Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	3/29/2014 _____
Chang Liu	Date

Impact of Pneumococcal Conjugate Vaccination of Children on Hospitalizations Resulting from Invasive Pneumococcal Diseases in the United States

By

Chang Liu

MPH

Biostatistics and Bioinformatics

_____

Dr. Michael J. Haber

_____

Dr. Howard H. Chang

Impact of Pneumococcal Conjugate Vaccination of Children on Hospitalizations Resulting from Invasive Pneumococcal Diseases in the United States

By

Chang Liu

B.S. in Statistics
Virginia Polytechnic Institute and State University
2012

Thesis Committee Chair: Michael J. Haber, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics and Bioinformatics
2014

Abstract

Impact of Pneumococcal Conjugate Vaccination of Children on Hospitalizations Resulting from Invasive Pneumococcal Diseases in the United States

By Chang Liu

The incidence of invasive pneumococcal disease (IPD) declined after the introduction of a seven-valent pneumococcal conjugate vaccine (PCV7) in the United Sates in 2000. This vaccine was administrated to children less than 5 years of age. Other studies have shown that colder temperature and existing influenza virus increase the susceptibility to IPD. Using data from 10 states, we estimated the impact of PCV7 coverage in both the vaccinated and unvaccinated populations adjusting for influenza and temperature using linear regression, Poisson regression, negative binomial regression, and generalized additive models. Negative binomial regression outperformed the other models in terms of AIC and the differences between observed and fitted values. From the negative binomial models, we found significant reductions of 84% and 73% in IPD hospitalizations among the vaccinated population, children <2 years of age and children 2-4 years of age, respectively. Due to the herd immunity (indirect) effects, IPD hospitalizations significantly declined in the unvaccinated age groups. We found significant reductions of 46% in children 5-17 years of age, 54% in adults 18-39 years of age, 36% in adults 40-64 year of age, and 48% in adults ≥65 years of age. Our result also indicate a significant association with increases in temperature and decreases in IPD hospitalizations in age groups, 2-4 and ≥ 64. There was also significant association with increases in influenza case and increases in IPD cases in age groups 18-39, 40-64, and ≥ 64.

Impact of Pneumococcal Conjugate Vaccination of Children on Hospitalizations Resulting from Invasive Pneumococcal Diseases in the United States

By

Chang Liu

B.S. in Statistics
Virginia Polytechnic Institute and State University
2012

Thesis Committee Chair: Michael J. Haber, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Biostatistics and Bioinformatics
2014

Acknowledgements

Foremost, I would like to express my sincere gratitude to my thesis advisor, Dr. Michael J. Haber for his valuable advice, effort, and support on my thesis research. Also, I would like to thank my thesis reader Dr. Howard H. Chang for spending his valuable time on meeting and reading my thesis. I must also acknowledge all the faculty, staff members, and graduate students in the Department of Biostatistics and Bioinformatics for their great help in my study and life over the past two year. Lastly but not the least, I would like to extend my deepest appreciation to my family for their unconditional love and support.

Table of Contents

List of Figures

List of Tables

Introduction

Definition, Incidence and Symptoms of Invasive Pneumococcal Disease

Invasive pneumococcal disease (IPD) is a result of infection caused by the bacterium *Streptococcus pneumonia* inside of a major organ or blood (Atkinson, 2006). It is estimated that more than 4,400 deaths from IPD occurred in 2007 in the United States (Wolfe 2012). An estimated 175,000 hospitalizations from pneumococcal pneumonia occurs annually in the Unites States (Atkinson, 2006). In the United States, an estimated 15-30 cases per 100,000 population of IPD are observed (ABC report 2009). More than 43,500 cases (14.3 cases per 100,000) and more than 5000 (1.6 cases per 100,000) deaths of IPD were observed in 2009 in the United States. (ABC report 2009). *Streptococcus pneumonia* is a leading cause of serious illness in children worldwide. About 70% of observed IPD occur in children under 2 years of age. It is most prevalent among children under 5 years of age, elderly, and those who have a weak immune system (Robinson et al. 2001). In the United States, *Streptococcus pneumonia* causes about 17,000 cases of IPD annually in children less than 5 years of age (Atkinson, 2006).

Streptococcus pneumonia bacteria are lancet-shaped, gram-positive, facultative anaerobic organism (Atkinson, 2006). *Streptococcus pneumonia* has more than 80 serotypes and most of them do cause serious diseases (Bogaert, D et al. 2004). The top 10 serotypes account for about 62% of invasive diseases worldwide (Atkinson, 2006). Serotypes prevalence differs in patients of different age groups and geographic regions. In the United States, the 7 most prevalent serotypes isolated from blood among those younger than 6 years of age account for 80% of infection (Hausdorff, W. P., et al. 2000).

Streptococcus pneumonia is a human pathogen; it is transmitted from person to person via respiratory droplets (Bogaert, D et al. 2004). *Streptococcus pneumonia* infects in the upper respiratory tract and can be isolated from the nasopharynx (Bogaert, D et al. 2004). The clinical syndromes of IPD include pneumonia, bacteremia, and meningitis (Robinson et al. 2001). Patients might experience fever, shaking chills, pleuritic chest pain, productive cough, dyspnea, tachypnea, and hypoxia. Diagnosis of infection with *Streptococcus pneumonia* result from isolation of organism from blood or other normally sterile body sites (Atkinson, 2006). Use of antibiotics in treating invasive pneumococcal disease is common; however, 40% of cases are resisted to penicillin in the United States (Kaplan and Mason 1998).

Several conditions increase the risk of IPD, including decreased immune function from disease or drugs, functional or anatomic asplenia, chronic cardio pulmonary problems such as asthma, liver or renal disease, smoking cigarettes, and cerebrospinal fluid, or CDF leak. (Atkinson, 2006) It has been documented that respiratory viral infection, particularly with influenza virus increases the susceptibility to IPD, specifically, invasive pneumococcal pneumonia (Talbot TR et al. 2004; Ampofo K. et al. 2008; Jansen AG. Et al. 2008; Murdoch and Jennings 2008; Walter ND et al. 2010). In addition, cold weather increases people's indoor activities thus increases the transmission of respiratory diseases through person-to-person contact (Murdoch DR and Jennings LC 2009; Dowell SF et al. 2003; Watson et al. 2006)).

Prevnar® or PCV7 vaccine

In the United States, most IPD was caused by 7 serotypes which are 4,14, 6B, 9V, 18C, 19F, and 23F in the 2000 (Hausdorff, W. P. et al. 2002). In response to that, a seven-valent conjugate pneumococcal vaccine (PCV7) was introduced in the year of 2000 (Poehling, K. A. et

al. 2006). Data from the Active Bacterial Core surveillance (ABCs) suggested that incidence of IPD of children under 5 years of age decreased from about 99 cases per 100,000 during 1998-1999 to 21 cases per 100,000 in 2008. The seven serotypes in the vaccine along with serotype 6A, provided cross-protection, are responsible for the decrease in IPD incidences (ACIP 2000).

The vaccine was administrated to infants at ages 2 months, 4months, 6months, and 12-15 months and also administrated to older infants and toddlers as catch-up in a 0.5-ml dose (ACIP 2000). Not only did PCV7 directly reduced incidence in children, it also indirectly reduced incidences in adults through induction of herd immunity (Whitney CG. Et al. 2003; Poehling KA. Et al. 2006; Kyaw MH et al. 2006, Pilishvili T. ea al. 2010; Pelton S. et al. 2004)). Herd immunity refers to the idea that by vaccinating a population widely, unvaccinated persons benefit from the reduction in number of cases among vaccinated. As more people are protected from a pathogen, there are less means of transmission of a pathogen. As non-vaccinated population indirectly protected by the presence and proximately of the vaccinated population, the transmission stopped prior to the infection of susceptible population (Isaacman et al. 2007).

Disease caused by non-vaccine serotypes begin to increase around 2005 (Moore MR et al. 2008). In response to that a 13-valent pneumococcal conjugate vaccine (PCV13) replace PCV7 in the US in March 2010. Nevertheless, data on IPD hospitalizations vs PCV7 coverage still provide valuable information on herd immunity effects.

Statement of Research Question

Development of a model is important to estimate vaccine impact, cost benefit, and how to introduce PCV7 into communities. We believe that PCV7 reduces IPD hospitalizations in both vaccinated children and adults through herd immunity. In addition, influenza increases

the susceptibility of IPD and cold weather aids the transmission of influenza. We want to evaluate the herd immunity effect of the vaccinated children on the incidence of IPD hospitalizations through several potential models.

1. Model the risk ratio of IPD hospitalizations though multiple linear regression.
2. Estimate IPD hospitalizations through Poisson regression.
3. Estimate IPD hospitalizations through negative binomial regression.
4. Estimate IPD hospitalizations through generalized additive Poisson regression.
5. Finally, evaluate the fit of the above models by comparing the estimated and the observed cases of IPD and AIC.

Data and Methods

Data Source and Study Design

IPD and influenza data were collected from Health Care Utilization Project State Inpatient Databases (SID) from 1996 to 2006 for 10 states in the United States. The SID contains 100% samples of IDC-9-coded hospitalization data for 10 states (Arizona, Colorado, Iowa, Massachusetts, New Jersey, New York, Oregon, Utah, Washington, and Wisconsin). Since the SID data has no personal identifiers, this study is exempt from Institutional Review Board consideration. The PCV7 immunization data was collected from the National Immunization Survey (NIS) from the Centers of Disease Control and Prevention (CDC) to estimate the growing portion of the under-5-year old age cohort that has been immunized in each of the 10 states and for each year between 2000 to 2006. U.S. Census Bureau database was used to estimate the yearly U.S. age group and state specific populations. (Simonsen et al. 2011) Climate data was collected from the monthly National Climatic Data Center from the National Oceanic and Atmospheric Administration (NOAA). We performed all statistical analyses using SAS statistical software version 9.3 (SAS Insititude Inc., Cary, NC) and R software.

Study Population

The study population included individuals of all ages during the influenza season between 1996 to 2006 from the 10 study states. Each influenza season was defined as July of one year through June of the next year. Their ages were categorized into 6 groups, which were <2, 2-4, 5-17, 18-39, 40-64, and ≥ 65.

Outcome Variable

Case-finding relies on International Classification Diseases ICD 9th Revision code of the medical records. For each state, year, and age group combination, the number of hospitalizations of IPD was calculated. IPD was defined as any hospitalizations with an IDC9 code 320.1 (Pneumococcal meningitis) or 038.2 (Pneumococcal septicemia [Streptococcus pneumoniae septicemia]) or as codes 320.8 (Meningitis due to other specified bacteria), 790.7 (Bacteremia), or 038.9 (Unspecified septicemia) and 041.2 (Pneumococcus). Baseline rates before the introduction of PCV7 on the year of 2000 were defined as the average yearly rates during the 1996-1997 through 1998-1999 flu seasons (Simonsen et al. 2011).

Primary Explanatory Variables

PCV7 vaccination coverage was estimated from the coverage rates by infant under 1 year of age for each year and then summed these to estimate the coverage rates for all children less than 5 years of age. All vaccinated children had received at least three or more doses, with the first occurring before 1 year of age, or one or more catch-up doses, with the first occurring at ≥1 years of age. (Simonsen et al. 2011).

Secondary Explanatory Variables

The number of cases of influenza were counted using ICD9 code 487 (Influenza) for a given state, year, and, age group combination. Temperature (in Fahrenheit) was calculated for each state and year by averaging the monthly temperature of all the stations (sites which are equipped with instrumentation for measuring meteorological information) within each state.

Epidemiology Measures

Number of IPD hospitalizations were converted into incidence rates, and rate ratios for modeling and graphical purposes.

Let y_{ijk} denote the IPD hospitalizations, i: year=0,1,…,7 (0 denotes baseline)

j: age group =1,…,6 k=1,…,10 states

Population sizes N_{ijk}

IPD hospitalization incidence rates: $R_{ijk} = \frac{y_{ijk}}{N_{ijk}} \times 100{,}000$

IPD hospitalization rate ratio: $Q_{ijk} = \frac{R_{ijk}}{R_{0jk}}$, i=1,…,7

Aggregate rates over states: $\frac{\sum_k y_{ijk}}{N_{ijk}} \times 100{,}000$

Log Linear Regression Modeling

Log linear regression models were constructed to assess the effect of the vaccinated children less than 5 years of age on the rate ratio of IPD hospitalization incidence per 100,000 for the years of 1999-2000 through 2005-2006 flu seasons for each vaccinated and unvaccinated age groups (<2, 2-4, 5-17, 18-39, 40-64, and ≥65) for a total of 6 models. The expected number of IPD hospitalizations were also calculated from the fit equations for each year and age group combination.

Model (1) for each unvaccinated age group (j):

$$\log(Q_{ijk}) = \beta_{0j} + \beta_1 PCV7_{ik} + \beta_3 Influenza_{ijk} + \beta_4 Temperature_{ik} + e_{ijk}$$

PCV7 is the vaccine coverage rate of children under five years of age. Influenza is the influenza incidence rate per 100,000 observed. Temperature is the average temperature for a

given state in a given flu season. For each age group the expected rate ratio for each increasing 10% increments in PCV7 coverage rates was calculated to evaluate the herd immunity effect. To assess the goodness of fit, histograms and normal probability plots were generated to evaluate the normal assumption. Leverage plots and Cook's D plots were generated to identify extreme outliers and influential points. Rstudent vs. predicted plots and residual vs. predicted plots were used to assess residual homoscedasticity and independence.

Poisson Regression Modeling

Poisson regression models were constructed to estimate impact of PCV7 vaccination of children less than 5 years of age on hospitalization incidence of IPD. The model used data from the 1999-2000 through 2005-2006 flu seasons. We constructed 6 models representing each age group (<2, 2-4, 5-17, 18-39, 40-64, and ≥65). For each age group the expected rate ratio for each increasing 10% increments in PCV7 coverage rates was calculated to evaluate the herd immunity effect. The expected number of IPD hospitalizations were also calculated from the fit equations for each year and age group combination. Scaled deviance divided by the degree of freedoms was used to detect any over-dispersion issues, any values significant differs than 1 signals over-dispersion. AIC was calculated to assess the goodness of fit, and to use as measures for model comparison purpose.

Model (2) for each age group (j):

$$\log(Y_{ijk}) = \log(N_{ijk}) + \beta_{0j} + \beta_1 PCV7_{ik} + \beta_2 Baseline_{ijk} + \beta_3 Influenza_{ijk} + \beta_4 Temperature_{ik}$$

Where for year i, and state k, Y_{ijk} is the count of IPD hospitalizations and N_{ijk} is the population offset. β_{0j} is the intercept, baseline is the average IPD hospitalizations incidence per 100,000 of the baseline years, PCV7 is the coverage rate of PCV7 vaccination in children ≤ 5 years of age, influenza is the influenza incidence rate per 100,000 observed, and temperature is the average temperature for a given state in a given flu season.

Negative Binomial Regression Modeling

Because of possible over-dispersion issues, negative binomial regression models were constructed to estimate impact of vaccinating children < 5 years of age on hospitalization incidence of IPD. Again, for each age group the expected rate ratio for each increasing 10% increments in PCV7 coverage rates was calculated to evaluate the herd immunity effect. The expected number of IPD hospitalizations were also calculated from the fit equations for each year and age group combination. Scaled deviance divided by the degree of freedoms and AIC were also calculated.

Model (3) for each age group (j):

$$\log(Y_{ijk}) = \log(N_{ijk}) + \beta_{0j} + \beta_1 PCV7_{ik} + \beta_2 Baseline_{ijk} + \beta_3 Influenza_{ijk} + \beta_4 Temperature_{ik}$$

Same model coefficients were used. However, Y follows a negative binomial distribution instead of a Poisson distribution. Negative binomial regression is a discrete probability distribution for the number of successes (no IPD hospitalizations) in a sequence of Bernoulli trials before a specified number of failures (IPD hospitalizations) occur.

Generalized Additive Modeling

It seems reasonable that the effect of a vaccine tappers off after a significant percentage of the population are vaccinated, so a generalized additive poison regression was used to model the nonlinearity of PCV7 coverage, influenza, and temperature. These models were performed using the gam package in R software.

Model (4) for each age group (j):

$$\log(Y_{ijk}) = \log(N_{ijk}) + \beta_0 + f_1 PCV7_{ik} + \beta_2 Baseline_{ijk} + f_4 Influenza_{ijk} + f_4 Temperature_{ik}$$

Where$f_1 PCV7_{ik}$, $f_4 Influenza_{ijk}$, $f_4 Temperature_{ik}$ denote non-linear smooth relationships between the response and the predictors.

Again, the expected number of IPD hospitalizations were calculated from the fit equations for each year and age group combination. AIC was calculated to assess the goodness of fit, and to use as measures for model comparison purpose.

Assessing the fit of the models

Two methods were used to compare the goodness of fit for Poisson, negative binomial, and generalized additive Poisson models. First was the Akaike information criterion (AIC), which provides penalties for adding predictors. AIC is useful for comparing the relative fit of two parametric generalized linear models to the same set of data, where the scale of the response variable (Y) is the same in each model. This is why we cannot use AIC to compare linear

regression models with the other 3 models, because the response variable in linear regression are rate ratios. In general, a smaller AIC yields a better model, and a differences of 6 is considered to be strong evidence of differing model fit. (Chang 2014).

