Original Article Pol J Public Health, Vol. 132 (2022): 67-72

MIECZYSŁAW SZYSZKOWICZ[1], EUGENIUSZ PORADA[2]

A common concentration-response function based on the results applying lags

Abstract

Introduction. Estimating the impact of short-term exposure on health outcomes needs knowledge of both the profile and magnitude of the relative risks. This motivates constructions of practical and reliable concentration-response functions (C-RFs).

Aim. To define a practical method of finding concentration-response parametric function whose adjustable parameters can be tuned by data-driven well established routines.

Material and methods. Mortality data for the period from 1987 to 2015 (10,592 consecutive days) in Montreal, Canada, are used for illustrative purposes. Exposure to ambient ozone measured by its concentration levels is considered health risk. Concentration-response function is built using statistical modelling, conditional Poisson regression, natural spline technique, and a rudimentary hierarchical data clustering. The case-crossover design is applied to fit the model of C-RF to the mortality data consisting of daily counts of non-accidental deaths.

Results. Log-linear models of the concentration-response functions were computed for the concentrations and cofactors data lagged by 0 to 7 days; the results were statistically significant within this range of lags. The effectiveness of fitting was confirmed by reliable statistical tests. Digital routines were created to perform all computational tasks; software codes (written for R software platform) are included. The C-RF specifying the current responses to the cumulative exposure in several previous days can be obtained from the responses to lagged exposures.

Conclusions. The proposed method of concentration-response function estimation appears practical and effective in producing reliable results. The constructed function is a parametric and monotonic non-decreasing.

Keywords: Ambient air pollution; concentration-response function; log-linear modelling.

DOI: 10.2478/pjph-2022-0014

INTRODUCTION

The dependency of health risks on short-term acute changes in the concentration of ambient air pollutants continues to be a subject of intensive research [1,2]. In the majority of such studies, the dependency is represented by concentration-response functions (C-RFs), understanding the response as the relative health risk. The C-RFs are useful in forming the data-driven public health policies. It is practical to have parametric C-RFs with data-adjustable shapes rather than responses to tabulated data of health risk factors [3-6].

The models of C-RFs suffer from the following often-observed phenomenon: they produce a risk which may decrease with an increase in the pollution level. It is primarily a consequence of lags in the cause-to-effect processes, particularly for health effects when only a relatively small number of days have high levels of air pollution. In these not frequent days, an increase in health problems could not be observed yet, if ever. In reality, the health conditions are mainly driven by acute changes in the pollution levels in the range close to the average levels. The problem can also reside in design, when not enough care is given to the non-linear dynamics of health effect versus air pollution.

In this context, it is important to find the generic shape of the functions associating the lagged factors to responses and to make the functions sufficiently flexible by means of parameterization. In this paper, as well as in the general practice, the generic shape is determined by the logarithmic dependence of health relative risks on the levels of air pollution. The extreme level occurrences are too rare to allow for consistent measurements and analysis. So in such a situation the logarithmic risks are weighted by logistic function becoming scaled and consequently are small, far away from average levels. In the scenario where concentrations are in the range close to the average, parameters are tuned to warrant a good fit to the reality described by data.

[1] Environmental Health Science and Research Bureau, Health Canada, Ottawa, Canada
[2] An independent consultant

MATERIAL AND METHODS

Data

The dataset consists of daily counts of all non-accidental deaths in Montreal, Canada, in the period from January 1987 to December 2015. In this period of 10,592 days, 441,272 deaths were recorded. This work does not conduct epidemiological analysis of the data. These data already have been extensively used in another analysis related to ozone and mortality [7]. Here, they are exploited for illustration of the proposed methodology of modeling the dose-response function. In the model, ground-level ozone (O_3) appears as ambient air pollutant; its concentrations were measured by the National Air Pollution Surveillance Program (NAPS, [8]) and reported as records of 8-hour daily maximum concentrations. Daily average of ambient temperature and of relative humidity figure in the model as weather factors.

Statistical Model

This study design is based on time-stratified case-crossover technique [9] and conditional Poisson regression [10,11] on the hierarchical clusters grouping the same days of week. The bottom clusters group the 4 or 5 days having the same hierarchical label <year:month:day of week>. They form the principal factor influencing the health response to the pollution and weather factors. In the syntax of the R software tool [12,13], the model of delayed health response to an air pollutant dose has the form FitModel = gnm(Mortality ~ AP + ns (T,3) + ns(RH,3), family=poisson, eliminate=factor(cluster)).

