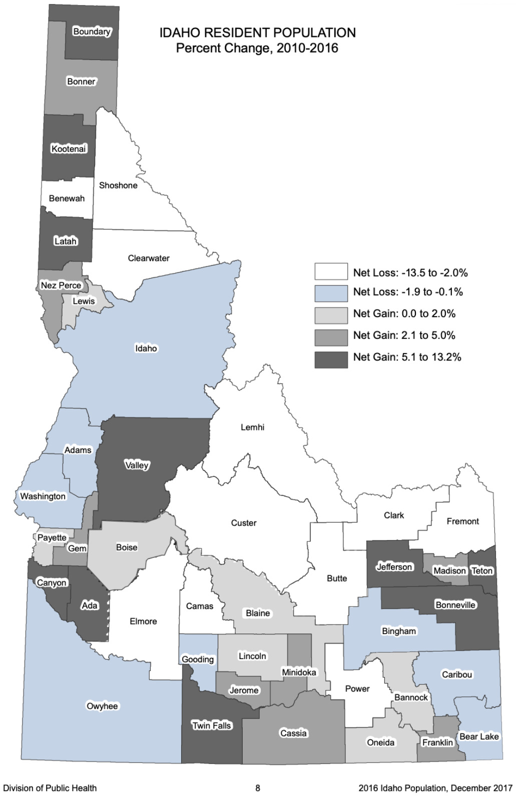


Figure 1.4: Idaho Resident Population
Source: IDHW, Division of Public Health

1.3 Calculating Suicide Rates

Suicide rates are calculated by aggregating over a five-year period the number of deaths by the Bureau of Vital Records and Health Statistics. Rates are calculated using crude rates and direct standardization [11]. A crude rate(age unadjusted) is the total number of deaths divided by the population at risk per 100,000. Crude rates do not take into account the age distribution and therefore are poor estimates of comparing among different groups or assessing mortality over time [18]. Direct standardization applies the age specific rates of the area of interest to the age distribution of the reference population to obtain the expected of deaths that would have been observed if the age distribution of the area of interest was same as the reference population. Direct standardization is normally labelled according to its application. Examples include ASIR (Age-Standardized Incidence Rate), the ADR

(Adjusted Death Rate) and the AAR (Age Adjusted Rate) [19]. For the purpose of this study, age adjusted rates are used. Age adjusted rates are estimates that show what the mortality rate would be if the age distribution of the area of interest does not change from year to year. As such it should not be viewed as actual measures of death rates but rather relative measures. Age adjusted rates are better indicators than crude rates, when the age distribution of the area of interest is changing over a period of time. The age specific death rate is the number of deaths per 100,000 population for a specific age distribution over a period of time (rate for a given age group) [20]. The population age distribution is categorized to the age groups used in the 2000 US standard population estimates. The steps used in the estimation of age adjusted rate are outlined below [11]:

$$\text{Age specific death rates (ASDR)} = \frac{\text{total deaths in an specific age group}}{\text{total population in specific age group}} \tag{1.1}$$

$$E_s = P_s \times \text{ASDR} \tag{1.2}$$

$$\text{Age adjusted death rates} = \sum_s E_s \tag{1.3}$$

- s represents the age specific groups.
- P_s represents the standard population for age specific groups.
- E_s represents the expected number of deaths for age specific groups.

Direct standardization is inappropriate when the age specific groups of the populations being compared are not consistent [18]. Therefore, direct standardization is not feasible when specific rates of the area of interest are not available [21]. Due to this issue, indirect standardization measured by standardized mortality ratio (SMR) is mostly preferred. Direct standardization converts counts to rates while with indirect standardization, the population of the area of interest is compared with a reference population [19]. SMR measures the observed number of deaths (O_i) to the expected number of deaths (E_i) in a general population or a specific geographic area, specifically on the basis of age and sex specific death rates [22]. Thus SMR measures the risk of an area with respect to the reference population [23]. Moreover, SMR compares the relative mortality of a population to another population but fails to adjust for mortality risk factors such as age or gender between the two populations [24]. An SMR of 1.0 means the observed number of deaths of the area of interest are the same as the expected number of deaths of the area; that is, the risk of the area is not different from that of the reference population. SMR of 1.2 implies there is 20 percent increase in the overall death rate of that specific area. Similarly SMR of 0.95 means a decrease of 5 percent overall death rate of that specific area [25].

The number of outcomes or population size of an area affects the SMR. Geographic areas with high number of outcomes or large population sizes yield stable estimates, however, with rare outcomes or the small population size, SMR tends to produce unstable estimates [26].

Confidence intervals for SMR can be computed to determine the reliability of estimates given the width of the interval. For events more than 100 deaths, equation (1.5) should be used, otherwise, use equation (1.6). The numerator of equation (1.6) is computed from the Poisson distribution based on the number of events observed. The Poisson distribution returns a lower and upper limits which is then divided by the expected number of cases to yield the appropriate confidence intervals [27].

$$\begin{aligned} \text{Lower Limit} &= \left[1 - \frac{1}{9 \cdot O_i} - \frac{z_{\alpha/2}}{3\sqrt{O_i}}\right]^3 \frac{O_i}{E_i} \\ \text{Upper Limit} &= \left[1 - \frac{1}{9 \cdot (O_i + 1)} - \frac{z_{\alpha/2}}{3\sqrt{O_i} + 1}\right]^3 \frac{O_i + 1}{E_i} \end{aligned} \tag{1.4}$$

$$\begin{aligned}\text{Lower Limit} =& \frac{LL}{E_i}\\ \text{Upper Limit} =& \frac{UL}{E_i}\end{aligned} \tag{1.5}$$

- $z_{\alpha/2}$ = a value obtained from the standard normal table where α is the confidence level.
- LL = lower limit obtained from the Poisson distribution.
- UP = Upper limit obtained from the Poisson distribution.

1.4 Challenges Faced Calculating Estimates in Small Area

A small area represents a small geographic area or sub populations such as town, county, age-sex-race of groups of people within a large area [28]. Small geographic areas present small populations and rarity of events which yield unreliable estimates [11]. The rate of an event is affected by the number of outcomes involved. A smaller number of events tend to yield rates with wider variability due to the large confidence intervals. Rates based on a smaller number of events are considered unstable and should not be used for comparative purposes or assessing the severity of an outcome [11][19][20]. Relevant to estimating rates of suicide, a small number of deaths in a geographic area with a small population size will tend to have high random variability. Therefore to estimate the rates, areas with small numbers are mostly aggregated over time period or areas with similarities are summed together [11]. SMRs also tend to be highly affected by small population sizes since the variance of SMR is proportional to $1/E_i$ [22]. Small E_i due to small population size will yield large variability [23][24]. SMRs result in estimates that are large due to relatively small changes in expected value (areas with small population sizes). Also, an almost zero expected value will yield large SMR [23].

Furthermore, SMRs could be misleading especially in areas where there are zero counts. Two geographic areas with zero counts will have the same observed value but different expected number of deaths. The calculated SMR (O_i/E_i) for both areas will be zero regardless of the population size . Another drawback of SMR is losing potential information for the areas of interest, thus the actual risk associated with each county is omitted by SMR. This results in unreliable estimates thereby making it difficult if not impossible for detailed interpretation of results. For example, an area with 1000 observed and expected value of 1000 will have an SMR of 1. Similarly, an area with 2 observed and expected value of 2 will also have SMR of 1, thereby making interpretation of relative risk estimates difficult [26][29]. Also, in disease mortality where estimates are mapped, estimates derived from SMRs are exaggerated usually in areas with few observed. Certain areas maybe presented as areas with high risk of an event, but this might not be necessarily true [30].

Moreover, SMRs fails to account for between and within area variability in estimating relative risk. The use of SMR may not be appropriate for small areas since it tends to yield estimates that do not account for high variability between different geographic areas [31][32]. This can be attributed to the fact SMRs are calculated independently from other geographic areas [33]. Due to the rarity of suicide in some counties associated with small population sizes, SMRs are unstable and unreliable [23]. Small area estimation specifically model-based methods helps resolve uncertainties associated with direct and indirect standardization [28].

1.5 Studies on Suicide rates

There have been lot of studies on suicide rates with different models in the world. Some of these studies include investigating the relationship between suicide rates and economic crisis. Examples include a study conducted to determine the overall increase in suicide estimates during an economic recession in 2008. There was an overall increase of 4884 deaths by suicide in 2008 than what was expected. Most countries experienced economic crisis in that year. There was an increase of 4.2 and 6.4 percent in suicide rates in men in 27 European countries and 18 countries across North and South America respectively. Negative binomial regression was used to model suicidal trend from 2000-2007. Spearman rank correlation was used to determine the relationship between suicide rates and unemployment. Suicide rates increased in countries with higher unemployment rates [34].

With rare studies of suicide rates at the county level, hierarchical Bayes model was used by the Integrated Nested Laplace Approximation (INLA) R package [35][36][37] to estimate suicide rates in 3140 counties in the US from 2005-2015. 87 percent of the counties showed 20 percent increase and 99 percent showed an overall increase of 10 percent in the posterior predicted county level mean estimates. Counties in rural settings had the highest increase in suicide rates. These counties were mostly found in Northern California, Colorado, Idaho, Montana, Nevada, North and South Dakota, Oregon and Wyoming [38].

Negative binomial regression model has also been used to model suicide rates to estimate relative risk of suicide by country and sex from 2001 to 2008. Out of 20 countries, 14 countries including the US had relative risk greater than 2.0 which indicates more people are likely to commit suicide in these countries. The study showed that suicide in adolescents and youth has significantly increased over these years. Furthermore, men had a higher relative risk compared to women. Men were 3.2 times likely to commit suicide than women [39].

Empirical Bayes was employed by the use of penalized quasi likelihood to estimate the relative risk of suicide in study of geographic distribution of suicide rates in England and Wales. Geographic areas with large population density, urban areas for both sexes, economically depressed district and rural areas for men showed high trend in relative risk of suicide [29].

Poisson regression model with random effect modelling between area and local variability has also been used to model suicide rates. The model was used to map age and sex specific rates of suicide in England and Wales. Remote coastal areas as well as the central parts of the cities showed high rates of suicide [33].

1.6 Purpose of Study

The purpose of this research is 1) to develop a reliable model to estimate relative risk of suicide, and 2) identify counties at high or low risk of suicide relative to the state of Idaho from 1999 to 2016. This study is continuation of a thesis submitted on May 2019 by Rosebella Capio. Her research was titled a Bayesian hierarchical Poisson approach to estimate county level suicide risks in Idaho [17].

CHAPTER 2

Methodology

2.1 Introduction

This chapter presents the objectives of the study, the data used for the study, the models utilized in small area estimation, and statistical methods used in the analysis and its implementation. The chapter also reports detailed explanations of the model employed in this study.

2.2 Motivation and Objectives

The motivation for this study stems from the alarming increase of suicide in Idaho with few or almost no studies in suicide rates in Idaho. The study aims to fit a model-based method to suicide data in Idaho to determine the relative risk of suicide associated with each county over an 18-year period 1999-2016 , to model the trend in time to investigate possible reasons that might be responsible for low or high relative risk associated with various counties in Idaho. This will aid local government officials and decision makers in the apportionment of government funds to counties that are prone to high risk of suicide.

2.3 Dataset

The overall data on counts of death by suicide in Idaho from 1999 to 2016 was obtained from the Idaho Department of Health and Welfare. The data consisted of the year of death, gender, county, age, manner of death, mechanism of death, drug involvement and the different kinds of drugs used. Death by suicide was then extracted with variables of interest, year, sex, county and age. Since adjusted and unadjusted age distribution was of interest in this study, the age distribution of suicide was categorized into eight specific age groups namely, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79 and 80+. In order to estimate expected values of each county for a specific year, population estimates from each of the 44 counties in Idaho from 1999 to 2016 were also obtained from Bureau of Vital Records and Health Statistics. The population data consisted of the following variables: year, age, sex, county, districts and population estimates. The specific age distribution of the suicide counts was categorized according to the specific age distribution of the population data.

2.4 Small Area Estimation

Due to large standard errors associated with small geographic areas, estimates derived are unstable and unreliable. Small area estimation helps provide reliable estimates for the parameter of interest by the use of model-based approaches. Model-based estimators are indirect estimators based on small areas that borrow strength from neighboring areas through linking models and supplementary data such as census and administrative records to estimate reliable measures. With model-based estimation, models can be justified by the use of sample data and best estimators can be derived [28][40]. Specifically, Bayesian models adjust small area estimates by combining the observed data with a prior distribution for all geographic areas under study thereby accounting for within and between area variability [41]. Model-based estimation methods of relative risk improves on SMRs by providing smoothing tools that help in reducing the high variability in SMRs [23]. Model-based approaches help resolve the issue of uncertainty with estimates by providing reliable estimates for small areas with zero observation and differences between counties [29].

2.5 Model Based Approaches Used in Small Area Estimation

Small area estimation methods include synthetic estimation, composite estimation, empirical best linear unbiased prediction (EBLUP), empirical Bayes and hierarchical Bayes etc. Empirical best linear unbiased prediction, empirical Bayes and hierarchical Bayes have various advantages over the others [40]. The parameter of interest for this study is relative risk.

Relative risk measures the risk of an outcome in one group compared to the outcome of the same event occurring in another group [42]. In this study, the interest is comparing the risk of suicide in counties to that of the state of Idaho. Models predominately used in estimating relative risk include empirical and hierarchical Bayes. Empirical and hierarchical Bayes by the use of linking models produce stable and reliable relative risk estimates compared to direct and indirect standardization [30][39]. For the purpose of this study, hierarchical Bayes approach is employed due to the following reasons: The specification of a model in hierarchical Bayes approach is straightforward and easily computes estimates using complex small area models. However, hierarchical Bayes is computationally expensive but the use of Markov chain Monte Carlo (MCMC) methods helps overcome this difficulty [43]. In a study conducted by Hulting et al., hierarchical Bayes approach produced reliable and meaningful estimates compared to the EBLUP with counts close to zero or no counts of the outcome [44]. Hierarchical Bayes approach is very flexible by accommodating different models which account for variation across time and space [45]. In contrast, in a study by Maiti 1997, he found that the empirical Bayes approach does not account for uncertainty in estimates. Moreover, the posterior variance used in the measurement of the reliability of the posterior mean fails to account for over-dispersion due to estimation of model parameters [31].

2.6 Hierarchical Bayes

In Bayesian modelling, inferences are made on the parameter of interest, which is estimated by the posterior mean and its precision is measured by the posterior variance. The posterior distribution of a small area is obtained by combining the likelihood thus the data and the prior distribution by the use of Bayes theorem. The posterior distribution is given as

$$p(\mu \mid y) \propto l(y \mid \mu) \times f(v) \tag{2.1}$$

- $p(u|y)$ is the posterior distribution of the parameter of interest.

- $l(y \mid \mu)$ is the data which is conditional on μ
- $f(v)$ is the prior distribution

In hierarchical Bayes modelling, a prior distribution must be specified on the model parameters. Parameters in the model are regarded as random, therefore parameters are assigned probability distribution to model their behavior. These probability distributions and parameters are known as hyperprior distributions and hyperparameters respectively [28]. The likelihood makes inferences about the parameters given the data. The priors are user chosen which assumes a distribution for the parameters based on prior beliefs or assumptions [23][46]. Normally when the size of the data are large, the data has great effect in the estimation of the relative risk compared to the prior distribution. For simple models, it is possible to obtain the closed form of the posterior distribution. However, when it is not possible to obtain closed form of the posterior distribution, samples are drawn from the posterior distribution by MCMC methods [22].

2.7 Model Description

As stated earlier, the data used in the study are suicide counts. Counts data with rare or low number of events in small areas with relatively large population sizes are best modelled by Poisson distribution [46]. Since the data are sparse, the Poisson model accounts for extra variation (over dispersion) by the use of random effects [47][48]. Suicide counts for each county for each year (y_{it}) is modelled by Poisson distribution. y_{it} is independently distributed as:

$$y_{it} \sim Poisson(\mu_{it}) \tag{2.2}$$

The likelihood is given as:

$$L(y \mid \mu) = \prod_{i=1}^{n} \mu_{it}^{y_{it}} \exp\left(\frac{-\mu_{it}}{y_{it}!}\right) \tag{2.3}$$

The mean (μ_{it}) consist of two components: relative risk and expected number of deaths.

$$\begin{aligned}
\mu_{it} &= \theta_{it} e_{it} \\
&= \theta_{it} \times \left(N_{it} \times \frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} N_{it}}\right) \\
&= \frac{\mu_{it}}{e_{it}} \times \left(N_{it} \times \frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} N_{it}}\right) \\
&= \frac{\mu_{it}}{\left(N_{it} \times \frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} N_{it}}\right)} \times \left(N_{it} \times \frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} N_{it}}\right) \\
&= \mu_{it}
\end{aligned} \tag{2.4}$$

- i represents each counties.
- n is the total number of counties(44).
- t denotes the 18-year period (1999-2016).
- N_{it} is the population of a county per year. The standard population used is Idaho state.
- λ_{it} is the relative risk of suicide per county for a specific year.
- e_{it} represents the expected number of suicides per county for a specific year. The expected deaths were computed for both adjusted and unadjusted age rates using age specific and crude rates respectively. Expected values are computed by indirect standardization. A standard population is chosen as a baseline for the event or outcome to be compared. The overall rates are computed, then applied to the population of each local area [48][49].

e_{it} for unadjusted age is given as:

$$R_{it} = \frac{\sum_{i=1}^{n} y_{it}}{\sum_{i=1}^{n} N_{it}} \tag{2.5}$$

$$e_{it} = R_{it} \times N_{it} \tag{2.6}$$

e_{it} for adjusted age is given as:

$$R_{jt} = \frac{\sum_{i=1}^{n} y_{ijt}}{\sum_{i=1}^{n} N_{ijt}} \tag{2.7}$$

$$e_{it} = R_{jt} \times N_{ijt} \tag{2.8}$$

- j represents the age specific group (10-19, . . . 80+).
- R_t is the overall rate per year.
- R_{jt} represents rate of a specific age group per year.

A scatter plot of suicide count by county revealed that the data varied from county to county. Also, auto correlation plots (Appendix E) showed no significant dependence. Since the data was random, Gaussian process was used due to its flexibility. Moreover, it does not impose unreasonable restrictions on the time trend. The longitudinal trend was modelled as Gaussian process (semi parametric), specifically Ornstein Uhlenbeck process with mean zero and an exponential covariance function [49]:

$$cov\{g(t,k)\} = \theta_1 \exp(-\theta_1 \mid t-k \mid) \tag{2.9}$$

A hyper prior distribution must be specified on the parameters of the covariance function of the Gaussian process which enables the model to discover properties of the data. Furthermore, Gaussian process is very useful if data properties cannot be met by distributions from parametric family [50].

In this study, the first stage in the hierarchical modelling is y_{it} which is conditional on the parameterλ_{it}(relative risk). y_{it} and which is assumed to follow a Poisson distribution with mean $(\lambda_{it}e_{it})$. The second stage λ_{it}, is conditional on the overall relative risk (α). The third stage a hyperprior distribution is specified on the hyperparameters, these are outlined below:

$$\theta \sim N(0, 1000) \tag{2.10}$$

$$\theta_1 \sim U(0, 100) \tag{2.11}$$

$$\theta_2 \sim U(0, 100) \tag{2.12}$$

2.8 Just Another Gibbs Sampler (JAGS)

JAGS is widely used in epidemiological context to efficiently model Bayesian multistage models [51]. To ensure effective and efficient computations, JAGS was used to obtain the posterior distribution. JAGS is a program written in C++ used for the analysis of Bayesian models which was developed to be in accordance with Bayesian inference using Gibbs sampler (BUGS). JAGS aim to modify BUGS by allowing computations of Bayesian multistage markov chain models and is flexible for users to write their own algorithms. Either the coda or rjags package can be used within R for JAGS. It uses markov chain Monte Carlo to make inferences in Bayesian modelling. The process involved in modeling thus generating samples from the posterior distribution of the parameters include: definition of the model, compilation, initialization, adaptation and monitoring [52].

Other summaries like summary statistics and diagnostic plots are obtained outside of JAGS by the use of coda package [53]. To ensure that JAGS reports estimates for the parameter of interest, λ_{it} was monitored.

2.9 JAGS Implementation

As stated earlier, the first stage in JAGS implementation is definition of the model. The model was specified in R software since R software was used to access JAGS. The model was specified in the following order:

$$\begin{aligned}
y_{it} &\sim Poisson(\mu_{it}) \\
\ln(\mu_{it}) &= \ln(\lambda_{it}) + \ln(e_{it}) \\
\lambda_{it} &= \exp(g_{it} + \alpha) \\
\mathbf{g}_i \mid \boldsymbol{\Sigma} &\sim MVN(\mathbf{0}, \boldsymbol{\Sigma}) \\
g_{it} &\ \textit{is an element in}\ \mathbf{g}_i \\
\boldsymbol{\Sigma}_{tt^*} &= \theta_1 \exp(-\theta_2 \mid t - t^* \mid) \\
\boldsymbol{\Sigma}_{tt^*} &\ \textit{is an element in}\ \boldsymbol{\Sigma} \\
\alpha &\sim N(0, 1000) \\
\theta_1 &\sim U(0, 100) \\
\theta_2 &\sim U(0, 100)
\end{aligned} \tag{2.13}$$

The data consisting of the suicide counts, expected number of the deaths and years were also imported into R to successfully compile and initialize the model. Burn-in is the period whereby the first n samples are discarded. A total of 110,000 iterations were used with a burn-in of 10,000. Gelman-Rubin diagnostics and traceplots were used to check the convergence of the model by running multiple chains [46]. The number of chains specified was three. Thinning is the process of discarding all samples except the kth iteration. In MCMC computations, samples of the posterior distribution are sometimes dependent. One way to correct this is by discarding samples (thinning) [46]. Coda.sample was used to draw samples from the posterior distribution at every 50th iteration since the samples from the posterior distribution were dependent. This was necessary in order to reduce auto correlation. In total, there were 795 parameters. 792 of the parameters were relative risk for a specific county per year. The remaining three were α, θ_1 and θ_2.

CHAPTER 3

Results and Discussions

3.1 Introduction

This chapter presents the results of the study. The relative risk of the 44 counties per year were estimated by markov chain Monte Carlo using JAGS. The parameter of interest was relative risk of suicide which is estimated by the posterior mean. A relative risk of 1.0 implies that the risk of suicide in that specific county is no different from expected based on rates across the state of Idaho. A relative risk above 1.0 implies the risk of suicide for the county is above the state of Idaho. Similarly, a relative risk below 1.0 implies the risk of suicide in the county is below that of the state of Idaho. Credible intervals were also obtained to determine if relative risk estimates were accurate and reliable. Relative risk of suicide for all counties using HB were summarised in plots (Appendix A). Model diagnostics such as trace plots and Gelman-Rubin plots were used to check if all chains in MCMC converged to a stationary distribution.

3.2 Adjusted Age Relative Risk of Suicide

Overall, Madison was the only county with a relative risk of suicide significantly (thus both upper and lower bound of the credible interval was below or above the state of Idaho) below the state of Idaho over the period. The relative risk of suicide was consistently higher than the state of Idaho in Lemhi, Custer, Shoshone, and Nez Perce for all the years under consideration. All reporting in this section is based on the relative risk of suicide (posterior mean) unless stated otherwise.

In Bannock, the relative risk of suicide was not different from the state-wide rates from 1999 to 2006, but in 2007 it began to increase steadily through 2016. The lower limit of the credible interval increased from 1999 to 2008 though it was clearly below the state of Idaho. From 2009 to 2016 the lower limit of the credible interval was not different from the state of Idaho. This indicates that rates in Bannock county truly increased faster than expected from 2009 to 2016.

The relative risk of suicide in Bear Lake declined gradually from 1999 to 2016. However, the relative risk of suicide was above the state of Idaho throughout the period. The lower limit of the credible interval was not different from the state of Idaho from 1999 to 2007, then began to decline from 2008 to 2016.

In Lemhi, the relative risk of suicide was significantly higher than the state of Idaho from 1999 to 2016. This indicates that rates increased faster than expected over the period in Lemhi county.

In Kootenai, the relative risk of suicide was higher than the state-wide rates from 1999 to 2006, but in 2007 it began to decline steadily through 2016. The lower limit of the credible interval was slightly below the sate-wide rates but began to decline from 2008 to 2016. This indicates that rates in Kootenai county decreased faster than expected from 2008 to 2016.

In Shoshone, the relative risk of suicide was consistently higher than the state-wide rates from 1999 to 2006. The relative risk of suicide increased steadily from 2005 to 2016. The lower limit of the credible interval was not different from the state of Idaho from 1999 to 2005, but from 2006 to 2016 it increased. This indicates that rates in Shoshone county truly increased faster than expected from 2006 to 2016.

The relative risk of suicide in Nez Perce was significantly higher than the state-wide rates from 1999 to 2015, but in 2016 the lower limit of the credible interval was not different from the state of Idaho. This indicates that rates in Nez Perce county were higher than those expected from 1999 to 2016.

In Custer, the relative risk of suicide was significantly above the state of Idaho from 2004 to 2016, indicating that rates in Custer county were higher than expected from 2004 to 2016.

The relative risk of suicide in Ada was consistently below the state-wide rates from 1999 to 2016. The upper limit of the credible interval was not different from the state-wide rates from 1999 to 2003 but decreased from 2004 to 2013. The lower limit of the credible interval was clearly below the state of Idaho over the period. This indicates that rates were lower than expected from 1999 to 2016 in Ada county.

In Minidoka county, the relative risk of suicide was above the state from 1999 to 2007. However in 2008, the relative risk of suicide declined and was the same as the state till 2016. The lower limit of the credible interval was clearly below the state-wide rates over the period.

The relative risk of suicide in Twin Falls was slightly higher than that of the state from 1999-2002, increased steadily from 2003 to 2016. The lower limit of the credible interval increased from 1999 to 2006 though it was clearly below the state of Idaho. But the lower limit of the credible interval was not different from the state-wide rates from 2007 to 2016.

3.3 SMR for Age Adjusted and Unadjusted

Age adjusted and unadjusted relative risk of suicide were computed for each county per year for the 18-year period. The plots are presented in Appendix B. Age adjusted relative risk obtained from the SMR had narrow confidence intervals compared to unadjusted relative risk. Age adjusted relative risk of suicide was high through out the time period in counties such as Adams, Butte, Boise, Camas, Custer, Lemhi, Lincoln, Clark, Oneida and Lewis.

3.4 HB and SMR

The age adjusted and unadjusted relative risk of HB had narrow credible intervals compared to both age adjusted, and unadjusted confidence intervals of relative risks obtained using SMR. Credible intervals of age adjusted relative risk were narrow compared to age unadjusted relative credible intervals. Age adjusted relative risk were the best estimates because differences in the age distribution of all 44 counties were adjusted thereby controlling for differences in population of all counties.

This study illustrates the weaknesses of using SMRs to estimate suicide. A community with more than 20,000 residents is considered urban and a community between 7,000 and

20,000 residents is considered rural. Counties like Ada, Kootenai, Bannock, Latah, Nez Perce, Canyon, Bannock, Twin Falls, Bonneville and Madison are considered urban based on the above criterion [54]. These urban counties had narrow confidence intervals (SMR) compared to the other counties. Bonner, Blaine, Mindoka, Cassia and Bingham were considered as rural based on criterion mentioned earlier [54]. Rural counties had wide confidence intervals in some years for both age adjusted and unadjusted (SMR). The confidence intervals as expected for counties with small population size was extremely wide. Least populous counties such as Clark, Camas, Butte, Lewis, Adams Custer and Oneida had extremely wide confidence intervals, suggesting that when the population size is small SMRs fail to provide reliable estimates. The confidence interval of Ada was very narrow as compared to other counties, which was reasonable since Ada had the highest population over the time period. Though Ada had fairly narrow confidence intervals (SMR), the intervals were wider than credible interval of Ada obtained from HB. Comparably, the standard error of estimates obtained by HB were smaller than SMR. Thus relative risk estimates obtained by HB had low standard error which indicates less variability in estimates. However, SMRs do not account for variation in estimates hence large standard error resulting in wide confidence intervals. It is obvious that model-based approaches specifically HB provide reliable and stable estimates compared to SMRs when the population size of the geographic area is small.