Another way to assess the goodness of fit was to calculate the expected number of IPD hospitalizations in a given age group and year from the model coefficients. Then calculate the Chi-squared test statistics using the following equation:

$$D = \frac{(observed\ cases - expected\ cases)^2}{expected\ cases}$$

D is approximately Chi-squared with one degree of freedom. This means that if the model fits well then about 95% of these ratios are expected to be less than 4. (Agresti 1990)

Lastly, we constructed the negative binomial regression model and generalized additive model using 1999-2000 through 2004-2005 flu seasons to predict the number of IPD hospitalizations for the 2005-2006 flu season. Again, we used the chi-squared test statistics to calculated D.

Result

Descriptive Statistics

A total of 720 IPD hospitalizations records were included in the study, where each recorded corresponded to a state-by-season combination. There were only 7 missing observations due to missing in influenza information (0.97%). Since there was no pattern in the missingness, we can conclude that the missingness was random and was not informative. IPD hospitalization incidence rates had decreased in all age groups and across all of the 10 study states from baseline years to flu season 2005-2006 (Table 1, Figure1). On average, IPD hospitalization incidence rate decreased from about 9 per 100000 in baseline to 5 per 100000 in 2005-2006 flu season. Children under 2 years of age and adults above 65 years of age had the highest IPD hospitalization incidence rates across all studied flu seasons compared to the other age groups (Figure 1). Children between 5 to 17 years of age had the lowest IPD hospitalization incidence rates. Iowa had the highest IPD hospitalization incidence rates and Oregon had the lowest throughout the study periods (Figure 2). In terms of IPD hospitalization rate ratio, it decreased to an average of 0.5 at the end of the study period for all age groups and states, with largest decreases in age groups <2 and 2-4 (Figure 3).

Over the entire study period, Arizona had the highest average temperature and Colorado had the lowest temperature. Adults above 65 years of age and children below 2 years of age had higher influenza incidence rates, and children between 5 to 17 years old had the lowest incidence of influenza (Figure 4).

PCV7 coverage rates had increased every year for the entire study period across all of the study states (Figure 5). Massachusetts had the highest coverage rate, and Colorado had the

lowest coverage rate. On average, PCV7 increased to about 83% across all studied states from vaccine initiation to the end of the study period. Again, PCV7 was only administrated to children less than 5 years of age.

Log Linear Regression Model

Result of the 6 linear regression models to assess the association of PCV7 vaccine coverage on logged IPD rate ratios for all age groups were summarized in Table 2. There was a significant association of increases in PCV7 vaccine coverage rate with reductions in IPD hospitalization rate ratios for the vaccinated children below 2 years of age and for the vaccinated children between 2 to 4 years of age (p-value<0.0001). Possibly due to the herd immunity effect, there was also a significant association of increases in PCV7 vaccine coverage rate with reductions in IPD hospitalization rate ratio for the other 4 unvaccinated age groups (p-value<0.005). The exponentiated model coefficients for the logged response in linear regression models are interpreted as changes in median IPD hospitalization rates. Using above interpretation, a 10% increase in PCV7 coverage was associated with a decrease of median IPD hospitalization rate by 0.93 in children between 5 to 17 years of age. Adults between 18 to 39 years of age experienced a rate decrease of 0.91, adults between 40 to 65 experienced a rate decrease of 0.94, and adults ≥ 65 experienced a rate decrease of 0.93.

The magnitude of IPD rate ratio reduction was different across age groups. To evaluate the effect of herd immunity, expected rate ratios per 10% point increase in PCV7 coverage were calculated (Table 3). The rate ratio for children between 5 to 17 years of age decreased from 92% to 50%, for 18 to 39 years of age the rate ratio decreased from 90% to 43%, the rate ratio for adults between 40 to 64 years of age decreased from 94% to 58%, and the rate ratio for adults

≥ 65 years of age decreased from 93% to 52%. The fitted plots for the 6 models also show reductions in rate ratios (Figure 6).

The linear regression does provide a good fit for the data. The histograms and the normal probability plots of all of the 4 regression models show the errors are normally distributed. There are no severe outliers or influential points from the leverage plots and Cook's D plots. The points are randomly scattered in the residual vs. predicted plots and the rstudent vs. predicted plots which means that the errors are independent of each other and does not form a pattern.

Poisson Regression

Poisson Regression was used evaluate the association between PCV7 coverage and IPD hospitalizations. Results of the 6 multivariable Poisson regression models were summarized in Table 2. Consistent with the result of the linear regression model, there was a significant association of increases in PCV7 vaccine coverage rate with reductions in IPD hospitalizations for the vaccinated children below 2 years of age and for the vaccinated children between 2 to 4 years of age (p-value<0.0001). Possibly due to the herd immunity effect, there was also a significant association of increases in PCV7 vaccine coverage rate with reductions in IPD hospitalizations in all for the other 4 unvaccinated age groups (p-value<0.005).

For children <2 years of age, there was a significant association of increases in influenza cases and increases in IPD hospitalizations (p-value<0.05). For children 2 to 4 years of age, there was a significant association of increases in temperature and decreases in IPD hospitalizations (p-value=0.0366). For adults 18-39 and 40-64 years of age there was a significant association between increases in influenza cases and increases in IPD hospitalizations

(p-value<0.05). For adults ≥65, there were significant associations of increases in flu and increases in IPD hospitalizations and increases in temperature and decreases in IPD hospitalizations (p-values<0.05).

Expected rate ratios per 10% point increase in PCV7 coverage were calculated to evaluate the effectiveness of the PCV7 vaccine for all age groups (Table 3). Reduction was highest in the vaccinated children who were under <2 and 2 to 4 years of age. Those groups experienced a significant rate reduction in IPD hospitalizations, about 85% to 74% respectively. Furthermore, the unvaccinated age groups also experienced significant amount of reductions, about 52% to 38%.

Goodness of fit statistics for Poisson regression indicate over-dispersion. The scaled deviances for all 6 models are severely above 1 as shown in Table 4. The scaled deviance range from 1.28 (Age group: 5-17) to 3.51(Age group: ≥65). This suggests the possibility that another model might provide a better fit. AIC values were also produced in Table 5 to use as a comparison for the later models.

Negative binomial Regression

To correct the over-dispersion problem in the Poisson models, 6 negative binomial models were used for the data (Table 2). Again, consistent with the previous results, there was a significant association of increases in PCV7 vaccine coverage rate with reduction in IPD hospitalizations for the vaccinated children below 2 years of age and for the vaccinated children between 2 to 4 years of age (p-value<0.0001). Possibly due to the herd immunity effect there was also a significant association of increases in PCV7 vaccine coverage rate with reductions in IPD hospitalizations in all for the other 4 unvaccinated age groups (p-value<0.0001).

The coefficient estimates for the negative binomial models were very similar with the coefficients estimates of the Poisson regression models. The standard errors for the coefficients estimates were slightly larger than the ones generated from Poisson models. Rate ratios per 10% point increase in PCV7 coverage were calculated again. The associations of influenza and temperature were also similar to the ones from the Poisson models, except in age group <2. The flu variable became insignificant; this might be due to the slightly larger standard error. The rate reductions were extremely similar with the ones generated from the Poisson models for all age groups.

Goodness of fit statistics for negative binomial regression seems better. The scaled deviances for all 6 models are around 1 as shown in (Table 4), which suggest there is no over-dispersion problems. The scaled deviance range from 1.06 (Age group: 40-64) to 1.27 (Age group: 2-4). AIC values were also produced in table 5 to use as a comparison for the later models.

Generalized Additive Poisson Model

Non-linear effect of PCV7 coverage, influenza, and temperature were examined for each age group for a total of 6 models. As expected, consistent with the previous results, there was a significant association of increases in PCV7 vaccine coverage rate with reduction in IPD hospitalizations for the vaccinated children below 2 years of age and for the unvaccinated age groups (p-value<0.0001) (Figure 7).

Model estimates are summarized in Table 2. For children under 2 years of age, influenza and temperature didn't have significant association with IPD hospitalizations. For children between 2 to 4 years of age, there was a significant non-linear association of increases in

temperature with decreases in IPD hospitalizations (P-value<0.05); the effect seems to level off when temperature was over 60F (Figure 7). For individuals between 5 to 17 years of age, there seemed to be a significant non-linear association in temperature and IPD hospitalizations (P-value<0.05) (Figure 7). For adults between 18-39, 40-64, and ≥ 65 years of age, there was a significant association with increases in influenza with increases in IPD hospitalizations (P-value<0.05), however the association was linear in age groups 18-39 and 40-64 (Figure 7). For adults over 65 years of age, there was a significant non-linear association with increases in temperature and decreases in IPD hospitalizations (P-value<0.05) (Figure 7).

The expected rate ratios per 10% point increase in PCV7 coverage for all age groups were summarized in Table 3. The rates were different than the ones generated from Poisson and negative binomial models for age groups 2-4 and ≥ 65

Assessing the fit of the models

Table 5 shows the AIC values from Poisson, negative binomial, and generalized additive Poisson model. For age groups 2-4 and 5-17, generalized additive models were significantly better, however, for all other 4 age groups negative binomial models performed better. As Table 6 shows, 9 out of 42 (21.4%) from the linear regression, 5 values out of 42 (11.9%) from the Poisson models, 6 values out of 42 (14.3%) from negative binomial models, and 9 values out of 42 (21.4%) from generalized additive Poisson models exceed 4 when we assessed the values of D= (observed cases-expected cases)^2/expected cases for each combination of age group and year. From the above evidence, I believe negative binomial is the best fit for our data. First, the models do not have an over-dispersion problem. Second, the models produce smaller AIC values. Third, the models generate more accurate fitted values. Lastly, the interpretation of the

coefficients from the negative binomial models are easier than the additive models, and it uses less degrees of freedom.

The predicted hospitalizations of IPD for 2005-2006 flu season is shown in table 7. Negative binomial, and additive models were used. Poisson model was not used because of over-dispersion issues. Again expected and observed IPD hospitalizations were calculated. All models do not perform well. From negative binomial models 4 D values out of 6 exceeded 4 (67%), and from additive models 3 D values out of 6 (50%) exceeded 4.

Discussion

Conclusion

The objective of this study is to fit and evaluate models that could describe the effect of PCV7 vaccination children and to estimate the herd immunity benefits among unvaccinated children and adults. Four modeling methods were used in this study, log linear regression, Poisson, negative binomial, and generalized additive Poisson. Negative binomial adds a dispersion parameter to the model and thus is a robust alternative to the Poisson model. The over-dispersion parameter in the negative binomial contributes to the slightly larger stand errors for the coefficient estimates. In general, both negative binomial and generalized additive Poisson models outperformed Poisson models. Negative binomial models were associated with smaller AIC values and more accurate fitted values than generalized additive Poisson models. In addition, negative binomial model uses less degree of freedoms and the model coefficient are easier to interpret than additive models. The above evidences led us to believe that negative binomial model provides the best fit for our data.

Negative binomial model results suggest that by administrating PCV7 vaccine to children less than 5 years of age significantly decreased the incidence of IPD in older children and all age groups of adults. For the vaccinated children <2 and children 2 to 4 years old, we found reductions of 84% and 73%, respectively. For the unvaccinated population, we found a reduction of 46% for children 5-17 years old, a reduction of 51% for adults 18-39 years old, a reduction of 36% in adults 40-64 years old, and a reduction of 48% in adults ≥ 65. Our results are similar to the study conducted by Simonse et al. This study adds to the growing literature on the herd immunity effect of PCV7 vaccine.

Our models also suggest that for the elderly (≥65) during high influenza period and colder periods were significantly more susceptible to IPD. Children between 2-4 years of age were significantly more susceptible to IPD during colder periods. In addition, adults 18-64 years of age were significantly more susceptible to IPD, during periods of high influenza activity.

Limitation and Future Work

There are several limitations in our study. First, we used ICD-9 codes of hospitalized individuals for case identifications. Our IPD data might therefore represent an underestimate of the true burden of this condition. We also had to assume that physicians correctly diagnosed their patients. Second, our IPD-9 coded hospitalization data cannot separate vaccine-type reductions from non-vaccine-type disease. Therefore, we do not know if the IPD was caused by the vaccine strand or not. Third, there is variability within influenza seasons, and it might be more appropriate to perform separate analysis for each influenza season. Fourth, the temperature variable is the average temperature of each flu season for each state; there might not be enough data points to capture the true relationship between temperature and IPD. A model with monthly IPD data might be better suited for such analysis. Lastly, this is an ecological study, hence we cannot claim that we found a causal association between the increase in PCV7 coverage and the decrease in IPD hospitalizations.

Our data do not have personal demographic information of the patients such as gender, race, and socioeconomic status. It will be interesting to see which gender, race, or socioeconomic groups benefited most from PCV7 vaccination. Since we only have a few data points for each state, wer were unable to evaluate the effect of PCV7 vaccination on each states

by using state identifiers as factor variables on the models. Thus, we cannot determine if states experience uniform IPD rate reductions due to PCV7 vaccination.

In May 2012, a new PCV13 vaccine replaced PCV7 in the US, due to the rise of non-vaccine serotypes. This vaccine includes the seven PCV7 serotypes and six additional serotypes. A recent study conducted by Simonsen el at. found a significant reduction in hospital admission for IPD in children and adults following the introduction of the new vaccine. Continuing monitoring of the PCV7 effectiveness should be done in the future to detect any increases in IPD incidence rates due to possible changes in the prevalent strains of *S. Pneumonae.*

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

Ampofo, K., Bender, J., Sheng, X., Korgenski, K., Daly, J., Pavia, A. T., & Byington, C. L. (January 01, 2008). Seasonal invasive pneumococcal disease in children: role of preceding respiratory viral infection. *Pediatrics, 122,* 2, 229-37.

Atkinson, W., Public Health Foundation (U.S.), & Centers for Disease Control and Prevention (U.S.). (2006). *Epidemiology and prevention of vaccine-preventable diseases: The pink book*. Wahsington, D.C: Public Health Foundation.

Bogaert, D., De, G. R., & Hermans, P. W. (January 01, 2004). Streptococcus pneumoniae colonisation: the key to pneumococcal disease. *The Lancet Infectious Diseases, 4,* 3, 144-54.

Centers for Disease Control and Prevention. 2009. Active Bacterial Core (ABC) Surveillance Report, Emerging Infections Program Network, *Streptococcus pneumoniae*, 2010.

Chang Howard. (2014) *Lecture6: Bayesian Model Comparison I* Personal Collection of (Howard Chang), Emory University, Atlanta, GA

Dowell, Scott F., Whitney, Cynthia G., Wright, Carolyn, Rose, Charles E., & Schuchat, Anne. (2003). *Seasonal Patterns of Invasive Pneumococcal Disease*. Centers for Disease Control and Prevention.

Haber, M., Barskey, A., Baughman, W., Barker, L., Whitney, C. G., Shaw, K. M., Orenstein, W., ... Stephens, D. S. (July 20, 2007). Herd immunity and pneumococcal conjugate vaccine: A quantitative model. *Vaccine, 25,* 29, 5390-5398.

Hausdorff, W. P., Bryant, J., Paradiso, P. R., & Siber, G. R. (January 01, 2000). Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America, 30,* 1, 100-21.

Hausdorff, W. P., Yothers, G., Dagan, R., Kilpi, T., Pelton, S. I., Cohen, R., Jacobs, M. R., ... Bryant, J. (January 01, 2002). Multinational study of pneumococcal serotypes causing acute otitis media in children. *The Pediatric Infectious Disease Journal, 21,* 11, 1008-16.

JANSEN, A. G. S. C., SANDERS, E. A. M., VAN DER ENDE, A., VAN LOON, A. M., HOES, A. W., & HAK, E. (n.d.). *Invasive pneumococcal and meningococcal disease: association with influenza virus and respiratory syncytial virus activity?*. Cambridge University Press.

Isaacman, D. J., Fletcher, M. A., Fritzell, B., Ciuryla, V., & Schranz, J. (March 01, 2007). Indirect effects associated with widespread vaccination of infants with heptavalent pneumococcal conjugate vaccine (PCV7; Prevnar). *Vaccine, 25,* 13, 2420-2427.

Kaplan, S. L., & Mason, E. O. J. (January 01, 1998). Management of infections due to antibiotic-resistant Streptococcus pneumoniae. *Clinical Microbiology Reviews, 11,* 4, 628-44.

Kyaw, M. H., Lynfield, R., Schaffner, W., Craig, A. S., Hadler, J., Reingold, A., Thomas, A. R., ... Active Bacterial Core Surveillance of the Emerging Infections Program Network. (January 01, 2006). Effect of introduction of the pneumococcal conjugate vaccine on drug-resistant Streptococcus pneumoniae. *The New England Journal of Medicine, 354,* 14, 1455-63.

Moore, M. R., Gertz, R. E. J., Woodbury, R. L., Barkocy-Gallagher, G. A., Schaffner, W., Lexau, C., Gershman, K., ... Beall, B. (January 01, 2008). Population snapshot of emergent Streptococcus pneumoniae serotype 19A in the United States, 2005. *The Journal of Infectious Diseases, 197,* 7, 1016-27.

Murdoch, D. R., & Jennings, L. C. (January 01, 2009). Association of respiratory virus activity and environmental factors with the incidence of invasive pneumococcal disease.*Journal of Infection, 58,* 1, 37-46.

