Supplemental Content

CONCENTRATION-RESPONSE FUNCTIONS FOR SHORT-TERM EXPOSURE AND AIR POLLUTION HEALTH EFFECTS

Mieczysław Szyszkowicz

Population Studies Division, Health Canada, Ottawa, Canada.

mietek.szyszkowicz@canada.ca

In this presentation, for illustrative purposes, mortality data in London, UK, are used. These data are publicly available (1) and are associated with the publication on the conditional Poisson regression (2). The considered exposure is ambient ozone concentration.

Here we summarize some statistical methods which, in general, are hybrid methods in relation to the standard case-crossover (CC) technique. We propose the combination of these methods for constructing new approaches.

In the domain of air pollution health effects there is the very important factor of time. In the standard case-crossover technique, the effect of time is controlled by the time-window, which is one month. In the time-stratified CC technique, the event and controls are in the same (one) month (3). In one modified approach, it was proposed to use a hierarchical structure of the form <year:month:day-of-week> and treat such clusters as repeated measurements for a given day-of-week (dow) (4). We can observe that the smallest possible cluster of such type has the following form <year:2-week:dow>. In such a case, we have only two data points. This is enough to determine the corresponding slope for two exposures and two counts. In this situation we group 14 days (2 weeks) in separate groups: (1,2), (3,4), (4,6), etc. Here the pair (1,2) means the first and second week in the first 14 days of the year, and so on. Thus we realize many regressions on the data points (d1,r1) and (d2,r2), where dose (d) and response (r) are sorted by the magnitude of the dose, but not by time. In such a constructed scheme there is no time to model. The time was "cut into pieces" by the hierarchical structure of the constructed strata. The pairs considered can be chained in such a way that the structures (1, 2), (2, 3), (3, 4) etc., can be build (5). In this case we don't have gaps between clusters. Codes 3 and 4 show such constructions for the London mortality data for the years 2002-2006.

The standard CC method also "cuts" the time into pieces. In the case of the time-stratified CC method one piece is one month. The authors of (2,6) proposed to use such a structure to realize the conditional Poisson models as an alternative of the case-crossover. In this situation, we also consider the counts rather than individual events.

In a standard CC method, the following model is considered: log(OR) = Beta*z + covariates, where OR is an odds ratio, Beta is the coefficient to be estimated, z is the concentration of air pollutant. In this study, we propose the controlled case-crossover (CCC) model of the form: log(OR) = Beta*g(z) + covariates, where g(z)=f(z)*LWF(mu, tau), f(z) is a transformation function, and LWF (mu, tau)=1/(1+exp[(mu-z)/(r*tau)]).

The data related to mortality in London (2002-2006) and ozone exposure were used. Firstly, we tested the CC methods (in the form of conditional Poisson) with three time-windows: month, 2-week, and 2-week chained. Table 1 summarizes the results of this approach.

Table 1. The estimated values for three different time-windows (month, 2-week, and 2-week chained). London (2002-2006) mortality data.

Window	Variables	Coefficient	Std. Error	z-value	Pr(>|z|)	AIC
Month	ozone10	0.0033849	0.0015969	2.120	0.034	15237
	temperature	0.0041932	0.0008042	5.214	1.85e-07	
2-week	ozone10	0.0022550	0.0020340	1.180	0.268	15512
	temperature	0.0062060	0.0011050	5.617	1.94e-08	
Chained	ozone10	0.0025296	0.0014287	1.771	0.077	31041
	temperature	0.0064092	0.0007869	8.145	<2e-16	

As we see, the results are different, even by qualifications. Only the CC method using one month as cluster results in statistically significant coefficient for ozone (Beta, p-value = 0.034). Two other methods put more weight on temperature (larger Beta for temperature). It should be noted that the AIC criterion cannot be used here to compare the results, since they were obtained on different numbers of measurement points. In the case of the chained clusters the amount of data is almost twice as large as in the original.

Table 2 summarizes the results for three time-windows: month, 2-week, and 2-week-chained.

Table 2. The estimated parameters of the concentration-response functions for three time-windows. London (2002-2006) mortality data.