Furthermore, least populous counties such as Adams, Butte, Oneida, Power, Lemhi, Lincoln, Clark, Custer and Camas with positive suicide counts had large relative risk of suicide indicating that this counties were more prone to suicide but that was not strictly the same in HB. The relative risk of suicide in some of these counties mentioned were high in HB, however the estimates were outrageous in SMR. Also, relative risk of suicide was zero in most years for the aforementioned counties using SMR due to zero counts. Implying no risk of suicide for these counties, which could be very misleading. However nonzero relative risks of suicide were obtained for all counties over the time period using HB.

3.5 MCMC Checks

Trace plots showed good mixing of the three chains for each parameter (age adjusted relative risk of suicide), indicating convergence of the posterior distribution. The Gelman Rubin statistic was approximately one for all parameters signifying convergence of the chains as well.

3.6 Discussion

In this study, data on suicide counts were examined to identify the trends in relative risk of suicide from 1999 to 2016. Overall, relative risk of suicide was high in most counties. All discussions below are based on age adjusted relative risk of suicide.

Expected number of suicide deaths and observed suicide counts were compared. Generally the observed number of death was higher than expected but not significantly so, based on our model and credible intervals for relative risk.

The standard errors of age adjusted relative risk of suicide were lower compared to age unadjusted relative of suicide leading to very narrow credible intervals for age adjusted relative risk. With age unadjusted relative risk of suicide, differences in the age distribution of the population are not considered. Therefore, an aging population might overestimate age unadjusted relative risk. Similarly, a population with more young people might underestimate age unadjusted relative risk of suicide. The population of Idaho state is changing over time, hence it was important to standardized the population in order to omit any differences that might arise as a result of the population changing over a period of time. Age adjusted relative risk estimated by a Poisson hierarchical Bayes model therefore provided more narrower credible interval for the estimates.

The consistently low relative risk of suicide in Madison county, along with narrow credible bands produced by our model, provide strong evidence that there is a lower risk of suicide in that county. County level attributes could contribute to this low risk. For instance Rexburg, the county seat is known as an American family community [55] and about 95 percent of residents are members of the Church of Latter Day Saints. Residents in the county might have a sense of belonging, feel loved and connected with each other [17] thereby easily reaching out when necessary. Madison was ranked as the healthiest county in Idaho from 2013 to 2015 [55]. County level estimates of factors such as education, income, housing transportation, access to health care and social connectivity that affects the health of a community [56] can be used as predictors in the model developed here to statistically evaluate relationships between relative risk of suicide and the factors mentioned above.

These predictions will help public health experts and government officials target counties with increasing trends in relative risk of suicide in order to intensify suicide preventive measures or proper allocation of resources. Furthermore, public health experts could use this model to better understand suicide trends.

In conclusion, a reliable model was developed to estimate relative of suicide. Credible intervals of age adjusted relative risk using Poisson hierarchical Bayes model were precise and stable compared to credible intervals of age unadjusted relative risk.

References

[1] Who. (2014). Preventing suicide: A global imperative. Stylus Publishing.

[2] American Foundation for Suicide Prevention, Retrieved from `https://afsp.org/about-afsp`.

[3] National Center for Injury Prevention and Control, Division of Violence Prevention.(September 5, 2019). Retrieved from `https://www.cdc.gov/violenceprevention/suicide/index.html`.

[4] Centers for Disease Control and Prevention Retrieved from `https://www.cdc.gov/vitalsigns/suicide/index.html`

[5] Suicide in Idaho: Fact Sheet (September 2018) Idaho Department of Health and Welfare. Retrieved from `https://healthandwelfare.idaho.gov/Portals/0/Users/145/93/2193/Fact`

[6] Suicide Prevention Resource Center, Retrieved from `http://www.sprc.org/hispanic-populations`

[7] Center for Disease Control and Prevention. (7,June 2018) Retrieved from `https://www.cdc.gov/vitalsigns/suicide/index.html`

[8] Center for Disease Control Retrieved from `https://www.cdc.gov/nchs/pressroom/states/idaho/idaho.htm`

[9] Facts at a Glance(2015) National Center for injury Prevention and Control. Retrieved from `https://www.cdc.gov/violenceprevention/pdf/suicide-datasheet-a.pdf`

[10] Suicide: Idaho 2017 Facts & Figures. American Foundation for suicide prevention Retrieved from `https://1035kissfmboise.com/files/2018/02/Idaho-Facts-2017.pdf`

[11] Statistics, H. (2016). Idaho Resident Suicide Deaths 2010-2014. June.

[12] Suicide in Idaho: Fact Sheet (February 2018) Idaho Department of Health and Welfare. Retrieved from `https://healthandwelfare.idaho.gov/Portals/0/Users/145/93/2193/Fact`

[13] US Census Bureau, Retrieved from `https://www.census.gov/quickfacts/fact/table/ID/PST040218`

[14] Idaho Department of Health and Welfare. Retrieved from `https://healthandwelfare.idaho.gov/Health/HealthDistricts/tabid/97/Default.aspx`

[15] United States Census Bureau, Idaho. Retrieved from `https://www.census.gov/quickfacts/ID`

[16] Division of Public Health, Idaho Department of Health and Welfare. Retrieved from `https://healthandwelfare.idaho.gov/Portals/0/Health/Statistics/2016-Reports/2016_Population.pdf`

[17] Rosebella Capio, A Bayesian Hierarchical Poisson Approach to Estimate County Level Suicide Risks in Idaho. (2019)

[18] Lester R. Curtin, Ph.D. and Richard J. Klein, M.P.H.,Direct Standardization. Retrieved from `https://ibis.health.state.nm.us/view/docs/PHStatistics/statnt06rv.pdf.`

[19] Namrata Bains,(March 2019). Standardization of Rates. Retrieved from `http://core.apheo.ca/resources/indicators/Standardization`

[20] Sherry L Murphy, J Xu, KD Kochanek, National Vital Statistics Reports, 2017. Retrieved from `https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_09\protect\discretionary{\char\hyphenchar\font}{}{}508.pdf`

[21] Fleiss, J. L., Levin, B., & Paik, M. C. (2013). Statistical methods for rates and proportions. john wiley & sons.

[22] Kelsey, J. L., Whittemore, A. S., Evans, A. S., & Thompson, W. D. (1996). Methods in observational epidemiology. Monographs in Epidemiology and Biostatistics.

[23] Lawson, A. B., Browne, W. J., & Rodeiro, C. L. V. (2003). Disease mapping with WinBUGS and MLwiN (Vol. 11). John Wiley & Sons.

[24] Emily Lauer, Matthew Hoghton, Health Disparities and Intellectual Disabilities in International Review of Research in Developmental Disabilities, 2015

[25] Techical notes, Kidney Epidemiology and Cost Center University of Michigan, School of Public Health(July 2018). Retrieved from `https://dialysisdata.org/sites/default/files/content/Methodology/Updated_DFC_Star_Rating_Methodology_for_October_2018_Release.pdf`

[26] Meza, J. L. (2003). Empirical Bayes estimation smoothing of relative risks in disease mapping. Journal of Statistical Planning and Inference, 112(1-2), 43-62.

[27] Washington State Department of Health. (2002). Guidelines for Using Confidence Intervals for Public Health Assessment

[28] Rao, J. N. (2014). Small area estimation. Wiley StatsRef: Statistics Reference Online, 1-8.

[29] Saunderson, T. R., & Langford, I. H. (1996). A study of the geographical distribution of suicide rates in England and Wales 1989-1992 using Empirical Bayes estimates. Social Science & Medicine, 43(4), 489-502.

[30] Datta, G., Ghosh, M., & Waller, L. A. (2000). Hierarchical and Empirical Bayes methods for environmental risk assessment. Handbook of Statistics, 18, 223-245.

[31] Tapabrata Maiti, Journal of Statistical Planning and Inference 69 (1998) 339-348

[32] Clayton, D., & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics, 671-681.

[33] Middleton, N., Sterne, J. A., & Gunnell, D. J. (2008). An atlas of suicide mortality: England and Wales, 1988-1994. Health & place, 14(3), 492-506.

[34] Chang, S. S., Stuckler, D., Yip, P., & Gunnell, D. (2013). Impact of 2008 global economic crisis on suicide: time trend study in 54 countries. Bmj, 347, f5239.

[35] Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the royal statistical society: Series b (statistical methodology), 71(2), 319-392.)

[36] Bivand, R., Gomez-Rubio, V., & Rue, H. (2015). Spatial data analysis with R-INLA with some extensions. American Statistical Association.

[37] Khana, D., Rossen, L. M., Hedegaard, H., & Warner, M. (2018). A Bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with R-INLA. Journal of data science: JDS, 16(1), 147.

[38] Rossen, L. M., Hedegaard, H., Khan, D., & Warner, M. (2018). County-level trends in suicide rates in the US, 2005-2015. American journal of preventive medicine, 55(1), 72-79.

[39] Quinlan-Davidson, M., Sanhueza, A., Espinosa, I., Escamilla-Cejudo, J. A., & Maddaleno, M. (2014). Suicide among young people in the Americas. Journal of Adolescent Health, 54(3), 262-268.

[40] Ghosh, M., & Rao, J. N. K. (1994). Small area estimation: an appraisal. Statistical science, 9(1), 55-76.

[41] Congdon, P. (1997). Bayesian models for spatial incidence: a case study of suicide using the BUGS program. Health & Place, 3(4), 229-247.

[42] 43. Zhang, J., & Kai, F. Y. (1998). What's the relative risk?: A method of correcting the odds ratio in cohort studies of common outcomes. Jama, 280(19), 1690-1691.

[43] Zhou, Q. M.,& You, Y. (2007). Hierarchical Bayes Small Area Estimation for the Canadian Community Health Survey. Statistics Canada, Household Survey Methods Division.

[44] Hulting, F. L., & Harville, D. A. (1991). Some Bayesian and non-Bayesian procedures for the analysis of comparative experiments and for small-area estimation: computational aspects, frequentist properties, and relationships. Journal of the American Statistical Association, 86(415), 557-568.

[45] Trevisani, M., & Torelli, N. (2007). Hierarchical Bayesian models for small area estimation with count data (pp. 1-30). Working Paper: Dipartimento di Scienze Economiche e Statistiche, UniversitÃ degli Studi di Trieste, Trieste, Italy.

[46] Lawson, A. B. (2013). Bayesian disease mapping: hierarchical modeling in spatial epidemiology. CRC press.

[47] Ugarte, M. D., Ibanez, B., & Militino, A. F. (2006). Modelling risks in disease mapping. Statistical methods in medical research, 15(1), 21-35.

[48] Jarup, L., Best, N., Toledano, M. B., Wakefield, J., & Elliott, P. (2002). Geographical epidemiology of prostate cancer in Great Britain. International journal of cancer, 97(5), 695-699.

[49] Branscum, A. J., Perez, A. M., Johnson, W. O., & Thurmond, M. C. (2008). Bayesian spatiotemporal analysis of foot-and-mouth disease data from the Republic of Turkey. Epidemiology & Infection, 136(6), 833-842.

[50] Bernardo, J., Berger, J., Dawid, A., & Smith, A. (1998). Regression and classification using Gaussian process priors. Bayesian statistics, 6, 475.

[51] Sharples L. Statistics in Medicine, 12:1155-1169, 1993.[6]

[52] Plummer, M. (2003, March). In Proceedings of the 3rd international workshop on distributed statistical computing (Vol. 124, No. 125.10, pp. 1-10).

[53] Gelman, A., & Rubin, D. B. (1992). Statistical science, 7(4), 457-472.

[54] Idaho Department of Health and welfare, Idaho Bureau of Vital Records and Health Statistics, CDC Wonder.

[55] `https://www.rwjf.org/en/cultureofhealth/what-were-learning/sentinel-communities/rexburg-idaho.html`

[56] National Association of Counties. Retrieved from `https://www.naco.org/resources/what-healthy-county`

Appendix A.1.1: Poisson Hierarchical Bayes

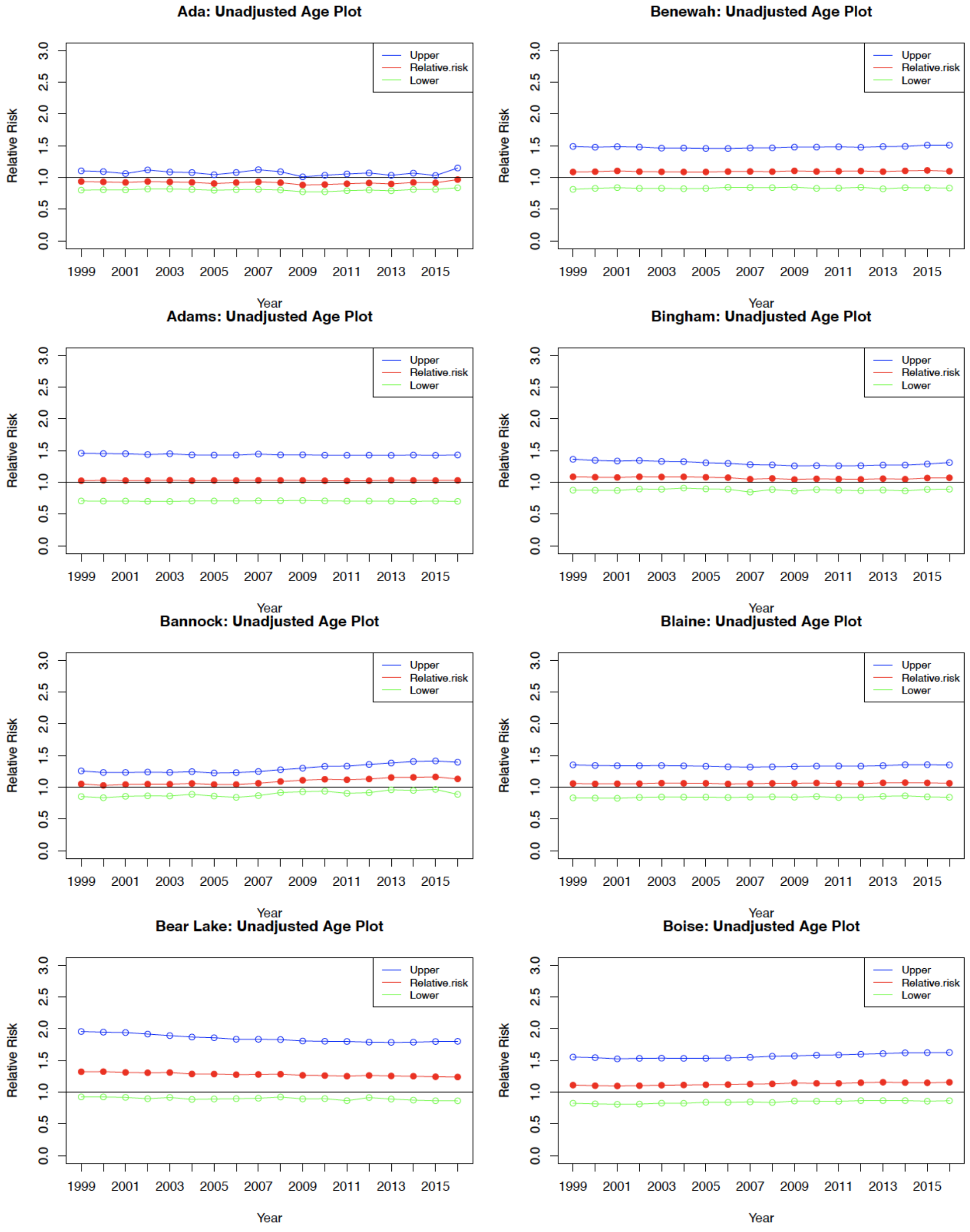


Figure 3.1: Poisson Hierarchical Bayes

Appendix A.1.2: Poisson Hierarchical Bayes

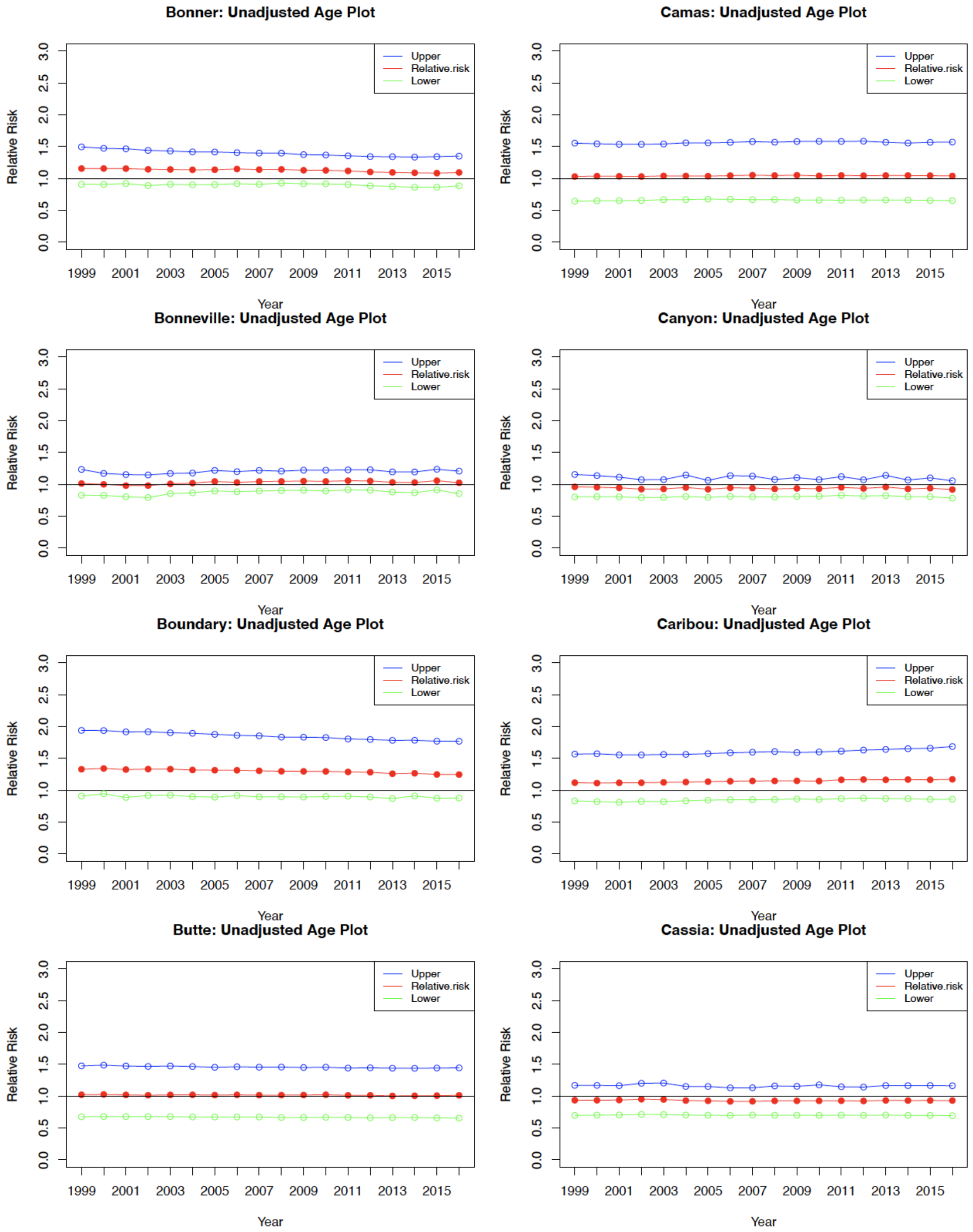


Figure 3.2: Poisson Hierarchical Bayes

Appendix A.1.3: Poisson Hierarchical Bayes

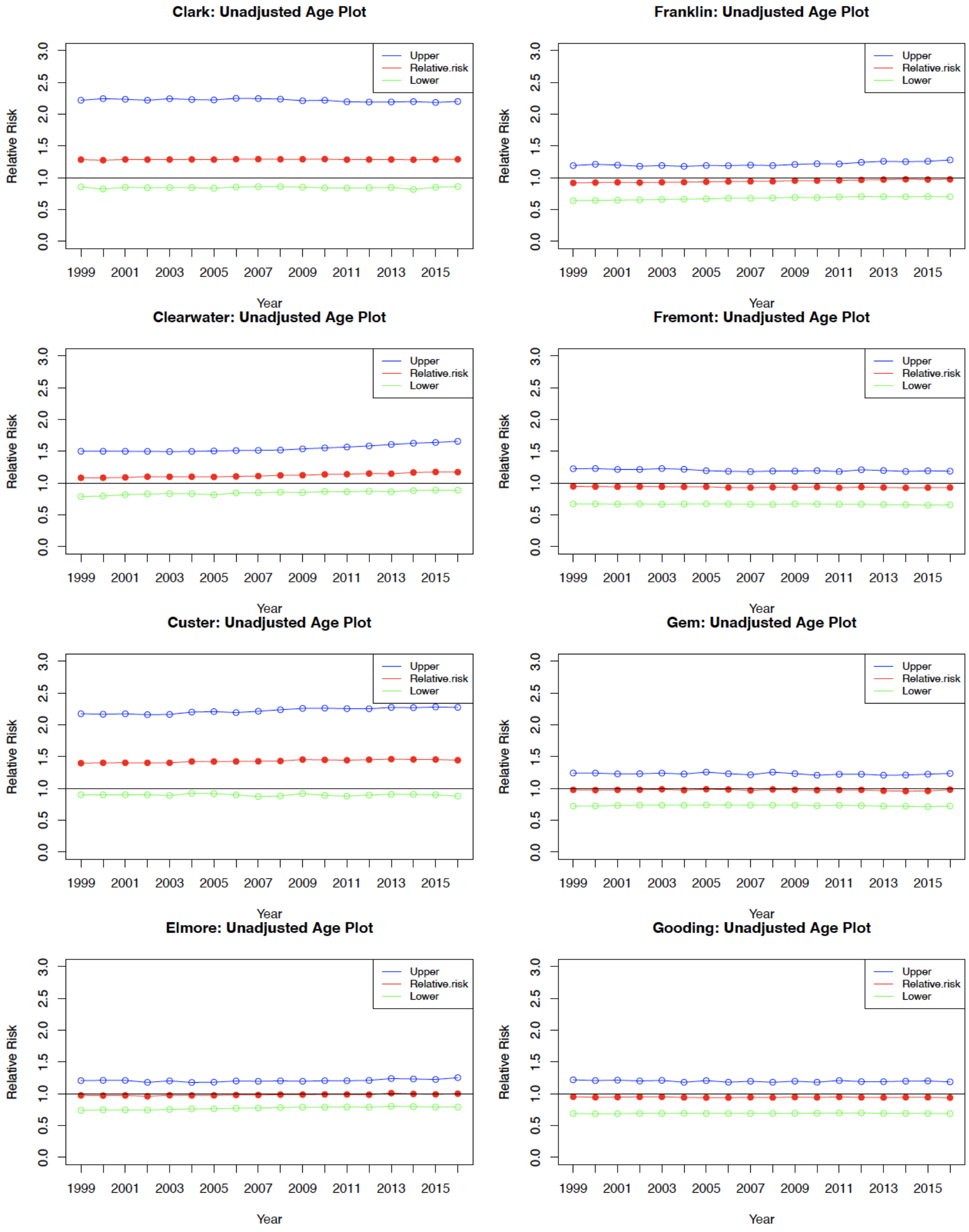
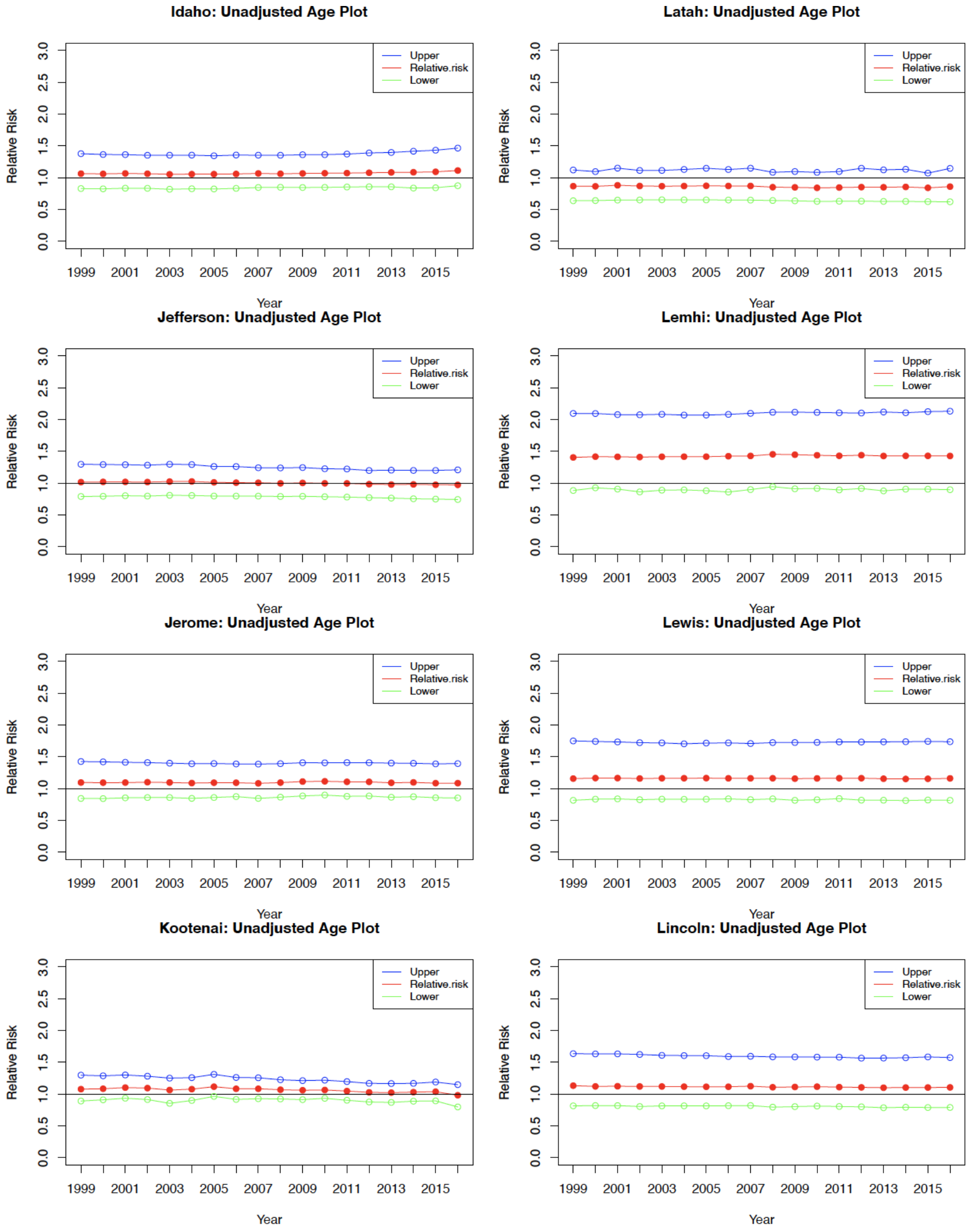


