Supplementary Material for

Long-term air pollution exposure impact on COVID-19 morbidity in China

Yihan Wu[1], Qingming Zhan[1*], Qunshan Zhao[2]

1 School of Urban Design, Wuhan University,8 Donghu South Road, Wuhan 430072, China.
2 Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, 7 Lilybank Gardens, Glasgow, G12 8RZ, UK.

Introduction of Poisson regression and negative binomial regression

Poisson regression method is often applied to model count data occurring in different geographic units, in a Poisson model, the probability that an area will experience y events is estimated by

$$P(Y=y)=exp(-\lambda)\frac{\lambda^y}{y!} \quad y=0,1,2,\ldots \quad \lambda>0 \tag{1}$$

Where, Y represents the event count at geographic unit i during a certain period of time, y denotes a realization of Y. λ is an event occurrence intensity parameter calculated using the total number of events divided by the total number of geographic units. In a Poisson process, it is assumed that the expected value and the variance of the response variable equal its mean, thus

$$Var(Y)=E(Y)=\lambda \tag{2}$$

Nonetheless, since events are usually unevenly distributed across a region, some studies suggest the intensity parameter λ should be adjusted for this spatial heterogeneity, thereby allowing each geographical unit to have its own intensity parameter λ_i (Cameron and Trivedi, 1986; Beck and Tolnay, 1995;). Under the conditional Poisson assumption, the probability that the ith area will experience y number of events is

$$P(Y_i=y \mid X)=exp(-\lambda_i)\frac{\lambda_i^y}{y!} \qquad y=0,1,2,\ldots\lambda_i>0 \tag{3}$$

And it is assumed λ_i is a function of multiple explanatory variables, X_{ij}, that describe each geographical unit

$$E(Y_i \mid X) = \lambda_i = exp\left(\sum_{j=1}^{k} \beta_j X_{ij}\right) \tag{4}$$

Where β_j is the estimated effect parameter for explanatory variable X_j.

Because in the presence of over-dispersed count variables, the standard errors of estimated coefficients predicted by a Poisson regression are often underestimated (Cameron and Trivedi, 1986; McCullagh and Nelder, 1989). A regression method that relaxes the mean=variance constraint is thus needed. Based on eq.(4), a negative binomial model can be expressed as

$$E(Y_i \mid X) = \tilde{\mu}_i = exp\left(\sum_{j=1}^{k} \beta_j X_{ij} + \varepsilon_i\right) = \mu_i\, G \tag{5}$$

Where $\tilde{\mu}_i$ is the true count for area i, μ_i is a function of various explanatory variables, and the effect coefficient β_j is estimated using a maximum likelihood approach. and G= exp(ε_i), which is a random variable that follows a Gamma distribution. The probability density function of G is

$$f_G(\kappa) = \frac{\kappa^{(1-\alpha)/\alpha} e^{-\kappa/\alpha}}{\alpha^{1/\alpha}\, \Gamma(1/\alpha)} \qquad \kappa > 0 \tag{6}$$

Where $1/\alpha$ and κ represent the shape parameter and the scale parameter of the Gamma distribution, respectively.

The distribution of Y_i around $\tilde{\mu}_i$ is negative binomial, therefore, in a negative binomial model, the probability that geographical unit i will experience y_i events is given by

$$P(Y_i = y_i) = \int_0^{\infty} P(Y_i = y_i \mid \tilde{\mu}_i) f(\kappa) d\kappa = \int_0^{\infty} \frac{(\mu_i \kappa)^{y_i} e^{-\mu_i \kappa}}{y_i!} \frac{\kappa^{(1-\alpha)/\alpha} e^{-\kappa/\alpha}}{\alpha^{1/\alpha}\, \Gamma(1/\alpha)} d\kappa$$

$$= \mu_i^{y_i} \frac{\Gamma(y_i + (1/\alpha))}{y_i!\, \Gamma(1/\alpha)} \frac{(1/\alpha)^{1/\alpha}}{(\mu_i + 1/\alpha)^{y_i + (1/\alpha)}} \qquad y = 0,1,2,3 \ldots, \alpha, \kappa > 0 \tag{7}$$

The negative binomial distribution has a mean of μ_i and a variance of

$$Var[Y_i] = \mu_i + \alpha \mu_i^2 \tag{8}$$

As can be seen from eq.(8), NB distribution approaches Poisson distribution as α gradually approaches 0.