Method	Beta	Std. Error	mu	tau	AIC
z/month	0.0630015	0.0057681	10.35	0.097	15127.16
Log(z)/month	0.3198311	0.0293258	10.63	0.094	15127.52
z/2-week	0.0220780	0.0027240	7.77	0.002	15446.78
Log(z)/2-week	0.1448650	0.0157920	8.83	0.002	15427.48
z/chained	0.0472555	0.0043454	9.78	0.079	30923.60
Log(z)/chained	0.2404422	0.0221763	10.10	0.080	30924.22

Two kinds of the models are tested, with f(z)=z and f(z)=log(z), and the parameters mu and tau are determined by minimalizing the AIC value (Code 2). Figure 1 shows the results, left-hand panel – for f(z)=z, right-hand panel - for f(z)=log(z), and rows a-c for three time-windows: month, 2-week, and 2-week-chained, respectively.

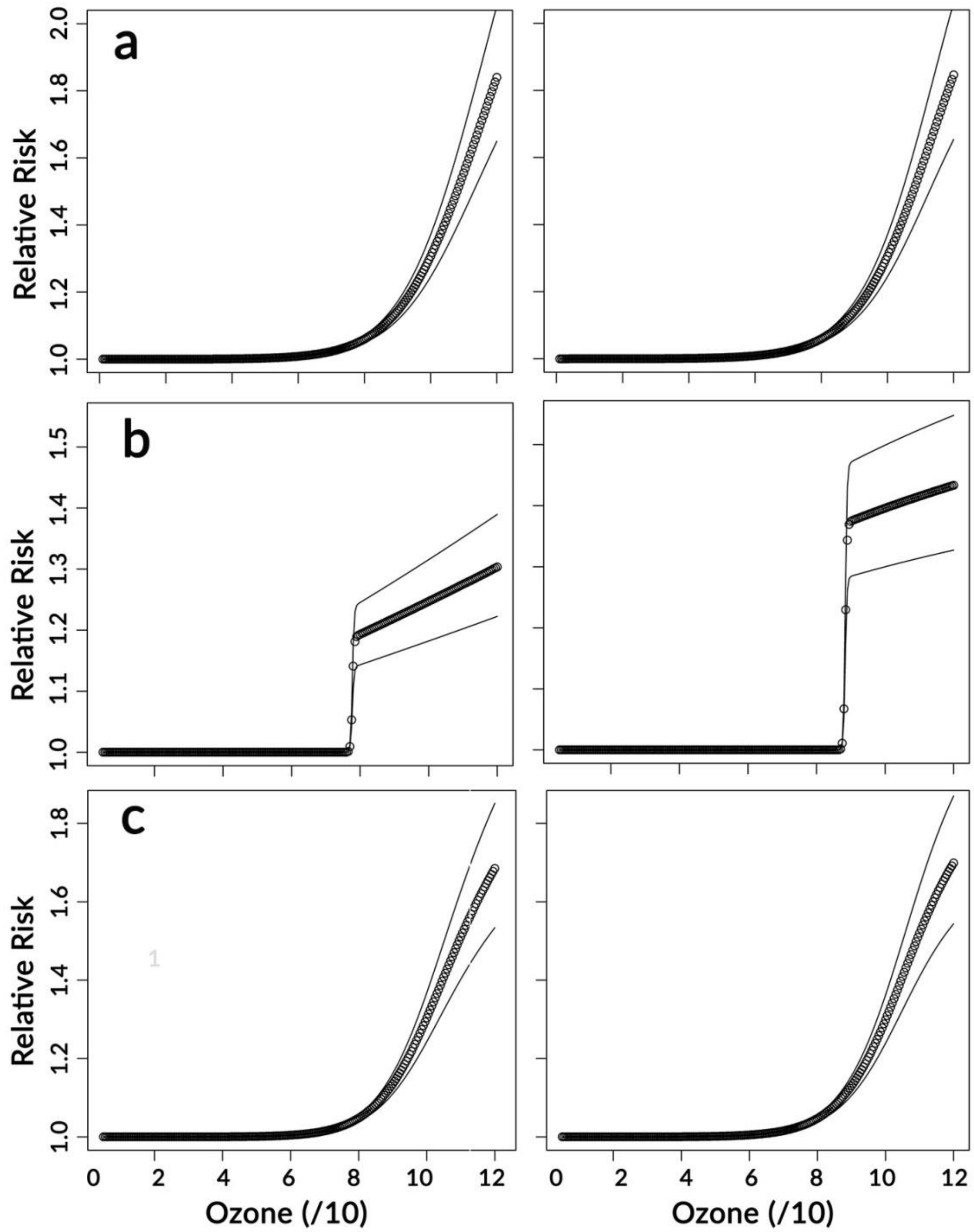
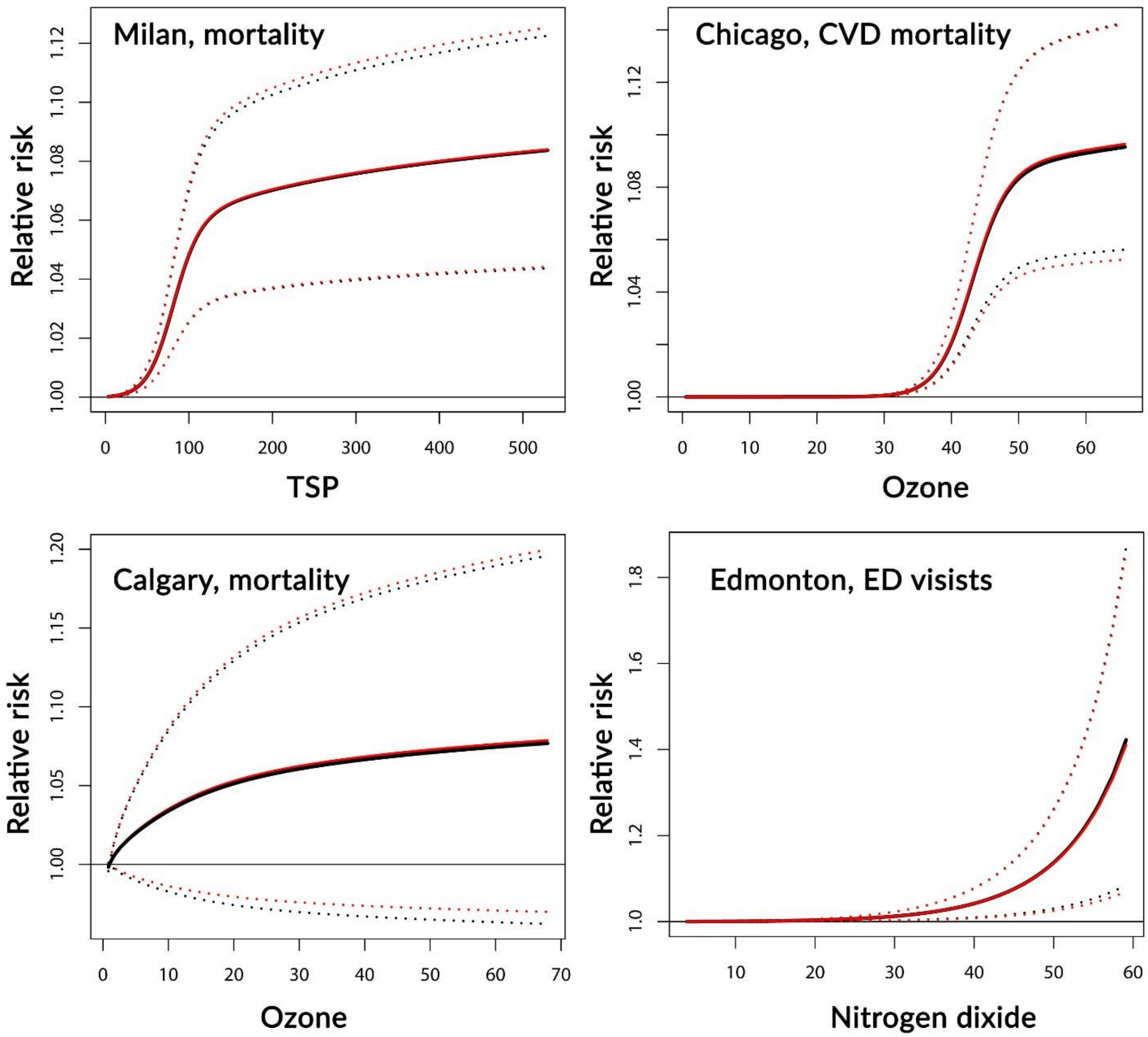


Figure 1. Concentration-response modelled by using the logistic weighting function and conditional Poisson with three strata (a: one month, b: 2 week, C: 2-week chained). London (2002-2006) mortality data.