Figure 3.3: Poisson Hierarchical Bayes

Appendix A.1.4: Poisson Hierarchical Bayes


Idaho: Unadjusted Age Plot

Latah: Unadjusted Age Plot

Jefferson: Unadjusted Age Plot

Lemhi: Unadjusted Age Plot

Jerome: Unadjusted Age Plot

Lewis: Unadjusted Age Plot

Kootenai: Unadjusted Age Plot

Lincoln: Unadjusted Age Plot

Figure 3.4: Poisson Hierarchical Bayes

Appendix A.1.5: Poisson Hierarchical Bayes

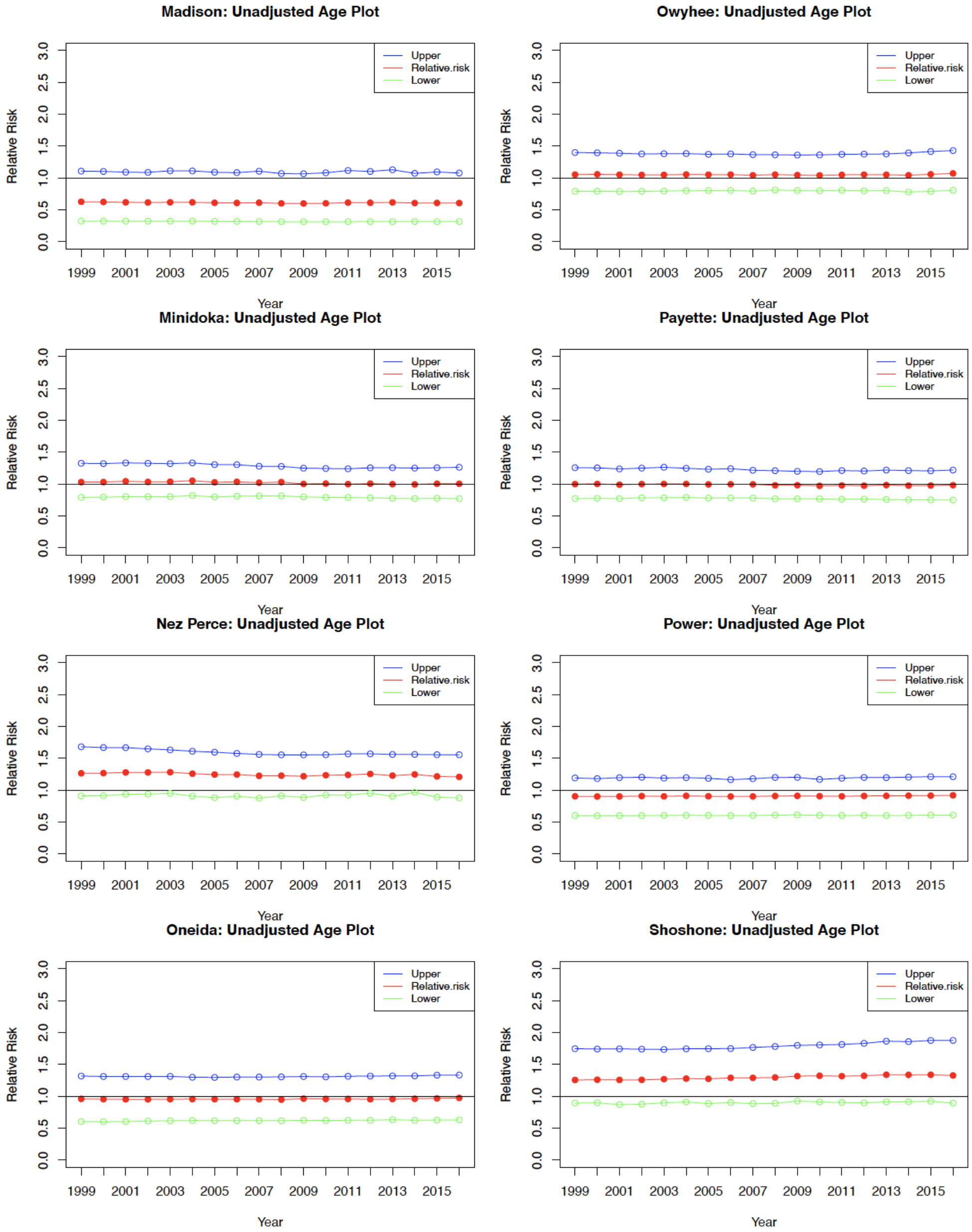


Figure 3.5: Poisson Hierarchical Bayes

Appendix A.1.6: Poisson Hierarchical Bayes

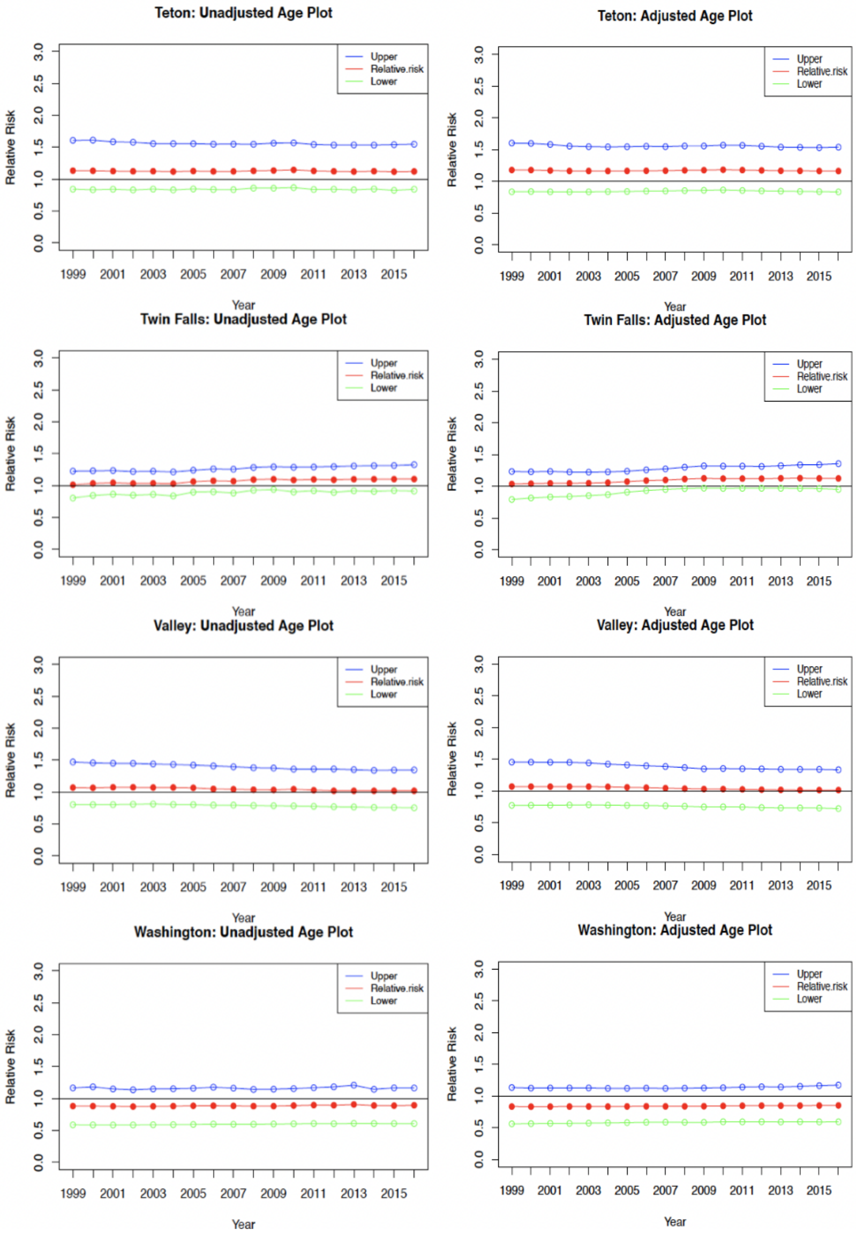


Figure 3.6: Poisson Hierarchical Bayes

Appendix A.1.7: Poisson Hierarchical Bayes

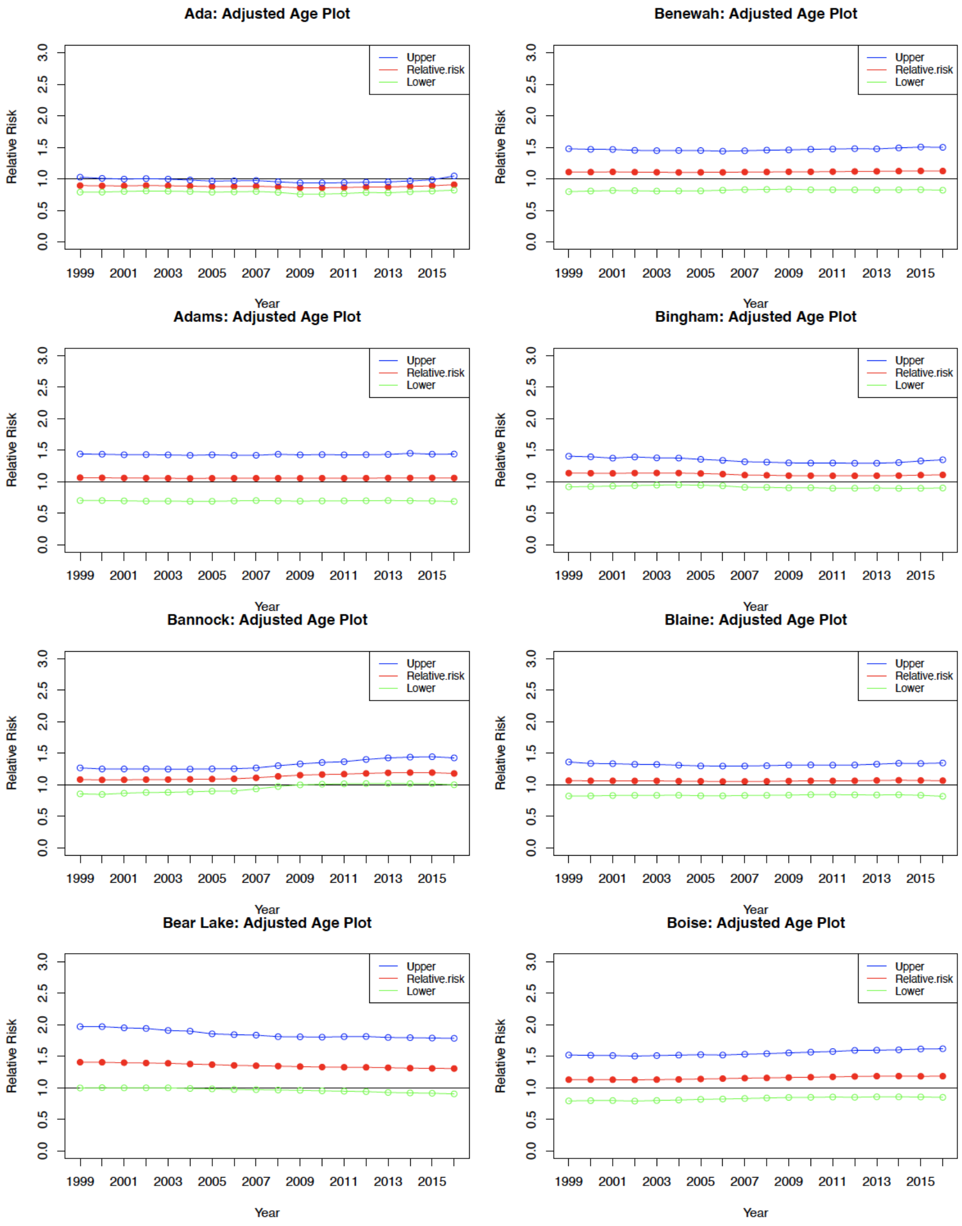


Figure 3.7: Poisson Hierarchical Bayes

Appendix A.1.8: Poisson Hierarchical Bayes

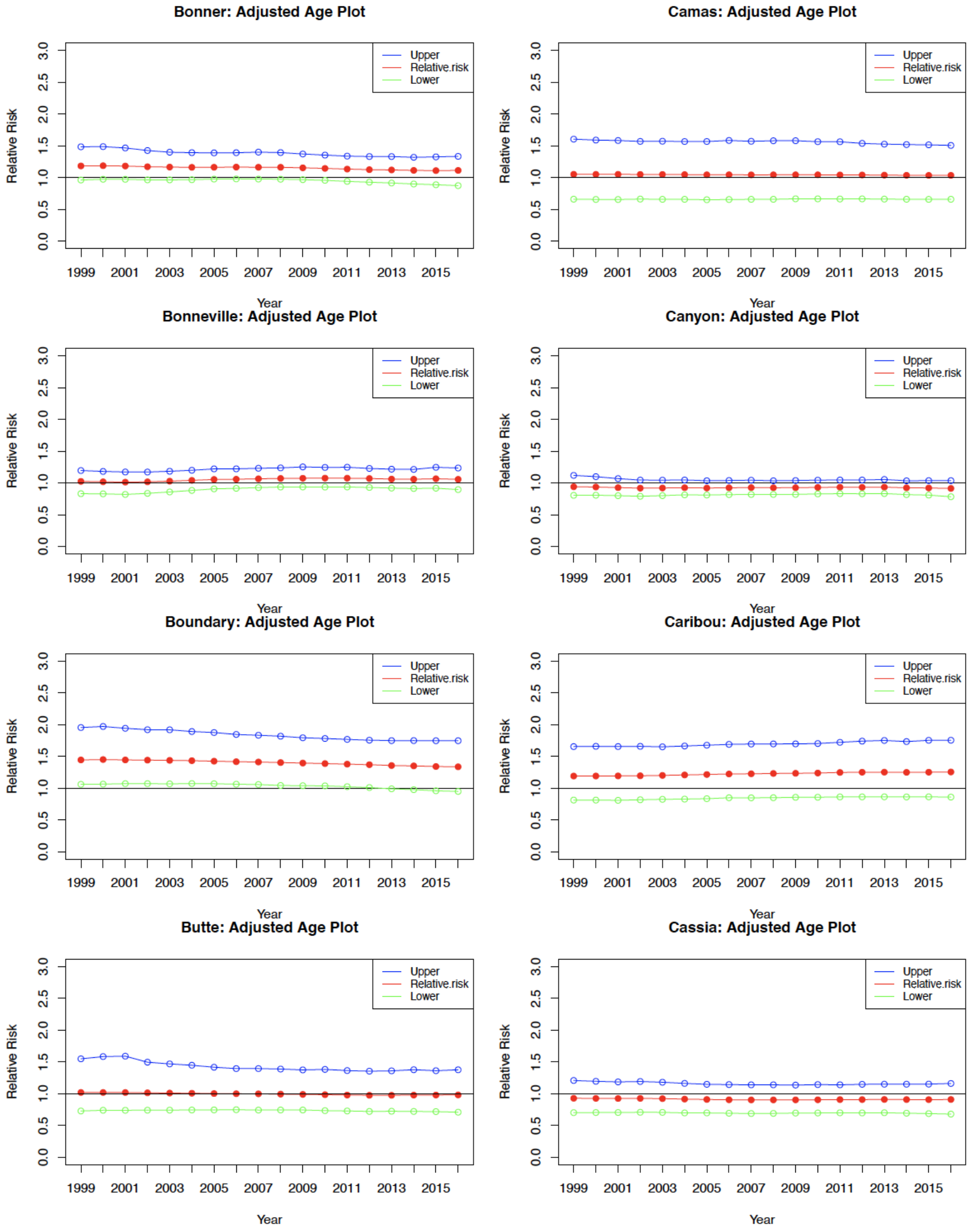
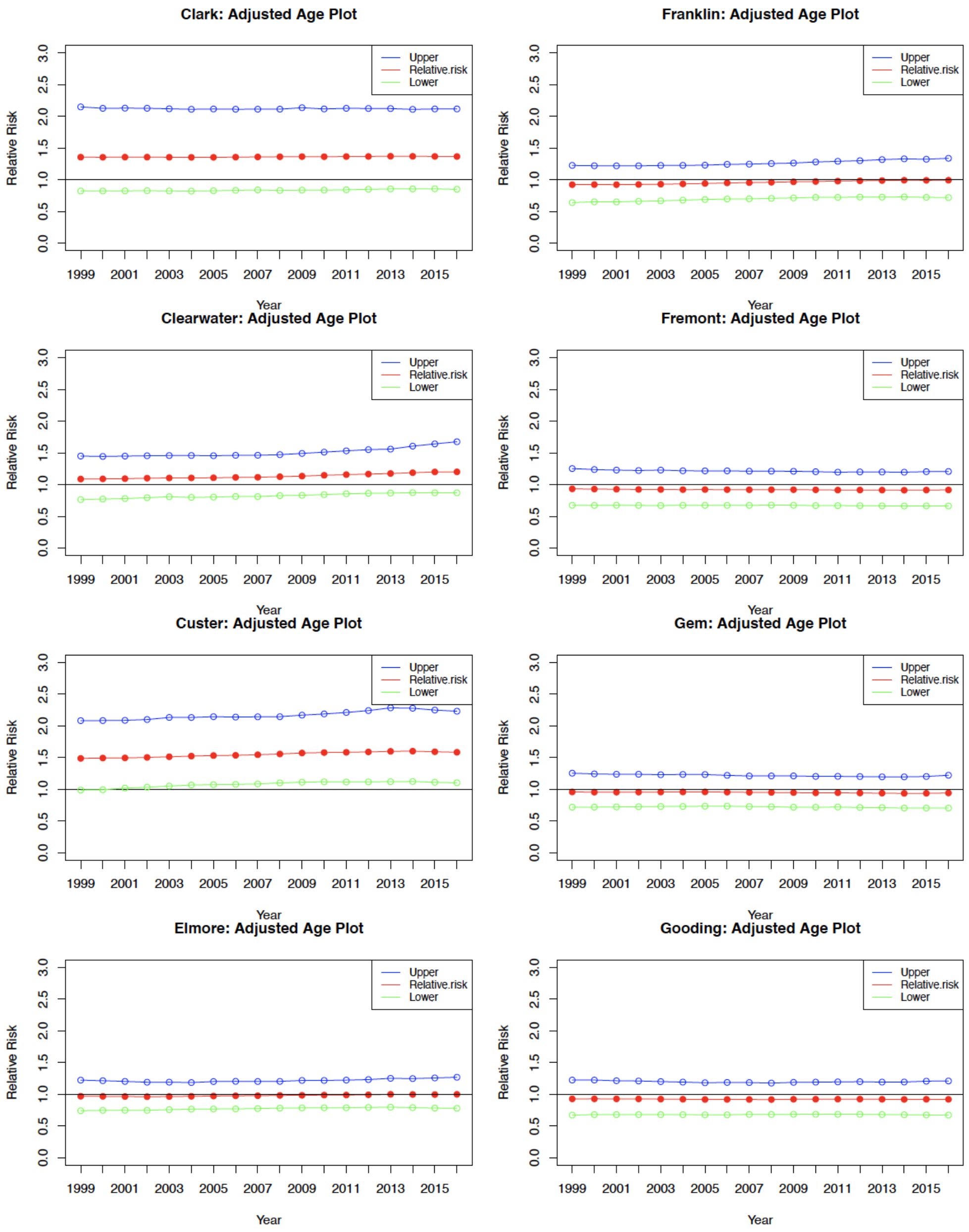