The model specifies the health variable Mortality representing daily count of deaths. AP (air pollution), T (temperature), and RH (relative humidity) enter the model lagged by the same number of days. The weather factors appear in the model as natural splines (with 3 degrees of freedom) fitting the weather variables. Application of the spline technique mediates influences of outliers in data and gives better-fitting models. The computations were performed by the R statistical software. For the model above, R implements the log-linear (LL) approach: the model fitting routine returns, among else, coefficient Beta, the slope of the logarithmic increase of relative risk (RR) with an increase in the pollution level Dose. In that notation, RR=exp(Beta*Dose); this is a popular scheme of estimating risks associated with changes in air pollution [14]. This paper follows the same scheme, but it adds flexibility to the concentration-response function and makes it more informative.

This paper also proposes taking advantage of knowing C-RFs for all relevant exposure lags and constructing the C-RF expressing the current response to the cumulative exposure in the last several days. The model above has been used to compute the response to air pollution caused by ground-level ozone (O_3); the variable representing the levels of O_3 is denoted Z. Traditionally, such a crude variable is submitted into a statistical model in order to obtain the potential relationships between air pollutants and health outcomes. Here, the following transformation of Z is performed in order to put in relief the shape of the response function around average ozone concentrations and tame the response at concentration extremities:

$$T(Z) = f(Z) * LWF(Z),$$

where $f(Z) = \log\left(1 + \frac{Z}{A}\right)$ and LWF is the logistic weighted function (LWF):

$$LWF(Z) = \frac{1}{\left(1 + e^{\frac{\mu - Z}{\tau * r}}\right)}$$

In the above formula, the parameters A, μ, τ control the shape of transformation T. Constant r is the range of Z, μ is said location parameter, and τ decides about the curvature of LWF(Z) [14]. Computing the model consists in determining the values of parameters in view of the goodness of fit of f(Z) to a linear combination of predictors; the case-crossover estimation technique is realized as conditional Poisson regression. The Akaike's information criterion (AIC, see Appendix, Program 1) of goodness of fit is applied. Since the model is computed in log-linear configuration, the final result is the concentration-response function

$$\text{C-}RF(Z) = \exp(\beta * T(Z))$$

parameter β is estimated in the model together with the parameters of the linear combination of predictors.

Once the series of concentration-response functions for lags giving positive statistically significant at p-value <0.05 results is produced, the functions can be aggregated into one C-RF of the same parametric profile. Here, a least square approximation can be applied. The new values of parameters A, μ, τ, and β describe a function associating the current health risk resulting to the cumulative exposure in recent days.

RESULTS

Within the period of the study (10,592 days), 441,272 non-accidental deaths occurred. Among the non-accidental deaths, the most frequent are deaths related to cardiac and respiratory health problems. Figure 1 shows the histograms of the cardiac and respiratory mortality and other non-accidental mortality.

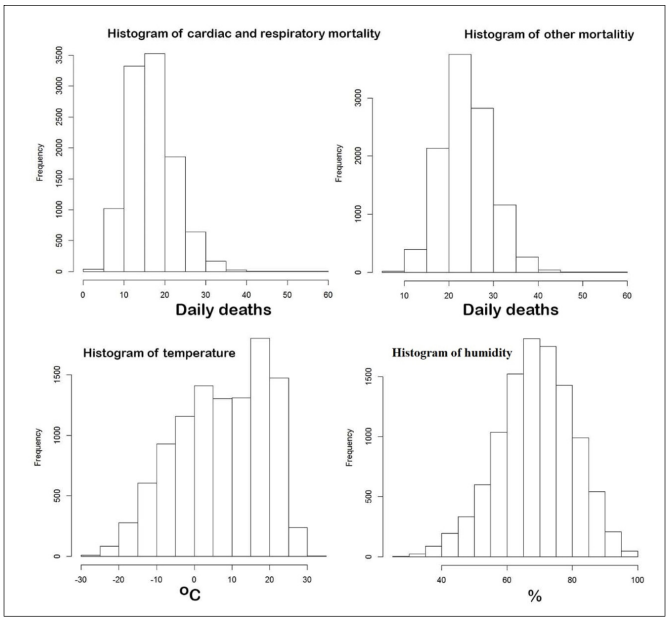

FIGURE 1. Histograms of mortality and two weather factors.