Pelton, S. I., Loughlin, A. M., & Marchant, C. D. (January 01, 2004). Seven valent pneumococcal conjugate vaccine immunization in two Boston communities: changes in serotypes and antimicrobial susceptibility among Streptococcus pneumoniae isolates.*The Pediatric Infectious Disease Journal, 23,* 11, 1015-22.

Pilishvili, T., Lexau, C., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., Reingold, A., ... Active Bacterial Core Surveillance/Emerging Infections Program Network. (January 01, 2010). Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *The Journal of Infectious Diseases, 201,* 1, 32-41.

Poehling, K. A., Talbot, T. R., Griffin, M. R., Craig, A. S., Whitney, C. G., Zell, E., Lexau, C. A., ... Schaffner, W. (January 01, 2006). Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. *Jama : the Journal of the American Medical Association, 295,* 14, 1668-74.

Preventing pneumococcal disease among infants and young children: recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR 2000;49(RR-9):1-35

Robinson, K. A., Baughman, W., Rothrock, G., Barrett, N. L., Pass, M., Lexau, C., Damaske, B., ... Active Bacterial Core Surveillance (ABCs)/Emerging Infections Program Network. (January 01, 2001). Epidemiology of invasive Streptococcus pneumoniae infections in the United States, 1995-1998: Opportunities for prevention in the conjugate vaccine era. *Jama : the Journal of the American Medical Association,285,* 13, 1729-35.

Simonsen, Lone, Taylor, Robert J., Young-Xu, Yinong, Haber, Michael, May, Larissa, & Klugman, Keith P. (2011). *Impact of Pneumococcal Conjugate Vaccination of Infants on Pneumonia and Influenza Hospitalization and Mortality in All Age Groups in the United States*. American Society of Microbiology.

Simonsen, L., Taylor, R. J., Schuck-Paim, C., Lustig, R., Haber, M., & Klugman, K. P. (March 01, 2014). Effect of 13-valent pneumococcal conjugate vaccine on admissions to hospital 2 years after its introduction in the USA: a time series analysis. *The Lancet Respiratory Medicine.*

Talbot, T. R., Poehling, K. A., Hartert, T. V., Arbogast, P. G., Halasa, N. B., Edwards, K. M., Schaffner, W., ... Griffin, M. R. (March 01, 2005). Seasonality of invasive pneumococcal disease: Temporal relation to documented influenza and respiratory syncytial viral circulation. *The American Journal of Medicine, 118,* 3, 285-291.

Walter, N. D., Taylor, T. H., Shay, D. K., Thompson, W. W., Brammer, L., Dowell, S. F., Moore, M. R., ... Active Bacterial Core Surveillance Team. (January 01, 2010). Influenza circulation and the burden of invasive pneumococcal pneumonia during a non-pandemic period in the United States. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America, 50,* 2, 175-83.

Watson, M., Gilmour, R., Menzies, R., Ferson, M., McIntyre, P., & New South Wales Pneumococcal Network. (January 01, 2006). The association of respiratory viruses, temperature, and other climatic parameters with the incidence of invasive pneumococcal disease in Sydney, Australia. *Clinical Infectious Diseases : an Official Publication of the Infectious Diseases Society of America, 42,* 2, 211-5.

Whitney, C. G., Farley, M. M., Hadler, J., Harrison, L. H., Bennett, N. M., Lynfield, R., Reingold, A., ... Active Bacterial Core Surveillance of the Emerging Infections Program Network. (January 01, 2003). Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *The New England Journal of Medicine, 348,* 18, 1737-46.

Wolfe, R. M. (January 01, 2012). Update on adult immunizations. *Journal of the American Board of Family Medicine : Jabfm, 25,* 4.)

Figures

Figure 1. IPD hospitalization Rate per 100,000 by age group

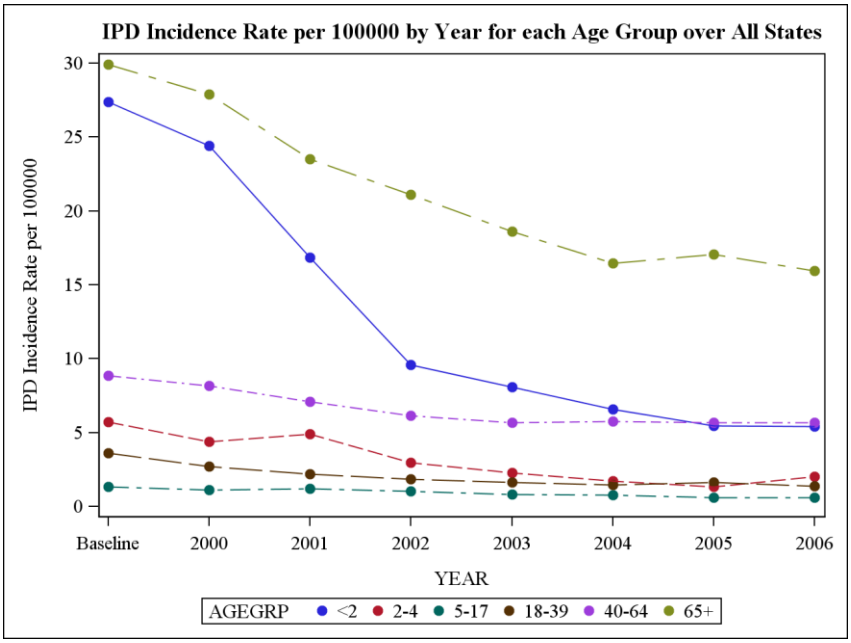


Figure 2. IPD hospitalization Rate per 100,000 by states

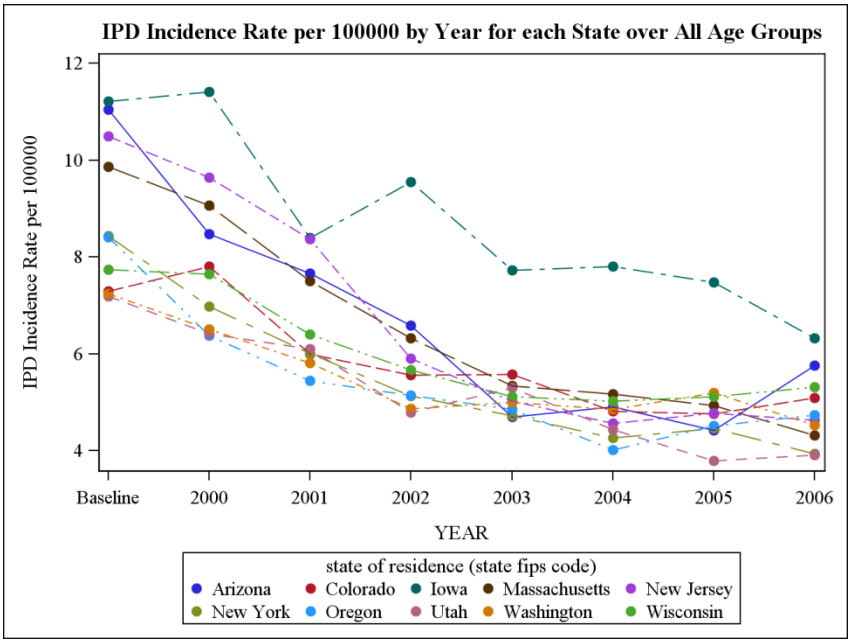


Figure 3. IPD hospitalization Rate Ratio by age groups

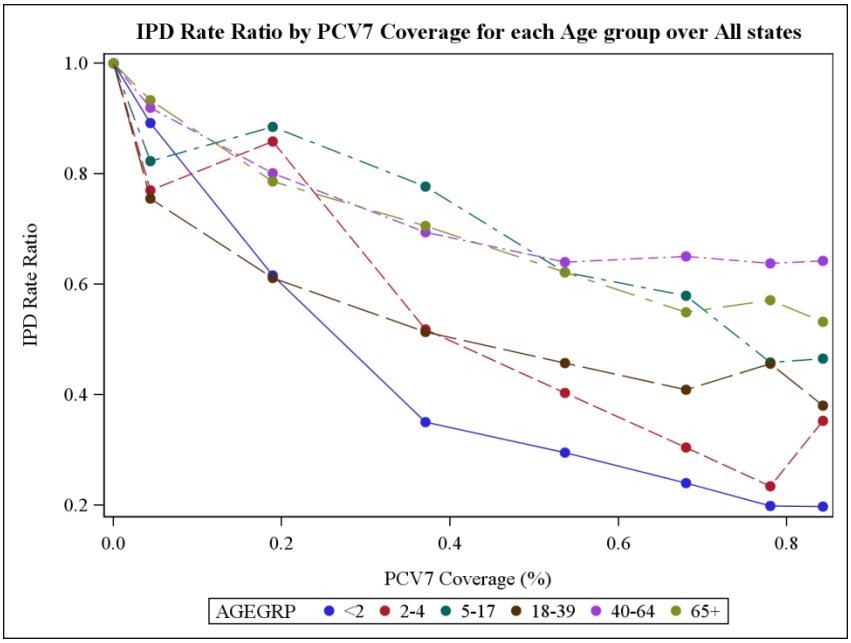


Figure 4. Influenza Incidence rate per 100,000 by age group

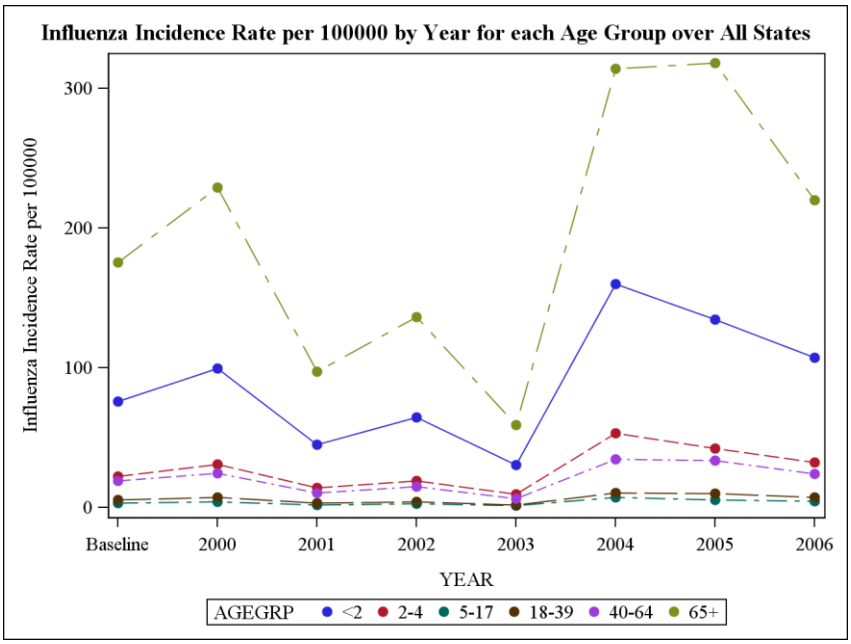


Figure 5. PCV7 coverage(%) by states

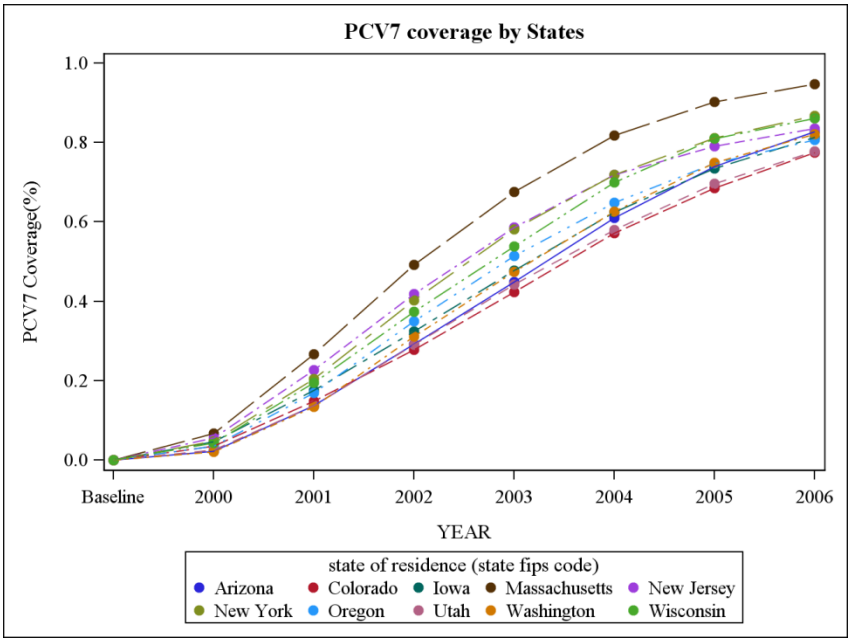


Figure 6. Fit and Confidence Band from Linear Regression models

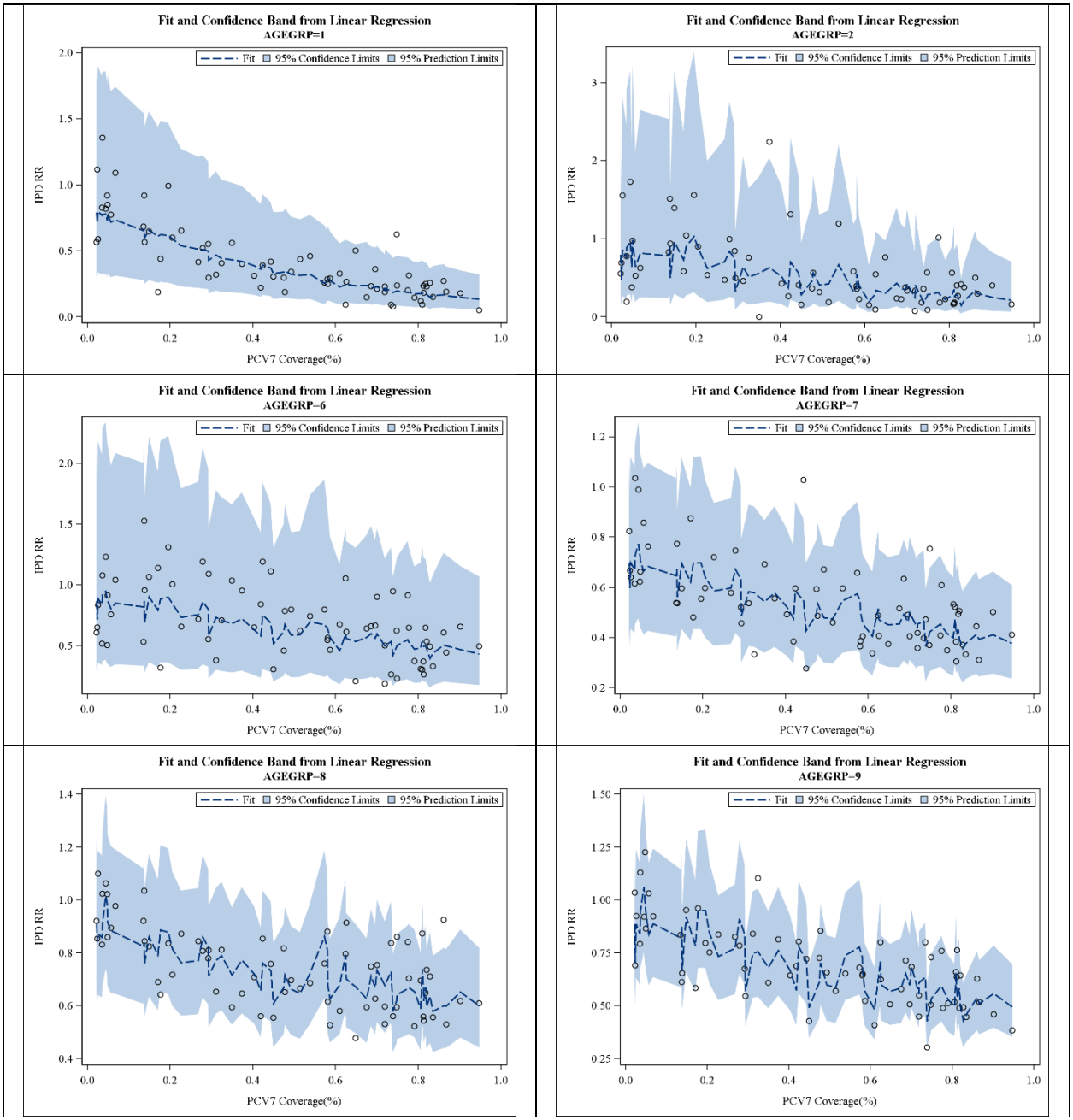
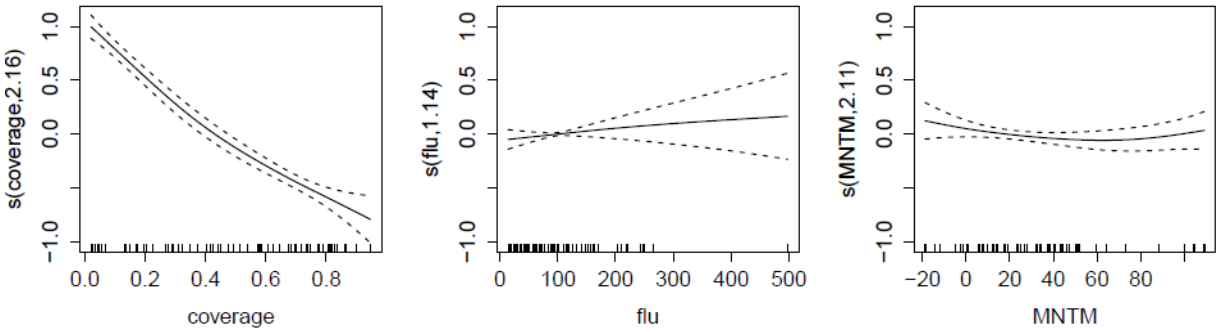
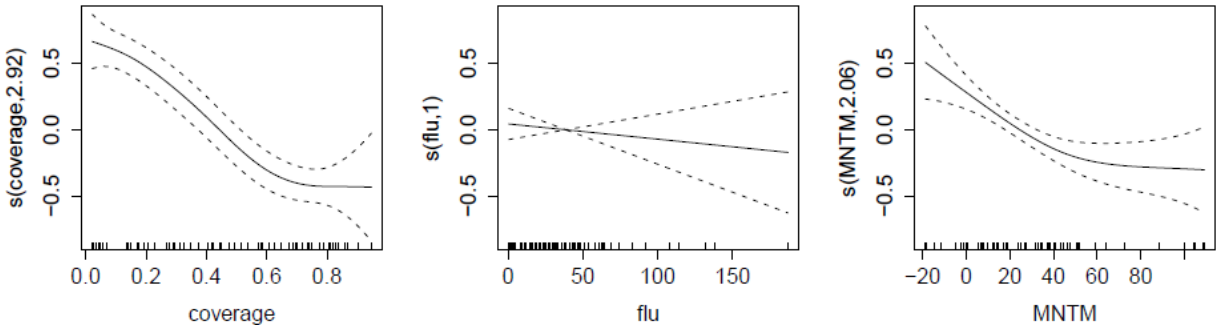


