The Report Committee for Ni Wang
Certifies that this is the approved version of the following report:

Multivariate GLS meta-analysis on Ambient air pollution and Congenital heart anomalies

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Susan N. Beretvas

John Hixon

Multivariate GLS meta-analysis on Ambient air pollution and Congenital heart anomalies

by

Ni Wang, B.S.; M.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2014

Acknowledgements

At first, I would love to express the deepest appreciation to Professor Beretvas for her valuable suggestions on my report and for being supportive and encouraging during the past years. Moreover, I also want to thank Dr. Hixon for being my reader and sharing his knowledge with me during this semester. At last but not least, my appreciation goes to my family, I know you are always there whenever I need you, thank you for being so supportive and caring. Love you.

Abstract

Multivariate GLS meta-analysis on

Ambient air pollution and Congenital heart anomalies

Ni Wang, M.S.Stat

The University of Texas at Austin, 2014

Supervisor: Susan N. Beretvas

The effects of air pollutants CO, NO_2, O_3, PM_{10} and SO_2 on congenital heart anomalies are represented by the odds ratio of each disease per unit increase in the concentration of each pollutant. In this study, the effects of air pollutants are summarized using multivariate GLS approach with correlation between outcomes being taken into account, where the correlations are sampled from uniform [-1,1]. Meta-analysis conducted here found no statistically significant increase in odds ratio of any disease.

This result is different from what Vrijheid et al. 2011 suggested when correlation is not considered using the same set of data. The difference in conclusions from the two meta-analysis indicates that correlation between outcomes may play an important role when synthesizing effect sizes. Thus, before conduct meta-analysis, a thorough consideration about whether to incorporate the correlation in synthesizing should be given.

Table of Contents

List of Tables viii

List of Figures x

INTRODUCTION 1

METHOD 3

Basic ideas of Meta-analysis 3

Multiple effect sizes within a study 3

Models for multivariate meta-analysis 4

Multivariate fixed effects meta-analysis 5

Multivariate random effects meta-analysis 6

Multivariate mixed effects meta-analysis 7

GLS methods for multivariate meta-analysis 7

Homogeneity test for model selection 9

Estimation of unknown within-study covariance matrix using simulation 11

DATA ANALYSIS AND RESULT 13

Source of Data 13

The effects of CO 14

Fixed effects model 14

Random effects model 18

Mixed effects model 21

The effects of NO2 22

Fixed effects model 23

Random effects model 24

Weighted regression model 24

The effects of O3 25

Fixed effects model 25

Random effects model 27

Mixed effects model 28

The effects of PM_{10} 29

Fixed effects model 29
Random effects model 30
Mixed effects model....32
The effects of SO2 32
Fixed effects model 33
Random effects model 34
Mixed effects model....35
CONCLUSION....37
DISCUSSION....38
REFERENCES 39

List of Tables

Table 2.1 Example giving for mixed effects model 9

Table 3.1. Estimated effect size of CO for each disease and corresponding 95% confidence interval generated from fixed effects model.....17

Table 3.2 Estimated effect size of CO for each disease and corresponding 95% confidence interval generated from random effects model.....20

Table 3.3 Estimated effect size of CO for each disease and corresponding 95% confidence interval generated from mixed effects model.....21

Table 3.4 Estimated effect size of NO2 for each disease and corresponding 95% confidence interval generated from fixed effects model.....23

Table 3.5 Estimated coefficients and their 95% confidence intervals obtained from weighted regression model 25

Table 3.6 Estimated effect size of O3 for each disease and its 95% confidence interval obtained from fixed effects model.....26

Table 3.7 Estimated effect size of O3 for each disease and its 95% confidence interval obtained from random effects model.....28

Table 3.8 Estimated coefficients and their 95% confidence intervals generated by mixed effects model.....28

Table 3.9 Estimated coefficients and their 95% confidence intervals generated by fixed effects model.....30

Table 3.10 Estimated effect size of PM10 for each disease and its 95% confidence interval given by random effects model.....31

Table 3.11 Estimated coefficients and their 95% confidence intervals from mixed effects model.....32

Table 3.12 Estimated effect size of SO2 for each disease and its 95% confidence interval given by fixed effects model.....33
Table 3.13 Estimated effect size of SO2 for each disease and its 95% confidence interval given by random effects model.....35
Table 3.14 Estimated coefficient and its 95% confidence interval given by mixed effects model.....36

List of Figures

Figure 3.1 Distribution of lnOR of CO on each outcome generated from fixed effects model.....17

Figure 3.2 Distribution of lnOR of CO on each outcome generated from random efects model. 19

Figure 3.4 Distribution of lnOR of NO2 on each outcome generated from fixed effects model. 23

Figure 3.6 Distribution of lnOR of O3 on each outcome generated from fixed effects model.....26

Figure 3.7 Distribution of lnOR of O3 on each outcome generated from random effects model. 27

Figure 3.9 Distribution of lnOR of PM10 on each outcome generated from fixed effects model. 29

Figure 3.10 Distribution of lnOR of PM10 on each outcome generated from random effects model. 31

Figure 3.12 Distribution of lnOR of SO2 on each outcome generated from fixed effects model. 33

Figure 3.13 Distribution of lnOR of SO2 on each outcome generated from random effects model. 35

INTRODUCTION

Severe ambient air pollution happening in China has recently drawn attention worldwide. The World's Health Organization (WHO) estimated that seven million people have died as a result of air pollution in 2012 and nearly six million of the deaths have been in South East Asia. Ambient air pollution is found to be mainly connected to cardiovascular disease which has become one of the biggest risk factor negatively impacting global health. More strikingly, it not only impacts the health of human beings today, but may also have a greater impact on human's future.