The figure also shows the histograms for temperature and relative humidity. They inform about weather conditions in the study area and period. Table 1 summarizes the results obtained when analysing the lagged exposures. For each considered lag, (here from lag 0 to lag 8) C-RF was fitted and shape parameters A, μ, τ , as well as response magnitude β with its standard error were estimated.

TABLE 1. Parametric descriptions of concentration-response functions for lagged exposure to ozone and the resulting non-accidental mortality in the period 1987-2015 in Montreal, Canada.

C-RF(Z)=exp(β(Z)), β(Z)=Beta*log(1+Z/A)*LWF(Z,τ,μ)					
Lag	**Beta**	**SE**	**μ**	**τ**	**A**
0	7.82	1.87	300.89	0.81	1932.32
1	3886.24	496.56	506.55	0.27	680.10
2	6.06	1.26	78.35	0.06	7040.00
4	0.10	0.03	54.66	0.05	146.56
5	0.18	0.07	99.99	0.25	150.03
6	0.16	0.07	99.95	0.61	150.05
7	0.19	0.07	100.02	0.36	150.03
8	-0.05	0.03	62.84	-0.02	149.26

There are no results for lag 3. Indeed, in some instances of log-linear models created using the case-crossover computational scheme it may happen that the convergence is not reached. This occurred for the case of lag 3.

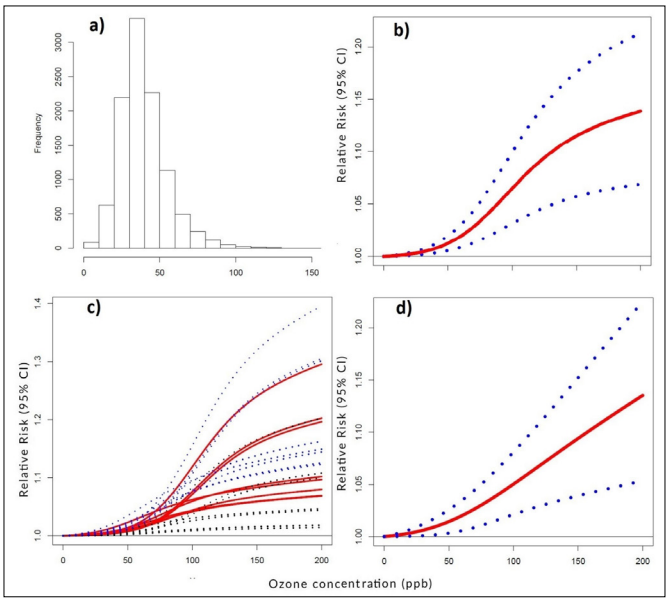

FIGURE 2. Histogram of ambient ozone and the C-RF profiles.

Panel a) in Figure 2. illustrates the histogram of daily concentrations of ground ozone in this study's period and setting. Panel b) shows the C-RF shape obtained when the exposure to ozone is lagged by 2 days. Panel c) shows all the shapes defined by the parameters listed in Table 1, Panel d) shows the C-RF function associating the current aggregated risk to the cumulative exposure in 7 last days. The parametric description of the function is given in Table 2.

TABLE 2. The parameters of the aggregated C-RF shape of associating cumulative exposure to ozone to non-accidental mortality in years 1987-2015 in Montreal, Canada.

Parameters	**Value**	**SE**	**Low Value**	**SE**	**Upper Value**	**SE**
Output	**Input: C-RF(Z)=exp(β(Z)), β(Z)=θ*log(1+Z/A)*LWF(Z,τ,μ)**					
θ	0.06	0.01	0.03	0.01	0.09	0.02
A	27.45	15.46	26.52	25.70	28.31	12.38
μ	84.19	3.46	87.90	5.44	82.81	2.79
τ	0.12	0.01	0.12	0.01	0.12	0.00

DISCUSSION

It is a well established fact that the relative probabilities of discrete outcomes in a Poisson process depend logarithmically on the strength of the stimuli guiding the process. In reality, the stimuli are usually random events forming a Bernoulli process, where strength means count of successes, and the logarithmic dependence is a mathematical fact. With a slight overstretch of the notion of discrete process, an air pollutant measured in ppb and influencing a public health outcome can be considered a Bernoulli process and the measured concentration becomes the strength of influence.