Figure 7. Non-linear effect of PCV7 coverage, influenza incidences, and temperature

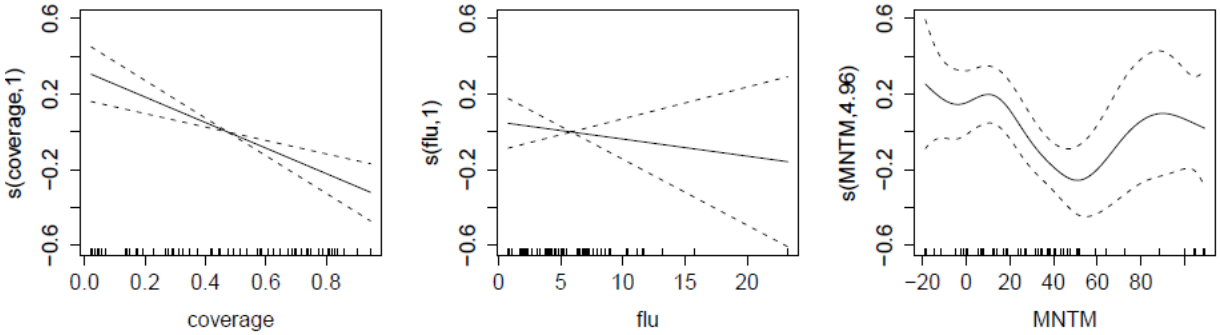
Age group <2:



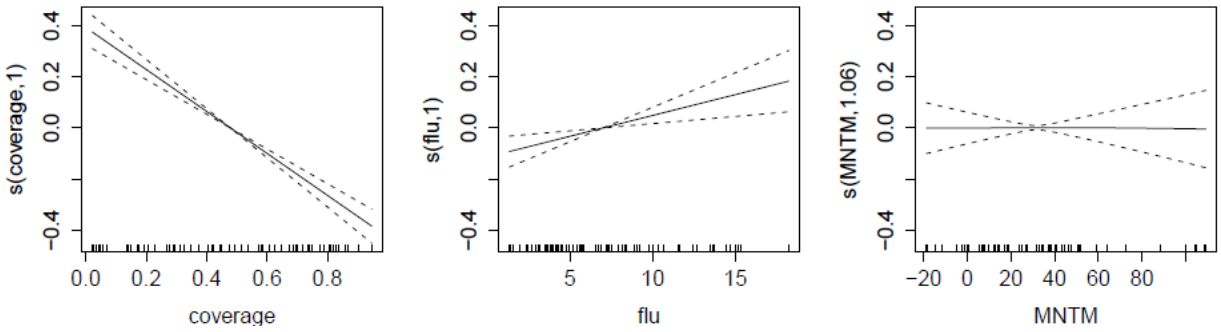
Age group 2-4:



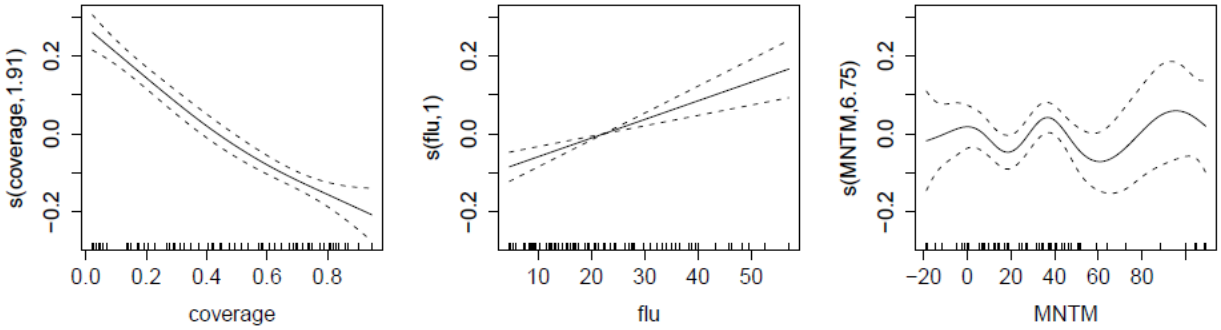
Age group 5-17:



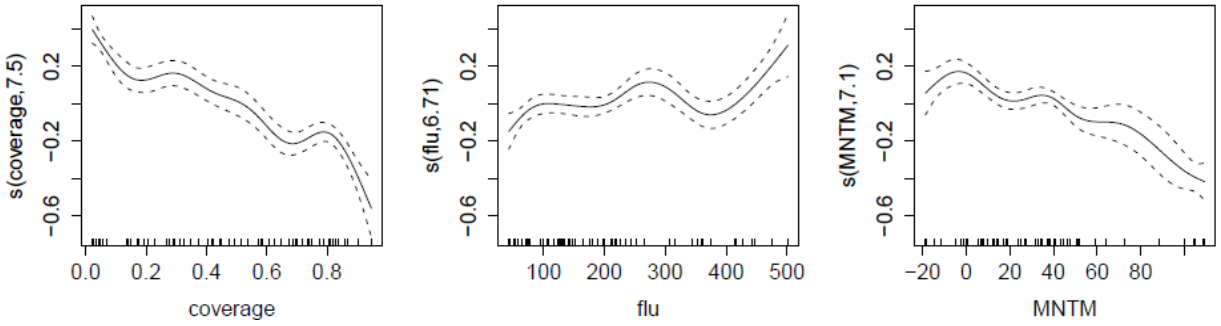
Age group 18-39:



Age group 40-64:



Age group ≥65:



*plots of X_j versus $S_j(X_j)$ shows the relationship between X_j and Y holding constant the other predictors in the model.

Tables

Table 1. IPD hospitalizations per 100,000 by year and age group

IPD Incidence per 100,000 by Year and Age Group								
Age(yr)	Baseline	2000	2001	2002	2003	2004	2005	2006
All	8.92	7.89	6.71	5.71	5.11	4.77	4.81	4.64
<2	27.37	24.42	16.87	9.59	8.08	6.58	5.45	5.41
2-4	5.71	4.40	4.90	2.96	2.30	1.74	1.34	2.02
5-17	1.35	1.11	1.19	1.05	0.84	0.78	0.62	0.63
18-39	3.62	2.73	2.21	1.86	1.66	1.48	1.65	1.38
40-64	8.87	8.15	7.10	6.15	5.67	5.76	5.66	5.70
≥65	29.91	27.90	23.51	21.10	18.59	16.45	17.07	15.94

Table 2. Coefficient estimates

Linear Regression model for log risk ratio

Age Group	Parameter	Estimate	Standard Error	P-Value
<2	Intercept	-0.11296	0.12203	0.358
	PCV7 Coverage	-1.91583	0.182	**<.0001**
	Flu	-0.00030625	0.00065458	0.6414
	Temperature	-0.00125	0.00163	0.4454
2-4	Intercept	0.25498	0.17028	0.1391
	PCV7 Coverage	-1.43282	0.2469	**<.0001**
	Flu	-0.00308	0.00204	0.136
	Temperature	-0.00803	0.00226	**0.0007**
5-17	Intercept	-0.00553	0.13459	0.9673
	PCV7 Coverage	-0.76118	0.20236	**0.0004**
	Flu	0.00183	0.01444	0.8994
	Temperature	-0.003	0.00172	0.0862
18-39	Intercept	-0.28872	0.0678	**<.0001**
	PCV7 Coverage	-0.70359	0.10419	**<.0001**
	Flu	0.00728	0.00711	0.3099
	Temperature	-0.00252	0.00089599	**0.0064**
40-64	Intercept	-0.12493	0.04352	**0.0055**
	PCV7 Coverage	-0.50069	0.06773	**<.0001**
	Flu	0.00497	0.0015	**0.0015**
	Temperature	-0.00186	0.00059441	**0.0026**
≥65	Intercept	-0.0147	0.04748	0.7578
	PCV7 Coverage	-0.7211	0.07412	**<.0001**
	Flu	0.00040363	0.00017248	**0.0223**
	Temperature	-0.00383	0.00063428	**<.0001**

Poisson Models

Age Group	Parameter	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > Chisq
<2	Intercept	-8.7979	0.1653	(-9.1219, -8.4739)	2832.22	**<.0001**
	PCV7 Coverage	-2.0551	0.1018	(-2.2547 -1.8555)	407.29	**<.0001**
	IPD baseline	0.0161	0.0063	(0.0037,0.0285)	6.47	**0.011**
	Flu	0.0009	0.0004	(0.0001,0.0017)	4.37	**0.0366**
	Temperature	-0.0004	0.0009	(-0.0022,0.0014)	0.16	0.688
2-4	Intercept	-10.0324	0.1712	(-10.3679,9.6969)	3434.07	**<.0001**
	PCV7 Coverage	-1.4933	0.1596	(-1.8062 -1.1805)	87.53	**<.0001**
	IPD baseline	0.0761	0.0265	(0.0242,0.1281)	8.24	**0.0041**
	Flu	-0.0012	0.0015	(-0.0041,0.0017)	0.67	0.4134
	Temperature	-0.0068	0.0017	(-0.01,-0.0035)	16.7	**<.0001**
5-17	Intercept	-11.6767	0.1586	(-11.9876,-11.3659)	5420.58	**<.0001**
	PCV7 Coverage	-0.6954	0.157	(-1.0030 -0.3878)	19.63	**<.0001**
	IPD baseline	0.356	0.0799	(0.1994,0.5126)	19.86	**<.0001**
	Flu	-0.0037	0.0124	(-0.0280,0.0206)	0.09	0.7648
	Temperature	-0.0025	0.0014	(-0.0052,0.0002)	3.38	0.0662
18-39	Intercept	-11.0057	0.1472	(-11.2943,10.7171)	5588.19	**<.0001**
	PCV7 Coverage	-0.8187	0.0707	(-0.9572 -0.6802)	134.19	**<.0001**
	IPD baseline	0.106	0.0449	0.0180 0.1940	5.58	**0.0182**
	Flu	0.0163	0.0053	(0.0058,0.0267)	9.3	**0.0023**
	Temperature	0	0.001	(-0.0019,0.0019)	0	0.9924
40-64	Intercept	-9.8769	0.1141	(-10.1006,9.6533)	7492.29	**<.0001**
	PCV7 Coverage	-0.5242	0.0385	(-0.5996 -0.4488)	185.56	**<.0001**
	IPD baseline	0.0379	0.014	(0.0104,0.0655)	7.3	**0.0069**
	Flu	0.005	0.0009	(0.0031,0.0068)	28.66	**<.0001**
	Temperature	0.0004	0.0005	(0.0006,0.0014)	0.57	0.4505
≥64	Intercept	-9.1361	0.0724	(-9.278,8.9942)	15923.8	**<.0001**
	PCV7 Coverage	-0.7475	0.0339	(-0.8141 -0.6810)	485.14	**<.0001**
	IPD baseline	0.0336	0.0026	(0.0285,0.0388)	165.4	**<.0001**
	Flu	0.0004	0.0001	(0.0002,0.0006)	20.23	**<.0001**
	Temperature	-0.0036	0.0004	(-0.0043,-0.0028)	83.91	**<.0001**

Negative Binomials

Age Group	Parameter	Estimate	Standard Error	Wald 95% Confidence Limits	Wald Chi-Square	Pr > Chisq
<2	Intercept	-8.7462	0.2049	(-9.1477,-8.3446)	1822.29	**<.0001**
	PCV7 Coverage	-2.011	0.1298	(-2.2655 -1.7566)	239.89	**<.0001**
	IPD baseline	0.0139	0.0077	(-0.0012,0.0289)	3.27	0.0707
	Flu	0.0008	0.0005	(-0.0003,0.0018)	2.15	0.1423
	Temperature	-0.0003	0.0012	(-0.0026,0.0020)	0.06	0.8034
2-4	Intercept	-10.0404	0.1777	(-10.3887,-9.6921)	3192.27	**<.0001**
	PCV7 Coverage	-1.4752	0.1777	(-1.8235 -1.1269)	68.92	**<.0001**
	IPD baseline	0.0752	0.0274	(0.0215,0.1289)	7.53	**0.0061**
	Flu	-0.0011	0.0015	(-0.0041,0.0019)	0.55	0.4577
	Temperature	-0.0067	0.0017	(-0.0101,-0.0033)	14.83	**0.0001**
5-17	Intercept	-11.677	0.174	(-12.0181,-11.3360)	4503.44	**<.0001**
	PCV7 Coverage	-0.6978	0.1743	(-1.0395 -0.3561)	16.02	**<.0001**
	IPD baseline	0.358	0.0879	(0.1857,0.5302)	16.6	**<.0001**
	Flu	-0.0036	0.0134	(-0.0299,0.0227)	0.07	0.7907
	Temperature	-0.0025	0.0015	(-0.0054,0.0004)	2.76	0.0964
18-39	Intercept	-10.9954	0.1597	(-11.3083,-10.6824)	4742.06	**<.0001**
	PCV7 Coverage	-0.7977	0.0822	(-0.9587 -0.6366)	94.24	**<.0001**
	IPD baseline	0.1032	0.0492	(0.0068,0.1996)	4.41	**0.0358**
	Flu	0.015	0.006	(0.0033,0.0267)	6.29	**0.0121**
	Temperature	0	0.0011	(-0.0021,0.0021)	0	0.998
40-64	Intercept	-9.8995	0.1541	(-10.2015,-9.5975)	4127.75	**<.0001**
	PCV7 Coverage	-0.5016	0.0577	(-0.6146 -0.3885)	75.61	**<.0001**
	IPD baseline	0.0415	0.019	(0.0043,0.0788)	4.77	**0.0289**
	Flu	0.0051	0.0013	(0.0025,0.0077)	15.12	**0.0001**
	Temperature	0.0001	0.0007	(-0.0013,0.0016)	0.02	0.8852
≥64	Intercept	-9.0022	0.1461	(-9.2884,-8.7159)	3798.4	**<.0001**
	PCV7 Coverage	-0.7362	0.0678	(-0.8690 -0.6033)	117.95	**<.0001**
	IPD baseline	0.0291	0.0051	(0.0191,0.0392)	32.34	**<.0001**
	Flu	0.0004	0.0002	(0.0001,0.0008)	6.87	**0.0088**
	Temperature	-0.0036	0.0007	(-0.005,-0.0022)	25.48	**<.0001**

Additive Models

Age Group				
<2	**Parametric**	Estimates	SE	P-Value
	Intercept	-9.73375	0.18701	**<0.0001**
	IPD baseline	0.01718	0.00658	**0.009**
	Smooth Terms	edf	P-value	
	PCV7 Coverage	2.16	**<0.0001**	
	Influenza	1.14	0.3	
	Temperature	2.11	0.31	
2-4	**Parametric**	Estimates	SE	P-Value
	Intercept	-11.09089	0.16846	**<0.0001**
	IPD baseline	0.08998	0.02752	**0.0011**
	Smooth Terms	edf	P-value	
	PCV7 Coverage	2.923	**<0.0001**	
	Influenza	1	0.454	
	Temperature	2.055	**<0.0001**	
5-17	**Parametric**	Estimates	SE	P-Value
	Intercept	-12.14204	0.12697	**<0.0001**
	IPD baseline	0.37405	0.08125	**0.0011**
	Smooth Terms	edf	P-value	
	PCV7 Coverage	1	**<0.0001**	
	Influenza	1	0.4844	
	Temperature	4.965	**0.0193**	
18-39	**Parametric**	Estimates	SE	P-Value
	Intercept	-11.2871	0.16473	**<0.0001**
	IPD baseline	0.10693	0.04507	**0.0177**
	Smooth Terms	edf	P-value	
	PCV7 Coverage	1	**<0.0001**	
	Influenza	1	**0.002**	
	Temperature	1.065	0.962	
40-64	**Parametric**	Estimates	SE	P-Value
	Intercept	-10.05361	0.13418	**<0.0001**
	IPD baseline	0.04367	0.01503	**0.0037**
	Smooth Terms	edf	P-value	
	PCV7 Coverage	1.911	**<0.0001**	
	Influenza	1	**<0.0001**	
	Temperature	6.753	0.165	
≥65	**Parametric**	Estimates	SE	P-Value
	Intercept	-9.655283	0.09408	**<0.0001**

