Distribution Agreement

Time series analysis of air pollution and health accounting for covariate-dependent overdispersion

By

Anqi Pan

Master of Science in Public Health

Biostatistics and Bioinformatics

Howard H Chang, PhD

(Thesis Advisor)

Stefanie Ebelt Sarnat, PhD

(Reader)

Time series analysis of air pollution and health accounting for covariate-dependent overdispersion

By

Anqi Pan

B.E.

South China University of Technology

2012

Thesis Committee Chair: Howard H Chang, PhD

Reader: Stefanie Ebelt Sarnat, PhD

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 Science in Public Health

in Biostatistics

2017

Abstract

Time series analysis of air pollution and health accounting for covariate-dependent overdispersion

By: Anqi Pan

The time series study design is routinely used to estimate short-term associations between various adverse health outcomes and exposures to ambient air pollutants. This is accomplished by analyzing daily air pollution concentrations and aggregated counts of adverse health events over a geographical region via a Poisson log-linear model under the assumption of constant overdispersion. In this paper, we develop covariate-dependent Bayesian generalized Poisson and negative binomial models to account for potential time-varying overdispersion. The proposed models are applied to a time series study of daily emergency department visits for respiratory diseases and ozone concentration in Atlanta, Georgia during the period 1999 to 2009. Allowing for covariate-dependent overdispersion results in a reduction in ozone effect standard error, while the ozone-associated relative risk remains robust to different model specifications. Through simulation studies, we similarly found that the standard quasi-Poisson approach can result in larger standard error for the air pollution effect estimate when the constant overdispersion assumption is violated. Our findings suggest that improved characterization of overdispersion may result in more accurate and precise health effect estimates in studies of short-term environmental exposures.

Time series analysis of air pollution and health accounting for covariate-dependent overdispersion

By

Anqi Pan

B.E.

South China University of Technology

2012

Thesis Committee Chair: Howard H Chang, PhD

Reader: Stefanie Ebelt Sarnat, PhD

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 Science in Public Health

in Biostatistics

2017

Acknowledgements

This study was partially supported by USEPA grant R834799 (Southern Center for Air Pollution & Epidemiology) and NIH grant R21ES022795-02. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the USEPA and NIH. Further, USEPA and NIH do not endorse the purchase of any commercial products or services mentioned in the publication.

Table of Contents

1. Introduction

Epidemiologic studies in the last two decades have consistently reported positive short-term associations between ambient air pollution and various adverse health outcomes, including mortality, hospital admissions, and emergency department visits [1–3]. The majority of these studies employed a time series study design to estimate associations between daily ambient air pollutant concentrations and aggregated counts of adverse health events via Poisson log-linear models. Overdispersion in the health outcome is common and is routinely accounted for by scaling the standard error of the health effect estimate when performing inference. However, to our knowledge, all previous studies assumed that the dispersion parameter to be constant across time.

Results from time series studies have provided crucial scientific evidence for setting regulatory air quality standards worldwide. Previous methodological work has focused mainly on confounder selection [4–5], exposure measurement error [6–7], and distributed lag effects [8–9]. The main objective of this paper is to investigate time-varying overdispersion in time series analysis of air pollution and health. A better characterization of overdispersion may further improve the accuracy and precision of health effect estimates. We are also interested in identifying variables that are associated with overdispersion that may offer additional insights on the role of environmental risk factors on mortality and morbidity. For example, the knowledge of when additional variation in adverse health outcome is expected can be used for health service management and emergency preparedness.

Several models have been widely utilized to accommodate overdispersion in count data. Examples include generalized Poisson, negative binomial, zero-inflated, and zero-altered models. Covariate-dependent dispersion modeling, also known as double generalized linear model (DGLM), was first proposed by Symth [10]. DGLM and its extensions have been utilized mainly in insurance claims applications [11–14]. This modeling approach has not been applied to air pollution and health studies, or more broadly in environmental epidemiology. In this paper, we investigate covariate-dependent Bayesian generalized Poisson and negative binomial models; these two models are attractive because they are frequently used in public health and both have a single dispersion parameter [15, 16]. We applied the proposed models to an analysis of ground-level ozone and emergency department (ED) visits for respiratory diseases in Atlanta, Georgia.