However, the concentration is also a measure of pollution dose taken by a person. The health effect of an increase of the dose depend not only on the increase, but also on the actual measure of the dose. In order to find the region of the dosage where increases have the highest impact, an added flexibility of concentration-response function is needed [15-22]. A variety of parametric modifications of the generic logarithmic shape can be considered, but weighing the profile by the logistic weighted function LWF appears to be the simplest, yet effective, modification. As a variation on a sigmoidal function is of interest the estimated parameters, μ (location) and τ (curvature), well define the C-RF in the region of pollutant concentrations range [14].

Acknowledgements

The authors acknowledge Environment and Climate Change Canada for providing the air pollution data from the National Air Pollution Surveillance (NAPS) network. Parts of this material are based on data and information compiled and provided by the Canadian Human Mortality Database (CHMD). However, the analyses, conclusions, opinions and statements expressed herein are not necessarily those of CHMD. The authors acknowledge Health Canada for supporting this study.

REFERENCES

1. Szyszkowicz M, de Angelis N. Ambient air pollution and emergency department visits in Toronto, Canada. Environ Sci Pollut Res Int. 2021;28(22):28789-96.
2. Szyszkowicz M, Schoen S, de Angelis N. Air Pollution and Emergency Department Visits for Disease of the Genitourinary System. Environ Health Insights. 2021;15:11786302211025360.
3. Hoek G, Boogaard H, Knol A, et al. Concentration response functions for ultrafine particles and all-cause mortality and hospital admissions: results of a European expert panel elicitation. Environ Sci Technol. 2010,44(1):476-82.

4. Pope CA 3rd, Cropper M, Coggins J, Cohen A. Health benefits of air pollution abatement policy: Role of the shape of the concentration-response function. J Air Waste Manag Assoc. 2015;65(5):516-22.
5. Burnett R, Chen H, Szyszkowicz M, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc Natl Acad Sci USA. 2018;115(38):9592-7.
6. Szyszkowicz M. Concentration – response functions for short-term exposure and air pollution health effects. Environ Epidemiol. 2018;2(2):p e011.
7. Vicedo-Cabrera AM, Sera F, Liu C, et al. Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries. BMJ. 2020;368:m108.
8. NAPS. Environment and Climate Change Canada. 2021. National Air Pollution Surveillance Program. Retrieved from http://maps-cartes.ec.gc.ca/rnspa-naps/data.aspx.
9. Janes H, Sheppard L, Lumley T. Case-crossover analyses of air pollution exposure data: referent selection strategies and their implications for bias. Epidemiology. 2005;16(6):717-26.
10. 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(4):224-7.
11. 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.
12. R Core Team. R: A Language and Environment for Statistical Computing; 2021. Retrieved from http://www.r-project.org/.
13. Turner H, Firth D. Generalized nonlinear models in R: An overview of the gnm package. For gnm version 1.1-1, 2020-02-02. [https://cran.r-project.org/web/packages/gnm/vignettes/gnmOverview.pdf.]
14. Nasari MM, Szyszkowicz M, Chen H, et al. A class of non-linear exposure-response models suitable for health impact assessment applicable to large cohort studies of ambient air pollution. Air Qual Atmos Health. 2016;9(8):961-72.
15. Atkinson RW, Yu D, Armstrong BG, et al. Concentration-response function for ozone and daily mortality: results from five urban and five rural U.K. populations. Environ Health Perspect. 2012;120(10):1411-7.
16. Yan M, Wilson A, Bell ML, et al. The Shape of the concentration-response association between fine particulate matter pollution and human mortality in Beijing, China, and its implications for health impact assessment. Environ Health Perspect. 2019;127(6):67007.
17. Ren M, Fang X, Li M, et al. Concentration-response relationship between PM2.5 and daily respiratory deaths in China: A systematic review and metaregression analysis of time-series Ssudies. Biomed Res Int. 2017:5806185.
18. Amoushahi S, Bayat R, Sanaei A, et al. Health and economic impacts of ambient fine particulate matter in Isfahan, Iran. Urban Climate. 2022;41:101048.
19. Gasparrini A. Distributed lag linear and non-linear models in R: The Package dlnm. J Stat Softw. 2011;43(8):1-20.
20. Szyszkowicz M, Burr WS. Distributed lag models: An analysis of Milan mortality data. J Poll Effects Control. 2014;02:1
21. Szyszkowicz M. The concentration-response functions for short-term exposure to ambient air pollution. Pol J Public Health. 2021;131:7-10.
22. Szyszkowicz M. Concentration-Response Functions as an Essence of the Results from Lags. Int J Environ Res Public Health. 2022;19(13):8116.