Animal toxicology literature suggests that adverse reproductive effects (Tsukue et al. 2001; Singh et al. 1988) and heritable gene mutations (Somers et al. 2002; Somers et al. 2004; Samet et al. 2004) can be produced by air pollution exposure. Outside the laboratory, the adverse effects of ambient air pollution on fetus and newborn from maternal prenatal exposure have been demonstrated in a growing number of epidemiology studies (Glinianaia et al. 2004; Maisoner et al. 2004; Sram et al. 2004; Vrijheid et al. 2011). Thus, new concerns are rising. It has been suggested that air pollution may play a critical role in causing congenital anomalies (Ritz et al. 2010; Vrijheid et al. 2011). Congenital anomalies are a main cause of infant mortality and remain the leading cause of disability worldwide (Ritz et al. 2002) which affect about 1 in 33 infants and result in approximately 3.2 million birth defect related disabilities every year. Congenital heart defects are the most common severe congenital anomalies, appearing in 4-8 of 1000 live births (van der Linde et al. 2013)

Major air pollutants that are most frequently studied are carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), Ozone (O_3) and particulate matter 10 (PM_{10}). During the past few decades, several epidemiological studies have focused on the

association between the major air pollutants and congenital heart anomalies. However, their results are not consistent (Ritz et al. 2002; Gilboa et al. 2005; Rankin et al. 2009; Strickland et al. 2009; Hansen et al. 2009; Dolk et al. 2010; Dadvand et al. 2011a; Dadvand et al. 2011b). Only one meta-analysis focused on air pollution and risk of congenital heart anomalies has been found in the literature. It was published by Vrijheid et al. in 2011. In this meta-analysis, the authors found evidence for an effect of ambient air pollution on congenital cardiac anomaly risk. They reported that nitrogen dioxide (NO_2) and sulfur dioxide (SO_2) exposure were related to increase in risk of Coarctation of the aorta and Tetralogy of Fallot, and PM_{10} exposure was related to an increased risk of Atrial septal defects.

When calculating the overall effect of each pollutant on a certain type of anomalies, the above meta-analysis did not consider the possible correlation between different types of congenital heart defects but treated them as independent effect sizes which may threaten the validity of the resulting conclusions. In the current meta-analysis, a new set of summary odds ratio were calculated with the correlations between different effect sizes were taken into account and Generalized least square (GLS) method was employed.

METHOD

BASIC IDEAS OF META-ANALYSIS

Different effect sizes are reported by different research papers for the same treatment but using different samples. Thus, it is difficult to tell what the true effect size might be for a certain treatment and a certain demographics. The major goal of a meta-analysis is to provide the estimated overall effect by combining the results of individual studies taking into account the precision of the study. For studies which are more precise than others, more weight will be given to associated effect size estimate. Instead of computing simple mean of the effect sizes, weighted mean is computed, with more weight given to some studies and less weight given to the others (Cooper et al. 2nd edition).

Multiple effect sizes within a study

If a single effect size is reported by each study, the pooled effect is easily calculated by plug in the formula below:

$$\bar{T}_{\cdot} = \frac{\sum_{i=1}^{n} w_i T_i}{\sum_{i=1}^{n} w_i} \qquad (1)$$

Where $\bar{T}_{\cdot}$ is the synthesized effect size, it can be any type of effect sizes. The common types of them are standardized mean difference, correlation and natural logarithm of odds ratio. And w_i is weight being assigned to the corresponding effect size which is simply the inverse of the variance of that effect size. At last, n is the number of studies under investigation.

For studies report multiple effect sizes, cases have to be treated differently.

If the effect sizes are independent, for example, if the treatment effect (e.g., drug versus placebo) varies by gender, then the treatment effect can be reported separately for

male and for female. The key feature is that the subgroups are independent of each other, so that each provides unique information. For this reason, we can treat each subgroup as separate study.

However, for some studies, the reported multiple effect sizes are not independent. For example, in the study of air pollution and congenital anomalies, for a specific pollutant, given CO, the effect sizes reported are the odds ratio of Coarctation of the aorta (CA), Tetralogy of Fallot (TF), and Atrial septal defects (ASD) for per 1ppm increase in CO. In this case, multiple measures yield more than one treatment versus control effect sizes. Because the effect sizes from the study share the same control group, the estimates of effect sizes are correlated. When synthesizing effect sizes from such studies, the possible dependency among effect sizes should be accounted for. In order to do this, correlations between effect sizes within a study has to be obtained (Cooper et al. 2nd edition, Gleser J et al. 1994, Raudenbush W et al. 1988, Timm H 1999, Dimitris 2011).

MODELS FOR MULTIVARIATE META-ANALYSIS

Multivariate methods for meta-analysis synthesis that recognize the correlation between estimated effect sizes within-study will be employed in this study (Cooper et al 2nd edition). There are two basic models used in multivariate meta-analysis, which include fixed effects model and random effects model. Fixed effects model assume that one true effect size are shared by all studies. The assumption is that variability in studies' effect size estimates are solely a function of sampling error.