The remainder of this paper is organized as follows. Section 2 describes the exposure and health data of the motivating Atlanta time series study. Section 3 describes the standard Poisson log-linear model, the proposed covariate-dependent overdispersion models, and estimation procedure. Section 4 describes two simulation studies examining impacts on health effect estimation when overdispersion is incorrectly assumed to be constant across time. Section 5 presents results of the Atlanta data analysis and discussion appears in Section 6.

2. Atlanta Emergency Department Visit and Air Quality Data

We obtained individual-level ED visit data from two sources for the 20-county Atlanta metropolitan area, GA, USA. For the period 1999 to 2004, patient records were obtained directly from 41 of 42 hospitals; and for the period 2005 to 2009, data from the same hospitals were obtained from the Georgia Hospital Association. Individual-level ED visits for respiratory diseases were aggregated to daily counts. Respiratory-related ED visits were identified using primary International Classification of Diseases 9th Revision (ICD-9) diagnostic codes of upper respiratory infection (ICD-9 codes: 460–465, 466.0, 477), bronchiolitis (ICD-9 codes: 466.1, 466.11, 466.19), pneumonia (ICD-9 codes: 480–486), chronic obstructive pulmonary disease (ICD-9 codes: 491, 492, 496), and asthma/wheeze (ICD-9 codes: 493, 786.07).

Daily population-level exposure to ambient ozone pollution (8-hour maximum average) was derived by combining measurements from 27 ozone monitors in Georgia and outputs from the numerical model Community Multi-Scale Air Quality (CMAQ) [17]. CMAQ is a chemical transport model that simulates ozone concentrations at a 12km gridded resolution using meteorology data, inventory of emission sources, and state-of-the-art knowledge on atmospheric physics and chemistry. Even though CMAQ provides complete spatial and temporal coverage, its simulations are known to exhibit bias and need to be calibrated by observations. We utilized results from a previously developed calibration method that optimally combines temporal variation in monitoring measurements and spatial variation in CMAQ simulations [18]. This calibration method has been shown to have high prediction performance in cross-validation experiment with an R^2 of 0.87. We used estimated ozone concentrations at a 12km gridded resolution to calculate daily population-weighted averages using tract-level population estimates from the US census in order to better reflect population-level exposure in a time-series design [19]. Local meteorological conditions from the Atlanta Hartsfield International Airport.

3. Methods

3.1 Time Series Analysis of Air Pollution and Health

We first describe the standard Poisson log-linear model to examine associations between daily levels of air pollution and daily counts of adverse health outcomes [20]. The basic form of the model for our ozone and ED visit analysis is given by:

$$Y_t \sim Poisson\ (\mu_t)$$

$$Log(\mu_t) = \alpha + \beta x_t + f(temp_t) + g(dewpt_t) + h(time_t)$$

where Y_t is the count of the outcome of interest on day t and x_t is the corresponding ozone exposure. The parameter of interest, β, represents the log relative risk of the short-term (acute) association between exposure and the health outcome. The model also includes smooth functions of temperature ($temp_t$) and dew-point temperature ($dewpt_t$) to control for short-term effects of meteorology and a smooth function of calendar date ($time_t$) to control for long-term and seasonal trends. Overdispersion is accounted by assuming $\mathrm{Var}(Y_t) = \phi\mu_t$ and inferences is conducted via a quasi-likelihood approach.

For our motivating analysis of ozone and ED visits for respiratory disease, the exposure of interest is defined as the population-weighed 3-day moving average (of 0-, 1-, and 2-day lags) of daily ozone concentration. Effects of meteorology were modeled using natural cubic splines with 6 equidistant knots for both moving averages (of 0-, 1-, and 2-day lags) of daily average temperature and dew point temperature. Time trend was modeled using natural cubic splines with monthly knots. The model also included the following additional covariates: indicator variables for day-of-the-week, holidays and the warm season (March to October), as well as indicators for the entry and exit of hospitals over time, which had influence on ED visit counts.