Figure 3.8: Poisson Hierarchical Bayes

Appendix A.1.9: Poisson Hierarchical Bayes



Clark: Adjusted Age Plot

Franklin: Adjusted Age Plot

Clearwater: Adjusted Age Plot

Fremont: Adjusted Age Plot

Custer: Adjusted Age Plot

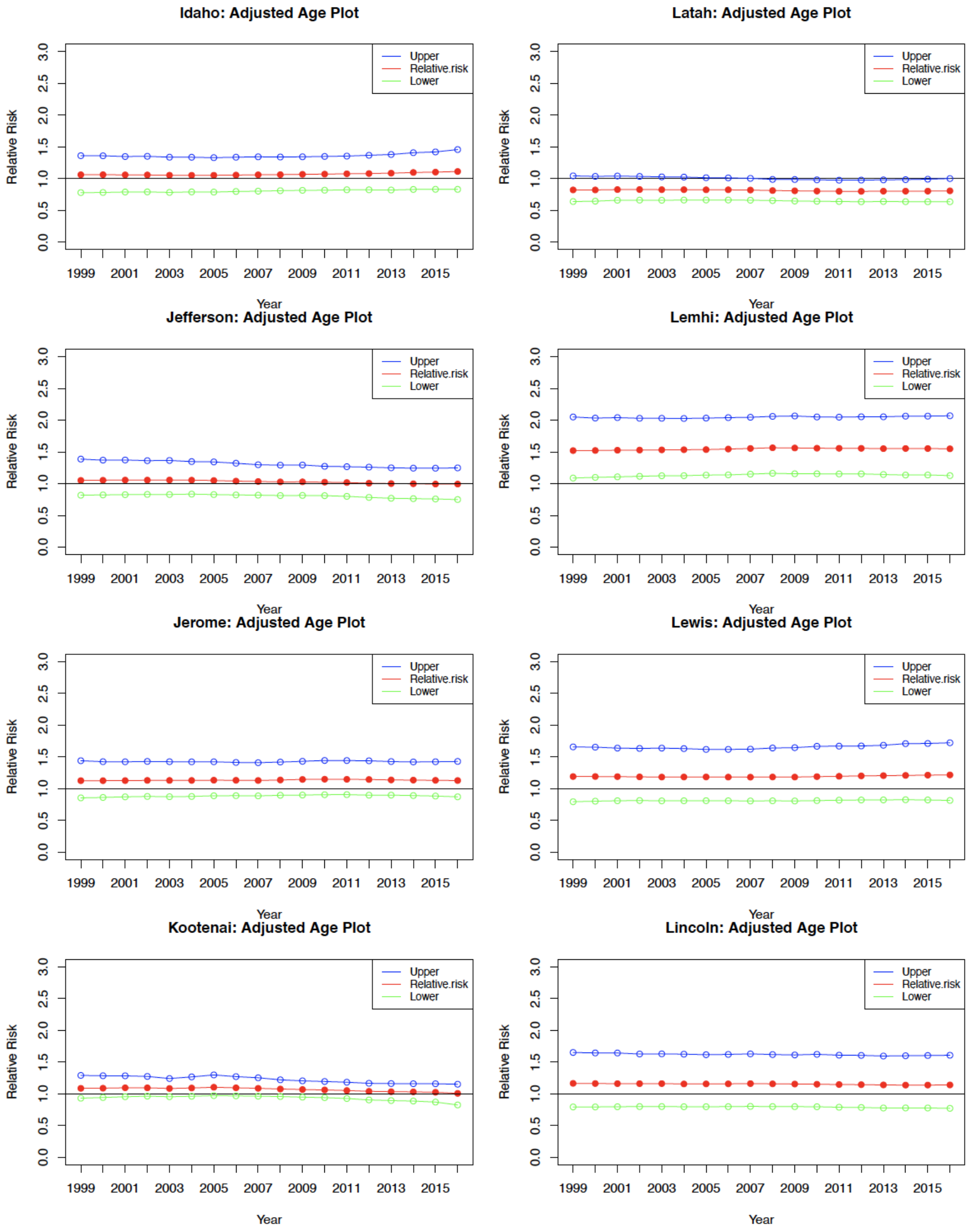
Gem: Adjusted Age Plot

Elmore: Adjusted Age Plot

Gooding: Adjusted Age Plot

Figure 3.9: Poisson Hierarchical Bayes

Appendix A.1.10: Poisson Hierarchical Bayes


Idaho: Adjusted Age Plot

Latah: Adjusted Age Plot

Jefferson: Adjusted Age Plot

Lemhi: Adjusted Age Plot

Jerome: Adjusted Age Plot

Lewis: Adjusted Age Plot

Kootenai: Adjusted Age Plot

Lincoln: Adjusted Age Plot

Figure 3.10: Poisson Hierarchical Bayes

Appendix A.1.11: Poisson Hierarchical Bayes

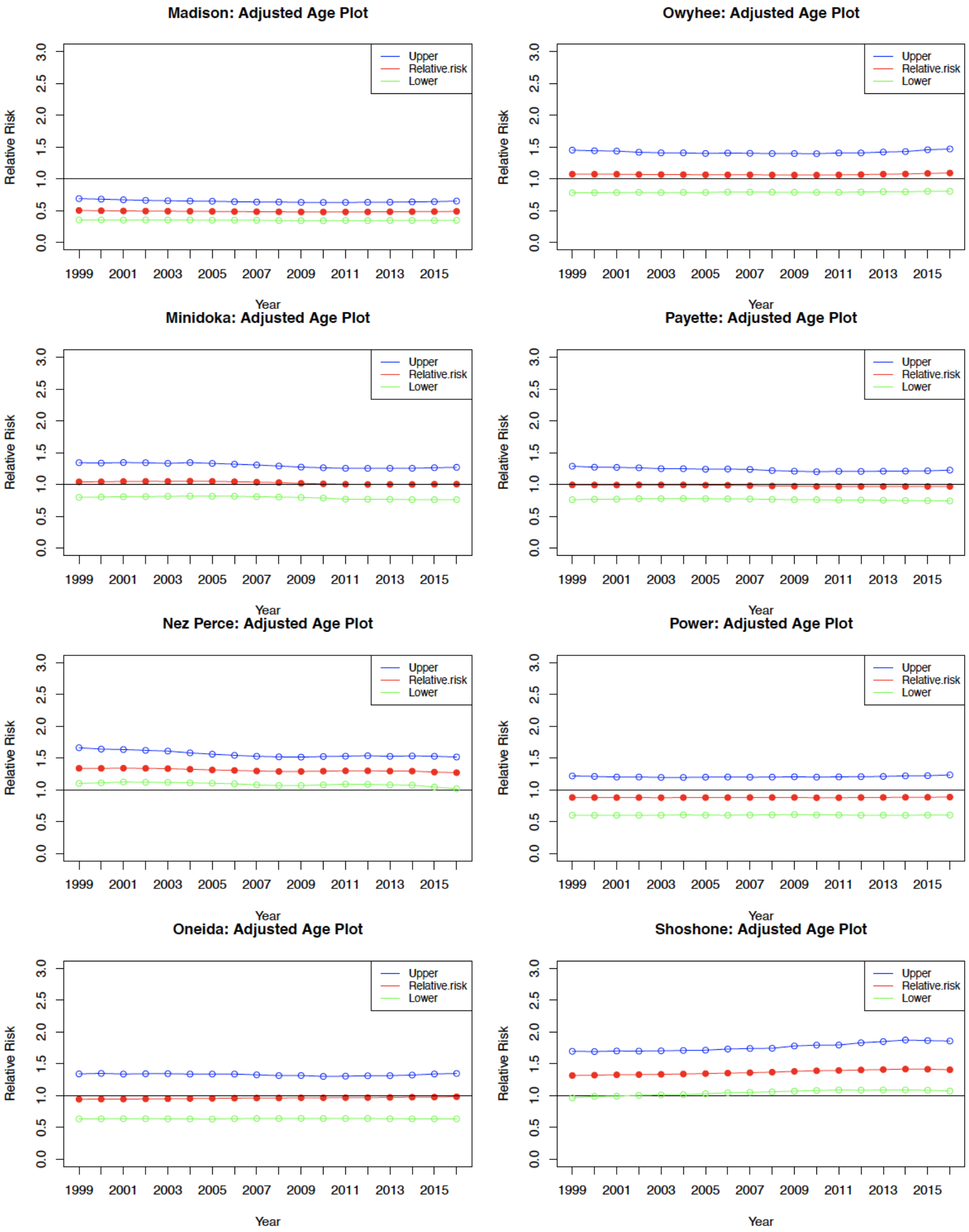


Figure 3.11: Poisson Hierarchical Bayes

Appendix B.1.1: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

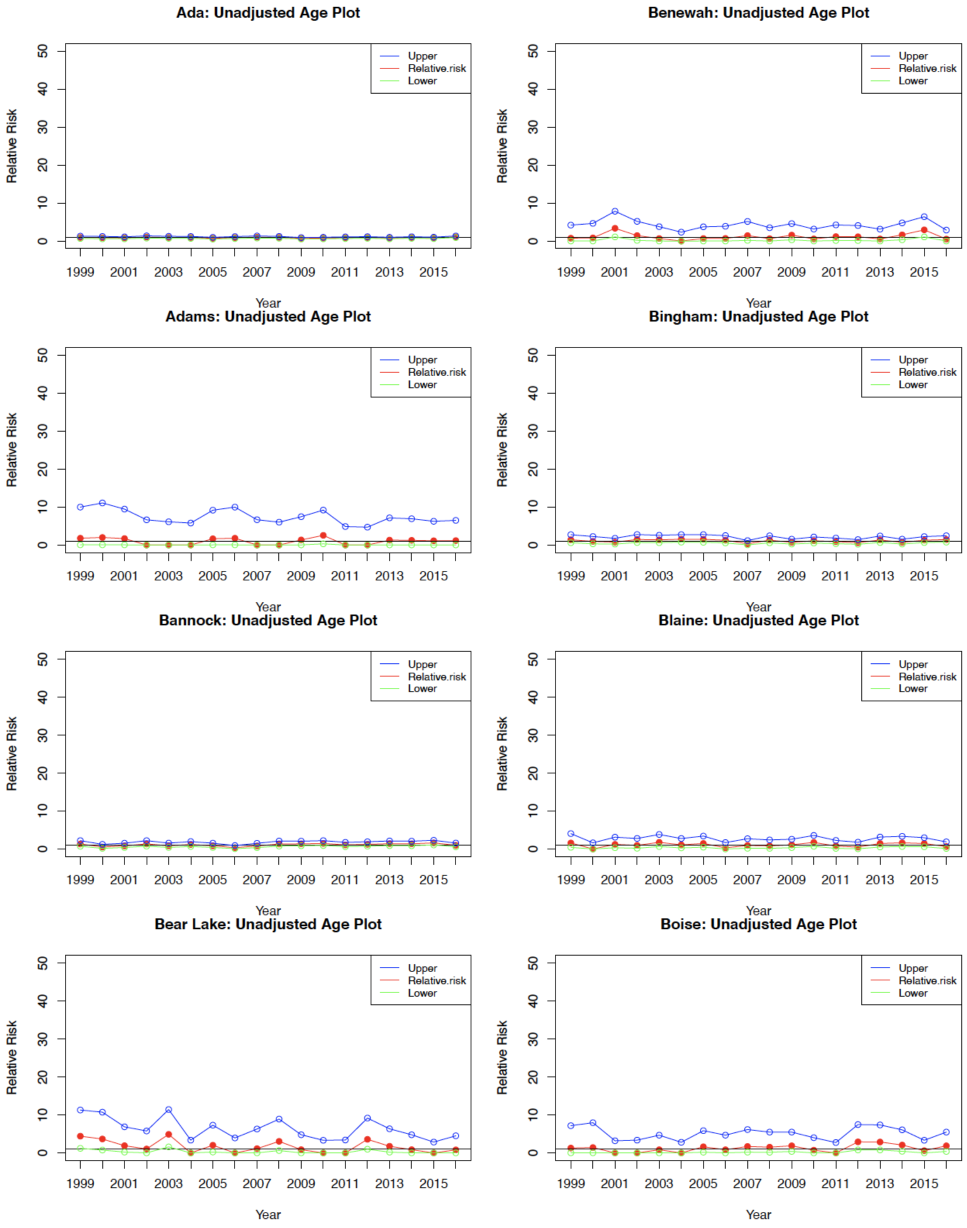
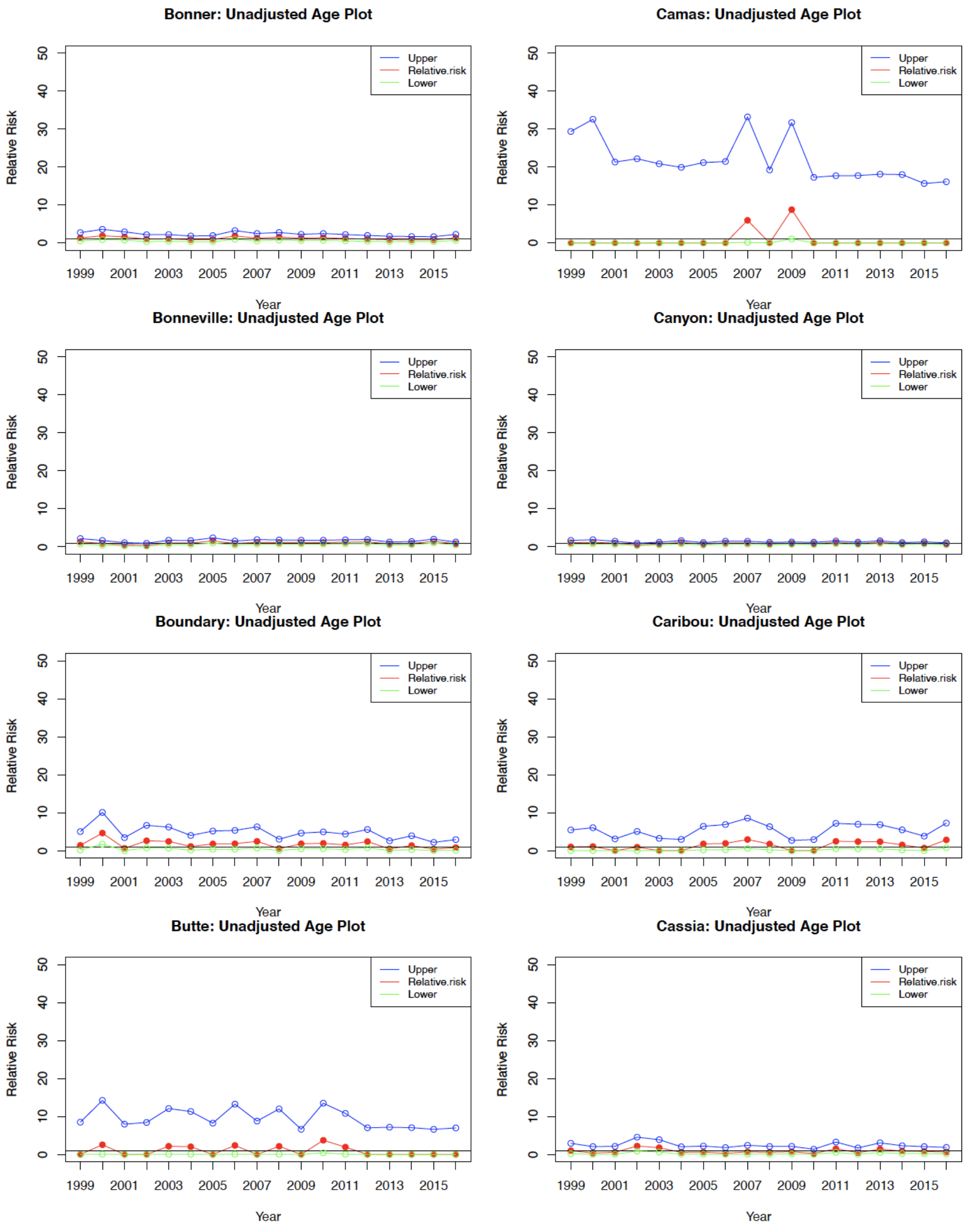


Figure 3.12: Standardized Mortality Ratio

Appendix B.1.2: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.



Bonner: Unadjusted Age Plot

Camas: Unadjusted Age Plot

Bonneville: Unadjusted Age Plot

Canyon: Unadjusted Age Plot

Boundary: Unadjusted Age Plot

Caribou: Unadjusted Age Plot

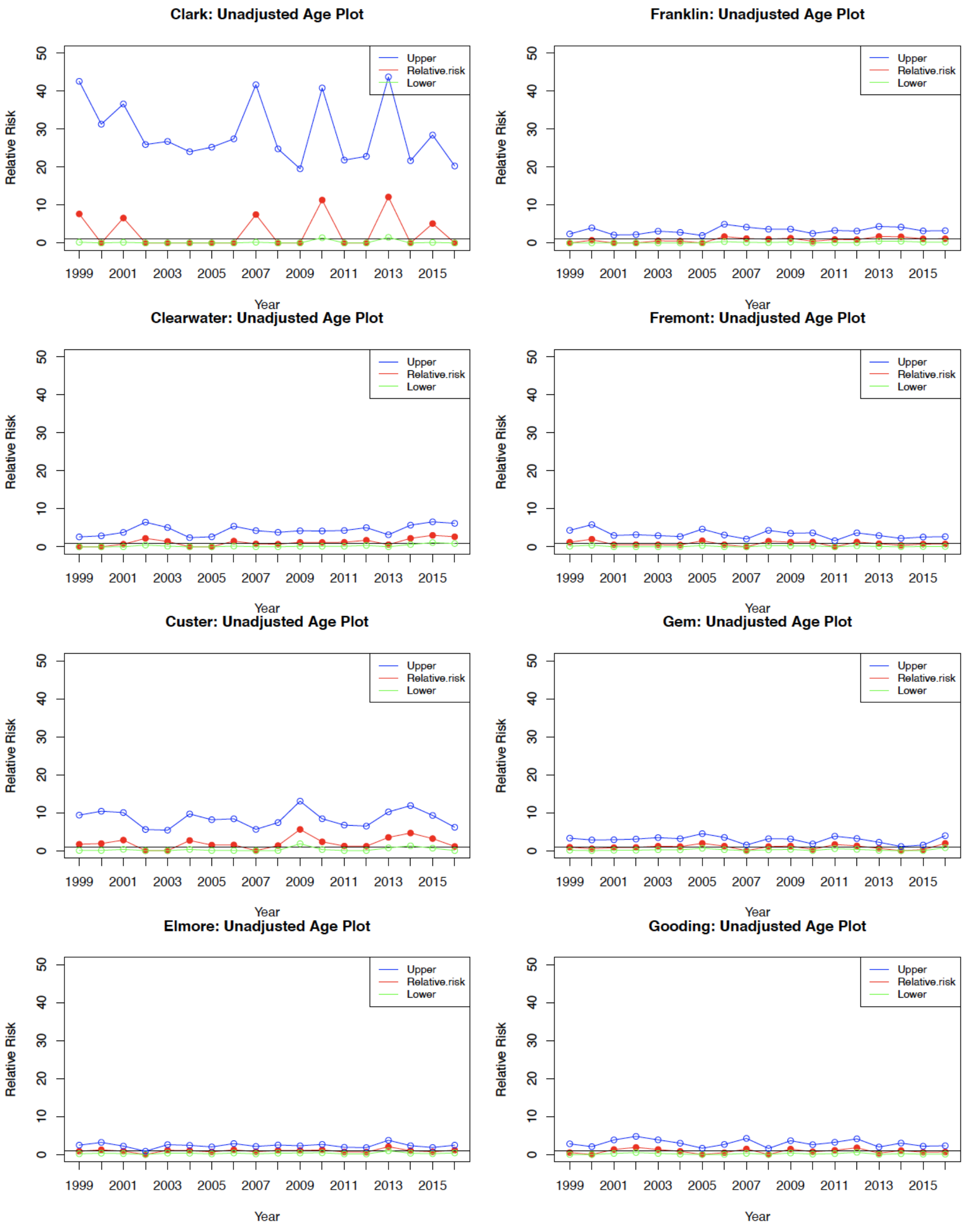
Butte: Unadjusted Age Plot

Cassia: Unadjusted Age Plot

Figure 3.13: Standardized Mortality Ratio

Appendix B.1.3: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.



Clark: Unadjusted Age Plot

Franklin: Unadjusted Age Plot

Clearwater: Unadjusted Age Plot

Fremont: Unadjusted Age Plot

Custer: Unadjusted Age Plot

Gem: Unadjusted Age Plot

Elmore: Unadjusted Age Plot

Gooding: Unadjusted Age Plot

Figure 3.14: Standardized Mortality Ratio

Appendix B.1.4. Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

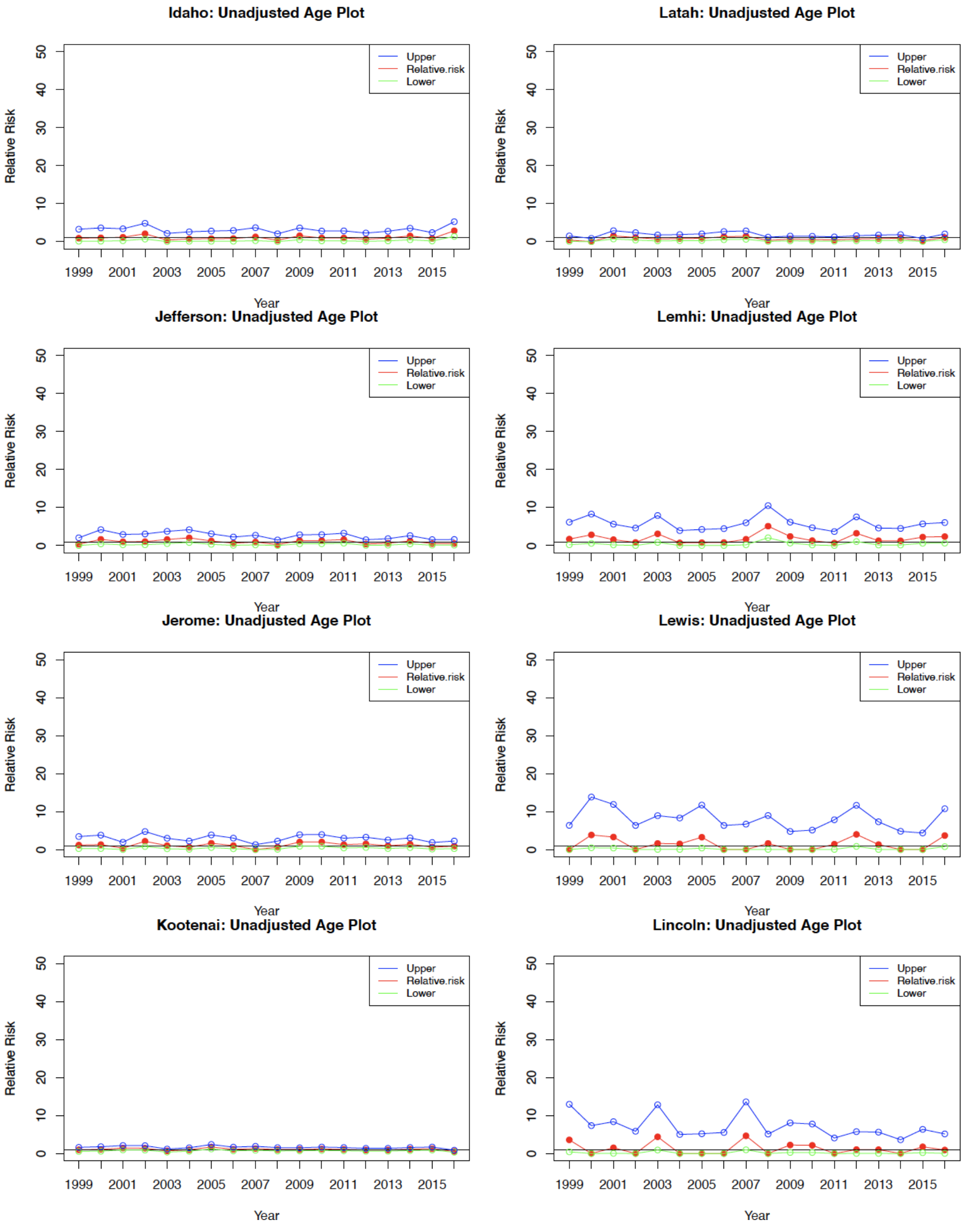
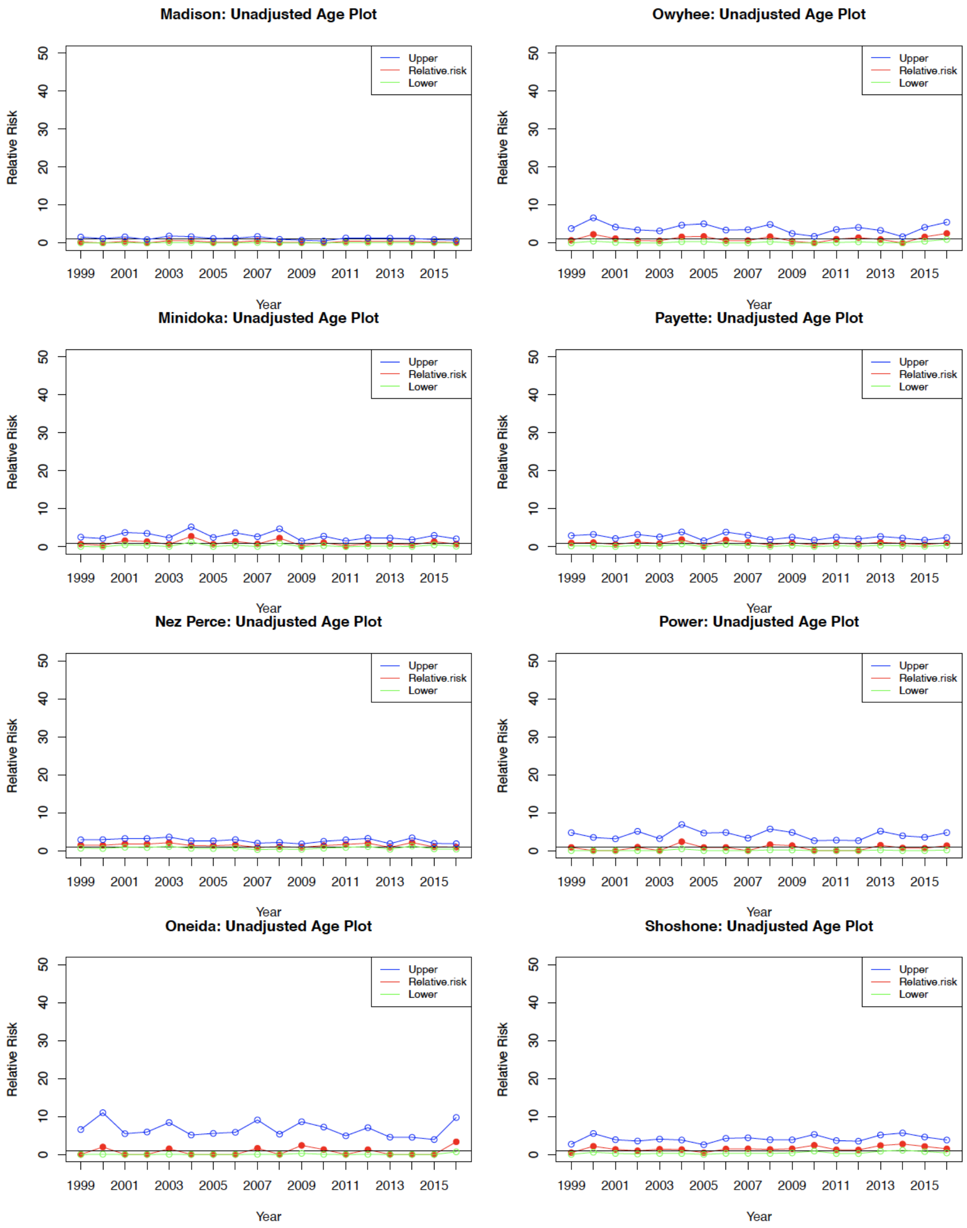


Figure 3.15: Standardized Mortality Ratio

Appendix B.1.5: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.



Madison: Unadjusted Age Plot

Owyhee: Unadjusted Age Plot

Minidoka: Unadjusted Age Plot

Payette: Unadjusted Age Plot

Nez Perce: Unadjusted Age Plot

Power: Unadjusted Age Plot

Oneida: Unadjusted Age Plot

Shoshone: Unadjusted Age Plot

Figure 3.16: Standardized Mortality Ratio

Appendix B.1.6: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

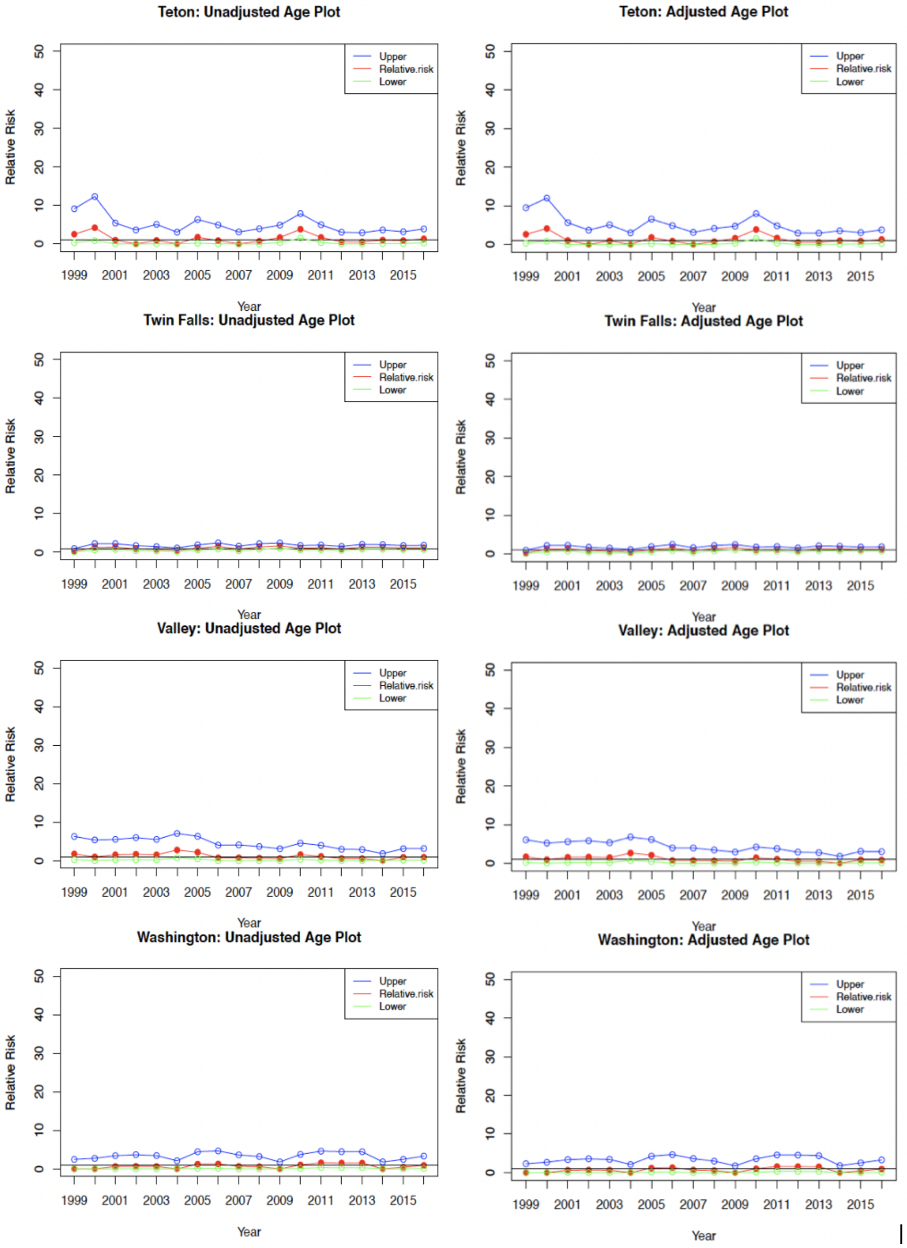
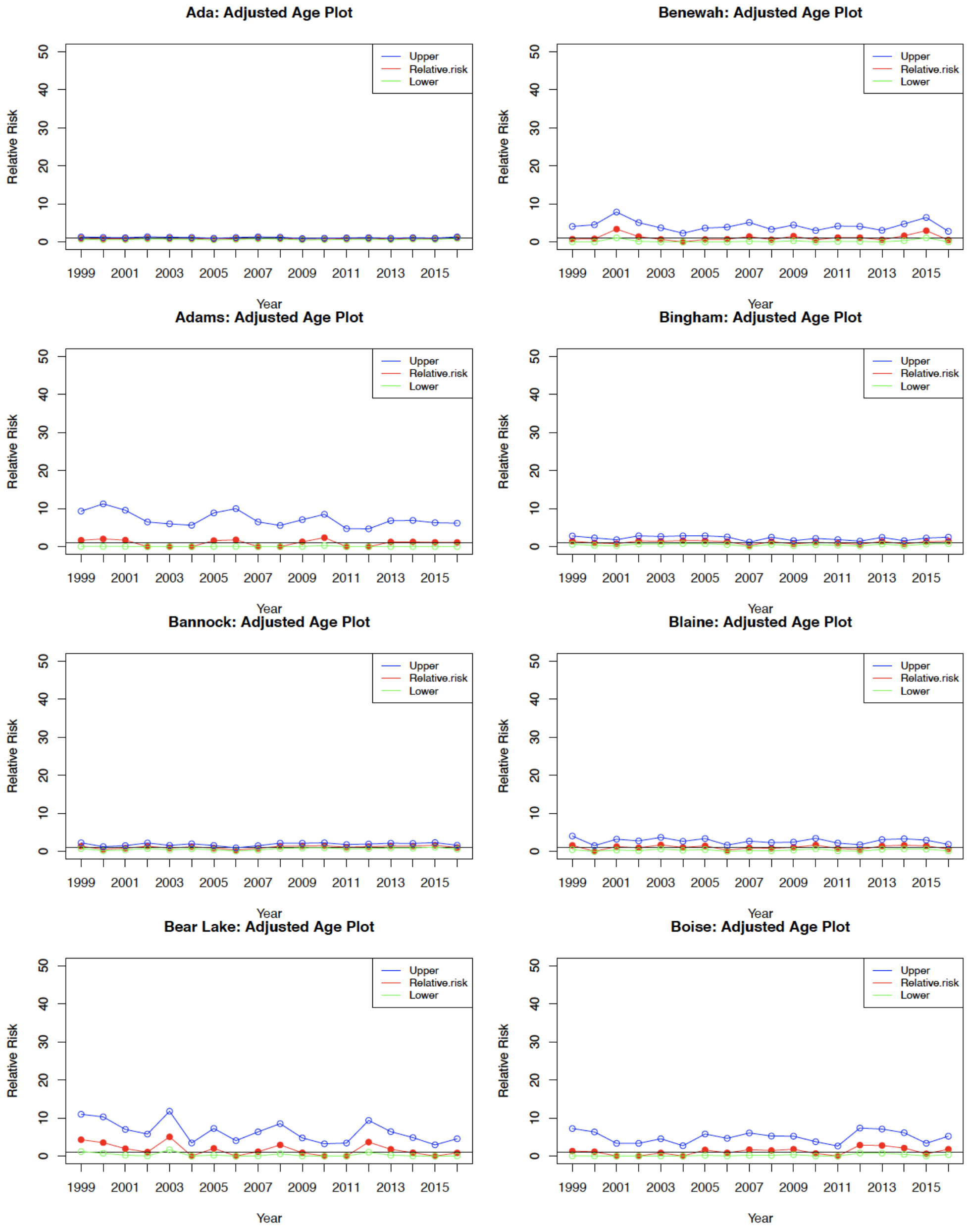


Figure 3.17: Standardized Mortality Ratio

Appendix B.1.7: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.