On the other hand, random effects model allows the true effect sizes to vary across studies, based on the assumption that the potential moderators underlying would inevitably vary among studies and generate different true effect sizes. Under random effects model, the effect size of each available study can be conceived as an item drawn

from a random sample of all possible effect sizes. The variance under random effects model consists two components. One component is random effects variance which represents the variance that arises as a result of the sampling of studies, each having a unique effect size. Another component is the estimation variance which arises because of the finite sample of subjects each effect size is based upon (Hedges 1983, DerSimonian 1986, Cooper et al. 2nd edition). Both models for multiple outcomes will be introduced in detail below. To make it simple, bivariate model will be given as examples for both multivariate fixed effects and random effects model.

Multivariate fixed effects meta-analysis

The following contents are cited from Mavridis et al. 2011 unless specified.

Suppose we obtain two estimate effect sizes y_{i1} and y_{i2} from each study i. In a fixed effects model, the study-specific estimated effect sizes follow a bivariate normal distribution. For odds ratio, in order to follow the normal distribution, it has to be converted to natural logarithm metric.

$$\binom{y_{i1}}{y_{i2}} \sim MVN\left(\binom{\mu_1}{\mu_2}, \begin{pmatrix} \sigma_{i1}^2 & \rho_i\sigma_{i1}\sigma_{i2} \\ \rho_i\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{pmatrix}\right) \quad (0)$$

Where $\mu=(\mu_1, \mu_2)'$ denotes the vector of means for each outcome, ρ_i is the correlation between outcome 1 and 2 within study i and the covariance matrix for study i is S_i

$$S_i = \begin{pmatrix} \sigma_{i1}^2 & \rho_i\sigma_{i1}\sigma_{i2} \\ \rho_i\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{pmatrix} \quad (3)$$

Fixed effects model can be represented by matrix notation equivalently as

$$y_i = \mu + e_i \quad (4)$$

where e_i denotes a vector of random sampling error.

Multivariate random effects meta-analysis

In random effects model, it assumes that both the overall variation and overall correlation come from two sources which are within-study and between-study. It suggests that

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \begin{pmatrix} \sigma_{i1}^2 & \rho_i \sigma_{i1} \sigma_{i2} \\ \rho_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2 \end{pmatrix} \right) \quad (5)$$

where $\theta=(\theta_{i1},\theta_{i2})'$ denotes the underlying study specific effect sizes which are also normally distributed as

$$\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho_\tau \tau_1 \tau_2 \\ \rho_\tau \tau_1 \tau_2 & \tau_2^2 \end{pmatrix} \right) \quad (6)$$

where is τ_1^2 the between-study variation for effect size 1 and τ_2^2 is the between-study variation for effect size 2. They are also known as heterogeneity or random effects variance. It is calculated based on method of moments suggested by Cooper et al. 2nd edition and other references (Hedges 1983; DerSimonian 1986).

The between-study covariance matrix is Δ, which is

$$\Delta = \begin{pmatrix} \tau_1^2 & \rho_\tau \tau_1 \tau_2 \\ \rho_\tau \tau_1 \tau_2 & \tau_2^2 \end{pmatrix} \quad (7)$$

In random effects model, the overall variation is the sum up of within-study variance σ_i^2 and between-study variance τ_i^2. Overall correlation is the sum up of within-study correlation and between-study correlation. Therefore, the overall covariance matrix for random effect model is

$$\begin{pmatrix} \sigma_{i1}^2 & \rho_i \sigma_{i1} \sigma_{i2} \\ \rho_i \sigma_{i1} \sigma_{i2} & \sigma_{i2}^2 \end{pmatrix} + \begin{pmatrix} \tau_1^2 & \rho_\tau \tau_1 \tau_2 \\ \rho_\tau \tau_1 \tau_2 & \tau_2^2 \end{pmatrix} = \begin{pmatrix} \tau_1^2+\sigma_{i1}^2 & \rho_\tau \tau_1 \tau_2+\rho_i \sigma_{i1} \sigma_{i2} \\ \rho_\tau \tau_1 \tau_2+\rho_i \sigma_{i1} \sigma_{i2} & \tau_2^2+\sigma_{i2}^2 \end{pmatrix} \quad (8)$$

Hence, generally the distribution of effect sizes under random effects model is the combining of equation 6 and 7 as shown below:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \tau_1^2 + \sigma_{i1}^2 & \rho_\tau \tau_1 \tau_2 + \rho_i \sigma_{i1} \sigma_{i2} \\ \rho_\tau \tau_1 \tau_2 + \rho_i \sigma_{i1} \sigma_{i2} & \tau_2^2 + \sigma_{i2}^2 \end{pmatrix}\right) \quad (9)$$

Random effects model can be represented by matrix notation as follows:

$$y_i = \mu + \delta_i + e_i \quad (10)$$

where δ_i is a vector of random effects associated with study i, $\delta_i \sim MVN(0,\Delta)$ and e_i is a vector of random sampling errors associated with study i, $e_i \sim MVN(0,S_i)$. Thus the covariance matrix of y_i is $\Delta+S_i$ (Hedges 1998).

Multivariate mixed effects meta-analysis

If covariates are included in the model, we will have mixed effects model. With $\mu=x_i * \beta$, the model becomes:

$$y_i = X_i\beta + \delta_i + e_i \quad (11)$$

where X_i is a design matrix with observed covariate values for coefficients and ones for intercept, it also contains columns of ones and zeroes which indicate the membership of studies, and β is the vector of coefficients for each of the study (Hedges 1992).