3.2. Time Series Analysis with Covariate-dependent Overdispersion

We propose to account for covariate-dependent overdispersion under either the generalized Poisson (GP) regression or the negative binomial (NB) regression setting. The likelihood function of Y_t for GP is given by:

$$f(y_t, \mu_t, \omega_t) = (1 - \omega_t)\lambda_t \frac{\{(1 - \omega_t)\mu_t + \omega_t y_t\}^{y_t - 1}}{y_t!} e^{-\{(1-\omega_t)\mu_t + \omega_t y_t\}}$$

where $\mu_t = E(Y_t)$ is specified similarly as in the Poisson model with a log link function. Parameter $\omega_t \in [0,1]$ captures time-varying overdispersion, which we parameterize using the cumulative distribution function of the standard normal distribution:

$$\omega_t = \Phi(\theta_t)$$

$$\theta_t = \boldsymbol{Z}_t \boldsymbol{\eta}$$

where θ_t is the Z score and $\boldsymbol{Z}_t$ is a vector of covariates (including an intercept) with corresponding regression coefficient vector $\boldsymbol{\eta}$. The likelihood function of Y_t for NB is given by:

$$f(y_t, \mu_t, \gamma_t) = \binom{y_t + \gamma_t - 1}{y_t} \left(\frac{\gamma_t}{\gamma_t + \mu_t}\right)^{\gamma_t} (1 - \frac{\gamma_t}{\gamma_t + \mu_t})^{y_t}$$

where $\mu_t = E(Y_t)$ and parameter $\gamma_t > 0$ describes overdispersion, which we transform using the natural log function to include time-varying covariates.

GP and NB can both account for overdispersion in count data. However, the amount of dispersion in GP is independent of the mean, while this is not true of NB. Specifically, for GP models, $Var(Y_t) = \mu_t(1 - \omega_t)^{-2}$, and the amount dispersion is proportional to ω_t. The Poisson distribution is a special case of GP with $\omega = 0$. For NB models, the variance is given by $Var(Y_t) = \mu_t\left(1 + \frac{\mu_t}{\gamma_t}\right)$. Hence overdispersion is inversely proportional to γ_t, and is a function of the mean.

3.3 Estimation

Parameter estimation was conducted via Markov chain Monte Carlo (MCMC) under a Bayesian framework in R version 3.3.1 [21]. For both GP and NB models, we assumed a normal distribution with mean zero and large variance (1×10^6) for each regression coefficient in the mean $E(Y_t)$ model and in the dispersion model. The vectors of regression coefficients in the mean and the dispersion model were updated as blocks iteratively using random walk Metropolis-Hastings algorithm where the acceptance parameters were tuned to be around 25%. For model comparison, we used the deviance information criterion (DIC). DIC is defined as DIC=$\overline{D} + p_D$, where $\overline{D}$ is a measure of model fit and p_D is a penalty for model complexity [22]. In Bayesian hierarchical models, p_D also represents the effective parameters. Smaller values of DIC suggest better model fit. We discarded the first 5000 MCMC samples and inference was conducted using a total of 20,000 posterior samples after burn-in.

4. Simulation Studies

We conducted two simulation studies to evaluate estimation performance of the log relative risk for air pollution (parameter β) when time-varying overdispersion is ignored in a time-series analysis. We repeatedly generated time series of ED visit counts following either the GP or the NB model with various covariate-dependent overdispersion specifications. Using the observed exposure and confounder variables, we assumed the true effect of ozone to have a log relative risk of 0.04 per inter quartile range (IQR) increase in concentration. Regression coefficients for all other confounders (e.g. meteorology and time splines) were set at the estimated values from real data. We then compared estimates of β obtained using the standard approach, i.e. via a quasi-Poisson model (QP), and the true data-generating model. Estimates were compared based on bias, root mean square error (RMSE), 95% confidence/posterior interval coverage rate, and average standard errors/posterior standard deviation across 2,000 simulations.