As the process is rather complex (minimization of nonlinear functions) in the case of 2-week clusters, the estimated tau is very small (tau=0.002). Consequently, we see a jump in the obtained curve-response function. In this situation the software failed. In this case, for fixed tau=0.1, and search done only to estimate the location parameter mu, the results are as follows: mu=10.6 and AIC= 15434.5. Thus, the obtained AIC value is smaller than previously estimated (AIC =15446.8), where mu and tau were both used to minimize AIC.

Figure S1. The figure illustrates the fitted models for the Milan, Chicago and Calgary mortality data, and for ED visits in Edmonton. (Please see the main text of this work for the details). The used red color indicates the values for the estimated models (RR and the 95% confidence intervals). In black colour are shown the results from the 1000 simulations. They are presented: mean value of the estimated values of RR, and two percentiles, 0.025 and 0.975.



References

1. London mortality data: https://github.com/gasparrini/2014_armstrong_BMCmrm_Codedata.

2. Armstrong BG, Gasparrini A, Tobias A. Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. BMC Med Res Methodol 2014; 14: 122. doi: 10.1186/1471-2288-14-122.

3. Szyszkowicz M, Tremblay N. Case-crossover design: air pollution and health outcomes. Int J Occup Med Environ Health. 2011;24(3):249-55. doi: 10.2478/s13382-011-0034-y.4. 4. 4.

Szyszkowicz M. Use of generalized linear mixed models to examine the association between air pollution and health outcomes. Int J Occup Med Environ Health 2006; 19: 224-7. https://doi.org/10.2478/v10001-006-0032-7

5. Szyszkowicz M, Burr WS. The use of chained two-point clusters for the examination of associations of air pollution with health conditions. Int J Occup Med Environ Health 2016; 29: 613-22. doi: 10.13075/ijomeh.1896.00379.

6. The codes in Stata and R: http://www.ag-myresearch.com/2014_armstrong_bmcmrm.html

7. Szyszkowicz M, Rowe B Respiratory Health Conditions and Ambient Ozone: A Case-Crossover Study. Insights in Chest Dis 2015; 1:9.

Supplemental Content (codes)

The examples of the used codes

CONCENTRATION-RESPONSE FUNCTIONS FOR SHORT-TERM EXPOSURE AND AIR POLLUTION HEALTH EFFECTS

Mieczysław Szyszkowicz

Population Studies Division, Health Canada, Ottawa, Canada.

mietek.szyszkowicz@canada.ca

1. The case-crossover algorithm

```
*******************************************************;
*** The program realizes the CC method using the subroutine phreg (in SAS);
*** The controls are defined as: +/- 7 days in the same (one) month;
*** The time-stratified design is realized. Day-of-week (dow) is adjusted by design.
*** Other controls +/- D, D=1,2,3, etc. and the variable dow should be in the model.
Reference 7.
*******************************************************;

/*
Two files of data are given:
 (1) HEALTH – records with the individual events (date, case, age, sex,...)
 example row: 20-May- 2014, 493, 35, female ... , i.e. date, asthma ICD-9 code, age, sex.
 (2) AIRPOLL - with air pollution and weather data
 example: 20-May-2014, 23.4, 7.5, 13.5,...  - date, ozone and PM values, temperature.
 CASEnum keeps an identification of each case and the corresponding controls
The HEALT and AIRPOLL data can be stored in one file.
*/

***CREATE CASE-CROSSOVER FILES;
***STEP 1: CASE PERIODS;
***For all cases CASE=1;
DATA  FCASE;
 SET HEALTH;
 RETAIN CASEnum;
 IF _N_=1 THEN CASEnum=0;
 CASEnum = CASEnum + 1; CASE = 1;
 KEEP CASEnum CCDate CASE;
*CCDate keeps event date;
*CCDate = mdy(month, day, year);
RUN:
*********************************;
***STEP 2: CONTROL PERIODS;
***Create referent intervals based on the time-stratified CC design;
***Time-stratified: day of week adjusted by design (+/-7 days);
```

```
***For all controls CASE=0;
DATA FCONT;
 SET FCASE; BY CASEnum;
 KDate = CCDate;
 CCmonth = MONTH(KDate);
 RETAIN CCDate CCMonth;
*** Create eight candidates as control days;
DO k=1 To 8;
  IF k=1 THEN CCDate=KDate+7;
  IF k=2 THEN CCDate=KDate+14;
  IF k=3 THEN CCDate=KDate+21;
  IF k=4 THEN CCDate=KDate+28;
  IF k=5 THEN CCDate=KDate-7;
  IF k=6 THEN CCDate=KDate-14;
  IF k=7 THEN CCDate=KDate-21;
  IF k=8 THEN CCDate=KDate-28;
        CASE=0;
*** CASE= 0 indicates control;
***include only the control days in the same month as case;
  IF MONTH(CCDate) = CCmonth THEN OUTPUT;
***maximum 4 control days;
 END;
KEEP CASEnum CCDate CASE;
RUN:

***STEP 3: BRING ALL CASES/CONTROLS TOGETHER;
DATA FCC;
 SET FCASE FCONT;
PROC SORT;
BY CCDate;
RUN;

***STEP 4: MERGE CASE-CONTROL data with POLLUTION/METO data;
DATA CCPOLL;
MERGE FCC AIRPOLL;
BY CCDate;
TIME=1;
IF CASE=0 THEN TIME=2;
RUN;


***STEP 5: Use the conditional logistic regression model (phreg);

PROC PHREG NOSUMMARY data=CCPOLL;
MODEL TIME*CASE(0)= APOLL Temperature Humidity
*** Usually Meto variables are represented by splines;
 /TIES=DISCRETE RL;
STRATA CASEnum;
RUN;
```