GLS METHODS FOR MULTIVARIATE META-ANALYSIS

Assuming that you are interested in doing a multivariate meta-analysis with p outcomes, the parameters of interest in your study are the synthesized effect sizes. In this case, the uncertainty in the estimated effect sizes represented by the within-study covariance matrix and the between-study covariance matrix should be considered. Various methods exist for estimating such parameters, one of them are Generalized Least Square (GLS) method and this is the one will be employed in my study.

Supposing n studies are investigated, with p outcomes from each. Each outcome can be modeled by a regression model, and the regression models are correlated. The vector of effect sizes y is a np X 1 vector, where the first p elements are from study 1 followed by p elements from next study etc. For fixed and random effects models, synthesized effect sizes are the parameters of interest, thus no covariates are added in the model. Design matrix X can be represented by combination of p x p identity matrix. S is the overall covariance matrix with each within study covariance matrix on diagonal. To simplify, n and p both equal to 2. Given no correlations exists between studies, X, y, and S can be represented below.

$$X = \begin{bmatrix} X1 \\ X2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad y = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix}$$

$$S = \begin{bmatrix} S1 & 0 & 0 \\ & 0 & 0 \\ 0 & 0 & S2 \\ 0 & 0 & \end{bmatrix} = \begin{bmatrix} \sigma_{11}^2 & \rho_1\sigma_{11}\sigma_{12} & 0 & 0 \\ \rho_1\sigma_{11}\sigma_{12} & \sigma_{12}^2 & 0 & 0 \\ 0 & 0 & \sigma_{21}^2 & \rho_2\sigma_{21}\sigma_{22} \\ 0 & 0 & \rho_2\sigma_{21}\sigma_{22} & \sigma_{22}^2 \end{bmatrix} \quad (12)$$

The summary effect size estimated by GLS method is:

$$\hat{\mu} = (X'S^{-1}X)^{-1}X'S^{-1}y \quad (13)$$

For fixed effects model, the covariance matrix S is as illustrated above in the example.

When talking about random effects model, between-study covariance matrix Δ should also be taken into account. The new covariance matrix in random effects model are showing below, again assuming no correlations exist between studies.

$$S = \begin{bmatrix} S1' & 0 & 0 \\ & 0 & 0 \\ 0 \quad 0 & & S2' \\ 0 \quad 0 & & \end{bmatrix} = \begin{bmatrix} \tau_1^2 + \sigma_{11}^2 & \rho_1\sigma_{11}\sigma_{12} & 0 & 0 \\ \rho_1\sigma_{11}\sigma_{12} & \tau_2^2 + \sigma_{12}^2 & 0 & 0 \\ 0 & 0 & \tau_1^2 + \sigma_{21}^2 & \rho_2\sigma_{21}\sigma_{22} \\ 0 & 0 & \rho_2\sigma_{21}\sigma_{22} & \tau_2^2 + \sigma_{22}^2 \end{bmatrix} \quad (14)$$

The way of finding out between-study variance τ_i^2 is clearly described in reference (Cooper et al. 2nd edition) and will not be discussed here.

If covariates are added in random effects model, mixed effects model is constructed. Under this model, the estimates coefficients from GLS are given (Berkey et al. 1998, Becker BJ and Wu MJ 2007, Mavridis 2011).

$$\hat{\beta} = (X'S^{-1}X)^{-1}X'S^{-1}y \quad (15)$$

The value of covariates needs to be added to the design matrix X of random effects model in order to fit the mixed effects model. A simple example about how to build the design matrix for mixed effects mode is given below.

Assuming the value of each covariate is:

	Study 1	**Study 2**
Outcome 1	0.5	0.5
Outcome 2	1.3	0.9

Table 2.1 Example giving for mixed effects model

The new design matrix used for mixed effects model is:

$$\mathrm{X} = \begin{bmatrix} 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 0.5 \\ 1 & 0 & 1.3 & 0 \\ 0 & 1 & 0 & 0.9 \end{bmatrix}$$

HOMOGENEITY TEST FOR MODEL SELECTION

The decision about whether a fixed effects model or a random effects model is more appropriate is made based on the conclusion of homogeneity test.

Homogeneity test is an analysis of variance which tests for the degree of variability among effect sizes. The null hypothesis assumes that the specific effect size estimates are homogeneous across studies, indicating the variability among them is purely from sampling error. If null hypothesis can't be rejected, a fixed effects model should be used, otherwise heterogeneity of effect sizes between studies should be taken into account and random effects model should be applied.

After effect sizes are estimated by fixed effects model, the Q statistics from homogeneity test can be calculated by equation 16:

$$Q = y'S^{-1}y - \mu'(X'y^{-1}X)\mu \qquad (16)$$

where μ denotes the vector of summarized effect sizes.

Q statistics follows a $\chi 2$ distribution, with degree of freedom equals to number of effect sizes included minus number of effect sizes estimated or number of outcomes. It tells whether that model specification is statistically adequate. Thus, it can be used as a diagnostic statistics (Dankmar et al. 2002; Pierre et al. 1989; Joseph F 1991). If the Q statistics is greater than the critical value for a given degree of freedom at $\alpha = 0.05$, the null hypothesis is rejected and random effects model applies.

When between-study correlation is considered, the covariance matrix for random effects model would be the one shown in equation 9. However, if the between-study correlation is not considered, then covariance matrix used for random effects model should be the one shown in equation 14.