In the first simulation study, we defined GP and NB overdispersion models based on the analysis of actual Atlanta ED visits and ozone data. Specifically, the covariate vector $\boldsymbol{Z}$ in dispersion models included an intercept, indicator for holidays, indicator for weekday, 3-day moving average temperature and a smooth function of calendar date modeled using natural cubic splines with 5 degrees for freedom. We also considered dispersion models with and without an ozone effect. Results for the first simulation study are shown in Table 1. To assess the amount of overall dispersion induced by the time-varying dispersion model across days, we used the dispersion parameter estimated from a QP model. We found that in the presence of time-varying overdispersion, estimates of ozone health effect from QP models were associated with larger bias, RMSE, and average standard error. The higher

RMSE is mainly driven by increase in standard error (about 20%). Hence, the common practice of scaling the variance of the estimated coefficient by a constant dispersion parameter may reduce statistical power considerably if time-varying overdispersion is present. In addition, we did not find evidence of poorer estimation performance if the exposure of interest (ozone concentration) also affects overdispersion.

In the second simulation study, we specified the dispersion model to induce various degrees of overdispersion using temperature as a covariate (i.e. $\theta_t = \eta + \kappa temp_t$). The goal was to examine whether larger degree of time-varying overdispersion impacts estimation performance more severely when the standard QP approach is used. Temperature was chosen because it is moderately correlated with ozone concentration (Pearson correlation = 0.68). We kept the intercept fixed and evaluated 3 different κ values for temperature in both GP and NB models. Results of the second simulation study are shown in Table 2. For GP models, the average standard error and RMSE for ozone estimates were consistently higher than when QP models were used; the relative difference also increases as dispersion increases. Differences in bias and coverage rate were minimal. Results for NB models are similar to GP models; however, we did not observe an increasing trend in relative difference in average standard error and RMSE as dispersion increases.

5. Application to Atlanta Emergency Department Visit and Ozone

Asthma and respiratory ED visits during the entire study period (1999-2009) totaled 1,536,907 with a mean of 384 visits per day. Daily ozone exposure, defined as the 3-day moving population-averaged, had a mean concentration of 43.8 ppb, a standard deviation of 17.4 ppb, and an IQR of 24.9 ppb.

We first examined various GP and NB covariate-dependent overdispersion models. Estimated log RR, 95% posterior interval, and DIC are given in Supplementary Table S1 and Table S2. For both GP and NB models, DIC is the smallest for an overdispersion model that includes the following covariates: indicator variable for holidays, indicator variable for weekday, 3-day moving average of temperature and a smooth function of calendar date modeled using cubic linear splines with 5 degrees of freedom. Including ozone concentration in both the GP and the NB overdispersion model further reduced DIC. We found that including covariates in the overdispersion resulted in a decrease in DIC value of 580 for the NB model and 300 for the GP model, indicating that modeling time-varying overdispersion resulted in better model fit. Overall, NB models outperformed GP models in terms of DIC.

Estimated associations between ED visits for respiratory diseases and 3-day moving average ozone concentration from different modeling approaches are given in Table 3. Associations are presented as log relative risk (RR) per IQR increase in ozone concentration. From the standard quasi-Poisson log-linear model, the estimated log RR is 0.042, with a standard error 0.007 and a 95% CI (0.028, 0.055). This estimate is nearly identical to those estimated from the Bayesian GP and NB model where overdispersion was assumed constant (i.e. intercept only models). Accounting for covariate-dependent overdispersion resulted in some smaller log relative risks, especially for GP models. Most notably, the standard error associated with the ozone effect reduced considerably (about 30% for the model with the smallest DIC). Table S3 gives the estimated regression coefficients and 95% posterior intervals for the overdispersion model. For both GP and NB models, we found that weekday, temperature, and ozone concentration were associated with a reduction in overdispersion, while holidays were associated with higher overdispersion. Figure S1 and Figure S2 show the estimated temporal trends in overdispersion. Both figures show slight general increasing trends in overdispersion from year of 1999 to 2009.