IPD baseline	0.037787	0.00309	**<0.0001**
Smooth Terms	edf	P-value	
PCV7 Coverage	7.503	**<0.0001**	
Influenza	6.71	**<0.0001**	
Temperature	7.096	**<0.0001**	

*Significant terms are bolded

Table 3. Expected rate reeducation for each 10% increase in PCV7 coverage

Linear Regression Model

Age Group	PCV7 Coverage								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5-17	0.92	0.86	0.79	0.74	0.68	0.63	0.58	0.54	0.50
18-39	0.90	0.82	0.75	0.68	0.62	0.57	0.52	0.47	0.43
40-64	0.94	0.89	0.84	0.79	0.74	0.70	0.66	0.62	0.58
≥65	0.93	0.86	0.80	0.74	0.69	0.64	0.60	0.55	0.52

Poisson

Age Group	PCV7 Coverage								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<2	0.81	0.66	0.54	0.44	0.36	0.29	0.24	0.19	0.15
2-4	0.86	0.74	0.63	0.55	0.47	0.4	0.35	0.3	0.26
5-17	0.93	0.87	0.81	0.76	0.71	0.66	0.61	0.57	0.53
18-39	0.92	0.85	0.78	0.72	0.66	0.61	0.56	0.52	0.48
40-64	0.95	0.9	0.85	0.81	0.77	0.73	0.69	0.66	0.62
≥65	0.93	0.86	0.8	0.74	0.68	0.64	0.59	0.55	0.51

Negative Binomial

Age Group	PCV7 Coverage								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<2	0.81	0.67	0.55	0.45	0.37	0.30	0.25	0.20	0.16
2-4	0.86	0.74	0.64	0.55	0.48	0.41	0.36	0.31	0.27
5-17	0.93	0.87	0.81	0.76	0.71	0.66	0.61	0.57	0.54
18-39	0.92	0.85	0.79	0.73	0.67	0.62	0.57	0.53	0.49
40-64	0.95	0.90	0.86	0.81	0.79	0.74	0.70	0.67	0.64
≥65	0.93	0.86	0.80	0.75	0.69	0.64	0.60	0.55	0.52

Additive

Age Group	PCV7 Coverage								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
<2	0.77	0.59	0.46	0.37	0.31	0.26	0.22	0.19	0.17
2-4	0.92	0.81	0.68	0.56	0.37	0.34	0.34	0.33	0.34
5-17	0.93	0.87	0.82	0.76	0.71	0.67	0.62	0.58	0.54
18-39	0.92	0.85	0.78	0.72	0.66	0.61	0.56	0.52	0.49
40-64	0.94	0.88	0.82	0.78	0.74	0.7	0.67	0.65	0.63
≥65	0.78	0.72	0.75	0.69	0.65	0.57	0.52	0.55	0.43

Table 4. Scaled deviance/df

	Poisson	Negative Binomial
Age Group	Scaled Deviance	Scaled Deviance
<2	1.66	1.14
2-4	1.32	1.27
5-17	1.28	1.07
18-39	1.30	1.10
40-64	2.08	1.06
≥65	3.51	1.10

Table 5. AIC values

	Poisson	Negative Binomial	Additive
Age Group	AIC	AIC	AIC
<2	430.04	425.27	424.03
2-4	336.79	338.72	331.74
5-17	329.03	329.89	322.11
18-39	462.18	463.14	462.17
40-64	602.19	583.85	598.38
≥65	709.73	636.75	646.59

Table 6. Expected and observed IPD incidence rate per 100,000

Model - Regression

Year	Age groups (years)											
	<2				2-4				5-17			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	413	350	63	11.27	112	115	-3	0.06	130	135	-5	0.19
2001	291	281	10	0.33	124	114	10	0.93	140	128	12	1.15
2002	167	196	-29	4.33	75	77	-2	0.07	123	107	16	2.39
2003	141	147	-6	0.23	59	71	-12	1.88	98	98	0	0.00
2004	116	107	9	0.70	45	48	-3	0.16	91	86	5	0.27
2005	96	89	7	0.55	35	42	-7	1.32	72	79	-7	0.61
2006	96	80	16	3.34	53	39	14	5.24	73	74	-1	0.01

Year												
	18-39				40-64				65+			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	550	502	48	4.50	1562	1534	28	0.53	2219	2123	96	4.30
2001	446	463	-17	0.61	1398	1417	-19	0.25	1881	1964	-83	3.52
2002	375	397	-22	1.17	1242	1321	-79	4.71	1699	1671	28	0.47
2003	333	358	-25	1.69	1171	1222	-51	2.14	1510	1524	-14	0.13
2004	297	337	-40	4.84	1217	1320	-103	7.97	1348	1494	-146	14.33
2005	331	311	20	1.25	1220	1271	-51	2.08	1414	1396	18	0.22
2006	278	290	-12	0.49	1251	1185	66	3.70	1336	1274	62	2.98

Model - Poisson

Year	Age groups (years)											
	<2				2-4				5-17			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	24.42	23.09	1.33	1.30	4.40	4.61	-0.22	0.26	1.11	1.18	-0.07	0.48
2001	16.87	16.41	0.46	0.22	4.90	4.32	0.58	1.98	1.19	1.13	0.06	0.45
2002	9.59	11.45	-1.86	5.27	2.96	2.98	-0.02	0.00	1.05	0.96	0.09	1.03
2003	8.08	7.93	0.15	0.05	2.30	2.57	-0.27	0.72	0.84	0.89	-0.05	0.29
2004	6.58	6.62	-0.04	0.00	1.74	1.88	-0.13	0.24	0.78	0.77	0.01	0.02
2005	5.45	5.26	0.19	0.12	1.34	1.61	-0.27	1.19	0.62	0.72	-0.10	1.56
2006	5.41	4.50	0.91	3.23	2.02	1.44	0.58	6.16	0.63	0.68	-0.05	0.46

Year												
	18-39				40-64				65+			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	2.73	2.65	0.08	0.53	8.15	8.08	0.08	0.13	27.90	26.86	1.04	3.24
2001	2.21	2.20	0.01	0.02	7.10	6.92	0.18	0.86	23.51	24.51	-0.99	3.14
2002	1.86	1.93	-0.07	0.52	6.15	6.47	-0.32	3.33	21.10	20.68	0.43	0.77
2003	1.66	1.62	0.03	0.15	5.67	5.66	0.01	0.00	18.59	18.57	0.03	0.01
2004	1.48	1.66	-0.18	3.95	5.76	6.06	-0.29	3.36	16.45	17.99	-1.54	10.33
2005	1.65	1.52	0.13	2.46	5.66	5.73	-0.07	0.30	17.07	16.59	0.48	1.32
2006	1.38	1.38	0.00	0.00	5.70	5.29	0.41	6.20	15.94	14.96	0.98	5.74

*Obs=observed IPD incidences rate per 100,000, Exp=expected IPD incidence rate per 100,000, $D=(obs-Exp)^2/Exp$

Model-Negative Binomial

Year	Age groups (years)											
	<2				2-4				5-17			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	24.42	22.80	1.33	1.93	4.40	4.59	-0.22	0.21	1.11	1.18	-0.07	0.53
2001	16.87	16.37	0.46	0.26	4.90	4.30	0.58	2.18	1.19	1.13	0.06	0.41
2002	9.59	11.51	-1.86	5.57	2.96	2.98	-0.02	0.00	1.05	0.96	0.09	0.97
2003	8.08	8.05	0.15	0.00	2.30	2.57	-0.27	0.71	0.84	0.89	-0.05	0.32
2004	6.58	6.68	-0.04	0.02	1.74	1.89	-0.13	0.30	0.78	0.77	0.01	0.01
2005	5.45	5.34	0.19	0.04	1.34	1.63	-0.27	1.30	0.62	0.72	-0.10	1.62
2006	5.41	4.60	0.91	2.51	2.02	1.45	0.58	5.83	0.63	0.68	-0.06	0.50

	18-39				40-64				65+			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	2.73	2.63	0.11	0.86	8.15	8.08	0.07	0.13	27.90	26.85	1.05	3.28
2001	2.21	2.20	0.01	0.02	7.10	6.98	0.12	0.42	23.51	24.54	-1.03	3.33
2002	1.86	1.94	-0.08	0.57	6.15	6.52	-0.37	4.34	21.10	20.75	0.35	0.57
2003	1.66	1.64	0.02	0.06	5.67	5.75	-0.07	0.21	18.59	18.67	-0.07	0.01
2004	1.48	1.66	-0.18	3.94	5.76	6.17	-0.41	5.93	16.45	18.11	-1.66	11.82
2005	1.65	1.52	0.13	2.31	5.66	5.85	-0.19	1.42	17.07	16.73	0.35	0.78
2006	1.38	1.39	-0.01	0.01	5.70	5.40	0.30	3.52	15.94	15.09	0.85	4.47

*Obs=observed IPD incidences rate per 100,000, Exp=expected IPD incidence rate per 100,000, D=$(obs-Exp)^2/Exp$

Model-Additive Poisson

Year	Age groups (years)											
	<2				2-4				5-17			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	24.42	23.26251	1.16	0.972595	4.40	4.21247	0.18	0.20228	1.08	0.977546	0.10	2.132695
2001	16.87	15.95489	0.92	0.905929	4.90	4.393198	0.51	1.4987	1.15	1.196069	-0.05	0.000215
2002	9.59	10.16738	-0.58	0.572152	2.96	2.733426	0.23	0.48386	1.13	0.84445	0.29	5.82276
2003	8.08	7.465307	0.62	0.887958	2.30	2.196262	0.11	0.137398	0.88	0.938624	-0.06	1.240779
2004	6.58	6.324329	0.26	0.187686	1.74	1.601047	0.14	0.324629	0.83	0.712976	0.12	0.771416
2005	5.45	5.389002	0.06	0.011273	1.34	1.532383	-0.19	0.617424	0.67	0.655904	0.02	0.229387
2006	5.41	4.823309	0.59	1.265385	2.02	1.481078	0.54	5.119802	0.65	0.594457	0.05	0.22871

Year												
	18-39				40-64				65+			
	Obs	Exp	Diff	D	Obs	Exp	Diff	D	Obs	Exp	Diff	D
2000	2.73	2.650916	0.08	0.503973	8.15	8.404764	-0.25	1.425891	27.90	29.63451	-1.73	8.061925
2001	2.21	2.199451	0.01	0.017436	7.10	6.703647	0.40	4.66622	23.51	22.71503	0.80	2.240094
2002	1.86	1.934319	-0.07	0.548472	6.15	6.624763	-0.47	6.745665	21.10	24.25903	-3.16	33.10414
2003	1.66	1.623483	0.03	0.138799	5.67	5.399369	0.27	2.870634	18.59	18.11828	0.48	1.014156
2004	1.48	1.663748	-0.18	4.033217	5.76	6.11417	-0.35	4.228765	16.45	17.62173	-1.17	6.347096
2005	1.65	1.516406	0.13	2.378553	5.66	5.949831	-0.29	3.051422	17.07	18.64701	-1.57	10.97998
2006	1.38	1.378754	0.00	3.17E-06	5.70	5.624137	0.08	0.222098	15.94	16.28426	-0.35	0.62151

*Obs=observed IPD incidences rate per 100,000, Exp=expected IPD incidence rate per 100,000, $D=(obs-Exp)^2/Exp$

Table 7. Model Prediction for 2005-2006 flu season

Negative Binomial Models

Age Group	Obs	Exp	Diff	D
<2	5.41	4.29	1	5.19
2-4	2.02	1.20	1	14.80
5-17	0.63	0.71	0	1.18
18-39	1.38	1.39	0	0.02
40-64	5.70	5.11	1	14.77
65+	15.94	14.42	2	13.35

Additive Models

Age Group	Obs	Exp	Diff	D
<2	5.41	4.56	1	2.82
2-4	2.02	1.17	1	16.34
5-17	0.63	0.51	0	3.23
18-39	1.38	1.40	0	0.07
40-64	5.70	5.12	1	14.28
65+	15.94	13.65	2	32.15

*Obs=observed IPD incidences rate per 100,000, Exp=expected IPD incidence rate per 100,000, $D=(obs-Exp)^2/Exp$