2. The algorithm to fit curve-response with the logistic weighting function.

```
# For the used notation and the data please see:
#https://github.com/gasparrini/2014_armstrong_BMCmrm_Codedata
# The software is presented in the R statistical language.
# FIT A CONDITIONAL POISSON MODEL WITH A YEAR X MONTH X DOW STRATA
# AND THE LOGISTIC WEIGHTING FUNCTION (find mu and tau).
funAIC <- function(param){
xs <- data$ozone10 # The exposure could be lagged: Lag(xs,M)
mu <- param[1]; tau <- param[2] #or keep tau=0.1 or 0.2
rtau = tau*diff(range(xs,na.rm=TRUE))
data$XT <- xs/(1+exp((mu-xs)/rtau))
#data$XT <- log(xs)/(1+exp((mu-xs)/rtau))  # if LOG is used

modGnm <- gnm(numdeaths ~ XT + temperature, data=data, family=poisson,
                              eliminate=factor(stratum))

 print(summary(modGnm))
 return(extractAIC(modGnm)[2]) }
#The end of the function definition

mu=0.0; tau=0.1  # initial values,
#M=2;  #Use M if needed lags of air pollution, temperature, etc.
nlminb(c(mu,tau), funAIC)  # Call minimization

#determine mu and tau, also beta and its SE
#print(summary(modGnm)) gives the estimation of beta
```

3. Create 2-week clusters

```
#2 week clusters (a,b),(c,d), etc.
weed365=c(rep(seq(1,26), each =14),27)
weed366=c(rep(seq(1,26), each =14),27,27)
day2w =c(weed365,weed365,weed366,weed365,weed365)
data$week2 <- day2w  #data is the used data file
#In the program replace month level by week2 (=2 weeks) strata
```

4. Create chained clusters

```
#Chained 2 week clusters (a,b),(b,c),(c,d), etc...
weed365=c(rep(seq(1,26), each =14),27)
weed366=c(rep(seq(1,26), each =14),27,27)
day2n1 =2* c(weed365,weed365,weed366,weed365,weed365)-1
data$week2 <- day2n1 # odd numbers(2n-1)

d365 <- c(rep(0,7),rep(seq(1,25), each=14),rep(26,7), 27)
d366 <- c(rep(0,7),rep(seq(1,25), each=14),rep(26,7), 27,27)
day2n = 2*c(d365,d365,d366,d365,d365)  # even numbers (2n)
keepd <- data
keepd$week2 <- day2n
data <- rbind(data, keepd) # It allows to have (a,b),(b,c),… clusters.
#In the program replace month level by week2 (=chain) strata
# The end of the Supplementary Information.
```

Use the presented algorithms on your own risk.