6. Discussion

We investigated covariate-dependent overdispersion under either the generalized Poisson (GP) regression or the negative binomial (NB) regression setting for conducting time series analysis of air pollution and health data. Through simulation experiments and real data analysis, we found that the standard quasi-Poisson approach can lead to larger standard errors when overdispersion is not constant in time. In our ozone and ED visit analysis, the standard quasi-Poisson approach resulted in a 40% increase in the log RR standard error. We chose this particular application due to the strong association observed in previous epidemiologic studies; however, statistical power may be impacted considerably for other pollutants and health outcomes. Environmental epidemiologic studies are typically observational and aim to estimate small health effects. As researchers continue to conduct large population-based time series studies to estimate acute associations between various environment exposures and health outcomes, our findings suggest that improved characterization of overdispersion may result in more accurate and precise health effect estimates. Recently, the time series approach has also been applied to estimate health effects of temperature [24], extreme rainfall [25], and pollen [26]. We chose to conduct Bayesian inference to fully account for various sources of uncertainty, particularly for the overdispersion parameters. In practice, these models can be implemented in software packages such as JAGS and WinBUGS, and are recommended when large overdispersion is observed.

Several aspects of modeling covariate-dependent overdispersion warrant further investigation, especially in the context of health effect estimation. First, we only considered count models that account for overdispersion. While, the generalized Poisson distribution can be parametrized to accommodate underdispersion, estimation becomes more challenging due to the constraints on the dispersion parameter. Second, GP and NB models can be viewed as Poisson mixture distributions, and other count data distributions maybe be of interest to better characterize various health data. For example, GP exhibits heavier tails when the first two moments of GP and NB are fixed [23]. Third, our time-varying overdispersion is completely driven by covariates. It is also possible to incorporate random effects in the dispersion model to better capture unexplained heterogeneity. Finally, we only compared different model specifications for the dispersion because the mean model is well established from previous health analyses. Joint selection of the mean and the dispersion model to adequately control for unmeasured confounders requires further investigation.

References

1. Castellsague, J., Sunyer, J., Saez, M., & Anto, J. M. (1995). Short-term association between air pollution and emergency room visits for asthma in Barcelona. *Thorax*, *50*(10), 1051-1056.
2. Ji, M., Cohan, D. S., & Bell, M. L. (2011). Meta-analysis of the association between short-term exposure to ambient ozone andrespiratory hospital admissions. *Environmental Research Letters*, *6*(2), 024006.
3. Peel, J. L., Tolbert, P. E., Klein, M., Metzger, K. B., Flanders, W. D., Todd, K., ... & Frumkin, H. (2005). Ambient air pollution and respiratory emergency department visits. *Epidemiology*, *16*(2), 164-174.
4. Dominici, F., McDermott, A., & Hastie, T. J. (2004). Improved semiparametric time series models of air pollution and mortality. *Journal of the American Statistical Association*, *99*(468), 938-948.
5. Peng, R. D., Dominici, F., & Louis, T. A. (2006). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(2), 179-203.
6. Peng, R. D., & Bell, M. L. (2010). Spatial misalignment in time series studies of air pollution and health data. *Biostatistics*, *11*(4), 720-740.
7. Chang, H. H., Peng, R. D., & Dominici, F. (2011). Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error. *Biostatistics*, kxr002.
8. Gasparrini, A. (2014). Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in medicine*, *33*(5), 881-899.
9. Heaton, M. J., & Peng, R. D. (2014). Extending distributed lag models to higher degrees. *Biostatistics*, *15*(2), 398-412.
10. Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47-60.
11. Boucher, J. P., & Davidov, D. (2011). On the importance of dispersion modeling for claims reserving: An application with the Tweedie distribution. *Variance*, *5*(2), 158-172.
12. Smyth, G. K., & Jørgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bulletin*, *32*(01), 143-157.
13. De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data* (Vol. 10). Cambridge: Cambridge University Press.
14. Frees, E. W., Lee, G., & Yang, L. (2016). Multivariate frequency-severity regression models in insurance. *Risks*, *4*(1), 4.
15. Ntzoufras, I. (2011). *Bayesian modeling using WinBUGS* (Vol. 698). John Wiley & Sons.
16. Joe, H., & Zhu, R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biometrical Journal*, *47*(2), 219-229.
17. Byun, D., & Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews*, *59*(2), 51-77.