SAS Code

```
libname it '\\dataserver.sph.emory.edu\CLIU55\My Documents\Thesis';

proc format;
      value yearb
            1999 = 'Baseline';
      value state
      4 ='Arizona'
      8 = 'Colorado'
      19 = 'Iowa'
      25= 'Massachusetts'
      34 = 'New Jersey'
      36 = 'New York'
      41= 'Oregon'
      49 = 'Utah'
      53 = 'Washington'
      55='Wisconsin';
      value agegroup
      1= '<2'
      2='2-4'
      6='5-17'
      7='18-39'
      8='40-64'
      9='65+';
run;

*list of  states in the dataset;
proc freq data =it.PCV7_yearly_04_19_10;
      tables statefips /nocum list;
run;

*keep only the relevant states;
data initial;
      set it.PCV7_yearly_04_19_10;
      where statefips = 4 or statefips =8 or statefips =19 or statefips =25
or statefips =34 or statefips =36 or statefips =41 or statefips =49 or
statefips =53 or statefips =55;

      if statefips=4 then state= 'Arizona       ';
      else if statefips=8 then state= 'Colorado';
      else if statefips=19 then state= 'Iowa';
      else if statefips=25 then state= 'Massachusetts';
      else if statefips=34 then state= 'New Jersey';
      else if statefips=36 then state= 'New York';
      else if statefips=41 then state= 'Oregon';
      else if statefips=49 then state= 'Utah';
      else if statefips=53 then state= 'Washington';
      else if statefips=55 then state= 'Wisconsin';

      if agegrp=1 or agegrp=2 then sum_pcv=pcv_total/100*POP;
            else sum_pcv=0; *calculate back from PCV coverage rate the actual
number of people that are covered;
run;
* recheck to see if the states specified are included in the dataset;
```

```
proc freq data =initial;
      tables statefips year /nocum list;
run;
*pull out dataset where year is between 2000-2006;
data pcvyears;
      set initial;
      where year ge 2000;
run;
*pull out baseline years 1997-1999;
data pcv_baseline;
      set initial;
      where year between 1997 and 1999 ;
run;

proc sort data=pcv_baseline;
      by year ;


data baseline_plot;
      set pcv_baseline;
      IPD_inc= IPD/POP*100000;
run;
/*
%macro lineplots (state=, name=);
proc sgplot data=baseline_plot (where= (statefips=&state)) ;
      scatter y=IPD_inc x=year/group = agegrp markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_inc x=year/ group= agegrp ;
      format agegrp agegroup.;
      yaxis label='IPD Incidence Rate per 100000';
      xaxis label='year';
      title 'IPD Incidence Rate per 100000 by year for each Age Group of
&name (baseline)';
run;
%mend lineplots;
%lineplots (state=4, name=Arizona);
%lineplots (state=8, name=Colorado);
%lineplots (state=19, name=Iowa);
%lineplots (state=25, name=Massachusetts);
%lineplots (state=34, name=New Jersey);
%lineplots (state=36, name=New York);
%lineplots (state=41, name=Oregon);
%lineplots (state=49, name=Utah);
%lineplots (state=53, name=Washington);
%lineplots (state=55, name=Wisconsin);
*/

*sum up all the years in baseline;
proc means data=pcv_baseline sum noprint;
      class STATEFIPS agegrp;
      output out=baselinesum sum=;
run;
data baselinesum ;
      set baselinesum (where= (_type_ = 3));
      if year= 5994 or year=3997 then year=1999;
      drop _type_ _freq_;
run;
```

```
*merge baseline and PCV years together;
data PCV;
set pcvyears baselinesum;
IPD_inc= IPD/POP*100000;                        *calculate the incidence rate;
run;
*check to see if years are correct;
proc freq data=pcv;
table year /nocum list;
run;

*******************************************************************************
************************************************;
*PCV coverage plots, plot by coverage and year by state;
ods rtf file = '\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\output\exploratory1.rtf';
proc sort data=pcv;
      by year statefips agegrp;
run;

data PCV_total;
      set PCV;
      by year statefips agegrp;
      if agegrp=1 or agegrp=2 ;
run;
proc means data=PCV_total sum noprint;
      var POP sum_pcv;
      class year statefips;
      output out=coverage_state sum= ;
run;
data coverage_state;
      set coverage_state;
      where _TYPE_=3;
      pcv_new=sum_PCV/POP;
run;
proc sgplot data=coverage_state ;
      scatter y=pcv_new x=year / group=statefips markerattrs=(size=7
symbol=circlefilled);
      series y=pcv_new x=year / group=statefips;
      format year yearb. statefips state.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='PCV7 Coverage(%)';
      title 'PCV7 coverage by States';
run;

*PCV overall plot;
proc means data=PCV_total sum noprint;
      var POP sum_pcv;
      class year;
      output out=coverage_all sum= ;
run;
data coverage_all;
      set coverage_all;
      if year ne .;
      pcv_new=sum_PCV/POP;
run;
```

```
proc sgplot data=coverage_all noautolegend ;            * overall plot by
coverage and year;
      scatter y=pcv_new x=year /markerattrs=(size=7 symbol=circlefilled);
      series y=pcv_new x=year ;
      format year yearb. statefips state.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='PCV7 Coverage(%)';
      title 'PCV7 coverage';
run;

*IPD incidence rate per 100000 for all states and age groups;
proc means data=pcv sum noprint;
      var POP IPD sum_pcv;
      class year;
      output out=IPD_year sum= ;
run;
proc means data=pcv mean ;
      var IPD;
      class year;
run;

data IPD_year (where= (year ne .));
      set IPD_year;
      IPD_inc= IPD/POP*100000;
      rr=IPD_inc/8.9206115483;
      new_sum_pcv=sum_pcv/POP;

run;

proc sgplot data=IPD_year NOAUTOLEGEND;
      scatter y=IPD_inc x=year/markerattrs=(size=7 symbol=circlefilled);
      series y=IPD_inc x=year;
      format year yearb.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 )max=8;
      yaxis label='IPD Incidence Rate per 100000';
      title 'IPD Incidence Rate per 100000 by Year';
run;

*IPD incidence rate per 100000 by all states for each age group;
proc means data=pcv sum noprint;
      var POP IPD;
      class year agegrp;
      output out=IPD_age sum=;
run;
data IPD_age1;
      set IPD_age(where= (_type_ = 3));
      drop _type_ _freq_;
      IPD_inc= IPD/POP*100000;
run;

proc sgplot data=IPD_age1;
      scatter y=IPD_inc x=year/ group=agegrp markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_inc x=year/ group=agegrp;
      format year yearb. agegrp agegroup.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 )max=8;
      yaxis label='IPD Incidence Rate per 100000';
```

```
      title 'IPD Incidence Rate per 100000 by Year for each Age Group over
All States';
run;


*IPD incidence rate per 100000 by each state and each age group;
proc sort data=pcv;
      by year statefips agegrp;
run;

%macro lineplots (state=, name=);
proc sgplot data=pcv (where= (statefips=&state)) ;
      scatter y=IPD_inc x=year/group = agegrp markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_inc x=year/ group= agegrp  ;
      format year yearb. agegrp agegroup.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='IPD Incidence Rate per 100000';
      title 'IPD Incidence Rate per 100000 by Year for each Age Group of
&name';
run;
%mend lineplots;
%lineplots (state=4, name=Arizona);
%lineplots (state=8, name=Colorado);
%lineplots (state=19, name=Iowa);
%lineplots (state=25, name=Massachusetts);
%lineplots (state=34, name=New Jersey);
%lineplots (state=36, name=New York);
%lineplots (state=41, name=Oregon);
%lineplots (state=49, name=Utah);
%lineplots (state=53, name=Washington);
%lineplots (state=55, name=Wisconsin);



*IPD incidence rate per 100000 by state for all age groups;
proc means data=pcv sum noprint;
      var POP IPD;
      class year statefips;
      output out=IPD_yearstate sum=;
run;
data IPD_yearstate1  ;
      set IPD_yearstate (where= (_type_ = 3));
      drop _type_ _freq_;
      IPD_inc= IPD/POP*100000;
run;
proc sort data=IPD_yearstate1;
      by year statefips;
run;
 /*
%macro lineplots (state=, name=);
proc sgplot data=IPD_yearstate1 (where= (statefips=&state)) ;
      scatter y=IPD_inc x=year / markerattrs=(size=7 symbol=circlefilled);
      series y=IPD_inc x=year ;
      format year yearb.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='IPD Incidence Rate per 100000';
```

```
      title 'IPD Incidence Rate per 100000 by Year for &name for all Age
Groups';
run;
%mend lineplots;
%lineplots (state=4, name=Arizona);
%lineplots (state=8, name=Colorado);
%lineplots (state=19, name=Iowa);
%lineplots (state=25, name=Massachusetts);
%lineplots (state=34, name=New Jersey);
%lineplots (state=36, name=New York);
%lineplots (state=41, name=Oregon);
%lineplots (state=49, name=Utah);
%lineplots (state=53, name=Washington);
%lineplots (state=55, name=Wisconsin);
*/
proc sgplot data=IPD_yearstate1 ;
      scatter y=IPD_inc x=year / group=statefips markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_inc x=year / group=statefips;
      format year yearb. statefips state.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='IPD Incidence Rate per 100000';
      title 'IPD Incidence Rate per 100000 by Year for each State over All
Age Groups';
run;


*******************************************************************************
*****************************************;

/*
%macro lineplots (state=);
proc sgplot data=PCV_total (where= (statefips=&state)) NOAUTOLEGEND ;
      scatter y=IPD_inc x=year / markerattrs=(size=7 symbol=circlefilled);
      series y=IPD_inc x=year ;
      format year yearb. statefips state.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='PCV7 Coverage (%)';
      title 'PCV7 coverage for statefips=&state';
run;
%mend lineplots;
%lineplots (state=4);
%lineplots (state=8);
%lineplots (state=19);
%lineplots (state=25);
%lineplots (state=34);
%lineplots (state=36);
%lineplots (state=41);
%lineplots (state=49);
%lineplots (state=53);
%lineplots (state=55);
*/



*1/7/2014;
```

```
*reformat the dataset to have the correct PCV coverage since PCV is for ages
0-5;
proc sort data= coverage_state;
by year statefips agegrp;
run;
data coverage_data;
      set coverage_state;
      keep pcv_new POP sum_PCV;
run;
data coverage_data;
      set coverage_data coverage_data coverage_data coverage_data
coverage_data coverage_data ;
      rename POP=pcv_pop;
      rename sum_pcv=sum_pcv1;
run;

proc sort data=PCV;
by agegrp;
run;

data PCV_coverage;
      set PCV;
      set coverage_data;
run;
proc sort data=PCV_coverage;
      by year statefips agegrp;
run;

*IPD rates vs coverage (each state, by age group);
%macro lineplots (state=, name=);
proc sgplot data=PCV_coverage (where= (statefips=&state)) ;
      scatter y=IPD_inc x=pcv_new/group = agegrp markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_inc x=pcv_new/ group= agegrp ;
      format agegrp agegroup.;
      yaxis label='IPD Incidence Rate per 100000';
      xaxis label='PCV7 Coverage (%)';
      title 'IPD Incidence Rate per 100000 by PCV7 Coverage for each Age
Group of &name';
run;
%mend lineplots;
%lineplots (state=4, name=Arizona);
%lineplots (state=8, name=Colorado);
%lineplots (state=19, name=Iowa);
%lineplots (state=25, name=Massachusetts);
%lineplots (state=34, name=New Jersey);
%lineplots (state=36, name=New York);
%lineplots (state=41, name=Oregon);
%lineplots (state=49, name=Utah);
%lineplots (state=53, name=Washington);
%lineplots (state=55, name=Wisconsin);

**********************************IPD Risk Ratio
dataset*****************************;
data baselinesum_rr1 baselinesum_rr2 baselinesum_rr3 baselinesum_rr4
baselinesum_rr5 baselinesum_rr6 baselinesum_rr7 baselinesum_rr8;
      set baselinesum;
```

```
      IPD_inc= IPD/POP*100000;
      rename pop=baseline_pop;
run;
data baselinesum_rr;
      set baselinesum_rr1 baselinesum_rr2 baselinesum_rr3 baselinesum_rr4
baselinesum_rr5 baselinesum_rr6 baselinesum_rr7 baselinesum_rr8;
      rename IPD_inc=IPD_inc_baseline;
      rename IPD=IPD_baseline;
run;
data baselinesum_rr;
      set baselinesum_rr;
      keep IPD_inc_baseline IPD_baseline baseline_pop;
run;
data IPD_rr;
      set PCV_coverage;
      set baselinesum_rr;
      IPD_rr= IPD_inc/IPD_inc_baseline;
run;

**********************************table of baseline with IPD
RR*********************************;
proc means data=IPD_rr (where= (year=1999)) ;
      class  statefips agegrp;
      var IPD_inc IPD;
run;

**********************************plots with IPD
RR*********************************;
*Risk ratio plots by year;
proc means data=IPD_rr sum noprint;
      var pcv_POP IPD sum_pcv1 POP;
      class year;
      output out=IPD_year_rr sum= ;
run;

data IPD_year_rr (where= (year ne .));
      set IPD_year_rr;
      IPD_inc= IPD/POP*100000;
      rr=IPD_inc/8.9206115483;
      new_sum_pcv=sum_pcv1/pcv_POP;
run;
proc sgplot data=IPD_year_rr NOAUTOLEGEND;
      scatter y=rr x=year/markerattrs=(size=7 symbol=circlefilled);
      series y=rr x=year;
      format year yearb.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 )max=8;
      yaxis label='IPD Rate Ratio';
      title 'IPD Rate Ratio by Year';
run;
* IPD RR vs Coverage by year;
proc sgplot data=IPD_year_rr NOAUTOLEGEND;
      scatter y=rr x=new_sum_pcv/markerattrs=(size=7 symbol=circlefilled);
      series y=rr x=new_sum_pcv;
      format year yearb.;
      *xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 )max=8;
      yaxis label='IPD Rate Ratio';
      xaxis label='PCV7 Coverage (%)';
```

```
      title 'IPD Rate Ratio by PCV7 Coverage';
run;
*IPD RR by state for all age groups;
proc means data=IPD_rr sum noprint;
      var POP IPD IPD_baseline sum_pcv1 pcv_pop baseline_pop;
      class year statefips ;
      output out=IPD_rr_yearstate sum=;
run;
data IPD_rr_yearstate1  ;
      set IPD_rr_yearstate (where= (_type_ = 3));
      drop _type_ _freq_;
      IPD_inc= IPD/POP*100000;
      IPD_inc_baseline=IPD_baseline/baseline_pop*100000;
      IPD_rr= IPD_inc/IPD_inc_baseline;
      new_sum_pcv=sum_pcv1/pcv_POP;
run;
proc sort data=IPD_rr_yearstate1;
      by year statefips;
run;
proc sgplot data=IPD_rr_yearstate1 ;
      scatter y=IPD_rr x=year / group=statefips markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_rr x=year / group=statefips;
      format year yearb. statefips state.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='IPD Rate Ratio';
      title 'IPD Rate Ratio by Year for each State over All Age Groups';
run;
proc sgplot data=IPD_rr_yearstate1 ;
      scatter y=IPD_rr x=new_sum_pcv / group=statefips markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_rr x=new_sum_pcv / group=statefips;
      format year yearb. statefips state.;
      yaxis label='IPD Rate Ratio';
      xaxis label='PCV7 Coverage (%)';
      title 'IPD Rate Ratio by PCV7 Coverage for each State over All Age
Groups';
run;
*IPD rate ratio by all states for each age group;
proc means data=IPD_rr sum noprint;
      var POP IPD IPD_baseline sum_pcv1 pcv_pop baseline_pop;
      class year agegrp ;
      output out=IPD_rr_age sum=;
run;
data IPD_rr_age1;
      set IPD_rr_age(where= (_type_ = 3));
      drop _type_ _freq_;
      IPD_inc= IPD/POP*100000;
      IPD_inc_baseline=IPD_baseline/baseline_pop*100000;
      IPD_rr= IPD_inc/IPD_inc_baseline;
      new_sum_pcv=sum_pcv1/pcv_POP;
run;
proc sgplot data=IPD_rr_age1;
      scatter y=IPD_rr x=new_sum_pcv/ group=agegrp markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_rr x=new_sum_pcv/ group=agegrp;
      format agegrp agegroup.;
```

```
      yaxis label='IPD Rate Ratio';
      xaxis label='PCV7 Coverage (%)';
      title 'IPD Rate Ratio by PCV7 Coverage for each Age group over All
states';
run;

proc sgplot data=IPD_rr_age1;
      scatter y=IPD_inc x=new_sum_pcv/ group=agegrp markerattrs=(size=7
symbol=circlefilled);
      series y=IPD_inc x=new_sum_pcv/ group=agegrp;
      format agegrp agegroup.;
      yaxis label='IPD Incidence Rate per 100000';
      xaxis label='PCV7 Coverage (%)';
      title 'IPD Incidence Rate by PCV7 Coverage for each Age group over All
states';
run;

ods rtf close;



libname it "\\dataserver.sph.emory.edu\CLIU55\My Documents\Thesis";

**************Cleaning Tempearture Data*****************************;
proc import datafile="\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\temperature.csv " dbms=csv    out=it.temperature_raw;
guessingrows=32767;
run;

data temp;
      set it.temperature_raw;
      state= trim(substr(station_name, length(station_name)-4,2) );
      year= trim(substr(date,length(date)-7,4) );
      month= trim(substr(date,length(date)-3,2) );
      if MNTM=-9999 then delete;
      if state="PL" then delete;
      if state="NV" then delete;
      if state="IL" then delete;
      if state="MN" then delete;
      if state="PA" then delete;
      if state="RP" then state="NY";
      if state="RT" then state="WI";
      if month=5 or month=6 or month=7 or month=8 or month=9 then delete;
      if year=1996 and (month=10 or month=11 or month=12) then flu_year=1997;
            else if year=1997 and (month=10 or month=11 or month=12) then
flu_year=1998;
            else if year=1998 and (month=10 or month=11 or month=12) then