The marginal effect of the jth covariate X_j on the expected outcome is given by

$$\delta_j = \left(\frac{1}{\tilde{\mu}_i}\right) \frac{\partial \tilde{\mu}_i}{\partial X_j} = \beta_j \tag{9}$$

It means if we increase one unit of X_j and holding other explanatory variables as constants, the relative rate of change of $\tilde{\mu}_i$ will be β_j .

And the incidence rate ratio (IRR) in a NB model can be derived from exp(β_j), which suggests with one unit increase of X_j from the average of X_j, we expect to see exp(β_j) times the incident events as those with the average of X_j. The reason why exp(β_j) can be interpreted as the incidence rate ratio is explained by UCLA Statistical Consulting Group.

Reference

Beck, E.M. and Tolnay, S.E. (1995). Analyzing Historical Count Data: Poisson and Negative Binomial Regression Models. *Hist. Method.* 28: 125-131. https://doi.org/10.1080/01615440.1995.9956360

Cameron, A.C. and Trivedi, P.K. (1986). Econometric Models Based on Count Data. Comparisons and Applications of Some Estimators and Tests. *J. Appl. Econom.* 1: 29-53. https://doi.org/10.1002/jae.3950010104

Introduction to SAS. UCLA: Statistical Consulting Group, https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/, Last Access: 30 June, 2020.

McCullagh, P. and Nelder, J.A. (1989). Generalized linear models. Chapman and Hall.

Table S1. The single-pollutant model, sensitivity analysis and Poisson model results for COVID-19 incidence rate ratio and CI(95%) associated with per μg/m^3 increase of $PM_{2.5}$, PM_{10}, NO_2, CO, SO_2, and O_3

Regression Model	Pollutant	Regression coefficient	p value	Incidence Rate Ratio(IRR)	95% CI
Single-pollutant model	$PM_{2.5}$	0.02	<0.001	1.02	(1.01-1.03)
	PM_{10}	0.01	<0.05	1.01	(1.00-1.01)
	SO_2	0.01	0.32	1.01	(0.99-1.02)
	NO_2	0.05	<0.001	1.05	(1.03-1.06)
	CO	0.1	0.68	1.11	(0.66-1.88)
	O_3	-0.02	<0.05	0.98	(0.96-0.99)
Exclude cities in Hubei	$PM_{2.5}$	0.01	<0.05	1.01	(1.00-1.02)
	PM_{10}	0	<0.05	1.00	(1.00-1.01)
	SO_2	0.02	<0.05	1.02	(1.00-1.03)
	NO_2	0.04	<0.001	1.04	(1.03-1.06)
	CO	0.21	0.35	1.24	(0.79-1.97)
	O_3	-0.01	0.06	0.99	(0.97-1.00)
Exclude number of doctors	$PM_{2.5}$	0.02	<0.001	1.02	(1.01-1.03)
	PM_{10}	0	0.08	1.00	(1.00-1.01)
	SO_2	0.01	0.42	1.01	(0.99-1.02)
	NO_2	0.04	<0.001	1.05	(1.03-1.06)
	CO	0.07	0.78	1.07	(0.65-1.82)
	O_3	-0.02	<0.05	0.98	(0.96-0.99)
Exclude per capita GDP and average wage	$PM_{2.5}$	0.02	<0.001	1.02	(1.01-1.03)
	PM_{10}	0	<0.05	1	(1.00-1.01)
	SO_2	0.01	0.45	1.01	(0.99-1.02)
	NO_2	0.04	<0.001	1.04	(1.03-1.06)
	CO	0.04	0.85	1.05	(0.65-1.70)
	O_3	-0.02	<0.05	0.98	(0.96-0.99)
Exclude all weather variables	$PM_{2.5}$	0.01	0.08	1.01	(1.00-1.02)
	PM_{10}	-0.01	<0.05	1	(0.99-1.00)
	SO_2	-0.01	0.29	0.99	(0.98-1.01)
	NO_2	0.03	<0.001	1.03	(1.01-1.05)
	CO	-0.32	0.20	0.73	(0.43-1.25)
	O_3	-0.03	<0.001	0.97	(0.95-0.98)
Exclude percentage of people over 65-year old	$PM_{2.5}$	0.02	<0.001	1.02	(1.01-1.03)
	PM_{10}	0	0.07	1.00	(1.00-1.01)
	SO_2	0.01	0.28	1.01	(0.99-1.03)
	NO_2	0.04	<0.001	1.04	(1.03-1.06)
	CO	0.07	0.78	1.07	(0.64-1.82)
	O_3	-0.02	<0.05	0.98	(0.96-0.99)
Exclude mobility	$PM_{2.5}$	0.02	<0.05	1.02	(1.01-1.04)
	PM_{10}	0	0.45	1.00	(1.00-1.01)