18. Friberg, M. D., Zhai, X., Holmes, H. A., Chang, H. H., Strickland, M. J., Sarnat, S. E., ... & Mulholland, J. A. (2016). Method for fusing observational data and chemical transport model simulations to estimate spatiotemporally resolved ambient air pollution. *Environmental science & technology*, *50*(7), 3695-3705.
19. Ivy, D., Mulholland, J. A., & Russell, A. G. (2008). Development of ambient air quality population-weighted metrics for use in time-series health studies. *Journal of the Air & Waste Management Association*, *58*(5), 711-720.
20. Peng, R. D., Dominici, F., & Louis, T. A. (2006). Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(2), 179-203.
21. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
22. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583-639.
23. Joe, H. & Zhu R. (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Journal of biopharmaceutical statistics*, *16*(4), 463-481.
24. Gasparrini, A., Guo, Y., Hashizume, M., Kinney, P.L., Petkova, E.P., Lavigne, E., Zanobetti, A., Schwartz, J.D., Tobias, A., Leone, M., Tong, S., Honda, Y., Kim, H., & Armstrong, G.B. (2015). emporal Variation in Heat-Mortality Associations: A Multicountry Study. *Environmental health perspectives*, 123 (11), 1200-7.
25. Carlton, E.J., Eisenberg, J.N.S., Goldstick, J., Cevallos, W., Trostle, J., & Levy, K. (2014). Heavy rainfall events and diarrhea incidence: the role of social and environmental factors. *American Journal of Epidemiology*, 179(3), 344-352.
26. Darrow, L.A., Hess, J., Rogers, C.A., Tolbert, P.E., Klein, M., & Sarnat, S.E. (2012). Ambient pollen concentrations and emergency department visits for asthma and wheeze. *Journal of Allergy and Clinical Immunology*, 130 (3), 630-638.

Table 1. Estimation results for the ozone log relative risk using standard quasi-Poisson (QP) model when data are simulated from generalized Poisson (GP) or negative binomial (NB) models with time-varying overdispersion that are either dependent or independent of ozone level.

Model	Ozone Independent				Ozone Dependent			
	GP		NB		GP		NB	
	True	QP	True	QP	True	QP	True	QP
Bias (per IQR × 100)	0.10	0.10	0.09	0.10	-0.02	-0.02	0.08	0.09
RMSE (per IQR × 100)	0.61	0.69	0.59	0.69	0.60	0.68	0.58	0.67
95% Interval Coverage	0.93	0.96	0.94	0.95	0.94	0.96	0.94	0.96
Average SE (per IQR × 100)	0.58	0.71	0.57	0.70	0.58	0.71	0.57	0.70
Average overdispersion	3.90		3.80		3.88		3.77	

The dispersion model includes indicator variables for weekday, indicator variable for holiday, 3-day moving average of temperature and a smooth function of calendar date with 5 degrees of freedom.

RMSE: root mean squared error; SE: standard error or posterior standard deviation, IQR: interquartile range

Table 2. Estimation results for the ozone log relative risk using standard quasi-Poisson (QP) model when data are simulated from generalized Poisson (GP) or negative binomial (NB) models with varying degrees of overdispersion due to daily temperature.

	Average overdispersion	1.16		4.46		14.22	
		True	**QP**	**True**	**QP**	**True**	**QP**
GP	Bias (per IQR × 100)	0.03	0.05	0.05	0.05	0.14	0.10
	RMSE(per IQR × 100)	0.36	0.39	0.67	0.73	1.22	1.30
	95% Interval Coverage	0.96	0.95	0.94	0.96	0.94	0.96
	Average SE(per IQR × 100)	0.37	0.39	0.65	0.76	1.20	1.36
	Average overdispersion	1.23		4.74		16.78	
		True	**QP**	**True**	**QP**	**True**	**QP**
NB	Bias (per IQR ×100)	-0.01	-0.01	0.02	0.03	-0.01	-0.01
	RMSE (per IQR × 100)	0.39	0.39	0.69	0.77	1.42	1.49
	95% Interval Coverage	0.93	0.96	0.94	0.96	0.93	0.95
	Average SE (per IQR × 100)	0.38	0.40	0.67	0.79	1.36	1.48

RMSE: root mean squared error; SE: standard error or posterior standard deviation, IQR: interquartile range.