flu_year=1999;
            else if year=1999 and (month=10 or month=11 or month=12) then
flu_year=2000;
            else if year=2000 and (month=10 or month=11 or month=12) then
flu_year=2001;
            else if year=2001 and (month=10 or month=11 or month=12) then
flu_year=2002;
            else if year=2002 and (month=10 or month=11 or month=12) then
flu_year=2003;
```

```
            else if year=2003 and (month=10 or month=11 or month=12) then
flu_year=2004;
            else if year=2004 and (month=10 or month=11 or month=12) then
flu_year=2005;
            else if year=2005 and (month=10 or month=11 or month=12) then
flu_year=2006;
            else flu_year=year;
      if flu_year=1997 or flu_year=1998 or flu_year=1999 then flu_year=1999;
      if state="AZ" then statefips=4;
      if state="CO" then statefips=8;
      if state="IA" then statefips=19;
      if state="MA" then statefips=25;
      if state="NJ" then statefips=34;
      if state="NY" then statefips=36;
      if state="OR" then statefips=41;
      if state="UT" then statefips=49;
      if state="WA" then statefips=53;
      if state="WI" then statefips=55;
run;

proc freq data=temp;
      table state year month flu_year statefips;
run;

proc sort data=temp;
      by statefips flu_year;
run;

proc means data=temp noprint;
      var MNTM;
      class statefips flu_year;
      output out=temp1 mean= ;
run;

data temp2;
      set temp1;
      where _TYPE_=3;
run;

proc sgplot data=temp2;            * overall plot by coverage and year;
      scatter y=MNTM x=flu_year /group =state markerattrs=(size=7
symbol=circlefilled);
      series y=MNTM x=flu_year /group = state;
      *format year yearb. statefips state.;
      xaxis values=(1999 2000 2001 2002 2003 2004 2005 2006 );
      yaxis label='Temperature (F)';
      title 'Temperature';
run;

proc freq data=temp;
      table state*flu_year/ list;
run;

**********merge data**********;
proc sort data=temp2;
      by flu_year statefips;
run;
```

```
proc sort data=IPD_rr;
by agegrp year statefips ;
run;

data t1 t2 t3 t4 t5 t6;
      set temp2;
run;
data temp3;
      set t1 t2 t3 t4 t5 t6;
      keep MNTM;
run;

data pcv_final;
      set IPD_rr;
      set temp3;
run;

***************keep permanant copy of merged data for modeling;
data it.pcv_final;
      set pcv_final;
      logpop=log(pop);
run;



libname it "\\dataserver.sph.emory.edu\CLIU55\My Documents\Thesis";

data pcv_final;
      set it.pcv_final;
      flu=excess_flu/pop*100000;
      if year ne 1999;
      rr=log(IPD_rr);
      coverage=pcv_total/100;
run;

proc sort data=rr_model; by agegrp; run;
proc means data=pcv_final;
var ipd_rr;
run;
proc print data=pcv_final;
where ipd_rr > 2;
run;

* regression model;
ods rtf file = '\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\output\logregg.rtf';
proc reg data=pcv_final outest=est1 noprint; by agegrp;
      model rr=coverage flu MNTM /aic bic ;
      output out=out1 predicted=preds l95m=llmean u95m=ulmean l95=llpred
u95=ulpred;
run;
ods rtf close;

*Evaluate fit;
proc means data=pcv_final noprint;
```

```
      var IPD IPD_baseline pcv_total Excess_flu MNTM pop sum_pcv1 pcv_pop
baseline_pop;
      class year agegrp;
      output out=fit1 sum= ;
run;

data m1;
      set fit1;
      where _type_=3;
      temperature=MNTM/10;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
      if agegrp=1 then e_r=exp((-0.11296-1.91583*new_sum_pcv-0.00030625*flu-
0.00125*temperature))*IPD_inc_baseline;
       else if agegrp=2 then e_r=exp((0.25498-1.43282*new_sum_pcv-
0.00308*flu-0.00803*temperature))*IPD_inc_baseline;
       else if agegrp=6 then e_r=exp((-0.00553-
0.76118*new_sum_pcv+0.00183*flu-0.00300*temperature))*IPD_inc_baseline;
       else if agegrp=7 then e_r=exp((-0.28872-
0.70359*new_sum_pcv+0.00728*flu-0.00252*temperature))*IPD_inc_baseline;
       else if agegrp=8 then e_r=exp((-0.12493-
0.50069*new_sum_pcv+0.00497*flu-0.00186*temperature))*IPD_inc_baseline;
       else if agegrp=9 then e_r=exp((-0.01470-
0.72110*new_sum_pcv+0.00040363*flu-0.00383*temperature))*IPD_inc_baseline;
            else e_r=IPD;
      e_r=pop*e_r/100000;
      diff_r=IPD-e_r;
      D_r=diff_r**2/e_r;
run;
proc sort data=m1;
by agegrp;
run;
data m1;
      set m1;
      where agegrp ge 6;
run;

proc sort data=m1; by agegrp; run;

data eff1;
      set fit1;
      where _type_=3;
      temperature=MNTM/10;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
      if agegrp=1 then e_b=exp((-0.11296-1.91583*0-0.00030625*flu-
0.00125*temperature))*IPD_inc_baseline;
       else if agegrp=2 then e_b=exp((0.25498-1.43282*0-0.00308*flu-
0.00803*temperature))*IPD_inc_baseline;
       else if agegrp=6 then e_b=exp((-0.00553-0.76118*0+0.00183*flu-
0.00300*temperature))*IPD_inc_baseline;
       else if agegrp=7 then e_b=exp((-0.28872-0.70359*0+0.00728*flu-
0.00252*temperature))*IPD_inc_baseline;
       else if agegrp=8 then e_b=exp((-0.12493-0.50069*0+0.00497*flu-
0.00186*temperature))*IPD_inc_baseline;
```

```
       else if agegrp=9 then e_b=exp((-0.01470-0.72110*0+0.00040363*flu-
0.00383*temperature))*IPD_inc_baseline;
      if agegrp=1 then e_a=exp((-0.11296-1.91583*0.1-0.00030625*flu-
0.00125*temperature))*IPD_inc_baseline;
       else if agegrp=2 then e_a=exp((0.25498-1.43282*0.1-0.00308*flu-
0.00803*temperature))*IPD_inc_baseline;
       else if agegrp=6 then e_a=exp((-0.00553-0.76118*0.1+0.00183*flu-
0.00300*temperature))*IPD_inc_baseline;
       else if agegrp=7 then e_a=exp((-0.28872-0.70359*0.1+0.00728*flu-
0.00252*temperature))*IPD_inc_baseline;
       else if agegrp=8 then e_a=exp((-0.12493-0.50069*0.1+0.00497*flu-
0.00186*temperature))*IPD_inc_baseline;
       else if agegrp=9 then e_a=exp((-0.01470-0.72110*0.1+0.00040363*flu-
0.00383*temperature))*IPD_inc_baseline;
            else e_a=IPD;
      eff=(1-(e_a/e_b))*100;
run;
proc sort data=eff1;
      by agegrp;
run;

*plots;
data out2;
      set out1;
      llmean1=exp(llmean);
      ulmean1=exp(ulmean);
      preds1=exp(preds);
      llpred1=exp(llpred);
      ulpred1=exp(ulpred);
run;

proc sort data=out2;
      by agegrp coverage preds1;
run;

ods rtf file = '\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\output\logregg_noint_plots.rtf';
proc sgplot data=out2; by agegrp;
  title "Fit and Confidence Band from Linear Regression";
  band x=coverage lower=llpred1 upper=ulpred1 /
       fillattrs=GraphConfidence2
       legendlabel="95% Prediction Limits" name="band2";
  band x=coverage lower=llmean1 upper=ulmean1 /
       legendlabel="95% Confidence Limits" name="band1";
  scatter x=coverage y=ipd_rr;
  series x=coverage y=preds1 / lineattrs=GraphPrediction
         legendlabel="Fit" name="series";
xaxis label='PCV7 Coverage(%)';
yaxis label='IPD RR';
  keylegend "series" "band1" "band2" / location=inside position=topright;
run;
ods rtf close;



libname it "\\dataserver.sph.emory.edu\CLIU55\My Documents\Thesis";
```

```
data pcv_final;
      set it.pcv_final;
      flu=excess_flu/pop*100000;
      if year ne 1999;
      coverage=pcv_total/100;
run;
proc means data=pcv_final;
      var MNTM;
run;



ods rtf close;
*negative binomial;
ods rtf file = '\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\output\negb4_2_22.rtf';
proc genmod data=pcv_final; by agegrp;
  model ipd = coverage IPD_inc_baseline flu MNTM / dist=NB offset=logpop
link=log;
    estimate "effect of 10% increase in coverage" coverage 0.7;
      estimate "effect of 10% increase in coverage" coverage 0.8;
      estimate "effect of 10% increase in coverage" coverage 0.9;
run;
ods rtf close;
*poisson;
ods rtf file = '\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\output\poisson4_2_22.rtf';
proc genmod data=pcv_final ; by agegrp;
  model ipd = coverage IPD_inc_baseline flu MNTM / dist=poisson offset=logpop
link=log;
  estimate "effect of 10% increase in coverage" coverage 0.1;
      estimate "effect of 10% increase in coverage" coverage 0.2;
      estimate "effect of 10% increase in coverage" coverage 0.3;
      estimate "effect of 10% increase in coverage" coverage 0.4;
          estimate "effect of 10% increase in coverage" coverage 0.5;
      estimate "effect of 10% increase in coverage" coverage 0.6;
      estimate "effect of 10% increase in coverage" coverage 0.7;
          estimate "effect of 10% increase in coverage" coverage 0.8;
      estimate "effect of 10% increase in coverage" coverage 0.9;
run;
ods rtf close;
*Evaluate fit for each year and each agegrp;
proc means data=pcv_final noprint;
      var IPD IPD_baseline pcv_total Excess_flu MNTM pop sum_pcv1 pcv_pop
baseline_pop;
      class year agegrp;
      output out=fit1 sum= ;
run;
data m4;
      set fit1;
      where _type_=3;
      temperature=MNTM/10;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
```

```
      if agegrp=1 then e_p=exp((-8.7979-
2.0551*new_sum_pcv+0.0161*IPD_inc_baseline+0.0009*flu-
0.0004*temperature)+log(pop));
       else if agegrp=2 then e_p=exp((-10.0324-
1.4933*new_sum_pcv+0.0761*IPD_inc_baseline-0.0012*flu-
0.0068*temperature)+log(pop));
       else if agegrp=6 then e_p=exp((-11.6767-
0.6954*new_sum_pcv+0.3560*IPD_inc_baseline-0.0037*flu-
0.0025*temperature)+log(pop));
       else if agegrp=7 then e_p=exp((-11.0057-
0.8187*new_sum_pcv+0.1060*IPD_inc_baseline+0.0163*flu-
0.0000*temperature)+log(pop));
       else if agegrp=8 then e_p=exp((-9.8769-
0.5242*new_sum_pcv+0.0379*IPD_inc_baseline+0.0050*flu+0.0004*temperature)+log
(pop));
       else if agegrp=9 then e_p=exp((-9.1361-
0.7475*new_sum_pcv+0.0336*IPD_inc_baseline+0.0004*flu-
0.0036*temperature)+log(pop));
            else e_p=IPD;
      if agegrp=1 then e_b=exp((-8.7462-
2.0110*new_sum_pcv+0.0139*IPD_inc_baseline+0.0008*flu-
0.0003*temperature)+log(pop));
       else if agegrp=2 then e_b=exp((-10.0404-
1.4752*new_sum_pcv+0.0752*IPD_inc_baseline-0.0011*flu-
0.0067*temperature)+log(pop));
       else if agegrp=6 then e_b=exp((-11.6770-
0.6978*new_sum_pcv+0.3580*IPD_inc_baseline-0.0036*flu-
0.0025*temperature)+log(pop));
       else if agegrp=7 then e_b=exp((-10.9954-
0.7977*new_sum_pcv+0.1032*IPD_inc_baseline+0.0150*flu+0.0000*temperature)+log
(pop));
       else if agegrp=8 then e_b=exp((-9.8995-
0.5016*new_sum_pcv+0.0415*IPD_inc_baseline+0.0051*flu+0.0001*temperature)+log
(pop));
       else if agegrp=9 then e_b=exp((-9.0022-
0.7362*new_sum_pcv+0.0291*IPD_inc_baseline+0.0004*flu-
0.0036*temperature)+log(pop));
            else e_b=IPD;
      diff_p=IPD-e_p;
      D_p=diff_p**2/e_p;
      diff_b=IPD-e_b;
      D_b=diff_b**2/e_b;
run;

proc sort data=m4;
by agegrp;
run;
*Evaluate fit for each agegrp;
proc means data=pcv_final noprint;
      var IPD IPD_baseline pcv_total Excess_flu MNTM pop sum_pcv1 pcv_pop
baseline_pop;
      class agegrp;
      output out=fit1 sum= ;
run;
data m5;
      set fit1;
      where _type_=1;
```

```
      temperature=MNTM/70;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
      if agegrp=1 then e_p=exp((-8.7979-
2.0551*new_sum_pcv+0.0161*IPD_inc_baseline+0.0009*flu-
0.0004*temperature)+log(pop));
       else if agegrp=2 then e_p=exp((-10.0324-
1.4933*new_sum_pcv+0.0761*IPD_inc_baseline-0.0012*flu-
0.0068*temperature)+log(pop));
       else if agegrp=6 then e_p=exp((-11.6767-
0.6954*new_sum_pcv+0.3560*IPD_inc_baseline-0.0037*flu-
0.0025*temperature)+log(pop));
       else if agegrp=7 then e_p=exp((-11.0057-
0.8187*new_sum_pcv+0.1060*IPD_inc_baseline+0.0163*flu-
0.0000*temperature)+log(pop));
       else if agegrp=8 then e_p=exp((-9.8769-
0.5242*new_sum_pcv+0.0379*IPD_inc_baseline+0.0050*flu+0.0004*temperature)+log
(pop));
       else if agegrp=9 then e_p=exp((-9.1361-
0.7475*new_sum_pcv+0.0336*IPD_inc_baseline+0.0004*flu-
0.0036*temperature)+log(pop));
            else e_p=IPD;
      if agegrp=1 then e_b=exp((-8.7462-
2.0110*new_sum_pcv+0.0139*IPD_inc_baseline+0.0008*flu-
0.0003*temperature)+log(pop));
       else if agegrp=2 then e_b=exp((-10.0404-
1.4752*new_sum_pcv+0.0752*IPD_inc_baseline-0.0011*flu-
0.0067*temperature)+log(pop));
       else if agegrp=6 then e_b=exp((-11.6770-
0.6978*new_sum_pcv+0.3580*IPD_inc_baseline-0.0036*flu-
0.0025*temperature)+log(pop));
       else if agegrp=7 then e_b=exp((-10.9954-
0.7977*new_sum_pcv+0.1032*IPD_inc_baseline+0.0150*flu+0.0000*temperature)+log
(pop));
       else if agegrp=8 then e_b=exp((-9.8995-
0.5016*new_sum_pcv+0.0415*IPD_inc_baseline+0.0051*flu+0.0001*temperature)+log
(pop));
       else if agegrp=9 then e_b=exp((-9.0022-
0.7362*new_sum_pcv+0.0291*IPD_inc_baseline+0.0004*flu-
0.0036*temperature)+log(pop));
            else e_b=IPD;
      diff_p=IPD-e_p;
      D_p=diff_p**2/e_p;
      diff_b=IPD-e_b;
      D_b=diff_b**2/e_b;
run;
proc print data=pcv_final;
where statefips=4;
run;

data try;
set pcv_final;
where agegrp=7 and statefips ne 4 and statefips ne 8 or statefips ne 19 ;
run;
*calculate effect of coverage on cases;
```

```
data effect;
      set fit1;
      where _type_=3;
      temperature=MNTM/10;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
      if agegrp=1 then e_b=exp((-8.7462-
2.0110*0+0.0139*IPD_inc_baseline+0.0008*flu-0.0003*temperature)+log(pop));
       else if agegrp=2 then e_b=exp((-10.0404-
1.4752*0+0.0752*IPD_inc_baseline-0.0011*flu-0.0067*temperature)+log(pop));
       else if agegrp=6 then e_b=exp((-11.6770-
0.6978*0+0.3580*IPD_inc_baseline-0.0036*flu-0.0025*temperature)+log(pop));
       else if agegrp=7 then e_b=exp((-11.0057-
0.0082*0+0.1060*IPD_inc_baseline+0.0163*flu-0.0000*temperature)+log(pop));
       else if agegrp=8 then e_b=exp((-9.8769-
0.0052*0+0.0379*IPD_inc_baseline+0.0050*flu+0.0004*temperature)+log(pop));
       else if agegrp=9 then e_b=exp((-9.0022-
0.7362*0+0.0291*IPD_inc_baseline+0.0004*flu-0.0036*temperature)+log(pop));
            else e_b=IPD;
      if agegrp=1 then e_a=exp((-8.7462-
2.0110*0.1+0.0139*IPD_inc_baseline+0.0008*flu-0.0003*temperature)+log(pop));
       else if agegrp=2 then e_a=exp((-10.0404-
1.4752*0.1+0.0752*IPD_inc_baseline-0.0011*flu-0.0067*temperature)+log(pop));
       else if agegrp=6 then e_a=exp((-11.6770-
0.6978*0.1+0.3580*IPD_inc_baseline-0.0036*flu-0.0025*temperature)+log(pop));
       else if agegrp=7 then e_a=exp((-10.9954-
0.7977*0.1+0.1032*IPD_inc_baseline+0.0150*flu+0.0000*temperature)+log(pop));
       else if agegrp=8 then e_a=exp((-9.8995-
0.5016*0.1+0.0415*IPD_inc_baseline+0.0051*flu+0.0001*temperature)+log(pop));
       else if agegrp=9 then e_a=exp((-9.0022-
0.7362*0.1+0.0291*IPD_inc_baseline+0.0004*flu-0.0036*temperature)+log(pop));
            else e_a=IPD;
      diff=e_a-e_b;
      eff=(1-(e_a/e_b))*100;
run;
proc sort data=effect;
by agegrp;
run;



data base;
      set it.pcv_final;
      flu=excess_flu/pop*100000;
      if year = 1999;
      coverage=pcv_total/100;
run;
proc means data=base noprint;
      var IPD IPD_baseline pcv_total Excess_flu MNTM pop sum_pcv1 pcv_pop
baseline_pop;
      class year agegrp;
      output out=basefit sum= ;
run;

data basefit1;
```

```
      set basefit;
      where _type_=3;
      temperature=MNTM/10;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
run;



libname it "\\dataserver.sph.emory.edu\CLIU55\My Documents\Thesis";

data pcv_final;
      set it.pcv_final;
      flu=excess_flu/pop*100000;
      if year ne 1999 and year < 2006;
      coverage=pcv_total/100;
run;

data predict;
      set it.pcv_final;
      flu=excess_flu/pop*100000;
      if year =2006;
      coverage=pcv_total/100;
run;

proc means data=pcv_final;
      var MNTM;
run;

*negative binomial;
ods rtf file = '\\dataserver.sph.emory.edu\CLIU55\My
Documents\Thesis\output\negb4_2_22.rtf';
proc genmod data=pcv_final; by agegrp;
  model ipd = coverage IPD_inc_baseline flu MNTM / dist=NB offset=logpop
link=log;
run;
ods rtf close;

*Evaluate fit-----------by agegrp;
proc means data=predict noprint;
      var IPD IPD_baseline pcv_total Excess_flu MNTM pop sum_pcv1 pcv_pop
baseline_pop;
      class year agegrp;
      output out=fit1 sum= ;
run;
data p1;
      set fit1;
      where _type_=3;
      temperature=MNTM/10;
      new_sum_pcv=sum_pcv1/pcv_POP;
      IPD_inc_baseline= IPD_baseline/baseline_pop*100000;
      flu=excess_flu/pop*100000;
      logpop=log(pop);
```

```
      if agegrp=1 then e_b=exp((-8.6778-
2.1118*new_sum_pcv+0.0126*IPD_inc_baseline+0.0007*flu-
0.0005*temperature)+log(pop));
       else if agegrp=2 then e_b=exp((-9.9520-
1.7525*new_sum_pcv+0.0807*IPD_inc_baseline-0.0019*flu-
0.0081*temperature)+log(pop));
       else if agegrp=6 then e_b=exp((-11.7360-
0.6267*new_sum_pcv+0.3972*IPD_inc_baseline-0.0090*flu-
0.0021*temperature)+log(pop));
       else if agegrp=7 then e_b=exp((-11.0536-
0.7897*new_sum_pcv+0.1231*IPD_inc_baseline+0.0143*flu-
0.0004*temperature)+log(pop));
       else if agegrp=8 then e_b=exp((-10.0024-
0.5891*new_sum_pcv+0.0559*IPD_inc_baseline+0.0055*flu-
0.0003*temperature)+log(pop));
       else if agegrp=9 then e_b=exp((-9.0789-
0.8182*new_sum_pcv+0.0321*IPD_inc_baseline+0.0005*flu-
0.0039*temperature)+log(pop));
            else e_b=IPD;
      diff_b=IPD-e_b;
      D_b=diff_b**2/e_b;
run;
```