Corresponding author
Mieczysław Szyszkowicz
Environmental Health Science and Research Bureau, Health Canada
251 Sir Frederick Banting Driveway
Ottawa, Ontario, Canada K1A 0K9
e-mail: mietek.szyszkowicz@hc-sc.gc.ca, mszyszkowicz@yahoo.ca

APPENDIX

```
############ PROGRAM 1 ################################
# Author: M. Szyszkowicz
# Program 1: Fits individual (by lags) C-RF shapes.
#######################################################
library(sme) # To have AIC (Akaike's IC)
library(gnm); library(splines); library(quantmod)
options(digits=6); options(na.action="na.exclude")
#Read the data: health and environmental
datSET <- read.table(„MontrealD.csv",header=TRUE,sep=",")
########################
# CREATE STRATA YR x MONTH x DOW
datSET$Y <- as.factor(datSET$year)
datSET$M <- as.factor(datSET$month)
datSET$W <- as.factor(datSET$dow)
##### Define hierarchical clusters.
datSET$Cluster <- as.factor(datSET$Y:datSET$M:datSET$W)
attach(datSET); nameF="RESLALL.txt"
#############################################################
# Define the function to minimize AIC
funLL <- function(param){
mu <- param[1]; tau <- param[2]; A <- param[3]
rtau= tau*diff(range(xs,na.rm=TRUE))
# Transformation T(Z)
# Other functions f(z): #XT<-sqrt(xs)/(1+exp((mu-xs)/rtau))
XT <- log(1+xs/A)/(1+ exp((mu-xs)/rtau))
#The used model:
modelG <- gnm(MORTALITY ~ XT  + TNS + HNS,
data=datSET, family=poisson, eliminate=factor(Cluster))
# Retrieve the coefficients: Beta, SE, and p-value.
B=unname(summary(modelG)$coefff[1,1])
SE= unname(summary(modelG)$coefff[1,2])
P=unname(summary(modelG)$coefff[1,4])
RES=c(B,SE,mu,tau,extractAIC(modelG)[2]);
# The results sent to the file: collect all iterations
sink(file=nameF); print(RES); sink()
return(extractAIC(modelG)[2])
}##### Function description
##### Define Lag=M; air pollutant (xs)
##### Represent Temperature and relative humidity as ns (*,df=3)
M=1; xs = Lag(O3H8,M)
```

```
T=Lag(TEMP,M);  H=Lag(RHUM,M)
TNS=ns(T,df=3);  HNS=ns(H,df=3)
# Search for the best fit = find the optimal parameters mu, tau, and A
nlminb(c(100,0.1,150.0), funLL); #initial values, function funLL – returns AIC
#### Retrieve the last recorded line of the result (RES)
datX=read.table(nameF);  resT=(tail(datX,1))
B=(resT[1,2]) ; SE=resT[1,3]; mu=resT[1,4]; tau=resT[1,5]
#######################################
### Calculate RR and 95%CI
# Table 1, Lag 1: mu=506.55, tau=0.27, A=680.10
Low=B-1.96*SE; Upp=B+1.96*SE; xp <- sort(O3H8)
rtau= tau*diff(range(xp,na.rm=TRUE))
# Other T(z): gg = sqrt(xp)/(1+exp((mu-xp)/rtau))
gg = log(1+xp/A)/(1+exp((mu-xp)/rtau))
yp=B*gg;  RR=exp(yp); Lyp= Low*gg;  Uyp= Upp*gg
LP = exp(Lyp);  UP= exp(Uyp)
##### Plot the C-RF for a specific lag
plot(xp,RR, ylim=c(min(RR,LP),max(RR,UP)),col="white")
lines(xp,RR, lwd=5,col="red"); abline (h=1)
lines(xp,LP, lwd=4,lty=3, col="blue"); lines(xp,UP, lwd=4,lty=3,col="blue")
########### The END###################################
####&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&&####
##**********************************************##
################## PROGRAM 2 #########################
# Author: M. Szyszkowicz
# Program 2: Fits the global C-RF shape.
#########################################################
library(MASS);  library(minpack.lm)
#Read the data (Table 1)
dataIN <- read.table(„ResLLogzA.csv",header=TRUE,sep=",",); attach(datIN)