The basic steps of model selection are summarized as follows:

1. Obtain synthesized effect sizes from fixed effects
2. Carry out homogeneity test

3. Apply random effects model to obtain new set of pooled effect sizes if null hypothesis of homogeneity test is rejected, otherwise report effect sized generated from fixed effects model
4. Perform homogeneity test with effect sizes given by random effects model
5. Employ mixed effects model with covariate added if null hypothesis of homogeneity test is rejected and report the coefficients obtained from mixed effects model as well as report pooled effect sized synthesized by random effects model.

The detail in calculating Q statistic is provided by Cooper et al. 2nd edition and won't be discussed here.

ESTIMATION OF UNKNOWN WITHIN-STUDY COVARIANCE MATRIX USING SIMULATION

A crucial feature in multivariate meta-analysis is to obtain the possible correlation between outcomes and the incorporation of this correlation into a statistical model. With original data, variance and covariance can be directly estimated. However, in meta-analyses of aggregated data, it is impossible to obtain the correlation between outcomes. To figure that out, one suggestion is to obtain the possible correlation from external research papers (Mavridis et al. 2011). Unfortunately, it is infeasible in my case. Instead, a simulation method is applied in this study. More specifically, assuming that the correlation ρ is a random variable that follows a uniform [-1,1] distribution. Draw ρ from that distribution for 10,000 times. Thus, 10,000 within-study covariance matrices for each study will be generated with the covariance terms obtained from the sampled ρs. Then 10,000 overall covariance matrices for multivariate meta-analysis will then be obtained by assigning each within-study covariance matrix on diagonal.

With the simulated ρ and corresponding covariance matrix, the distribution of summary effect size for each outcome can be obtained. From which the estimated mean of each effect size and its 95% confidence interval (95% CI) are obtained.

DATA ANALYSIS AND RESULT

SOURCE OF DATA

In this study, I am interested in summarizing the odds ratio (OR) of a certain type of congenital heart anomalies per unit increase in continuous pollutant concentration assuming that the natural logarithm of odds ratio is linearly correlated with the concentration of air pollutant. The correlation between effect sizes within-study is considered in my model, but not that of between-study. The studies included in this meta-analysis were conducted by researchers all over the world across last decades, from which it is reasonable to assume independence of the studies' results. Therefore, the covariance matrix in my study is similar to equation 14 provided in the chapter of method.

There are two difficulties in summarizing correlated effect sizes directly from original papers. The first one is that the unit of measurement for each pollutant is different, thus it is not appropriate to combine the odds ratio for per unit increase of a certain pollutant directly. The second one is that congenital heart anomalies are categorized and named differently in different papers. Thanks to Vrijheid etc., they figured both out by unifying the unit of each pollutant and converting the corresponding effect sizes to same metric as well as listing the selected effect sizes commonly provided by included studies. Thus the effect sizes investigated in my study are solely taken from the appendix 2 from the supplements of Vriheid et.al 2011.

The unified units they converted to is 10μg/m3 for PM10, 10 ppb for O_3 and NO, 1 ppm for CO and 1 ppb for SO_2. And the common congenital heart anomalies they summarized are Atrial septal defects (ASD), Conotruncal defects (CD), Ventricular septal defects (VSD), Coarctation of the aorta (CA) and Tetralogy of Fallot (TF).

In the current meta-analysis, I summarized the effects of each air pollutant on a couple of congenital heart anomalies using GLS multivariate meta-analysis with R statistical software. The results are shown below.

THE EFFECTS OF CO

The effects of CO on 3 diseases are examined from 5 studies. The effect size studied is the odds ratio of a specific disease per unit increase in the concentration of CO. The examining diseases are Atrial septal defects (ASD), Ventricular septal defects (VSD) and Conotruncal defects (CD). Among the 5 studies, 4 of them reported the odds ratio of all the 3 outcomes and 1 of them only contained odds ratio of the first two outcomes. Such studies have situations where certain diseases are not measured. However, such "missingness" is designed on purpose, it is not the kind of missing data usually concerned in literature (Little and Rubin 1987, Little and Schenker 1994) and can be dealt with method described in Becker 1990 and Powers and Rock 1998, 1999. One thing need to keep in mind is that the odds ratio cannot be used in calculation directly. It needs to be converted to its natural logarithm metric in order for being used in meta-analysis involved calculation. Thus all the values in matrix below are in natural logarithm metric unless indicated.

Fixed effects model

First, the summarized effect sizes are estimated from fixed effects model. The design matrix X, vector of effect sizes y and covariance matrix S1 through S5 for fixed effects model are shown below.

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad y = \begin{bmatrix} 0.009950331 \\ 0.285178942 \\ -0.020202707 \\ -0.941608540 \\ 0.165514438 \\ 0.418710335 \\ -0.314710745 \\ -0.820980552 \\ -1.609437912 \\ -0.274436846 \\ -0.072570693 \\ 0.131028262 \\ -0.150822890 \\ 1.091923301 \end{bmatrix}$$