Ada: Adjusted Age Plot

Benewah: Adjusted Age Plot

Adams: Adjusted Age Plot

Bingham: Adjusted Age Plot

Bannock: Adjusted Age Plot

Blaine: Adjusted Age Plot

Bear Lake: Adjusted Age Plot

Boise: Adjusted Age Plot

Figure 3.18: Standardized Mortality Ratio

Appendix B.1.8: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

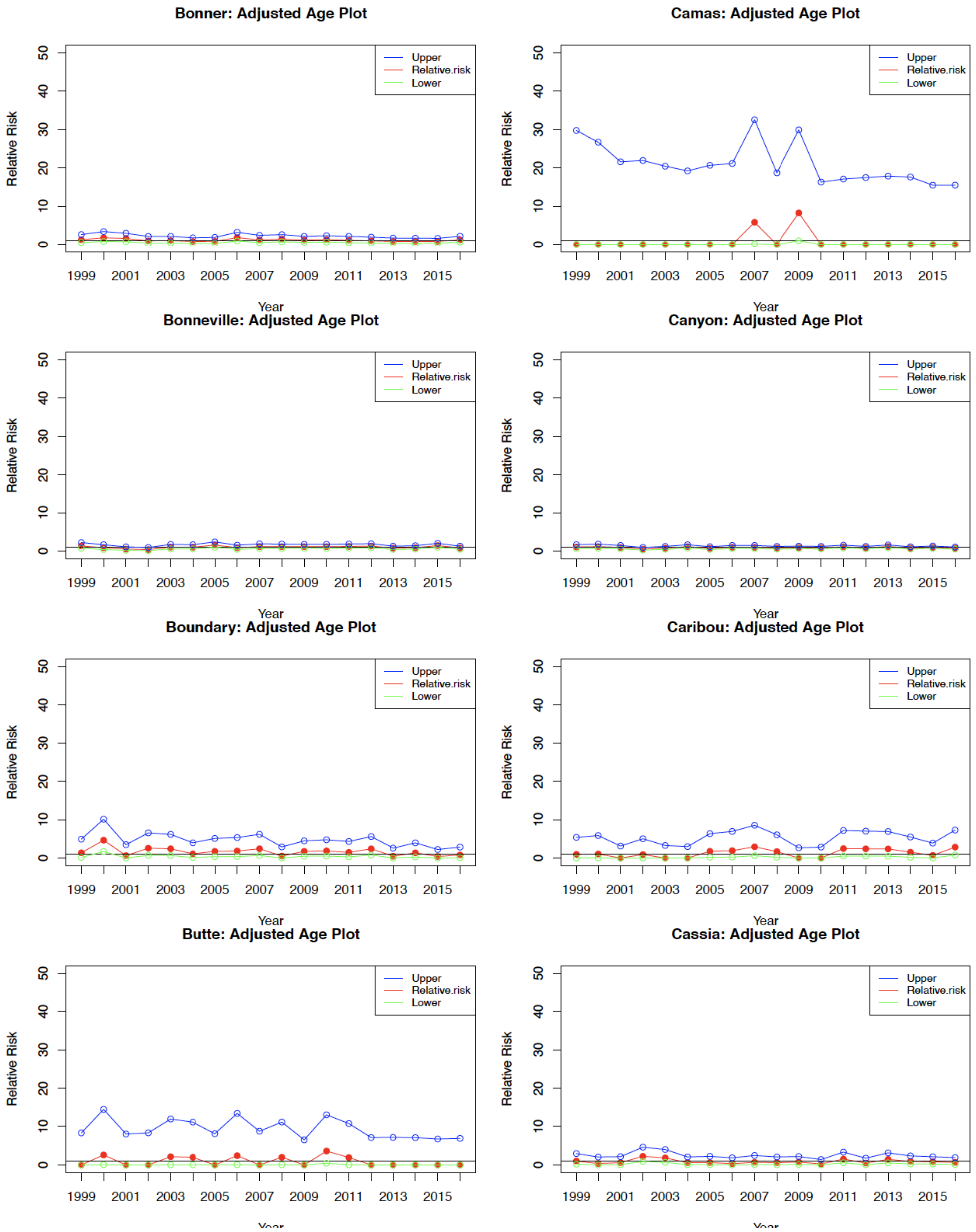


Figure 3.19: Standardized Mortality Ratio

Appendix B.1.9: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

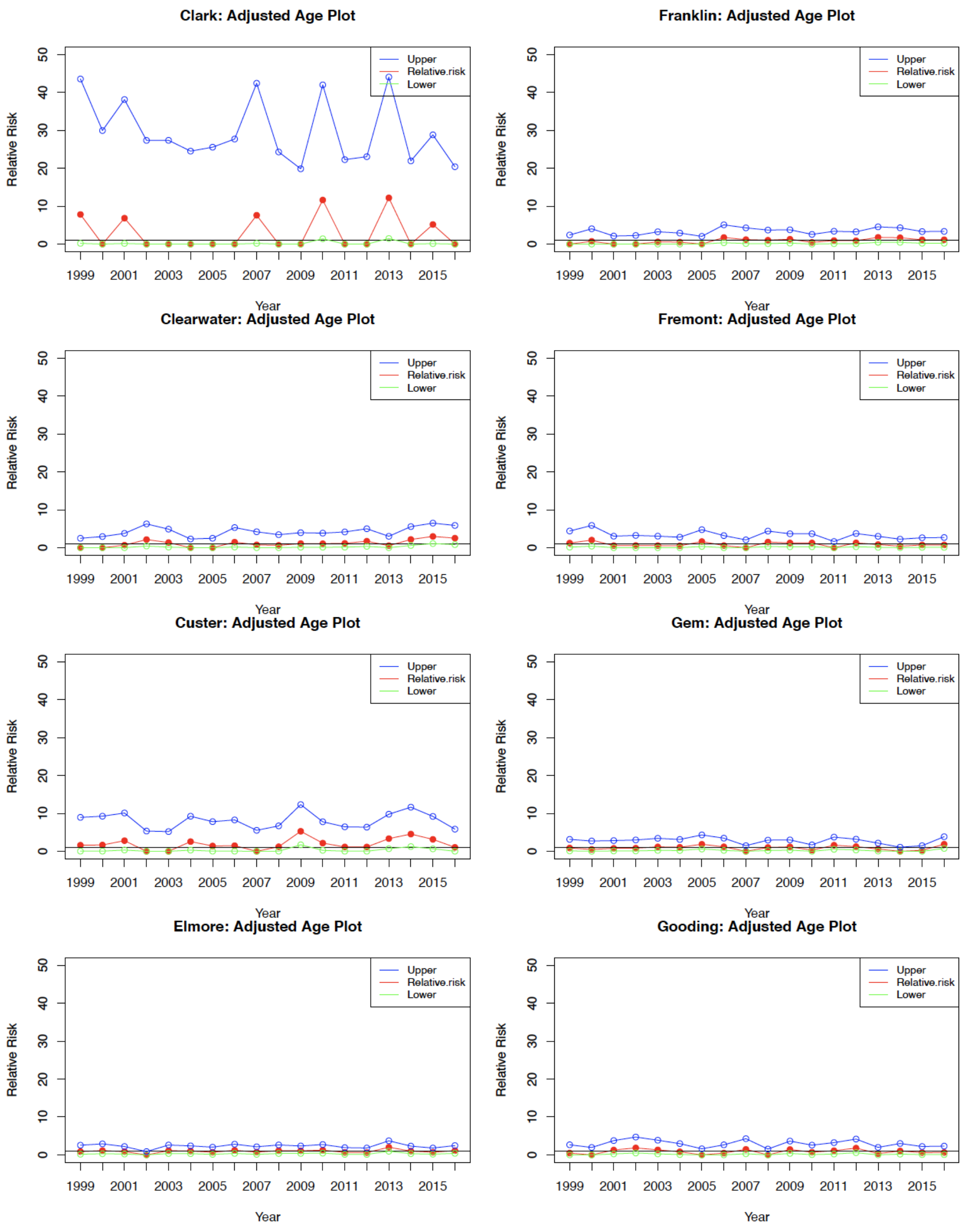


Figure 3.20: Standardized Mortality Ratio

Appendix B.1.10: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

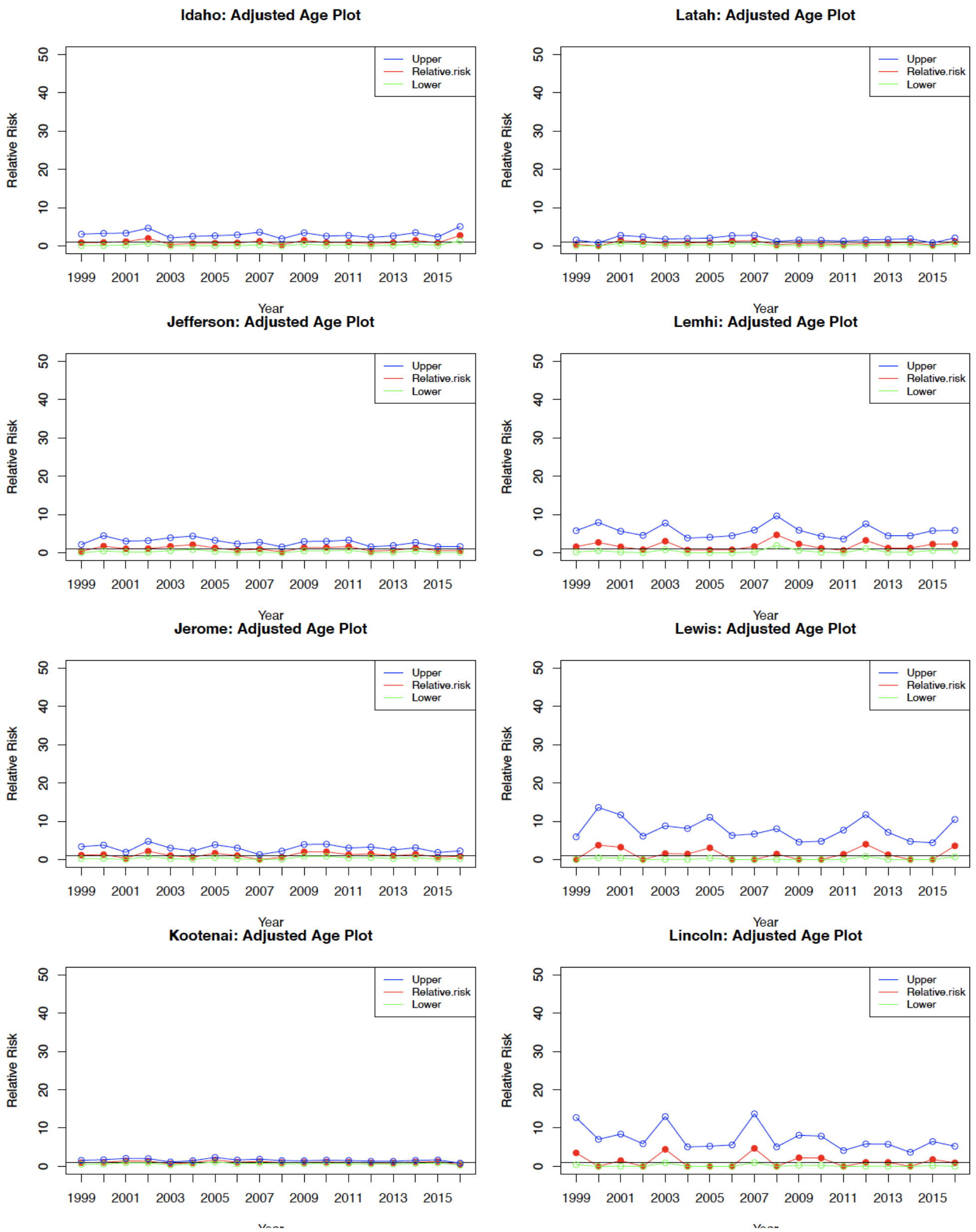


Figure 3.21: Standardized Mortality Ratio

Appendix B.1.11: Standardized Mortality Ratio

Note : Axis is not same as Hierarchical Bayes plots.