Table 3. Estimated ozone log relative risk per interquartile range increase in daily emergency department visits for asthma and other respiratory diseases from different modeling approaches.

Model	Covariate in Dispersion	**Ozone Estimate**	**SE**	**95% CI**	**DIC**
QP	NA	0.042	0.007	(0.028 - 0.055)	NA
	Intercept only	0.041	0.007	(0.026 - 0.056)	40315
GP	No Ozone*	0.037	0.006	(0.026 - 0.049)	39806
	With Ozone*	0.037	0.005	(0.027 - 0.048)	39788
	Intercept only	0.041	0.007	(0.026 - 0.056)	39959
NB	No Ozone*	0.038	0.005	(0.028 - 0.049)	39697
	With Ozone*	0.041	0.005	(0.030 - 0.052)	39670

Model for dispersion also includes indicator variable for weekday, indicator variable for holiday, 3-day moving average of temperature and a smooth function for calendar date.

SUPPLEMENTARY MATERIALS

Figure S1. Temporal effect in Overdispersion for GP

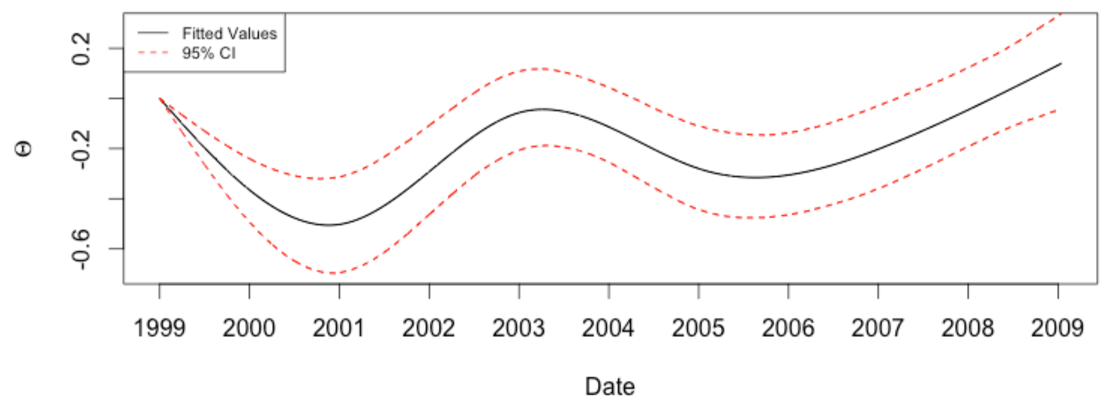

Figure S2. Temporal effect in Overdispersion for NB

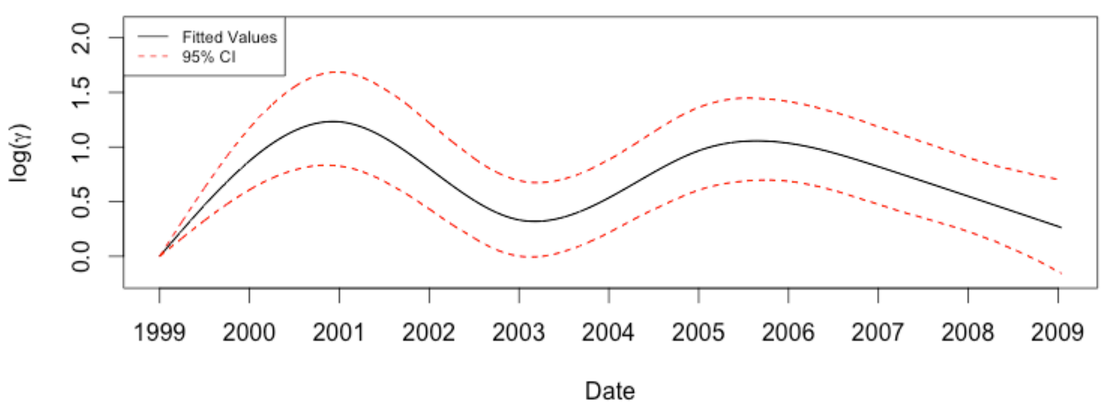

Table S1. Posterior summaries for the ozone log relative risk estimated by generalized Poisson regression with different overdispersion models.