With 5 studies in total, there are 5 within-study covariance matrices. The within-study covariance matrices for study 1 through 5 are:

$$S_1 = \begin{bmatrix} \sigma_{11}^2 & \rho_{i12}\sigma_{11}\sigma_{12} & \rho_{i13}\sigma_{11}\sigma_{13} \\ \rho_{i12}\sigma_{11}\sigma_{12} & \sigma_{12}^2 & \rho_{i23}\sigma_{12}\sigma_{13} \\ \rho_{i13}\sigma_{11}\sigma_{13} & \rho_{i23}\sigma_{12}\sigma_{13} & \sigma_{13}^2 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} \sigma_{21}^2 & \rho_{i12}\sigma_{21}\sigma_{22} & \rho_{i13}\sigma_{21}\sigma_{23} \\ \rho_{i12}\sigma_{21}\sigma_{22} & \sigma_{22}^2 & \rho_{i23}\sigma_{22}\sigma_{23} \\ \rho_{i13}\sigma_{21}\sigma_{23} & \rho_{i23}\sigma_{22}\sigma_{23} & \sigma_{23}^2 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} \sigma_{31}^2 & \rho_{i12}\sigma_{31}\sigma_{32} & \rho_{i13}\sigma_{31}\sigma_{33} \\ \rho_{i12}\sigma_{31}\sigma_{32} & \sigma_{32}^2 & \rho_{i23}\sigma_{32}\sigma_{33} \\ \rho_{i13}\sigma_{31}\sigma_{33} & \rho_{i23}\sigma_{32}\sigma_{33} & \sigma_{33}^2 \end{bmatrix}$$

$$S_4 = \begin{bmatrix} \sigma_{41}^2 & \rho_{i12}\sigma_{41}\sigma_{42} & \rho_{i13}\sigma_{41}\sigma_{43} \\ \rho_{i12}\sigma_{41}\sigma_{42} & \sigma_{42}^2 & \rho_{i23}\sigma_{42}\sigma_{43} \\ \rho_{i13}\sigma_{41}\sigma_{43} & \rho_{i23}\sigma_{42}\sigma_{43} & \sigma_{43}^2 \end{bmatrix}$$

$$S_5 = \begin{bmatrix} \sigma_{51}^2 & \rho_{i12}\sigma_{51}\sigma_{52} \\ \rho_{i12}\sigma_{51}\sigma_{52} & \sigma_{52}^2 \end{bmatrix}$$

The big covariance matrix S for multivariate meta-analysis is a 14 x 14 matrix as follows:

$$\begin{bmatrix}
\begin{bmatrix} \sigma_{11}^2 & \rho_{i12}\sigma_{11}\sigma_{12} & \rho_{i13}\sigma_{11}\sigma_{13} \\ \rho_{i12}\sigma_{11}\sigma_{12} & \sigma_{12}^2 & \rho_{i23}\sigma_{12}\sigma_{13} \\ \rho_{i13}\sigma_{11}\sigma_{13} & \rho_{i23}\sigma_{12}\sigma_{13} & \sigma_{13}^2 \end{bmatrix} & \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \\
\begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \begin{bmatrix} \sigma_{21}^2 & \rho_{i12}\sigma_{21}\sigma_{22} & \rho_{i13}\sigma_{21}\sigma_{23} \\ \rho_{i12}\sigma_{21}\sigma_{22} & \sigma_{22}^2 & \rho_{i23}\sigma_{22}\sigma_{23} \\ \rho_{i13}\sigma_{21}\sigma_{23} & \rho_{i23}\sigma_{22}\sigma_{23} & \sigma_{23}^2 \end{bmatrix} \begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \\
\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} & \begin{bmatrix} \sigma_{31}^2 & \rho_{i12}\sigma_{31}\sigma_{32} & \rho_{i13}\sigma_{31}\sigma_{33} \\ \rho_{i12}\sigma_{31}\sigma_{32} & \sigma_{32}^2 & \rho_{i23}\sigma_{32}\sigma_{33} \\ \rho_{i13}\sigma_{31}\sigma_{33} & \rho_{i23}\sigma_{32}\sigma_{33} & \sigma_{33}^2 \end{bmatrix} \begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \\
\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} & \begin{bmatrix} \sigma_{41}^2 & \rho_{i12}\sigma_{41}\sigma_{42} & \rho_{i13}\sigma_{41}\sigma_{43} \\ \rho_{i12}\sigma_{41}\sigma_{42} & \sigma_{42}^2 & \rho_{i23}\sigma_{42}\sigma_{43} \\ \rho_{i13}\sigma_{41}\sigma_{43} & \rho_{i23}\sigma_{42}\sigma_{43} & \sigma_{43}^2 \end{bmatrix} \begin{matrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{matrix} \\
\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} & \begin{bmatrix} \sigma_{51}^2 & \rho_{i12}\sigma_{51}\sigma_{52} \\ \rho_{i12}\sigma_{51}\sigma_{52} & \sigma_{52}^2 \end{bmatrix}
\end{bmatrix}$$

where $\rho_i \sim$ uniform[-1,1] , it is sampled 10,000 times, thus 10,000 big covariance matrix is generated, from which, 10000 summarized effect sizes for each outcome are calculated using equation 13 given before :

$$\hat{\mu} = (X'S^{-1}X)^{-1}X'S^{-1}y \qquad (13)$$

The above variance covariance matrix can also be represented as following for simplicity.

$$S = \begin{bmatrix} S1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S5 \end{bmatrix} \equiv Diag(S1, \ldots, S5)$$

Distribution of lnOR

From simulation, the distribution of natural logarithm of each outcome generated from fixed effects model is obtained. Three diseases ASD, VSD and CD are named as 1, 2 and 3 in the following figures for convenient.