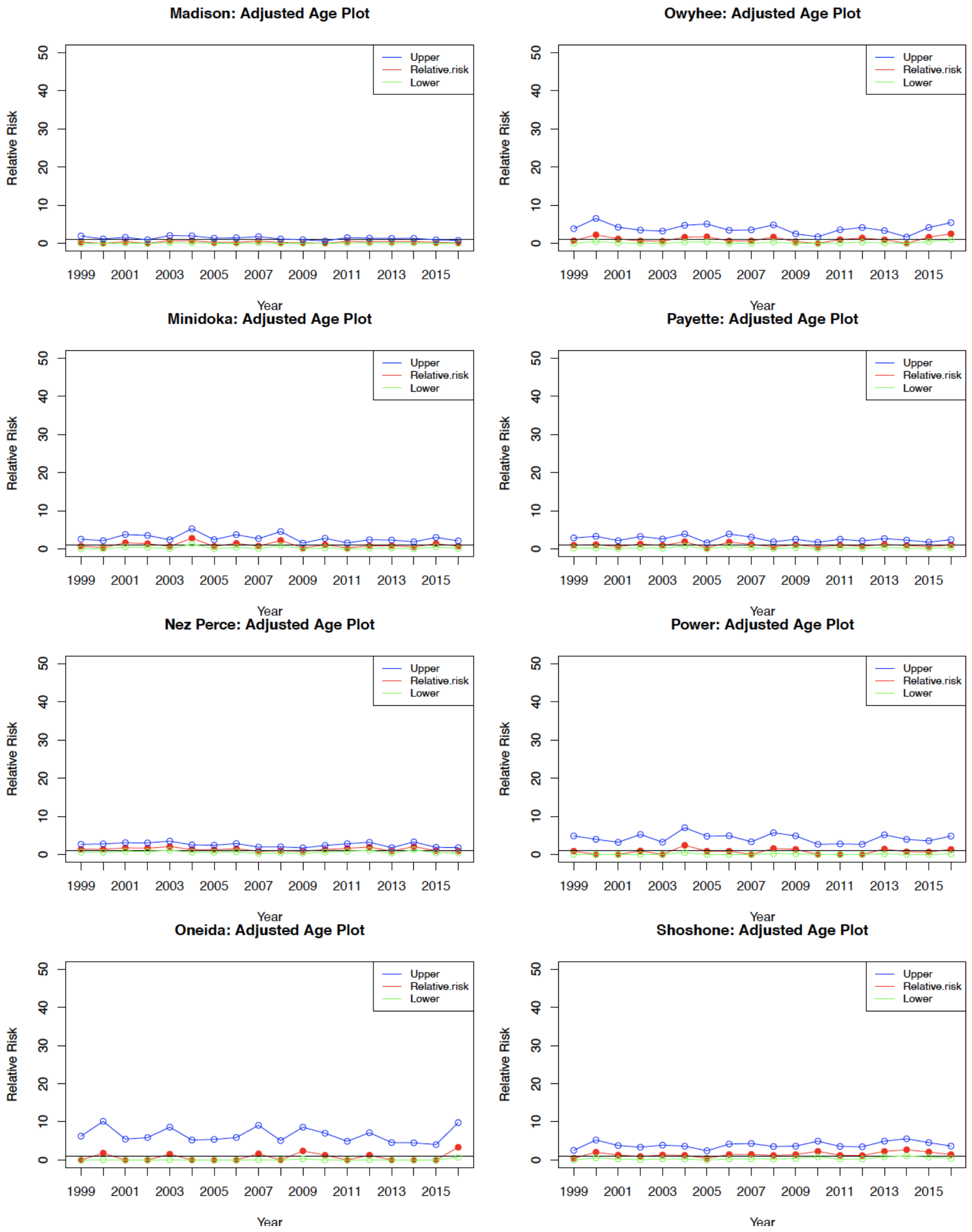


Figure 3.22: Standardized Mortality Ratio

Appendix C.1.1: Expected and Observed: Age Unadjusted

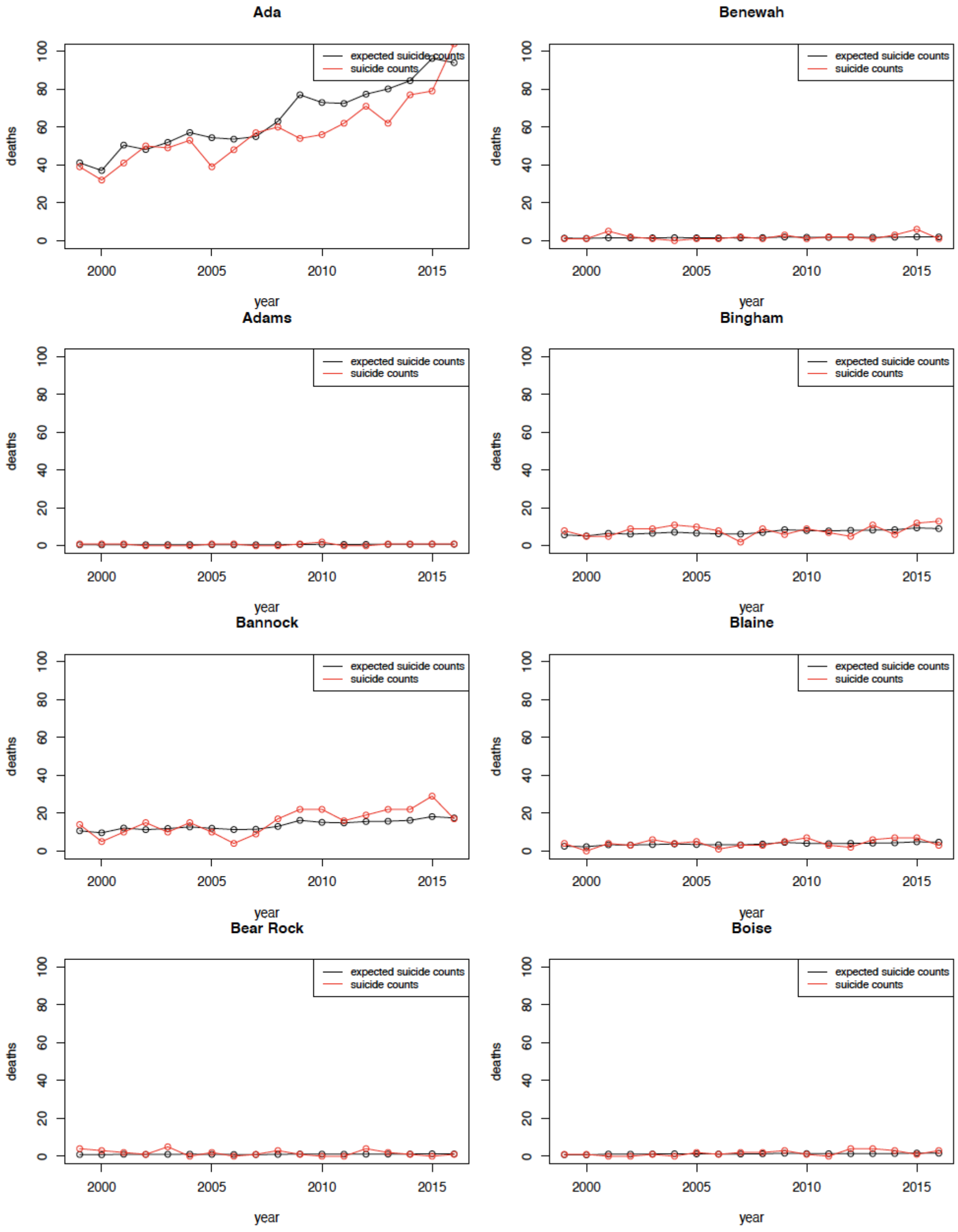


Figure 3.23: Expected and Observed

Appendix C.1.2: Expected and Observed: Age Unadjusted

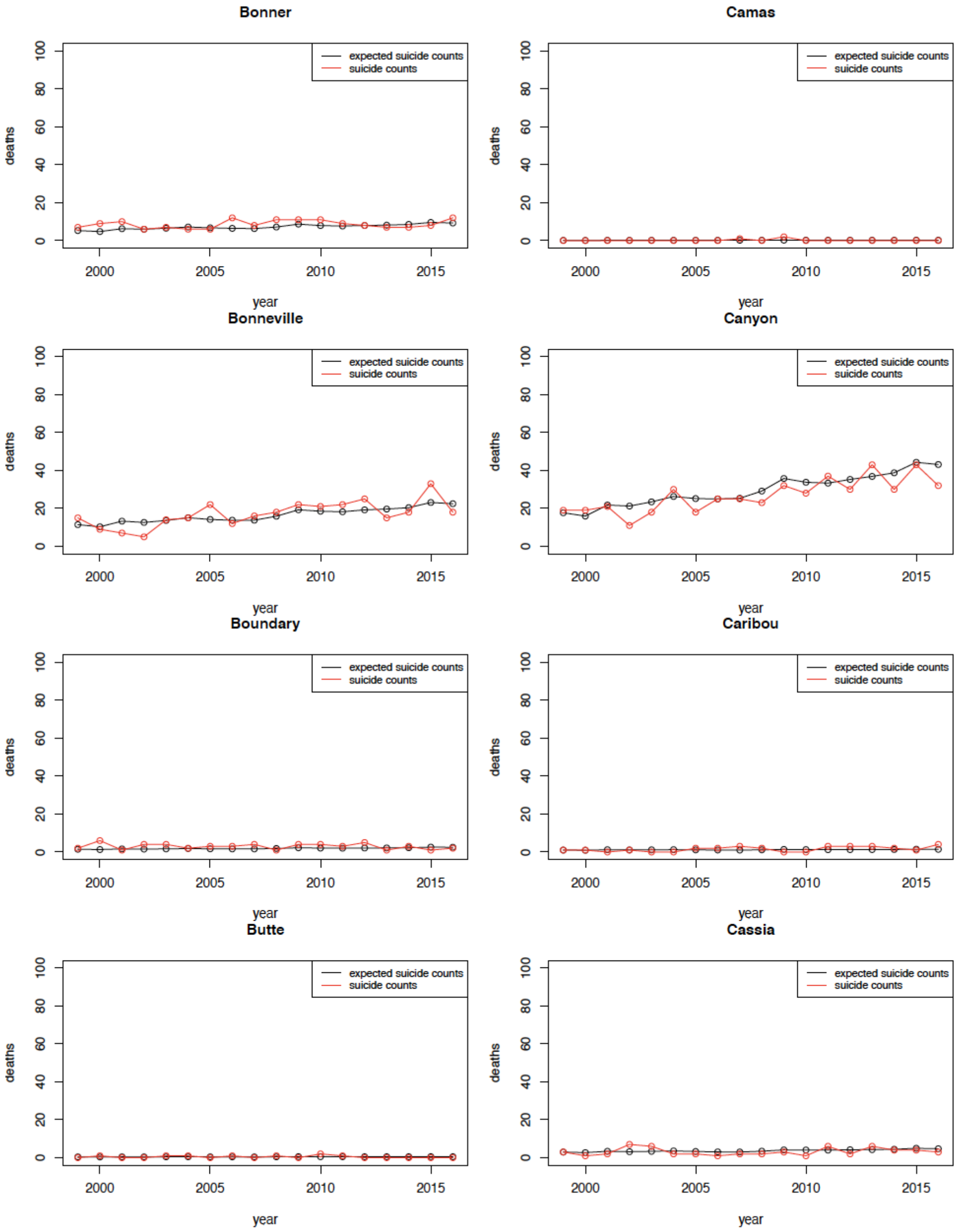


Figure 3.24: Expected and Observed

Appendix C.1.3: Expected and Observed: Age Unadjusted

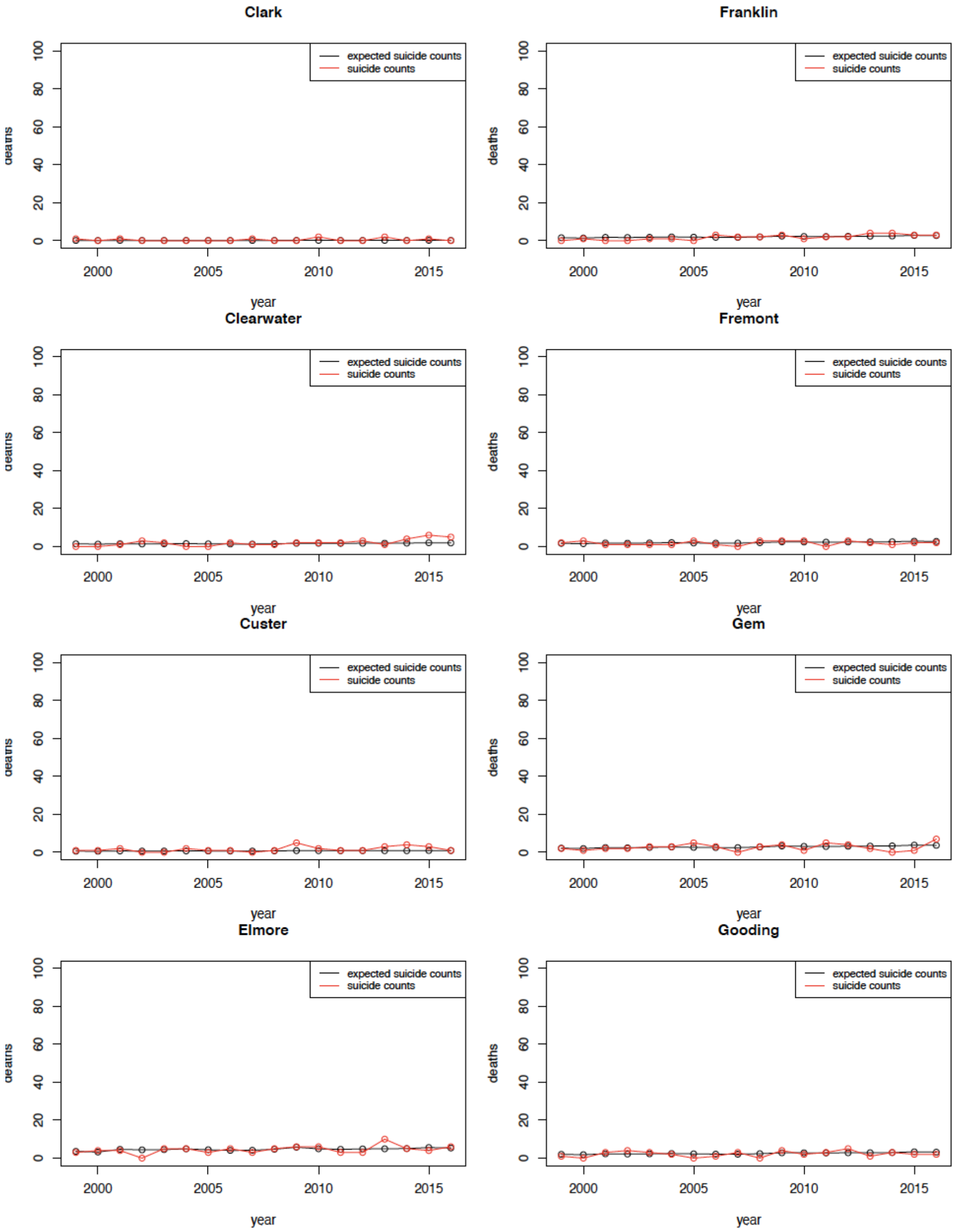


Figure 3.25: Expected and Observed

Appendix C.1.4: Expected and Observed: Age Unadjusted

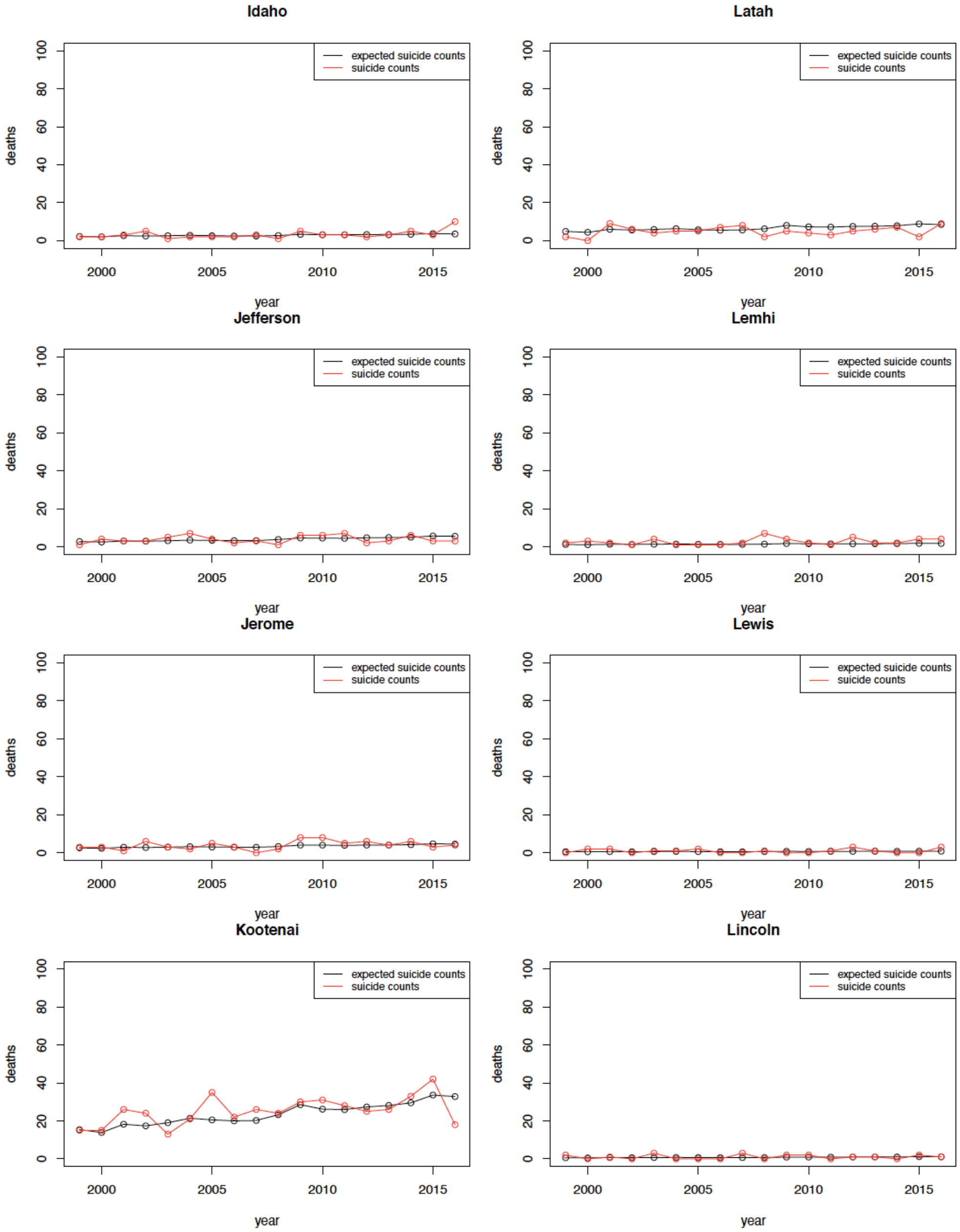


Figure 3.26: Expected and Observed

Appendix C.1.5: Expected and Observed: Age Unadjusted

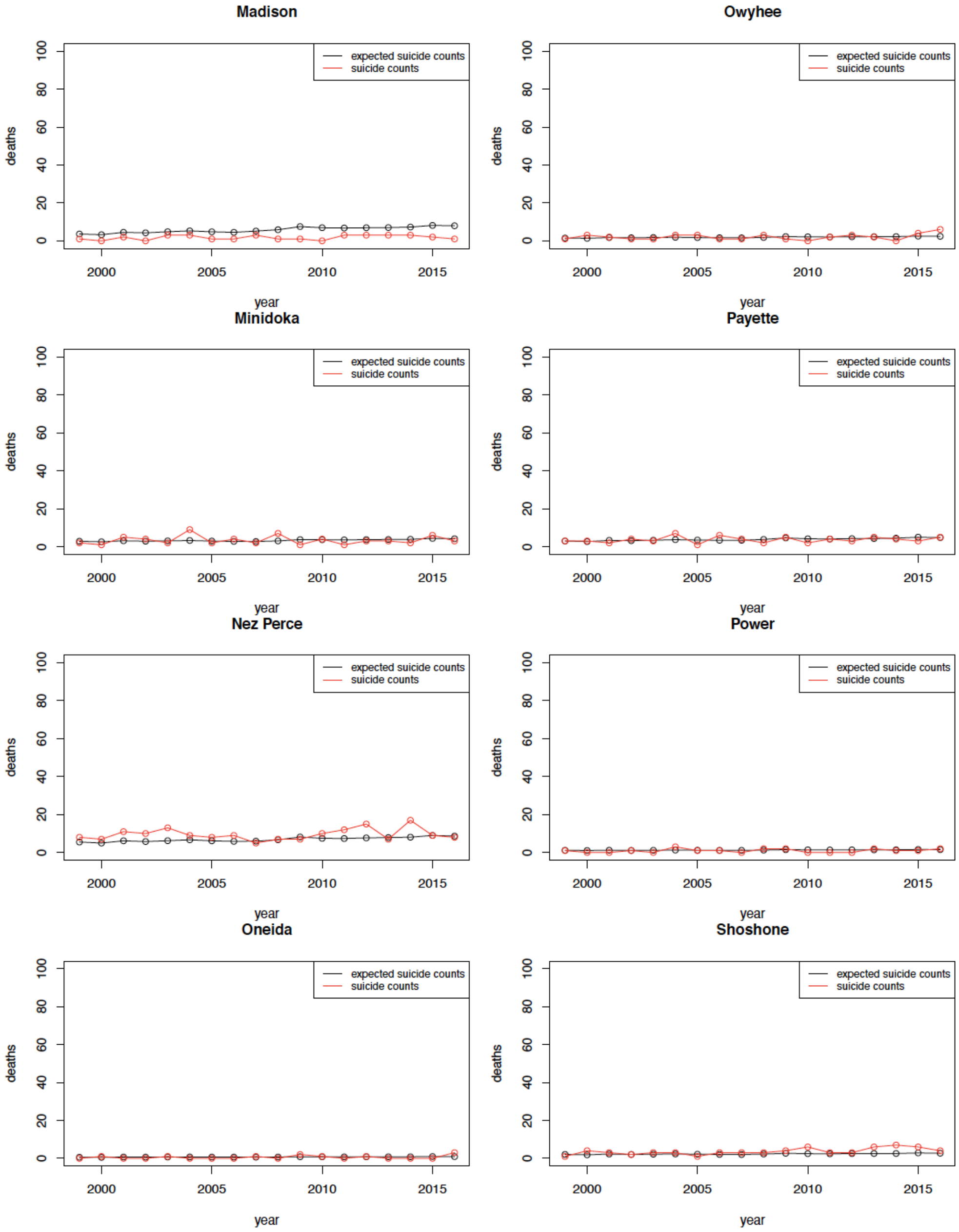


Figure 3.27: Expected and Observed

Appendix C.1.6: Expected and Observed

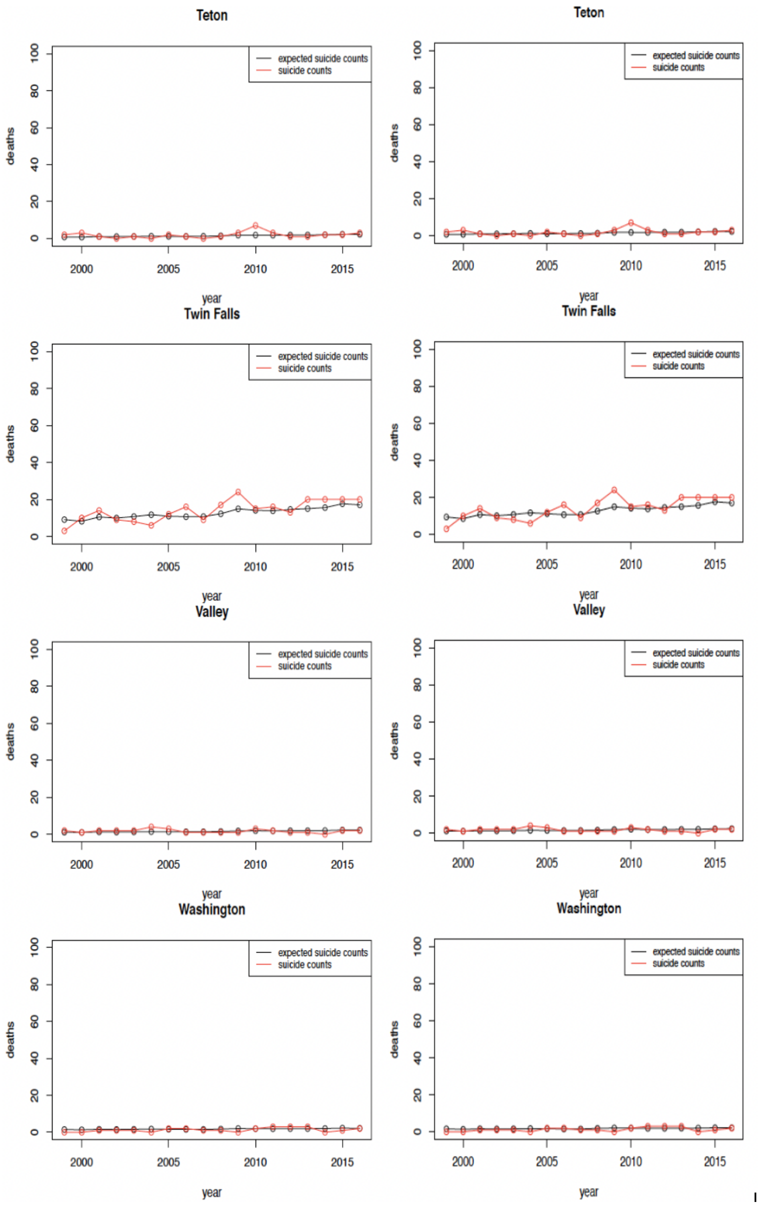


Figure 3.28: Unadjusted(left) , Adjusted(right)

Appendix C.1.7: Expected and Observed: Age Adjusted

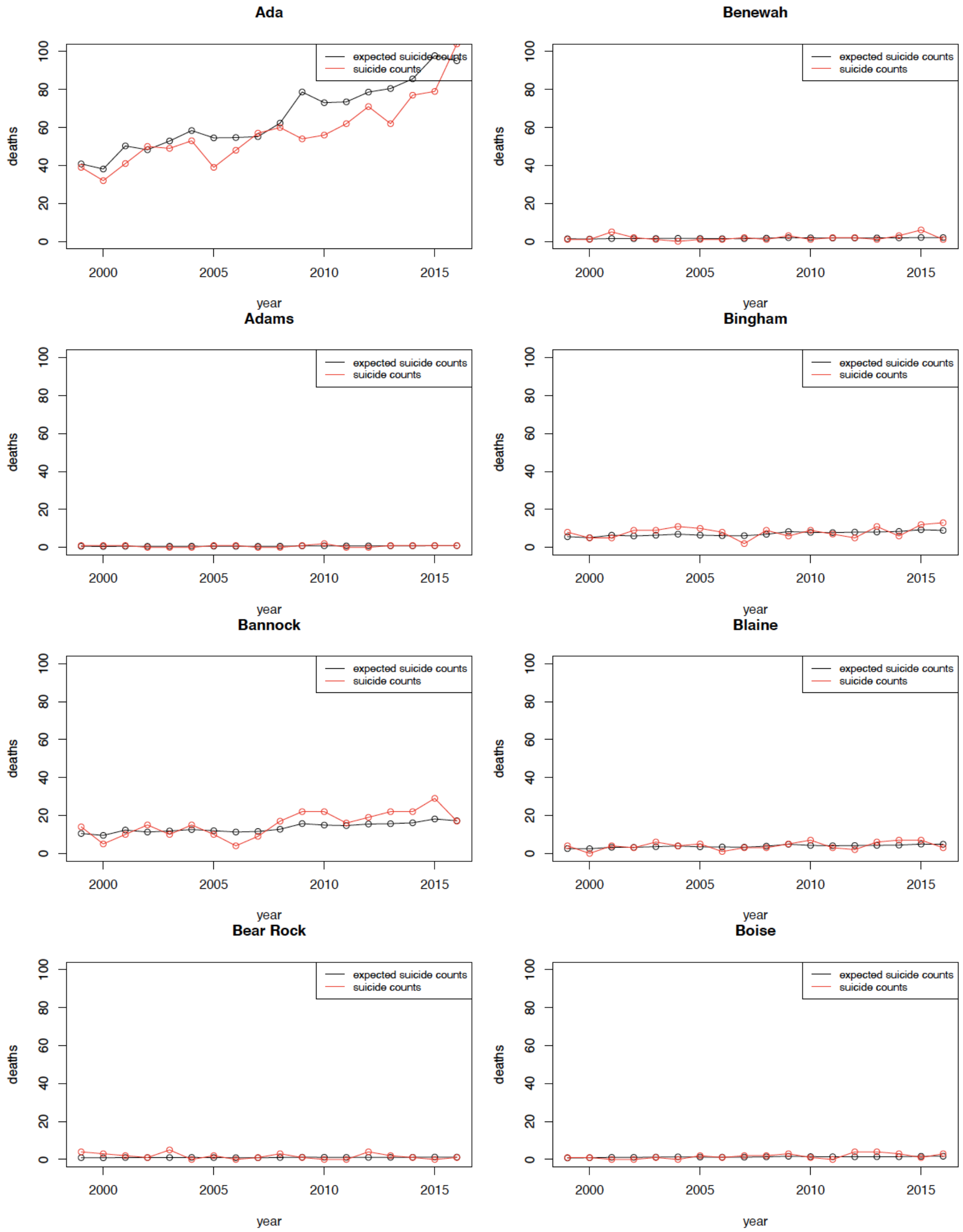


Figure 3.29: Expected and Observed

Appendix C.1.8: Expected and Observed: Age Adjusted

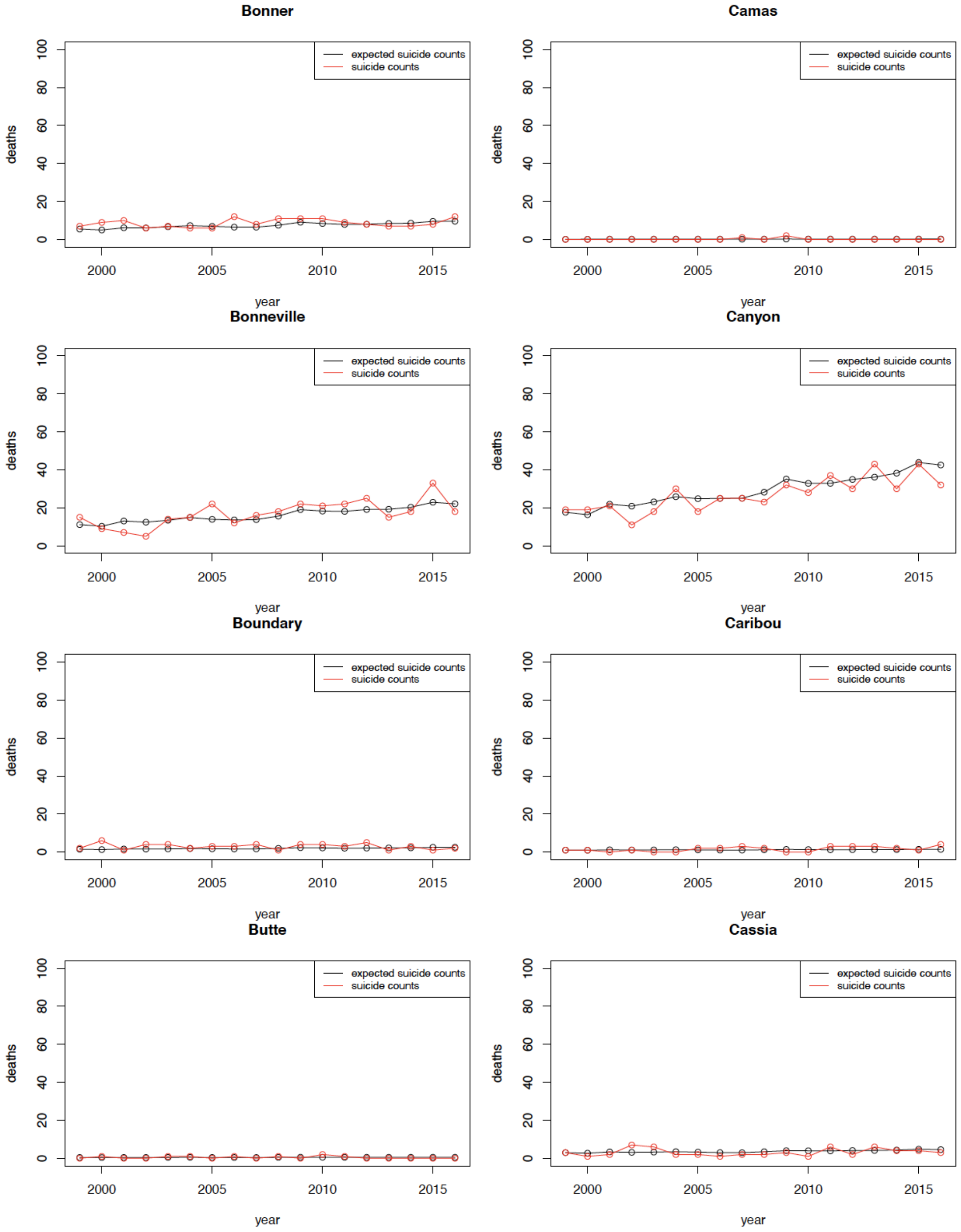


Figure 3.30: Expected and Observed

Appendix C.1.9: Expected and Observed: Age Adjusted

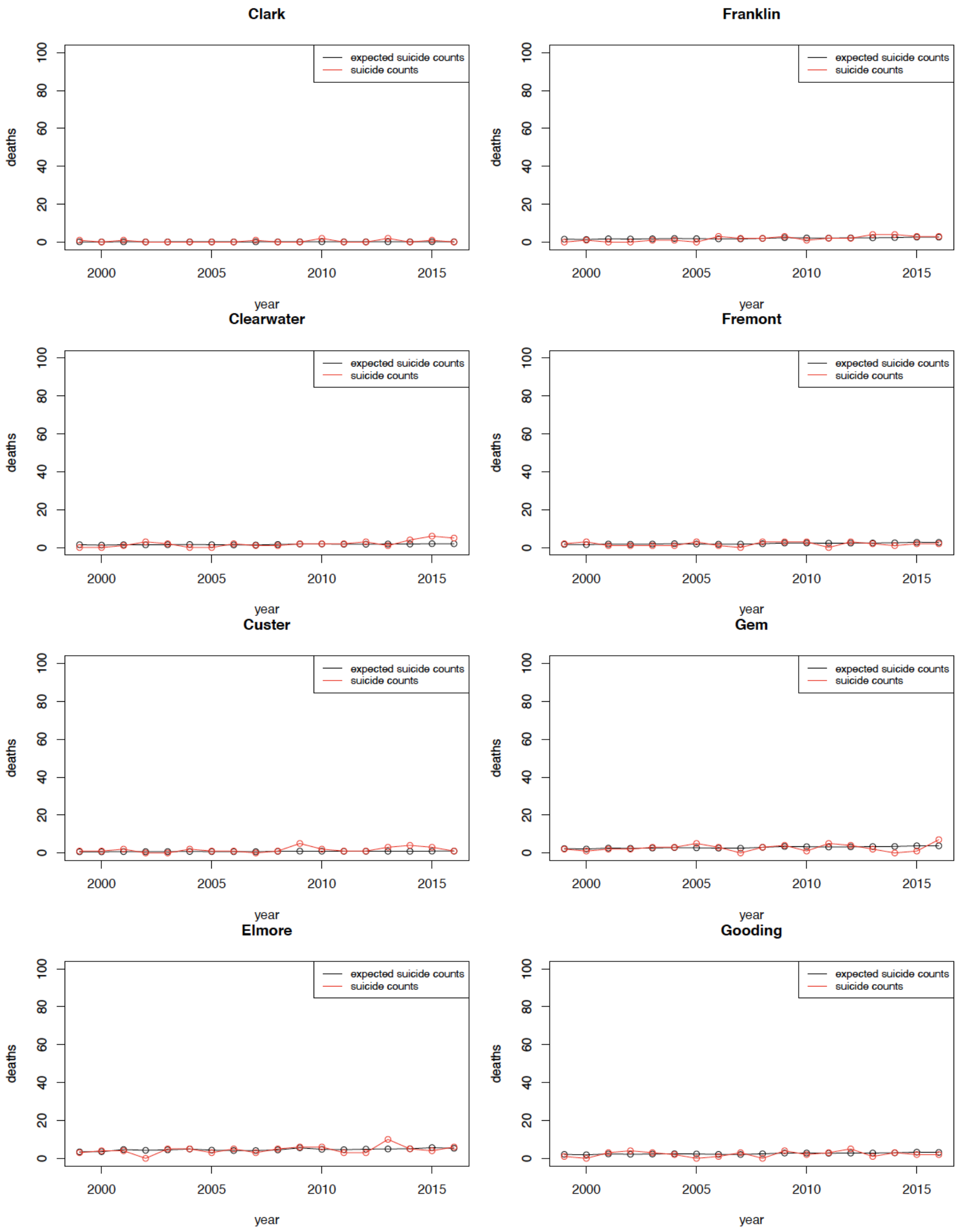


Figure 3.31: Expected and Observed

Appendix C.1.10: Expected and Observed: Age Adjusted

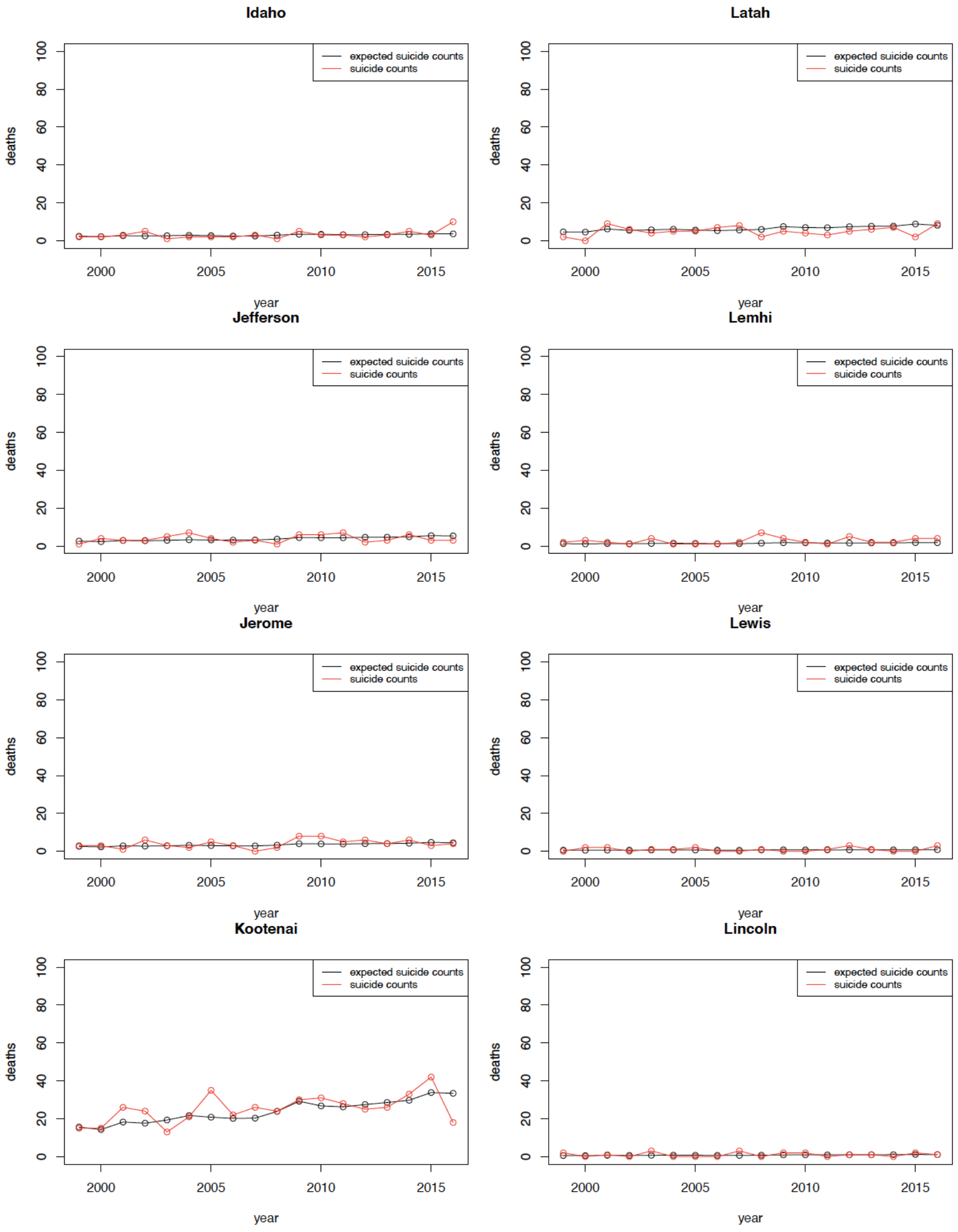


Figure 3.32: Expected and Observed

Appendix C.1.11: Expected and Observed: Age Adjusted

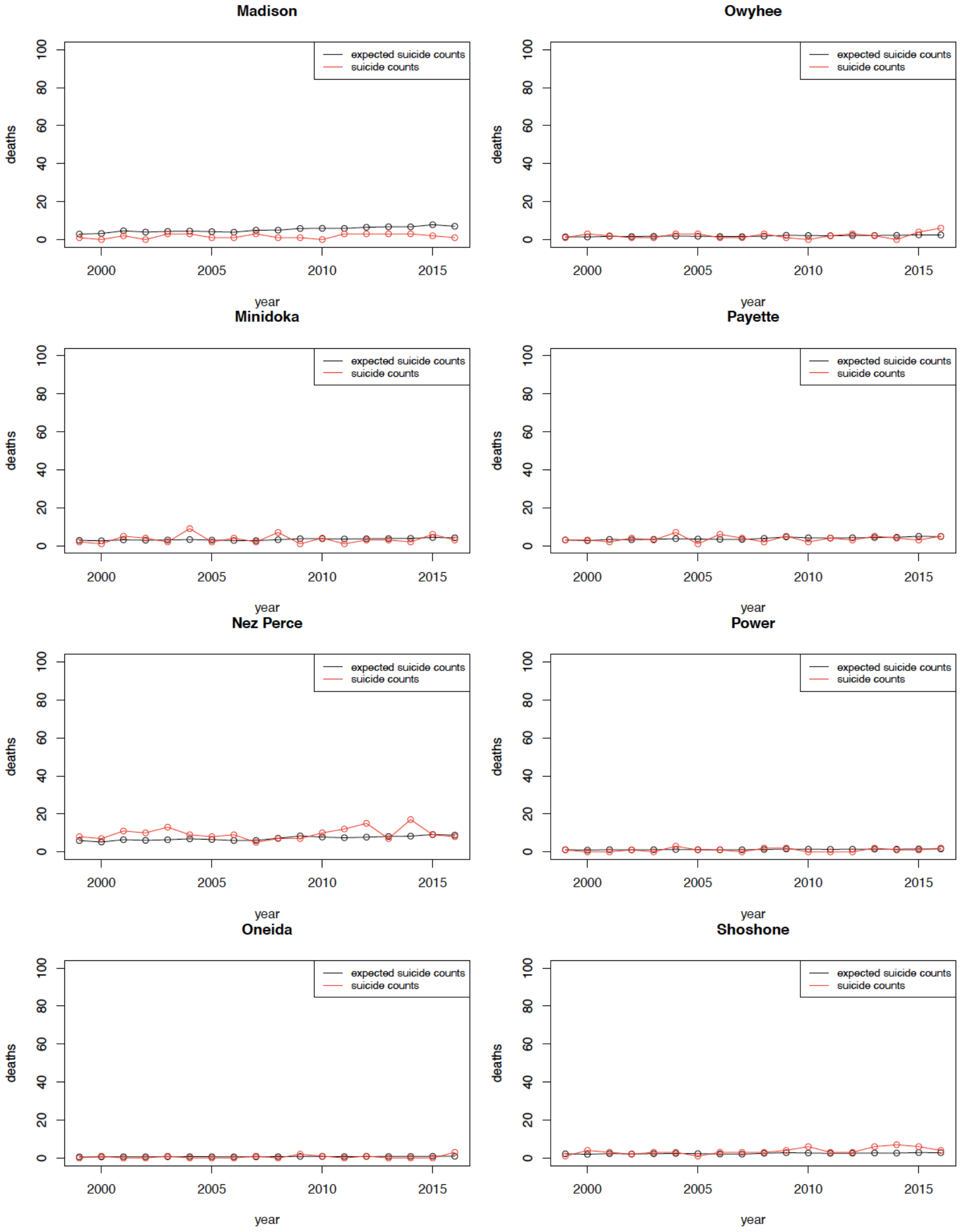
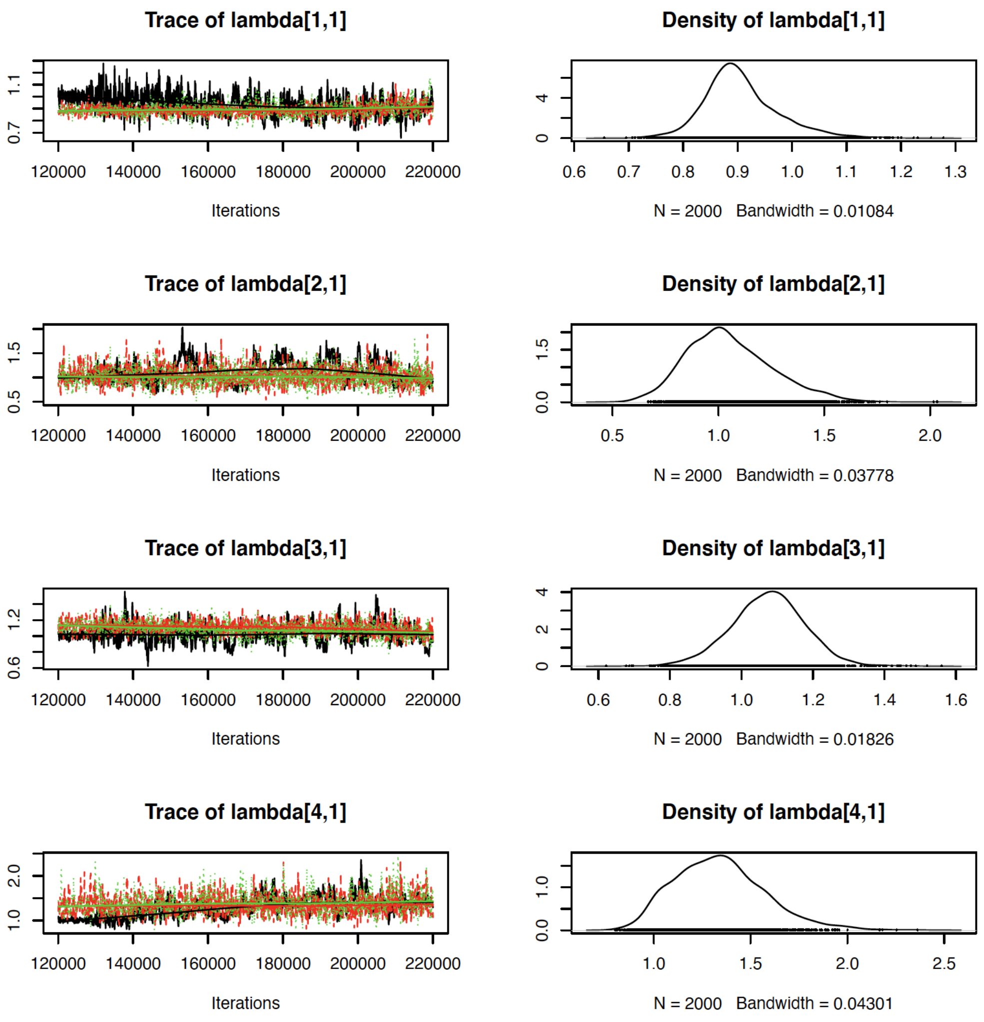


Figure 3.33: Expected and Observed

Appendix D.1.1: Trace plots: Age Unadjusted

The figure below is a sample of trace plots from the model.