variables	SO_2	-0.03	<0.001	0.97	(0.95-0.98)
	NO_2	0.04	<0.001	1.04	(1.02-1.06)
	CO	0.10	0.79	1.11	(0.52-2.36)
	O_3	-0.01	0.43	0.99	(0.97-1.01)
Poisson Model	$PM_{2.5}$	0.04	<0.001	1.04	(1.04-1.04)
	PM_{10}	0.01	<0.001	1.01	(1.01-1.02)
	SO_2	-0.02	<0.001	0.98	(0.98-0.98)
	NO_2	0.05	<0.001	1.05	(1.05-1.05)
	CO	0.16	<0.001	1.18	(1.10-1.26)
	O_3	-0.02	<0.001	0.98	(0.98-0.98)

Table S2. The two-pollutant model results for COVID-19 incidence rate ratio and CI(95%) associated with per μg/m^3 increase of $PM_{2.5}$, PM_{10}, NO_2, CO, SO_2, and O_3

	Pollutant	coefficient	p value	IRR	95% CI
$PM_{2.5}$	PM_{10}	0.04	<0.001	1.05	(1.02-1.07)
	SO_2	0.02	<0.001	1.02	(1.01-1.03)
	NO_2	-0.01	0.26	0.99	(0.98-1.01)
	CO	0.02	<0.001	1.03	(1.01-1.04)
	O_3	0.02	<0.001	1.02	(1.01-1.03)
PM_{10}	$PM_{2.5}$	-0.02	<0.05	0.99	(0.97-1.00)
	SO_2	0.01	<0.05	1.01	(1.00-1.01)
	NO_2	-0.01	<0.05	0.99	(0.98-1.00)
	CO	0.01	<0.05	1.01	(1.00-1.01)
	O_3	0.01	<0.05	1.01	(1.00-1.01)
SO_2	$PM_{2.5}$	0	0.83	1.00	(0.98-1.02)
	PM_{10}	0	0.70	1.00	(0.99-1.02)
	NO_2	-0.01	0.31	0.99	(0.98-1.01)
	CO	0.01	0.38	1.01	(0.99-1.03)
	O_3	0.01	0.38	1.01	(0.99-1.02)
NO_2	$PM_{2.5}$	0.05	<0.001	1.05	(1.03-1.08)
	PM_{10}	0.06	<0.001	1.06	(1.04-1.08)
	SO_2	0.05	<0.001	1.05	(1.03-1.07)
	CO	0.06	<0.001	1.06	(1.04-1.08)
	O_3	0.04	<0.001	1.04	(1.03-1.06)
O_3	$PM_{2.5}$	-0.02	<0.05	0.98	(0.97-1.00)
	PM_{10}	-0.02	<0.05	0.98	(0.97-1.00)
	SO_2	-0.02	<0.05	0.98	(0.97-1.00)
	CO	-0.02	<0.05	0.98	(0.96-0.99)
	NO_2	-0.01	0.46	0.99	(0.98-1.01)
CO	$PM_{2.5}$	-0.48	0.10	0.62	(0.35-1.11)
	PM_{10}	-0.15	0.58	0.86	(0.48-1.55)
	SO_2	0	1.00	1.00	(0.57-1.79)
	O_3	-0.07	0.79	0.93	(0.55-1.59)
	NO_2	-0.73	<0.05	0.48	(0.28-0.82)

Table S3. A comparison of negative binomial regression and Poisson regression results.

	Regression coefficient (NB model)	Regression coefficient (Poisson model)
Intercept	2.54***	1.56***
pol	0.02***	0.04***
den	0.15*	0.04***
col	0.02	0.43***
old	-0.16*	-0.16***
mal	-0.07	-0.05***
Lgdp	-0.10	0.26***
Lwag	0.14	-0.18***
doc	0.30***	-0.05***
smo	0.06	0.27***
sat	-0.39	-1.68***
wat	-0.05	0.63***
sdt	1.91***	3.37***
wdt	-0.83*	-1.28***
sws	0.03	0.62***
wws	0.07	-0.77***
mov	1.08***	0.35***
icm	-0.01	-0.08***
AIC	2906	22294

*:p-value <0.05, **: p-value<0.01, ***: p-value<0.001

Figure S1. The annual mean concentrations of $PM_{2.5}$, PM_{10}, SO_2, NO_2, and O_3 between 2015 and 2019 in mainland China