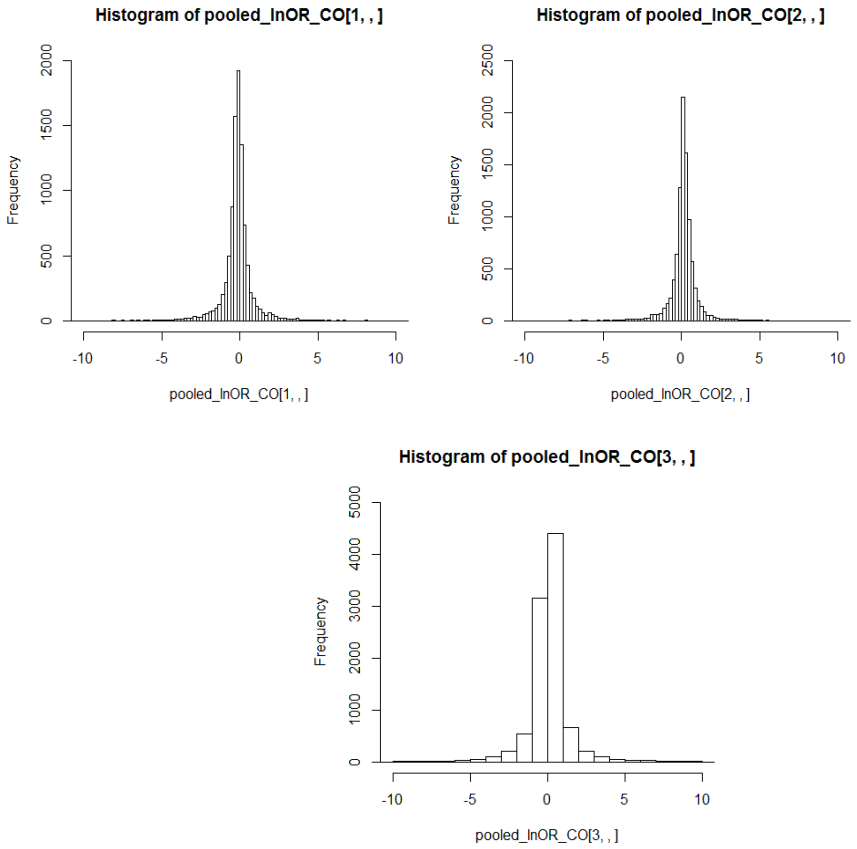

Figure 3.1 Distribution of lnOR of CO on each outcome generated from fixed effects model.

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	0.565	[0.014, 29.346]
VSD	1.071	[0.055, 29.781]
CD	1.508	[0.006, 90.051]

Table 3.2. Estimated effect size of CO for each disease and corresponding 95% confidence interval generated from fixed effects model

Homogeneity test

With 10,000 ρs sampled from uniform [-1,1], 10000 homogeneity tests are carried out. For each homogeneity test, the number of effect sizes provided is 14 and the number of effect sizes estimated is 3, thus the degree of freedom (df) is 11. The critical Q statistics is 19.695 with df equals to 11 and α is 0.05. Comparing each Q statistics from 10,000 homogeneity tests with the critical value 19.695, 7965 of the datasets led to rejection of the null hypothesis. The proportion of rejection is then 79.56%. Thus random effects model should be applied.

Random effects model

The estimated between-study variance τ_i^2 for each outcome is calculated based on method of moments mentioned before. Positive between-study variance suggests the existing of random effects among studies. The calculation of estimated between-study variance is thoroughly discussed in the reference mentioned above and all the formulas associated can be found there.

Among the 3 diseases, only disease 3 has positive between-study variance. This suggests that the random effects between-study is mainly from disease 3. The between-study variance for disease 3 is added to the original covariance matrix and the variance for disease 3 in each study is now replaced by new variance:$v_{i3} = \tau_3^2 + \sigma_{i3}^2$. Within-study covariance matrix for study 5 doesn't change since it doesn't contain effect size from disease 3.

The new within-study covariance matrices for study 1 through 4 as are listed below:

$$S_1' = \begin{bmatrix} \sigma_{11}^2 & \rho_{i12}\sigma_{11}\sigma_{12} & \rho_{i13}\sigma_{11}\sigma_{13} \\ \rho_{i12}\sigma_{11}\sigma_{12} & \sigma_{12}^2 & \rho_{i23}\sigma_{12}\sigma_{13} \\ \rho_{i13}\sigma_{11}\sigma_{13} & \rho_{i23}\sigma_{12}\sigma_{13} & \tau_3^2 + \sigma_{13}^2 \end{bmatrix}$$

$$S_2' = \begin{bmatrix} \sigma_{21}^2 & \rho_{i12}\sigma_{21}\sigma_{22} & \rho_{i13}\sigma_{21}\sigma_{23} \\ \rho_{i12}\sigma_{21}\sigma_{22} & \sigma_{22}^2 & \rho_{i23}\sigma_{22}\sigma_{23} \\ \rho_{i13}\sigma_{21}\sigma_{23} & \rho_{i23}\sigma_{22}\sigma_{23} & \tau_3^2 + \sigma_{23}^2 \end{bmatrix}$$

$$S_3' = \begin{bmatrix} \sigma_{31}^2 & \rho_{i12}\sigma_{31}\sigma_{32} & \rho_{i13}\sigma_{31}\sigma_{33} \\ \rho_{i12}\sigma_{31}\sigma_{32} & \sigma_{32}^2 & \rho_{i23}\sigma_{32}\sigma_{33} \\ \rho_{i13}\sigma_{31}\sigma_{33} & \rho_{i23}\sigma_{32}\sigma_{33} & \tau_3^2 + \sigma_{33}^2 \end{bmatrix}$$

$$S_4' = \begin{bmatrix} \sigma_{41}^2 & \rho_{i12}\sigma_{41}\sigma_{42} & \rho_{i13}\sigma_{41}\sigma_{43} \\ \rho_{i12}\sigma_{41}\sigma_{42} & \sigma_{42}^2 & \rho_{i23}\sigma_{42}\sigma_{43} \\ \rho_{i13}\sigma_{41}\sigma_{43} & \rho_{i23}\sigma_{42}\sigma_{43} & \tau_3^2 + \sigma_{43}^2 \end{bmatrix}$$

Again, the new big covariance matrix can be represented as following for simplicity.

$$S' = \begin{bmatrix} S1' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S5 \end{bmatrix} \equiv Diag(S1', \ldots, S5)$$

Using GLS method, the distribution of each estimated effect size as well as the 95% CI of them are obtained and pasted below.