Trace of lambda[1,1]

Density of lambda[1,1]

Trace of lambda[2,1]

Density of lambda[2,1]

Trace of lambda[3,1]

Density of lambda[3,1]

Trace of lambda[4,1]

Density of lambda[4,1]

Figure 3.34: Trace plots: Age Unadjusted

Appendix D.1.2: Trace plots: Age Unadjusted

The figure below is a sample of trace plots from the model.

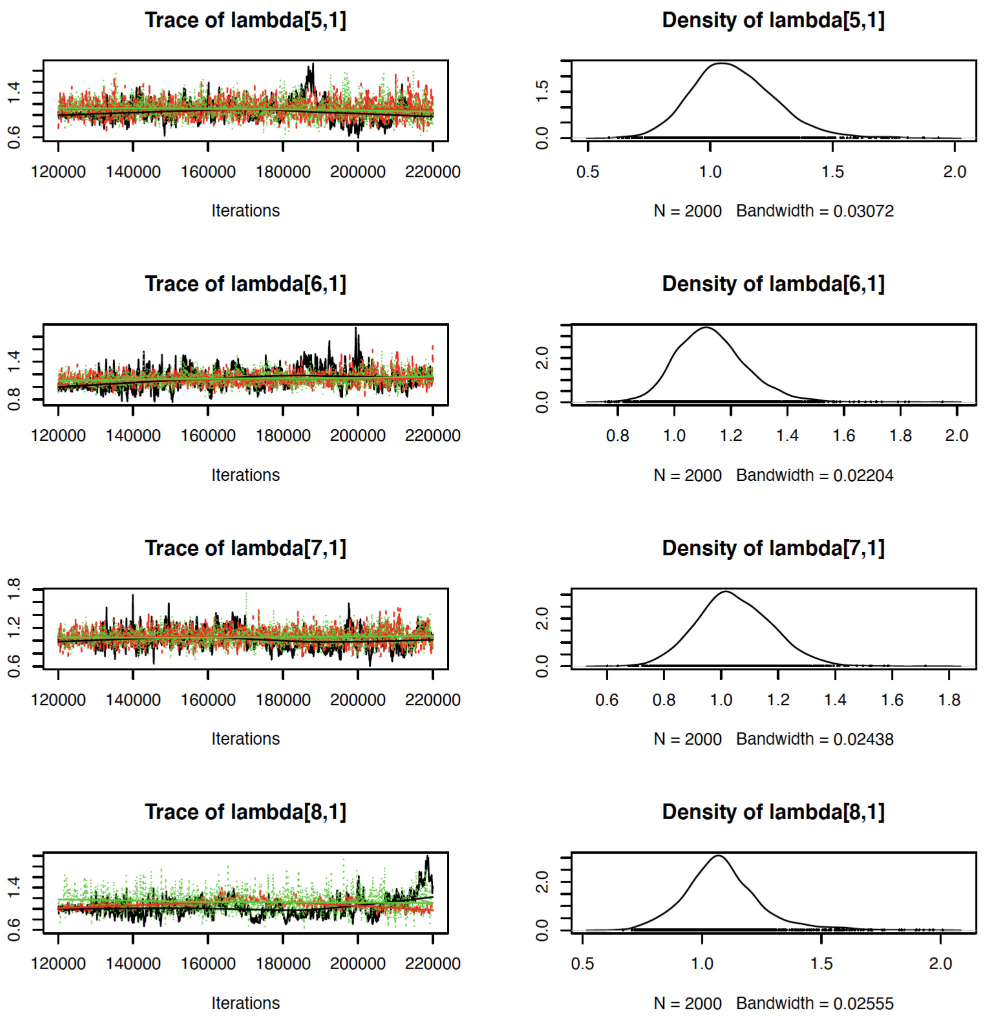
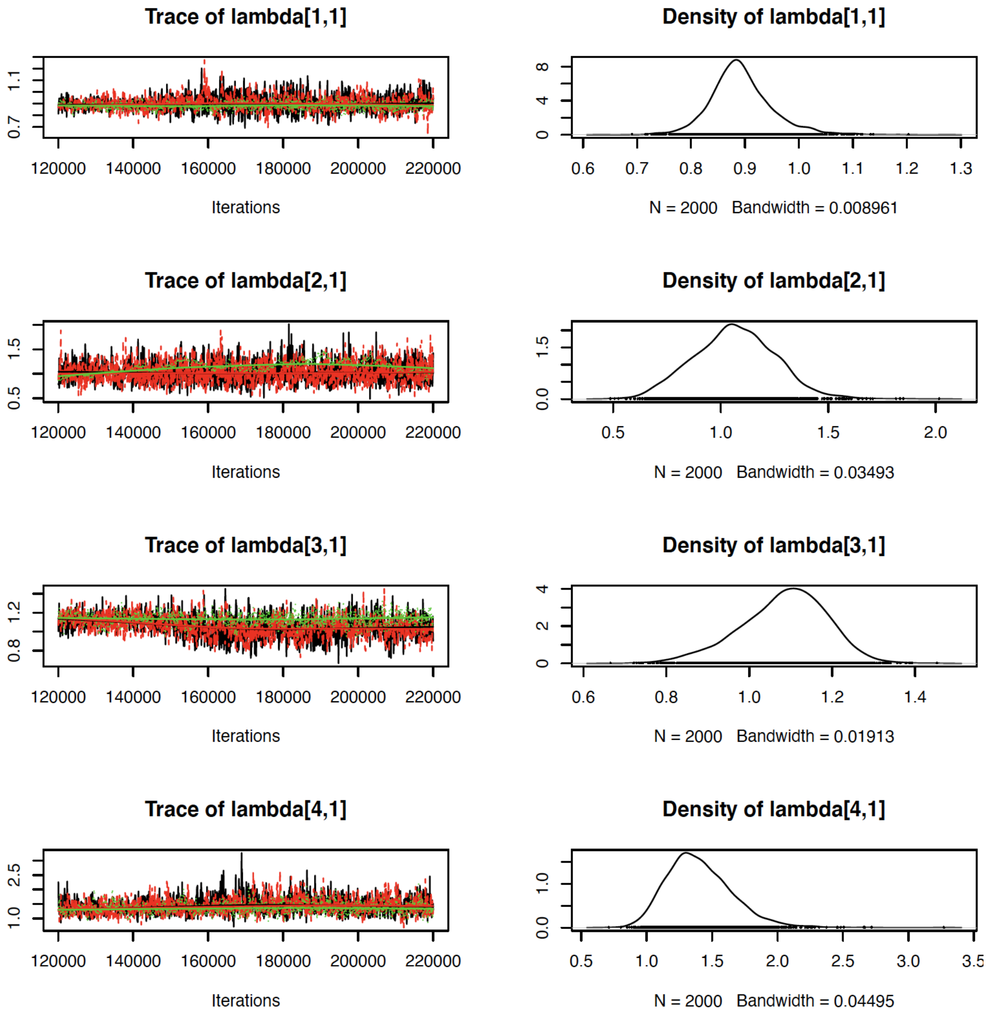


Figure 3.35: Trace plots: Age Unadjusted

Appendix D.1.3: Trace plots: Age Adjusted

The figure below is a sample of trace plots from the model.



Trace of lambda[1,1]

Density of lambda[1,1]

Trace of lambda[2,1]

Density of lambda[2,1]

Trace of lambda[3,1]

Density of lambda[3,1]

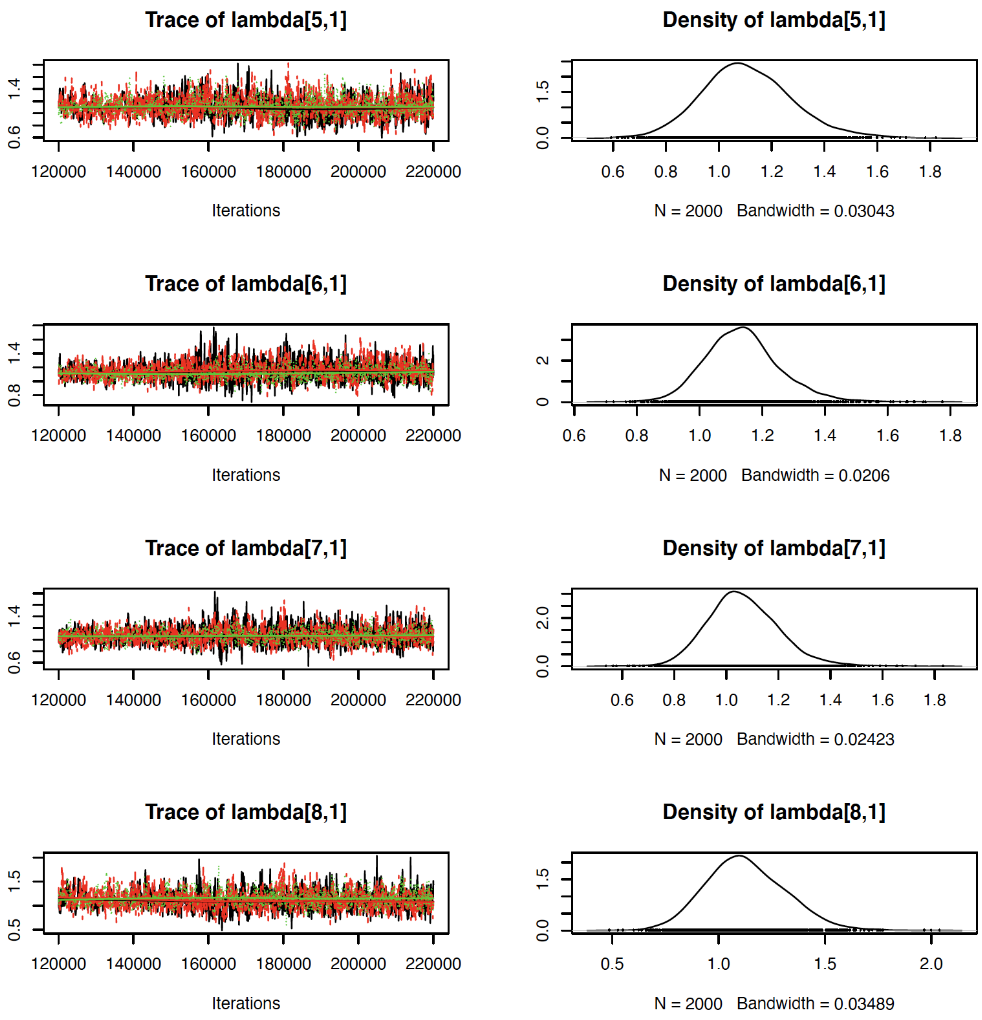
Trace of lambda[4,1]

Density of lambda[4,1]

Figure 3.36: Trace plots: Age Adjusted

Appendix D.1.4: Trace plots: Age Adjusted

The figure below is a sample of trace plots from the model.



Trace of lambda[5,1]

Density of lambda[5,1]

Trace of lambda[6,1]

Density of lambda[6,1]

Trace of lambda[7,1]

Density of lambda[7,1]

Trace of lambda[8,1]

Density of lambda[8,1]

Figure 3.37: Trace plots: Age Adjusted

Appendix E.1.1: Gelman Rubin Plot: Age Undjusted

The figure below is a sample of the Gelman Rubin plots from the model.

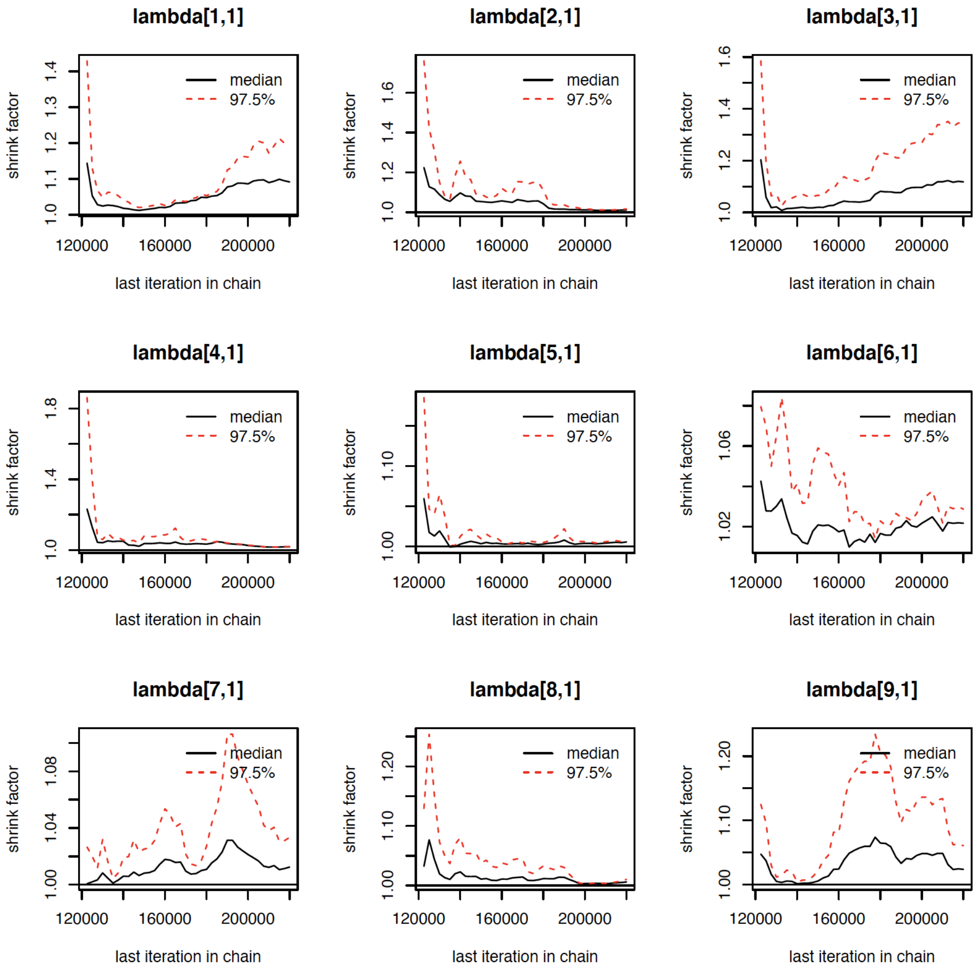


Figure 3.38: Gelman Rubin Plot: Age Unadjusted

Appendix E.1.2: Gelman Rubin Plot: Age Unadjusted

The figure below is a sample of the Gelman Rubin plots from the model.

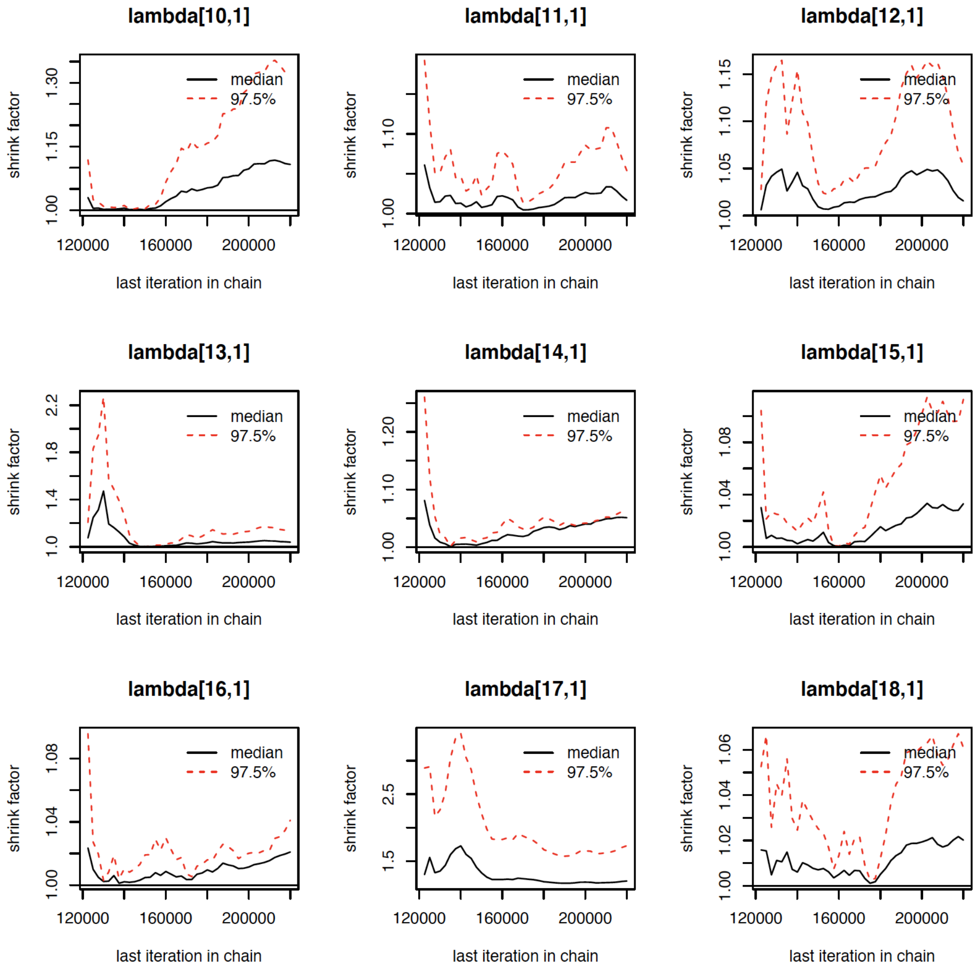


Figure 3.39: Gelman Rubin Plot: Age Unadjusted

Appendix E.1.3: Gelman Rubic Statistic: Age Adjusted

The figure below is a sample of the Gelman Rubin plots from the model.

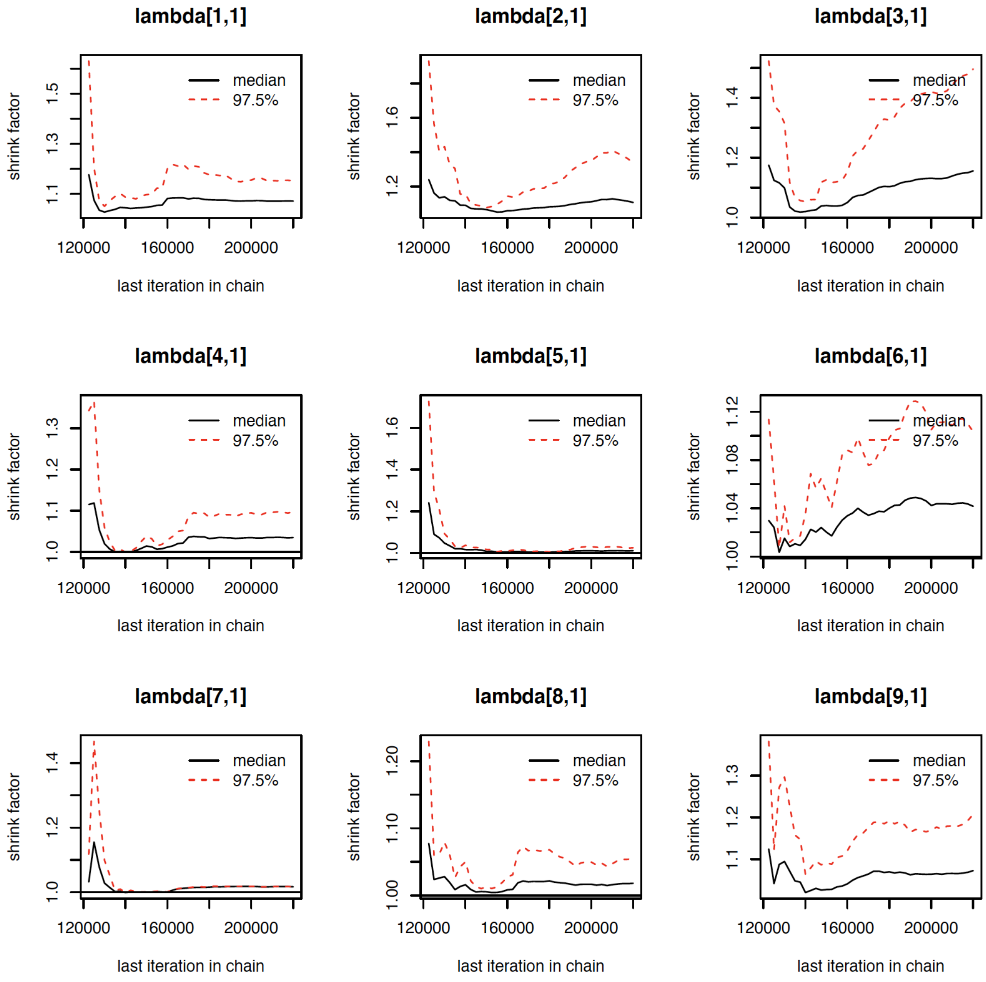


Figure 3.40: Gelman Rubin Plot: Age Adjusted

Appendix E.1.4: Gelman Rubic Statistic: Age Adjusted

The figure below is a sample of the Gelman Rubin plots from the model.

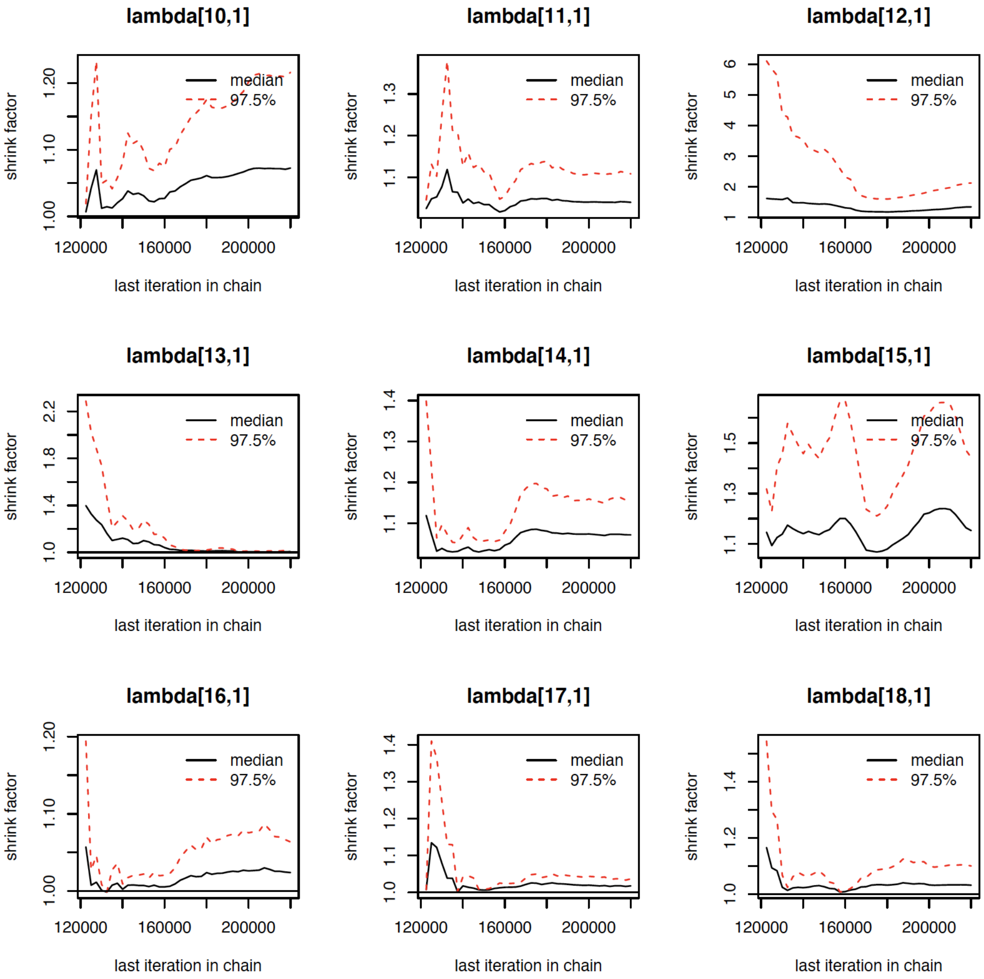


Figure 3.41: Gelman Rubin Plot: Age Adjusted

The figure below is a sample of the autocorrelation plots using the data.