# Creat the X-axis; ozone from 0 to 200, Z=O3
xp=seq(0, 200, by=0.1);  x=xp; xa <- xp
N=7; N1=N+1; MIN=2; MAX=-2
#Prepare the data frame by lags 0-7
for (k in 1:N) {
if (k==1){
X <- xa;  TRR=as.vector(X)
TLR=TRR; TUR=TRR}  # for k=1
# The names in dataIN should be as used (TAU, etc.)
rtau= TAU[k]*diff(range(X,na.rm=TRUE))
logit=1/(1+exp((MU[k]-X)/rtau))
XT=log(X/AP[k]+1)*logit; y=BET[k]*XT
yl=(BET[k]-1.96*SEB[k])*XT; yu=(BET[k]+1.96*SEB[k])*XT
rr=exp(y);  rl=exp(yl);  ru=exp(yu)
TRR=cbind(TRR,rr);  TLR=cbind(TLR,rl);  TUR=cbind(TUR,ru)
# Find minimum and maximum values for plot
mi=min(rl)
if( mi<MIN) {MIN=mi}
mx=max(ru)
if (mx>MAX) {MAX=mx}
if (k==1)
{xa=x;ya=rr;yL=rl;yU=ru}
else{
xa=c(xa,x); ya=c(ya,rr); yL=c(yL,rl); yU=c(yU,ru) }
}# # Define frame with X, RR, RRlower, and RRupper (for each lag)
dfram=data.frame(xa,ya); dframe=dfram[order(xa),]
dframL=data.frame(xa,yL); dframLe=dframL[order(xa),]
dframU=data.frame(xa,yU); dframUe=dframU[order(xa),]
################Define plot area (white) #############
plot(X,rr, ylim=c(MIN,MAX), col="white"); abline(h=1)
################ Fit a common C-RF for RRs ######################
#Start values:
r= diff(range(xp,na.rm=TRUE))
A=3.0; T=0.3; M=1; P= 0.1
fitM = nlsLM(dframe$ya~  ((dframe$xa+A)/A)^(T/(1+exp(-(dframe$xa-M)/(r*P)))),
start = list(A=A, T=T, M=M,P=P), data=dframe)
lines(dframe$xa,fitted(fitM), lwd=6,col="red")
# summary(fitM)
#########Fit a common C-RF for lower RRs ##########################
A=3.0; T=0.3; M=1; P= 0.1
fitL = nlsLM(dframLe$yL~  ((dframLe$xa+A)/A)^(T/(1+exp(-(dframLe$xa-M)/(r*P)))),
start = list( A=A,T=T, M=M,P=P), data=dframLe  )
lines(dframe$xa,fitted(fitL),  lwd=6,lty=3,col="blue" )
# summary(fitL)
####################Fit a common C-RF for upper RRs ###################
A=3.0;  T=0.3;  M=1;  P= 0.1
fitU = nlsLM(dframUe$yU~  ((dframUe$xa+A)/A)^(T/(1+exp(-(dframUe$xa-M)/(r*P)))),
```

```
start = list(A=A,T=T, M=M,P=P), data=dframUe  )
lines(dframe$xa,fitted(fitU),  lwd=6,lty=3,col="blue" )
# summary(fitU)
##### To see individual curves on a common plot
for (k in 1:N+1){
lines(X, TRR[,k], lwd=4, col="red") }
abline(h=1)
#####################################################################
for (k in 1:N+1){
lines(X, TLR[,k], lwd=3,lty=3,col="black") }
#####################################################################
for (k in 1:N+1){
lines(X, TUR[,k],lwd=3,lty=3, col="blue") }
##################The obtained C-RF coefficients #### Output from R ########
Middle RR - Parameters:
Estimate Std. Error t value Pr(>|t|)
A 27.447727  15.460105   1.775   0.0759 .
T  0.061867   0.014751   4.194 2.75e-05 ***
M 84.193094   3.464141  24.304  < 2e-16 ***
P  0.118445   0.005193  22.810  < 2e-16 ***
###########################################
Lower RR - Parameters:
Estimate Std. Error t value Pr(>|t|)
A 26.521316  25.703732   1.032   0.3022
T  0.031278   0.012706   2.462   0.0138 *
M 87.895953   5.435640  16.170   <2e-16 ***
P  0.115488   0.007861  14.692   <2e-16 ***
###########################################
Upper RR - Parameters:
Estimate Std. Error t value Pr(>|t|)
A 28.310907  12.378476   2.287   0.0222 *
T  0.093303   0.017420   5.356 8.61e-08 ***
M 82.807825   2.789874  29.682  < 2e-16 ***
P  0.119251   0.004325  27.572  < 2e-16 ***
################### The END ########################
```