Model	Overdispersion Model	Mean	SD	95% Posterior interval	DIC
GP	Intercept only	0.0017	0.0003	(0.0011 - 0.0024)	40315
	Weekday	0.0015	0.0003	(0.0010 - 0.0020)	40319
	Warm season	0.0014	0.0002	(0.0009 - 0.0019)	39910
	Holiday	0.0016	0.0002	(0.0012 - 0.0020)	40281
	Temperature	0.0013	0.0002	(0.0009 - 0.0017)	39952
	Dew-point Temperature	0.0015	0.0003	(0.0009 - 0.0020)	40047
	Date	0.0018	0.0002	(0.0014 - 0.0020)	40227
	Ozone	0.0014	0.0003	(0.0010 - 0.0020)	40006
	Weekday, Holiday, Temperature, Date	0.0015	0.0002	(0.0011 - 0.0020)	39806
	Weekday, Holiday, Temperature, Date, Ozone	0.0015	0.0002	(0.0011 - 0.0019)	39788

Covariates above denotes the following, Weekday: indicator variable for weekday, Warm season: indicator variable for warm season in Atlanta, Holiday: indicator variable for federal holidays, Temperature: 3-day moving average of daily average temperature, Dew point Temperature: 3-day moving average for dew point temperature, Date: a smooth function for calendar date modeled using cubic linear splines with 5 degrees of freedom, and Ozone: 3-day moving average of population-weighted ozone.
Mean: Posterior mean; SD: Posterior standard deviation.

Table S2. Posterior summaries for the ozone log relative risk estimated by Negative Binomial regression with different overdispersion models.

Model	Overdispersion Model	Mean	SD	95% Posterior interval	DIC
NB	Intercept only	0.0017	0.0003	(0.0011 - 0.0021)	39959
	Weekday	0.0016	0.0003	(0.0011 - 0.0021)	39957
	Warm season	0.0015	0.0003	(0.0010 - 0.0019)	39774
	Holiday	0.0017	0.0002	(0.0013 - 0.0022)	39938
	Temperature	0.0015	0.0002	(0.0010 - 0.0019)	39824
	Dew-point Temperature	0.0016	0.0002	(0.0011 - 0.0020)	39858
	Date	0.0018	0.0003	(0.0011 - 0.0024)	39902
	Ozone	0.0016	0.0003	(0.0012 - 0.0021)	39850
	Weekday, Holiday, Temperature, Date	0.0015	0.0002	(0.0011 - 0.0020)	39697
	Weekday, Holiday, Temperature, Date, Ozone	0.0016	0.0002	(0.0012 - 0.0021)	39670

Covariates above denotes the following, Weekday: indicator variable for weekday, Warm season: indicator variable for warm season in Atlanta, Holiday: indicator variable for federal holidays, Temperature: 3-day moving average of daily average temperature, Dew point Temperature: 3-day moving average for dew point temperature, Date: a smooth function for calendar date modeled using cubic linear splines with 5 degrees of freedom, and Ozone: 3-day moving average of population-weighted ozone.
Mean: Posterior mean; SD: Posterior standard deviation.

Table S3. Posterior regression estimates and 95% posterior intervals for the defined GP and NB models.

	GP		NB	
Covariate	**Estimate**	**95% Posterior interval**	**Estimate**	**95% Posterior interval**
Intercept	1.428	(1.232, 1.632)	2.364	(1.936, 2.779)
Weekday	-0.108	(-0.174, -0.040)	0.288	(0.140, 0.436)
Holiday	0.497	(0.356, 0.631)	-1.094	(-1.435, -0.769)
Temperature	-0.017	(-0.021, -0.014)	0.026	(0.019, 0.033)
Ozone	-0.006	(-0.009, -0.003)	0.007	(0.001, 0.014)

Covariates above denotes the following, Weekday: indicator variable for weekday, Holiday: indicator variable for federal holidays, Temperature: 3-day moving average of daily average temperature, and Ozone: 3-day moving average of population-weighted ozone.