Appendix F: R Code

Rcode used in data cleaning and analysis.

```
library(readxl)

library(plyr)

library(dplyr)

library(ggplot2)

library(shiny)

library(rmarkdown)

devtools::session_info()

#Cleaning suicide Data

my_data <- read_excel("WiestData.xlsx")

#Filter to select suicide deaths only

death_fil=filter(my_data,my_data$‘Manner of death‘ =="Suicide")

data1=death_fil[,1:5]

names(data1)=c("Year","Sex","County","Age","Manner")

attach(data1)

sui=count(data1, vars = c("Year","County"))

sui_sex=count(data1, vars = c("Year","County","Sex"))

su=count(data1, vars = c("County"))select$freq=rep(1,4654)

new.data  = data.frame(Year,County,AgeGroups,freq)

n_su=aggregate(freq~County+AgeGroups+Year,data=new.data,sum)

###### loop to extract each year.

  listofdfs <- list()

  for(i in 1:length(year)){

    my_su01=data.frame(sui[sui$Year ==  year[i], ])

    new01=merge(x = ner, y = my_su01, by  = "County", all = TRUE)
```

```
    new01[is.na(new01)] <- 0
    Year=rep(year[i],44)
    new01=new01[,-c(2,3)]
    new_01=data.frame(Year,new01)
    listofdfs[[i]] <-  new_01
  }
data_sui=rbind(listofdfs[[1]], listofdfs[[2]],listofdfs[[3]],listofdfs[[4]],
listofdfs[[5]],listofdfs[[6]],
listofdfs[[7]],listofdfs[[8]],listofdfs[[9]],listofdfs[[10]],listofdfs[[11]],
listofdfs[[12]],listofdfs[[13]],listofdfs[[14]],listofdfs[[15]],
listofdfs[[16]],listofdfs[[17]],listofdfs[[18]])
#having some fun with the data to check if some counties are correlted
plot(data_sui$County,data_sui$freq)
ada=data.frame(sui[sui$County ==  "Ada", ])
adams=data.frame(data_sui[data_sui$County ==  "Adams", ])
plot(ada$freq,adams$freq)
canyon=data.frame(data_sui[data_sui$County ==  "Canyon", ])
plot(ada$freq,canyon$freq, main="Plot of ada and canyon")
elmore=data.frame(data_sui[data_sui$County ==  "Elmore", ])
plot(ada$freq,elmore$freq, main="Plot of ada and elmore")
boise=data.frame(data_sui[data_sui$County ==  "Boise", ])
plot(ada$freq,boise$freq, main="Plot of ada and boise")
gem=data.frame(data_sui[data_sui$County ==  "Gem", ])
plot(ada$freq,gem$freq, main="Plot of ada and boise")
owyhee=data.frame(data_sui[data_sui$County ==  "Owyhee", ])
plot(ada$freq,owyhee$freq, main="Plot of ada and owyhee")
#Testing Autocorrelation
```

```
a_year <- read.csv("specif.csv",header = TRUE)
year=seq(1999,2016,1)
unadje <- read.csv("unadje.csv",header = TRUE)
hello=unadje[,-1]
colnames(hello)=year
ma=t(hello)
lambdas=a_year[1:44,1]
colnames(ma)=lambdas
for (i in 1:44) {
  mab=ma[,i]
  acf(mab,lag.max = 20,main=paste0(lambdas[i], ": ACF Plot" ))
}
##################finding the sum of population
new99=aggregate(value~county, data = suci99, sum)
new00=aggregate(value~county, data = suci00, sum)
new01=aggregate(value~county, data = suci01, sum)
new02=aggregate(value~county, data = suci02, sum)
new03=aggregate(value~county, data = suci03, sum)
new04=aggregate(value~county, data = suci04, sum)
new05=aggregate(value~county, data = suci05, sum)
new06=aggregate(value~county, data = suci06, sum)
new07=aggregate(value~county, data = suci07, sum)
new08=aggregate(value~county, data = suci08, sum)
new09=aggregate(value~county, data = suci09, sum)
new10=aggregate(value~county, data = suci10, sum)
new11=aggregate(value~county, data = suci11, sum)
new12=aggregate(value~county, data = suci12, sum)
```

```
new13=aggregate(value~county, data = suci13, sum)
new14=aggregate(value~county, data = suci14, sum)
new15=aggregate(value~county, data = suci15, sum)
new16=aggregate(value~county, data = suci16, sum)
pop=rbind(new99,new00,new01,new02,new03,new04,
new05,new06,new07,new08,new09, new10,new11,new12,new13,new14,new15,new16)
colnames(pop)=c("county","pop")
nasd=data.frame(sui_d,pop$pop)
rate_num=sum(nasd$freq)
rate_den=sum(nasd$pop.pop)
rate.new=rate_num/rate_den
nasd$exp=rate.new*nasd$pop.pop
masf=matrix(nasd$exp,nrow=44)
write.csv(masf,"expud.csv")
########calculating rates per year
su99=sui_d[sui_d$Year=="1999",]
dd=data.frame(su99,new99)
rate_num0=sum(dd$freq)
rate_den0=sum(dd$value)
rate.new0=rate_num0/rate_den0
dd$exp=rate.new0*dd$value
su00=sui_d[sui_d$Year=="2000",]
dd1=data.frame(su00,new00)
rate_num1=sum(dd1$freq)
rate_den1=sum(dd1$value)
rate.new1=rate_num1/rate_den1
dd1$exp=rate.new1*dd$value
```

```
su01=sui_d[sui_d$Year=="2001",]

dd2=data.frame(su01,new01)

rate_num2=sum(dd2$freq)

rate_den2=sum(dd2$value)

rate.new2=rate_num2/rate_den2

dd2$exp=rate.new2*dd2$value

su02=sui_d[sui_d$Year=="2002",]

dd02=data.frame(su02,new02)

rate_num02=sum(dd02$freq)

rate_den02=sum(dd02$value)

rate.new02=rate_num02/rate_den02

dd02$exp=rate.new02*dd02$value

su03=sui_d[sui_d$Year=="2003",]

dd03=data.frame(su03,new03)

rate_num03=sum(dd03$freq)

rate_den03=sum(dd03$value)

rate.new03=rate_num03/rate_den03

dd03$exp=rate.new03*dd03$value

su04=sui_d[sui_d$Year=="2004",]

dd04=data.frame(su04,new04)

rate_num04=sum(dd04$freq)

rate_den04=sum(dd04$value)

rate.new04=rate_num04/rate_den04

dd04$exp=rate.new04*dd04$value

su05=sui_d[sui_d$Year=="2005",]

dd05=data.frame(su05,new05)

rate_num05=sum(dd05$freq)
```

```
rate_den05=sum(dd05$value)
rate.new05=rate_num05/rate_den05
dd05$exp=rate.new05*dd05$value
su06=sui_d[sui_d$Year=="2006",]
dd06=data.frame(su06,new06)
rate_num06=sum(dd06$freq)
rate_den06=sum(dd06$value)
rate.new06=rate_num06/rate_den06
dd06$exp=rate.new06*dd06$value
su07=sui_d[sui_d$Year=="2007",]
dd07=data.frame(su07,new07)
rate_num07=sum(dd07$freq)
rate_den07=sum(dd07$value)
rate.new07=rate_num07/rate_den07
dd07$exp=rate.new07*dd07$value
su08=sui_d[sui_d$Year=="2008",]
dd08=data.frame(su08,new08)
rate_num08=sum(dd08$freq)
rate_den08=sum(dd08$value)
rate.new08=rate_num08/rate_den08
dd08$exp=rate.new08*dd08$value
su09=sui_d[sui_d$Year=="2009",]
dd09=data.frame(su09,new09)
rate_num09=sum(dd09$freq)
rate_den09=sum(dd09$value)
rate.new09=rate_num09/rate_den09
dd09$exp=rate.new09*dd09$value
```

```
su10=sui_d[sui_d$Year=="2010",]

dd10=data.frame(su10,new10)

rate_num10=sum(dd10$freq)

rate_den10=sum(dd10$value)

rate.new10=rate_num10/rate_den10

dd10$exp=rate.new10*dd10$value

su11=sui_d[sui_d$Year=="2011",]

dd11=data.frame(su11,new11)

rate_num11=sum(dd11$freq)

rate_den11=sum(dd11$value)

rate.new11=rate_num11/rate_den11

dd11$exp=rate.new11*dd11$value

su12=sui_d[sui_d$Year=="2012",]

dd12=data.frame(su12,new12)

rate_num12=sum(dd12$freq)

rate_den12=sum(dd12$value)

rate.new12=rate_num12/rate_den12

dd12$exp=rate.new12*dd12$value

su13=sui_d[sui_d$Year=="2013",]

dd13=data.frame(su13,new13)

rate_num13=sum(dd13$freq)

rate_den13=sum(dd13$value)

rate.new13=rate_num13/rate_den13

dd13$exp=rate.new13*dd13$value

su14=sui_d[sui_d$Year=="2014",]

dd14=data.frame(su14,new14)

rate_num14=sum(dd14$freq)
```

```
rate_den14=sum(dd14$value)
rate.new14=rate_num14/rate_den14
dd14$exp=rate.new14*dd14$value
su15=sui_d[sui_d$Year=="2015",]
dd15=data.frame(su15,new15)
rate_num15=sum(dd15$freq)
rate_den15=sum(dd15$value)
rate.new15=rate_num15/rate_den15
dd15$exp=rate.new15*dd15$value
su16=sui_d[sui_d$Year=="2016",]
dd16=data.frame(su16,new16)
rate_num16=sum(dd16$freq)
rate_den16=sum(dd16$value)
rate.new16=rate_num16/rate_den16
dd16$exp=rate.new16*dd16$value
expected=data.frame(dd$exp,dd1$exp,dd2$exp,dd02$exp,dd03$exp,
dd04$exp,dd05$exp,dd06$exp,dd07$exp,dd08$exp,dd09$exp,dd10$exp,
dd11$exp,dd12$exp,dd13$exp,dd14$exp,dd15$exp,dd16$exp)
msh=as.matrix(expected)
#Age Adjusted
suid_final$freq=rep(1,4654)
attach(suid_final)
new.data  = data.frame(Year,County,AgeGroups,freq) #counts of cases in each age
group
n_su=aggregate(freq~County+AgeGroups+Year,data=new.data,sum)
s99=n_su[n_su$Year==1999,]
df = merge(s99,ne99,all=TRUE)
```

```
s00=n_su[n_su$Year==2000,]

s01=n_su[n_su$Year==2001,]

s02=n_su[n_su$Year==2002,]

s03=n_su[n_su$Year==2003,]

s04=n_su[n_su$Year==2004,]

s05=n_su[n_su$Year==2005,]

s06=n_su[n_su$Year==2006,]

s07=n_su[n_su$Year==2007,]

s08=n_su[n_su$Year==2008,]

s09=n_su[n_su$Year==2009,]

s10=n_su[n_su$Year==2010,]

s11=n_su[n_su$Year==2011,]

s12=n_su[n_su$Year==2012,]

s13=n_su[n_su$Year==2013,]

s14=n_su[n_su$Year==2014,]

s15=n_su[n_su$Year==2015,]

s16=n_su[n_su$Year==2016,]

#1999

ne99=aggregate(value~county+Age.groups, data = suci99, sum)

colnames(ne99)=c("County","AgeGroups","pop")

ne99[ne99$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"

ne99[ne99$AgeGroups=="80-89" ,"AgeGroups"] ="80+"

mmd=merge(s99,ne99,by=c("County","AgeGroups"),all.y=TRUE)

mmd[is.na(mmd)] <- 0

#Calculating Rate Yearly

lab=labs[-1]

rate=c()
```

```
for (i in lab) {
  num=sum(mmd[which(mmd[,2]==i),4])
  den=sum(mmd[which(mmd[,2]==i),5])
  rate[i]=num/den
}
rate=rep(rate,44)
bob=data.frame(mmd,rate)
bob$exp=bob$rate*bob$pop
#2000
ne00=aggregate(value~county+Age.groups, data = suci00, sum)
colnames(ne00)=c("County","AgeGroups","pop")
ne00[ne00$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne00[ne00$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd1=merge(s00,ne00,by=c("County","AgeGroups"),all.y=TRUE)
mmd1[is.na(mmd1)] <- 0
rate1=c()
for (i in lab) {
  num=sum(mmd1[which(mmd1[,2]==i),4])
  den=sum(mmd1[which(mmd1[,2]==i),5])
  rate1[i]=num/den
}
rate00=rep(rate1,44)
bob1=data.frame(mmd1,rate00)
bob1$exp=bob1$rate00*bob1$pop
#2001
ne01=aggregate(value~county+Age.groups, data = suci01, sum)
colnames(ne01)=c("County","AgeGroups","pop")
```

```
ne01[ne01$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne01[ne01$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd2=merge(s01,ne01,by=c("County","AgeGroups"),all.y=TRUE)
mmd2[is.na(mmd2)] <- 0
rate01=c()
for (i in lab) {
  num=sum(mmd2[which(mmd2[,2]==i),4])
  den=sum(mmd2[which(mmd2[,2]==i),5])
  rate01[i]=num/den
}
rate1=rep(rate01,44)
bob01=data.frame(mmd2,rate1)
bob01$exp=bob01$rate1*bob01$pop
#2002
ne02=aggregate(value~county+Age.groups, data = suci02, sum)
colnames(ne02)=c("County","AgeGroups","pop")
ne02[ne02$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne02[ne02$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd3=merge(s02,ne02,by=c("County","AgeGroups"),all.y=TRUE)
mmd3[is.na(mmd3)] <- 0
rate02=c()
for (i in lab) {
  num=sum(mmd3[which(mmd3[,2]==i),4])
  den=sum(mmd3[which(mmd3[,2]==i),5])
  rate02[i]=num/den
}
rate2=rep(rate02,44)
```

```
bob02=data.frame(mmd3,rate2)
bob02$exp=bob02$rate2*bob02$pop
#2003
ne03=aggregate(value~county+Age.groups, data = suci03, sum)
colnames(ne03)=c("County","AgeGroups","pop")
ne03[ne03$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne03[ne03$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd4=merge(s03,ne03,by=c("County","AgeGroups"),all.y=TRUE)
mmd4[is.na(mmd4)] <- 0
rate03=c()
for (i in lab) {
  num=sum(mmd4[which(mmd4[,2]==i),4])
  den=sum(mmd4[which(mmd4[,2]==i),5])
  rate03[i]=num/den
}
rate03=rep(rate03,44)
bob03=data.frame(mmd4,rate03)
bob03$exp=bob03$rate03*bob03$pop
#2004
ne04=aggregate(value~county+Age.groups, data = suci04, sum)
colnames(ne04)=c("County","AgeGroups","pop")
ne04[ne04$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne04[ne04$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd5=merge(s04,ne04,by=c("County","AgeGroups"),all.y=TRUE)
mmd5[is.na(mmd5)] <- 0
rate04=c()
for (i in lab) {
```

```
  num=sum(mmd5[which(mmd5[,2]==i),4])
  den=sum(mmd5[which(mmd5[,2]==i),5])
  rate04[i]=num/den
}
rate04=rep(rate04,44)
bob04=data.frame(mmd5,rate04)
bob04$exp=bob04$rate04*bob04$pop
#2005
ne05=aggregate(value˜county+Age.groups, data = suci05, sum)
colnames(ne05)=c("County","AgeGroups","pop")
ne05[ne05$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne05[ne05$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd6=merge(s05,ne05,by=c("County","AgeGroups"),all.y=TRUE)
mmd6[is.na(mmd6)] <- 0
rate05=c()
for (i in lab) {
  num=sum(mmd6[which(mmd6[,2]==i),4])
  den=sum(mmd6[which(mmd6[,2]==i),5])
  rate05[i]=num/den
}
rate05=rep(rate05,44)
bob05=data.frame(mmd6,rate05)
bob05$exp=bob05$rate05*bob05$pop
#2006
ne06=aggregate(value˜county+Age.groups, data = suci06, sum)
colnames(ne06)=c("County","AgeGroups","pop")
ne06[ne06$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
```

```
ne06[ne06$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd7=merge(s06,ne06,by=c("County","AgeGroups"),all.y=TRUE)
mmd7[is.na(mmd7)] <- 0
rate06=c()
for (i in lab) {
  num=sum(mmd7[which(mmd7[,2]==i),4])
  den=sum(mmd7[which(mmd7[,2]==i),5])
  rate06[i]=num/den
}
rate06=rep(rate06,44)
bob06=data.frame(mmd7,rate06)
bob06$exp=bob06$rate06*bob06$pop
#2007
ne07=aggregate(value~county+Age.groups, data = suci07, sum)
colnames(ne07)=c("County","AgeGroups","pop")
ne07[ne07$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne07[ne07$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd07=merge(s07,ne07,by=c("County","AgeGroups"),all.y=TRUE)
mmd07[is.na(mmd07)] <- 0
rate07=c()
for (i in lab) {
  num=sum(mmd07[which(mmd07[,2]==i),4])
  den=sum(mmd07[which(mmd07[,2]==i),5])
  rate07[i]=num/den
}
rate07=rep(rate07,44)
bob07=data.frame(mmd07,rate07)
```

```
bob07$exp=bob07$rate07*bob07$pop
#2008
ne08=aggregate(value~county+Age.groups, data = suci08, sum)
colnames(ne08)=c("County","AgeGroups","pop")
ne08[ne08$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne08[ne08$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd8=merge(s08,ne08,by=c("County","AgeGroups"),all.y=TRUE)
mmd8[is.na(mmd8)] <- 0
rate08=c()
for (i in lab) {
  num=sum(mmd8[which(mmd8[,2]==i),4])
  den=sum(mmd8[which(mmd8[,2]==i),5])
  rate08[i]=num/den
}
rate08=rep(rate08,44)
bob08=data.frame(mmd8,rate08)
bob08$exp=bob08$rate08*bob08$pop
#2009
ne09=aggregate(value~county+Age.groups, data = suci09, sum)
colnames(ne09)=c("County","AgeGroups","pop")
ne09[ne09$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne09[ne09$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd9=merge(s09,ne09,by=c("County","AgeGroups"),all.y=TRUE)
mmd9[is.na(mmd9)] <- 0
rate09=c()
for (i in lab) {
  num=sum(mmd9[which(mmd9[,2]==i),4])
```

```
  den=sum(mmd9[which(mmd9[,2]==i),5])
  rate09[i]=num/den
}
rate09=rep(rate09,44)
bob09=data.frame(mmd9,rate09)
bob09$exp=bob09$rate09*bob09$pop
#2010
ne10=aggregate(value~county+Age.groups, data = suci10, sum)
colnames(ne10)=c("County","AgeGroups","pop")
ne10[ne10$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne10[ne10$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd10=merge(s10,ne10,by=c("County","AgeGroups"),all.y=TRUE)
mmd10[is.na(mmd10)] <- 0
rate10=c()
for (i in lab) {
  num=sum(mmd10[which(mmd10[,2]==i),4])
  den=sum(mmd10[which(mmd10[,2]==i),5])
  rate10[i]=num/den
}
rate10=rep(rate10,44)
bob10=data.frame(mmd10,rate10)
bob10$exp=bob10$rate10*bob10$pop
#2011
ne11=aggregate(value~county+Age.groups, data = suci11, sum)
colnames(ne11)=c("County","AgeGroups","pop")
ne11[ne11$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne11[ne11$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
```

```
mmd11=merge(s11,ne11,by=c("County","AgeGroups"),all.y=TRUE)
mmd11[is.na(mmd11)] <- 0
rate11=c()
for (i in lab) {
  num=sum(mmd11[which(mmd11[,2]==i),4])
  den=sum(mmd11[which(mmd11[,2]==i),5])
  rate11[i]=num/den
}
rate11=rep(rate11,44)
bob11=data.frame(mmd11,rate11)
bob11$exp=bob11$rate11*bob11$pop
#2012
ne12=aggregate(value~county+Age.groups, data = suci12, sum)
ne12=aggregate(value~county+Age.groups, data = suci12, sum)
colnames(ne12)=c("County","AgeGroups","pop")
ne12[ne12$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne12[ne12$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd12=merge(s12,ne12,by=c("County","AgeGroups"),all.y=TRUE)
mmd12[is.na(mmd12)] <- 0
rate12=c()
for (i in lab) {
  num=sum(mmd12[which(mmd12[,2]==i),4])
  den=sum(mmd12[which(mmd12[,2]==i),5])
  rate12[i]=num/den
}
rate12=rep(rate12,44)
bob12=data.frame(mmd12,rate12)
```

```
bob12$exp=bob12$rate12*bob12$pop
#2013
ne13=aggregate(value~county+Age.groups, data = suci13, sum)
colnames(ne13)=c("County","AgeGroups","pop")
ne13[ne13$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne13[ne13$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd13=merge(s13,ne13,by=c("County","AgeGroups"),all.y=TRUE)
mmd13[is.na(mmd13)] <- 0
rate13=c()
for (i in lab) {
  num=sum(mmd13[which(mmd13[,2]==i),4])
  den=sum(mmd13[which(mmd13[,2]==i),5])
  rate13[i]=num/den
}
rate13=rep(rate13,44)
bob13=data.frame(mmd13,rate13)
bob13$exp=bob13$rate13*bob13$pop
#2014
ne14=aggregate(value~county+Age.groups, data = suci14, sum)
colnames(ne14)=c("County","AgeGroups","pop")
ne14[ne14$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne14[ne14$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd14=merge(s14,ne14,by=c("County","AgeGroups"),all.y=TRUE)
mmd14[is.na(mmd14)] <- 0
rate14=c()
for (i in lab) {
  num=sum(mmd14[which(mmd14[,2]==i),4])
```

```
  den=sum(mmd14[which(mmd14[,2]==i),5])
  rate14[i]=num/den
}
rate14=rep(rate14,44)
bob14=data.frame(mmd14,rate14)
bob14$exp=bob14$rate14*bob14$pop
#2015
ne15=aggregate(value~county+Age.groups, data = suci15, sum)
colnames(ne15)=c("County","AgeGroups","pop")
ne15[ne15$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne15[ne15$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
mmd15=merge(s15,ne15,by=c("County","AgeGroups"),all.y=TRUE)
mmd15[is.na(mmd15)] <- 0
rate15=c()
for (i in lab) {
  num=sum(mmd15[which(mmd15[,2]==i),4])
  den=sum(mmd15[which(mmd15[,2]==i),5])
  rate15[i]=num/den
}
rate15=rep(rate15,44)
bob15=data.frame(mmd15,rate15)
bob15$exp=bob15$rate15*bob15$pop
#2016
ne16=aggregate(value~county+Age.groups, data = suci16, sum)
colnames(ne16)=c("County","AgeGroups","pop")
ne16[ne16$AgeGroups=="10--19" ,"AgeGroups"]  ="10-19"
ne16[ne16$AgeGroups=="80-89" ,"AgeGroups"]  ="80+"
```

```
mmd16=merge(s16,ne16,by=c("County","AgeGroups"),all.y=TRUE)
mmd16[is.na(mmd16)] <- 0
rate16=c()
for (i in lab) {
  num=sum(mmd16[which(mmd16[,2]==i),4])
  den=sum(mmd16[which(mmd16[,2]==i),5])
  rate16[i]=num/den
}
rate16=rep(rate16,44)
bob16=data.frame(mmd16,rate16)
bob16$exp=bob16$rate16*bob16$pop
ad99=aggregate(exp~County, data = bob, sum)
ad00=aggregate(exp~County, data = bob1, sum)
ad01=aggregate(exp~County, data = bob01, sum)
ad02=aggregate(exp~County, data = bob02, sum)
ad03=aggregate(exp~County, data = bob03, sum)
ad04=aggregate(exp~County, data = bob04, sum)
ad05=aggregate(exp~County, data = bob05, sum)
ad06=aggregate(exp~County, data = bob06, sum)
ad07=aggregate(exp~County, data = bob07, sum)
ad08=aggregate(exp~County, data = bob08, sum)
ad09=aggregate(exp~County, data = bob09, sum)
ad10=aggregate(exp~County, data = bob10, sum)
ad11=aggregate(exp~County, data = bob11, sum)
ad12=aggregate(exp~County, data = bob12, sum)
ad13=aggregate(exp~County, data = bob13, sum)
ad14=aggregate(exp~County, data = bob14, sum)
```

```
ad15=aggregate(exp~County, data = bob15, sum)
ad16=aggregate(exp~County, data = bob16, sum)
adjted=rbind(ad99,ad00,ad01,ad02,ad03,ad04,
ad05,ad06,ad07,ad08,ad09,ad10,ad11,
ad12,ad13,ad14,ad15,ad16)
adjustedage=matrix(adjted[,2],nrow=44)
#Jags Model
model{
for (i in 1:N)
{
for (t in 1:n)
{
y[i,t]~dpois(mu[i,t])
log(mu[i,t])=log(E[i,t])+alpha+g[i,t]
lambda[i,t]=exp(alpha+g[i,t])
}
}
for(i in 1:N){
g[i,1:n]~dmnorm(mu0[1:n],tau[1:n,1:n])
}
for (t in 1:n){
for (k in 1:n){
sigma[t,k]=theta1*exp(-theta2*abs(t1[t]-t1[k]))
}
}
tau=inverse(sigma)
#Priors
```

```
alpha~dnorm(0,0.0001)
theta1~dunif(0,100)
theta2~dunif(0,100)
}
#data-undajusted
mase=adje[,-1]
mas=unadje[,-1]
t1=seq(1,2,1)
t1=seq(1999,2016,1)
mu0=rep(0,18)
d <- list(y = mas, E = mase, N = nrow(mas),mu0=mu0,t1=t1,n=length(t1))
new_un <- jags.model(file = "cod.txt", d,n.chains = 3,n.adapt = 110000)
update(new_un,10000)
samp_un <- rjags::coda.samples(new_un, c( "lambda","mu"),  100000)
#adjusted
mase=adjust[,-1]
mas=unadje[,-1]
t1=seq(1,2,1)
t1=seq(1999,2016,1)
mu0=rep(0,18)
d <- list(y = mas, E = mase, N = nrow(mas),mu0=mu0,t1=t1,n=length(t1))
new_ad <- jags.model(file = "cod.txt", d,n.chains = 3,n.adapt = 110000)
update(new_ad,10000)
samp_ad <- rjags::coda.samples(new_ad, c( "lambda","mu"), thin=50, 100000)
result=summary(samp_ad)
nha_un1=data.frame(result[["statistics"]])
write.csv(nha_un1,"age.new1.csv")
```

```
nha1_un1=data.frame(result[["quantiles"]])
write.csv(nha1_un1,"age.new2.csv")
pdf()
gelman.plot(samp_ad)
dev.off()
pdf()
plot(samp_ad)
dev.off()
#hierarchical Bayes plots
a_year <- read.csv("new_final.csv",header = TRUE)
 year=seq(1999,2016,1)
# meh=data.frame(a_year[a_year$County=="Ada",],year)
 a=1:18
 lambdas=a_year[1:44,1]
 for(lambdas in a_year$County[1:44] )
meh=data.frame(a_year[a_year$County==lambdas,],year)
 plot(a, meh$Relative.risk, ylim=range(c(0,3)), axes=FALSE,pch=19,xlab = "Year",ylab
= "Relative Risk", main=paste0(lambdas, ": Adjusted Age Plot" ),type = "o",col="red")
 abline(h=1)
 lines(meh$Lower,type="o",col="green")
 lines(meh$Upper,type="o",col="blue")
axis(2)
axis(1, at=seq_along(meh$Relative.risk),labels=year)
legend("topright",lty=c(1,1),col=c("blue","red","green")
,legend = c("Upper","Relative.risk","Lower"),cex=.8 )
  box()}
```