R Code

```
dat1 <- read.csv(file="pcv.csv",head=TRUE,sep=",")
str(dat1)
library(lme4)
library(mgcv)
dat1$coverage=dat1$pcv_total/100


dat1<-subset(dat1,YEAR!=1999)

agegrp1<- subset(dat1, AGEGRP == 1)
str(agegrp1)
agegrp2<- subset(dat1, AGEGRP == 2)
str(agegrp2)
agegrp6<- subset(dat1, AGEGRP == 6)
agegrp6 <- na.omit(agegrp6)
agegrp7<- subset(dat1, AGEGRP == 7)
agegrp8<- subset(dat1, AGEGRP == 8)
agegrp9<- subset(dat1, AGEGRP == 9)
par(mfrow=c(2,3))
fit1 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp1)
plot(fit1)
summary(fit1)
fit2 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp2)
plot(fit2)
summary(fit2)
fit3 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp6)
plot(fit3)
summary(fit3)
fit4 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp7)
plot(fit4)
summary(fit4)
fit5 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp8)
plot(fit5)
summary(fit5)
fit6 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp9)
plot(fit6)
summary(fit6)
AIC(fit1,fit2,fit3,fit4,fit5,fit6)
BIC(fit1,fit2,fit3,fit4,fit5,fit6)
```

```
x<-fit1$fitted
x
agegrp1["fitted"] <- x
write.table(agegrp1, "H:/My Documents/Thesis/scratch/agegrp1.txt", sep="\t", row.names=FALSE)

x<-fit2$fitted
x
agegrp2["fitted"] <- x
write.table(agegrp2, "H:/My Documents/Thesis/scratch/agegrp2.txt", sep="\t", row.names=FALSE)

x<-fit3$fitted
x
agegrp6["fitted"] <- x
write.table(agegrp6, "H:/My Documents/Thesis/scratch/agegrp6.txt", sep="\t", row.names=FALSE)

x<-fit4$fitted
x
agegrp7["fitted"] <- x
write.table(agegrp7, "H:/My Documents/Thesis/scratch/agegrp7.txt", sep="\t", row.names=FALSE)

x<-fit5$fitted
x
agegrp8["fitted"] <- x
write.table(agegrp8, "H:/My Documents/Thesis/scratch/agegrp8.txt", sep="\t", row.names=FALSE)

x<-fit6$fitted
x
agegrp9["fitted"] <- x
write.table(agegrp9, "H:/My Documents/Thesis/scratch/agegrp9.txt", sep="\t", row.names=FALSE)

dat1 <- read.csv(file="pcv.csv",head=TRUE,sep=",")
str(dat1)
library(lme4)
library(mgcv)



dat1$coverage=dat1$pcv_total/100
dat1<-subset(dat1,YEAR!=1999)
pred<-subset(dat1,YEAR==2006)
dat1<-subset(dat1,YEAR!=2006)


agegrp1<- subset(dat1, AGEGRP == 1)
agegrp2<- subset(dat1, AGEGRP == 2)
agegrp6<- subset(dat1, AGEGRP == 6)
```

```
#agegrp6 <- na.omit(agegrp6)
agegrp7<- subset(dat1, AGEGRP == 7)
agegrp8<- subset(dat1, AGEGRP == 8)
agegrp9<- subset(dat1, AGEGRP == 9)

agegrp1_p<- subset(pred, AGEGRP == 1)
agegrp2_p<- subset(pred, AGEGRP == 2)
agegrp6_p<- subset(pred, AGEGRP == 6)
#agegrp6 <- na.omit(agegrp6)
agegrp7_p<- subset(pred, AGEGRP == 7)
agegrp8_p<- subset(pred, AGEGRP == 8)
agegrp9_p<- subset(pred, AGEGRP == 9)

par(mfrow=c(1,3))
fit1 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp1)
plot(fit1)
summary(fit1)
x<-fit1$fitted
xfit2 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp2)
plot(fit2)
summary(fit2)
fit3 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp6)
plot(fit3)
summary(fit3)
fit4 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp7)
plot(fit4)
summary(fit4)
fit5 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp8)
plot(fit5)
summary(fit5)
fit6 <- gam (IPD ~ s(coverage)+IPD_inc_baseline+s(flu)+s(MNTM)+offset(logpop), family=poisson ,
data=agegrp9)
plot(fit6)
summary(fit6)

ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=27.36811592,flu=107.18101532,MNTM=37.422483373,logpop=14.3890675
71)
x1<-predict(fit1, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.389067571)
est1
```

```
ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=5.7116454942,flu=32.485882161,MNTM=37.42248337,logpop=14.7808794
13)
x1<-predict(fit2, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.780879413)
est1

ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=1.350006019,flu=4.805486315,MNTM=37.42248337,logpop=16.26752618)
x1<-predict(fit3, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.26752618)
est1

ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=3.619432579,flu=7.350833509,MNTM=37.42248337,logpop=16.81925956)
x1<-predict(fit4, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81925956)
est1

ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=8.866203335,flu=24.25622814,MNTM=37.42248337,logpop=16.90423223)
x1<-predict(fit5, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.90423223)
est1

ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=29.91429636,flu=220.073431,MNTM=37.42248337,logpop=15.94172964)
x1<-predict(fit6, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.94172964)
est1

age1.predict<-predict(fit1,agegrp1_p,type="response")
agegrp1_p["pred"] <- age1.predict

write.table(agegrp1_p, "H:/My Documents/Thesis/scratch/agegrp1_p.txt", sep="\t", row.names=FALSE)

age2.predict<-predict(fit2,agegrp2_p,type="response")
age2.predict
agegrp2_p["pred"] <- age2.predict
write.table(agegrp2_p, "H:/My Documents/Thesis/scratch/agegrp2_p.txt", sep="\t", row.names=FALSE)

age3.predict<-predict(fit3,agegrp6_p,type="response")
agegrp6_p["pred"] <- age3.predict
```

```
write.table(agegrp6_p, "H:/My Documents/Thesis/scratch/agegrp6_p.txt", sep="\t", row.names=FALSE)

age4.predict<-predict(fit4,agegrp7_p,type="response")
agegrp7_p["pred"] <- age4.predict
write.table(agegrp7_p, "H:/My Documents/Thesis/scratch/agegrp7_p.txt", sep="\t", row.names=FALSE)

age5.predict<-predict(fit5,agegrp8_p,type="response")
agegrp8_p["pred"] <- age5.predict
write.table(agegrp8_p, "H:/My Documents/Thesis/scratch/agegrp8_p.txt", sep="\t", row.names=FALSE)

age6.predict<-predict(fit6,agegrp9_p,type="response")
ageagegrp9_p["pred"] <- age6.predict
write.table(agegrp9_p, "H:/My Documents/Thesis/scratch/agegrp9_p.txt", sep="\t", row.names=FALSE)

b <-
data.frame(coverage=0.0434630382,IPD_inc_baseline=27.368115919,flu=99.686749467,MNTM=41.806
147342,logpop=14.341006839)
x2<-predict(fit1, b,type="lpmatrix",scale=)
x2
a <-data.frame(coverage=0,IPD_inc_baseline=50,flu=0,MNTM=0,logpop=2000)
x1<-predict(fit6, a,type="lpmatrix")
x2<-predict(fit6, b,type="lpmatrix")
est1=x1 %*% coef(fit1)
est2=x2 %*% coef(fit1)
est2
ratio=exp(est2)/exp(est1)
ratio


#agegrp1
ag00 <-data.frame(coverage=0.043463038
          ,IPD_inc_baseline=27.36811592,flu=99.68674947,MNTM=41.80614734,logpop=14.34100684)
x1<-predict(fit1, ag00,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.34100684)
est1
ag01 <-data.frame(coverage=0.189440278
          ,IPD_inc_baseline=27.36811592,flu=45.27730465,MNTM=22.31003991,logpop=14.36069471)
x1<-predict(fit1, ag01,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.36069471)
est1
ag02 <-data.frame(coverage=0.370438569
          ,IPD_inc_baseline=27.36811592,flu=64.77163691,MNTM=36.30414188,logpop=14.37025909)
x1<-predict(fit1, ag02,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.37025909)
est1
```

```
ag03 <-data.frame(coverage=0.536582239
            ,IPD_inc_baseline=27.36811592,flu=30.60731446,MNTM=23.10746461,logpop=14.37208229)
x1<-predict(fit1, ag03,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.37208229)
est1
ag04 <-data.frame(coverage=0.679961494
            ,IPD_inc_baseline=27.36811592,flu=159.9430607,MNTM=30.42697247,logpop=14.38189028)
x1<-predict(fit1, ag04,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.38189028)
est1
ag05 <-data.frame(coverage=0.780295098
            ,IPD_inc_baseline=27.36811592,flu=134.6040134,MNTM=32.48344489,logpop=14.38207699)
x1<-predict(fit1, ag05,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.38207699)
est1
ag06 <-data.frame(coverage=0.842879822
            ,IPD_inc_baseline=27.36811592,flu=107.1810153,MNTM=37.42248337,logpop=14.38906757)
x1<-predict(fit1, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit1)
est1<-exp(est1+14.38906757)
est1

#agegrp2
ag00 <-data.frame(coverage=0.043463038
             ,IPD_inc_baseline=5.711645494,flu=31.00281144,MNTM=41.80614734,logpop=14.75088052)
x1<-predict(fit2, ag00,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.75088052)
est1

ag01 <-data.frame(coverage=0.189440278
             ,IPD_inc_baseline=5.711645494,flu=14.15674442,MNTM=22.31003991,logpop=14.7432673)
x1<-predict(fit2, ag01,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.7432673)
est1

ag02 <-data.frame(coverage=0.370438569
             ,IPD_inc_baseline=5.711645494,flu=19.31205637,MNTM=36.30414188,logpop=14.74455837)
x1<-predict(fit2, ag02,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.74455837)
est1

ag03 <-data.frame(coverage=0.536582239
```

```
            ,IPD_inc_baseline=5.711645494,flu=9.727189448,MNTM=23.10746461,logpop=14.75545336)
x1<-predict(fit2, ag03,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.75545336)
est1

ag04 <-data.frame(coverage=0.679961494
            ,IPD_inc_baseline=5.711645494,flu=53.13998435,MNTM=30.42697247,logpop=14.76402063)
x1<-predict(fit2, ag04,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.76402063)
est1

ag05 <-data.frame(coverage=0.780295098
            ,IPD_inc_baseline=5.711645494,flu=42.36641543,MNTM=32.48344489,logpop=14.77417012)
x1<-predict(fit2, ag05,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.77417012)
est1

ag06 <-data.frame(coverage=0.842879822
            ,IPD_inc_baseline=5.711645494,flu=32.48588216,MNTM=37.42248337,logpop=14.78087941)
x1<-predict(fit2, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit2)
est1<-exp(est1+14.78087941)
est1


#agegrp4
ag00 <-data.frame(coverage=0.043463038
            ,IPD_inc_baseline=3.619432579,flu=7.292984318,MNTM=41.80614734,logpop=16.81766884)
x1<-predict(fit4, ag00,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81766884)
est1

ag01 <-data.frame(coverage=0.189440278
            ,IPD_inc_baseline=3.619432579,flu=3.200770011,MNTM=22.31003991,logpop=16.81878437)
x1<-predict(fit4, ag01,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81878437)
est1

ag02 <-data.frame(coverage=0.370438569
            ,IPD_inc_baseline=3.619432579,flu=4.388751867,MNTM=36.30414188,logpop=16.81833765)
x1<-predict(fit4, ag02,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81833765)
```

```
est1

ag03 <-data.frame(coverage=0.536582239
            ,IPD_inc_baseline=3.619432579,flu=2.010255988,MNTM=23.10746461,logpop=16.81607827)
x1<-predict(fit4, ag03,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81607827)
est1

ag04 <-data.frame(coverage=0.679961494
            ,IPD_inc_baseline=3.619432579,flu=10.70488982,MNTM=30.42697247,logpop=16.81405163)
x1<-predict(fit4, ag04,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81405163)
est1

ag05 <-data.frame(coverage=0.780295098
            ,IPD_inc_baseline=3.619432579,flu=10.05268885,MNTM=32.48344489,logpop=16.81395595)
x1<-predict(fit4, ag05,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81395595)
est1

ag06 <-data.frame(coverage=0.842879822
            ,IPD_inc_baseline=3.619432579,flu=7.350833509,MNTM=37.42248337,logpop=16.81925956)
x1<-predict(fit4, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit4)
est1<-exp(est1+16.81925956)
est1

#agegrp3
ag00 <-data.frame(coverage=0.043463038
            ,IPD_inc_baseline=1.350006019,flu=4.247547042,MNTM=41.80614734,logpop=16.27517384)
x1<-predict(fit3, ag00,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.27517384)
est1

ag01 <-data.frame(coverage=0.189440278
            ,IPD_inc_baseline=1.350006019,flu=1.971071833,MNTM=22.31003991,logpop=16.2767657)
x1<-predict(fit3, ag01,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.2767657)
est1

ag02 <-data.frame(coverage=0.370438569
            ,IPD_inc_baseline=1.350006019,flu=2.678583463,MNTM=36.30414188,logpop=16.27703035)
x1<-predict(fit3, ag02,type="lpmatrix",scale=)
```

```
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.27703035)
est1

ag03 <-data.frame(coverage=0.536582239
          ,IPD_inc_baseline=1.350006019,flu=1.31808926,MNTM=23.10746461,logpop=16.27369491)
x1<-predict(fit3, ag03,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.27369491)
est1

ag04 <-data.frame(coverage=0.679961494
          ,IPD_inc_baseline=1.350006019,flu=7.198623156,MNTM=30.42697247,logpop=16.27005379)
x1<-predict(fit3, ag04,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.27005379)
est1

ag05 <-data.frame(coverage=0.780295098
          ,IPD_inc_baseline=1.350006019,flu=5.570599906,MNTM=32.48344489,logpop=16.26776901)
x1<-predict(fit3, ag05,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.26776901)
est1

ag06 <-data.frame(coverage=0.842879822
          ,IPD_inc_baseline=1.350006019,flu=4.805486315,MNTM=37.42248337,logpop=16.26752618)
x1<-predict(fit3, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit3)
est1<-exp(est1+16.26752618)
est1

#agegrp5
ag00 <-data.frame(coverage=0.043463038
          ,IPD_inc_baseline=8.866203335,flu=24.64140105,MNTM=41.80614734,logpop=16.76806155)
x1<-predict(fit5, ag00,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.76806155)
est1

ag01 <-data.frame(coverage=0.189440278
          ,IPD_inc_baseline=8.866203335,flu=10.61787228,MNTM=22.31003991,logpop=16.79530616)
x1<-predict(fit5, ag01,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.79530616)
est1

ag02 <-data.frame(coverage=0.370438569
```

```
            ,IPD_inc_baseline=8.866203335,flu=14.98908274,MNTM=36.30414188,logpop=16.82026972)
x1<-predict(fit5, ag02,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.82026972)
est1

ag03 <-data.frame(coverage=0.536582239
            ,IPD_inc_baseline=8.866203335,flu=6.652076027,MNTM=23.10746461,logpop=16.84274988)
x1<-predict(fit5, ag03,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.84274988)
est1

ag04 <-data.frame(coverage=0.679961494
            ,IPD_inc_baseline=8.866203335,flu=34.7132296,MNTM=30.42697247,logpop=16.86539895)
x1<-predict(fit5, ag04,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.86539895)
est1

ag05 <-data.frame(coverage=0.780295098
            ,IPD_inc_baseline=8.866203335,flu=33.92987566,MNTM=32.48344489,logpop=16.88617515)
x1<-predict(fit5, ag05,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.88617515)
est1

ag06 <-data.frame(coverage=0.842879822
            ,IPD_inc_baseline=8.866203335,flu=24.25622814,MNTM=37.42248337,logpop=16.90423223)
x1<-predict(fit5, ag06,type="lpmatrix",scale=)
est1=x1 %*% coef(fit5)
est1<-exp(est1+16.90423223)
est1


#agegrp6
ag6 <-data.frame(coverage=0.043463038
            ,IPD_inc_baseline=29.91429636,flu=229.0949592,MNTM=41.80614734,logpop=15.88906405)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.88906405)
est1
ag6 <-data.frame(coverage=0.189440278
            ,IPD_inc_baseline=29.91429636,flu=97.13775499,MNTM=22.31003991,logpop=15.89494947)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.89494947)
est1
```

```
ag6 <-data.frame(coverage=0.370438569
                 ,IPD_inc_baseline=29.91429636,flu=229.0949592,MNTM=36.30414188,logpop=15.9014053
6)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.90140536)
est1
ag6 <-data.frame(coverage=0.536582239
                 ,IPD_inc_baseline=29.91429636,flu=59.29129722,MNTM=23.10746461,logpop=15.9099542
5)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.90995425)
est1
ag6 <-data.frame(coverage=0.679961494
                 ,IPD_inc_baseline=29.91429636,flu=313.9924942,MNTM=30.42697247,logpop=15.9187749
4)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.91877494)
est1
ag6 <-data.frame(coverage=0.780295098
                 ,IPD_inc_baseline=29.91429636,flu=318.1387639,MNTM=32.48344489,logpop=15.9295095
5)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.92950955)
est1
ag6 <-data.frame(coverage=0.842879822
              ,IPD_inc_baseline=29.91429636,flu=220.073431,MNTM=37.42248337,logpop=15.88906405)
x1<-predict(fit6, ag6,type="lpmatrix",scale=)
est1=x1 %*% coef(fit6)
est1<-exp(est1+15.88906405)
est1
```