Distribution of lnOR

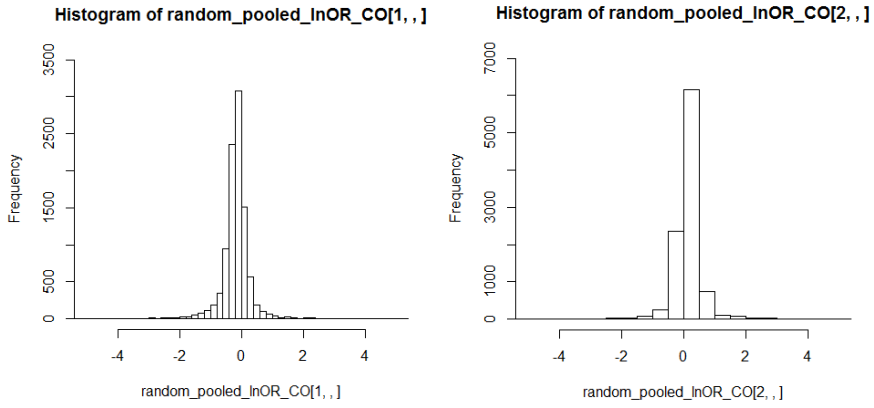

Figure 3.2 Distribution of lnOR of CO on each outcome generated from random effects model

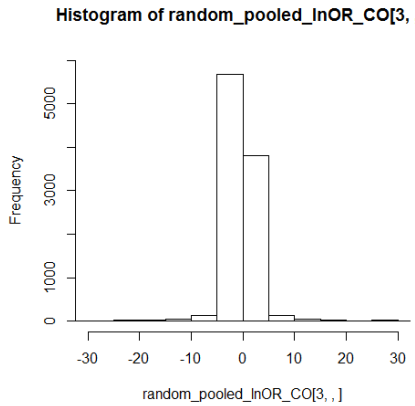

Figure 3.2 continued

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	0.699	[0.248, 2.527]
VSD	0.893	[0.383, 2.818]
CD	0.024	[0.005, 141.530]

Table 3.2 Estimated effect size of CO for each disease and corresponding 95% confidence interval generated from random effects model.

Homogeneity test

Homogeneity test is conducted with summarized effect sizes obtained from random effects model. Out of 10,000 tests, 9478 of them are rejected. Indicating the variation among effect sizes for each outcome cannot be explained by sampling error and random effects. Therefore, covariates need to be added to the model in order to interpret the variation. The value of covariate incorporated is the mean concentration of CO from each study. With covariate considered, the model becomes mixed effects model and will estimate the corresponding coefficients by applying GLS method.

Mixed effects model

The covariance matrix used in mixed effects model is the same as that of the random effects model. However, new design matrix is constructed with the concentration of CO added.

$$Design\ matrix\ X = \begin{bmatrix} 1 & 0 & 0 & 1.765 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1.765 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1.765 \\ 1 & 0 & 0 & 0.55 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.55 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.55 \\ 1 & 0 & 0 & 0.85 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.85 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.85 \\ 1 & 0 & 0 & 0.30 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.30 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.30 \\ 1 & 0 & 0 & 0.515 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.515 & 0 \end{bmatrix}$$

The mixed effects models are constructed for this multivariate meta-analysis. In the model, 2 coefficients for each outcome are estimated. One of the estimated coefficients is the intercept and the other one is coefficient for covariate which is the mean concentration of CO in this case. From equation 15 listed previously, the coefficients can be obtained.

$$\hat{\beta} = (X'S^{-1}X)^{-1}X'S^{-1}y \tag{15}$$

Point estimate and confidence interval

	β_0	95% CI of β_0	β_1	95% CI of β_1
ASD	-0.786	[-3.375, 1.943]	0.442	[-0.991, 1.726]
VSD	0.169	[-0.783, 0.994]	-0.032	[-0.660, 0.544]
CD	-0.400	[-4.602, 3.750]	-0.159	[-1.480, 1.309]

Table 3.3 Estimated effect size of CO for each disease and corresponding 95% confidence interval generated from mixed effects model

THE EFFECTS OF NO2

Models, related matrices and procedures of analysis about CO were shown in detail above. To avoid redundancy, I will not list the detail of analysis about other pollutants since they are following the same way as that of CO. Instead, I will only show the result of meta-analysis for the rest of the pollutants.

The effects of NO2 on 4 diseases from 5 studies are analyzed. The 4 diseases discussed are ASD, VSD, Coarctation of the aorta (CA) and Tetralogy of Fallot (TF). These four diseases are renamed as 1, 2, 3 and 4 for conveniently shown in figures below.

Fixed effects model

Distribution of lnOR

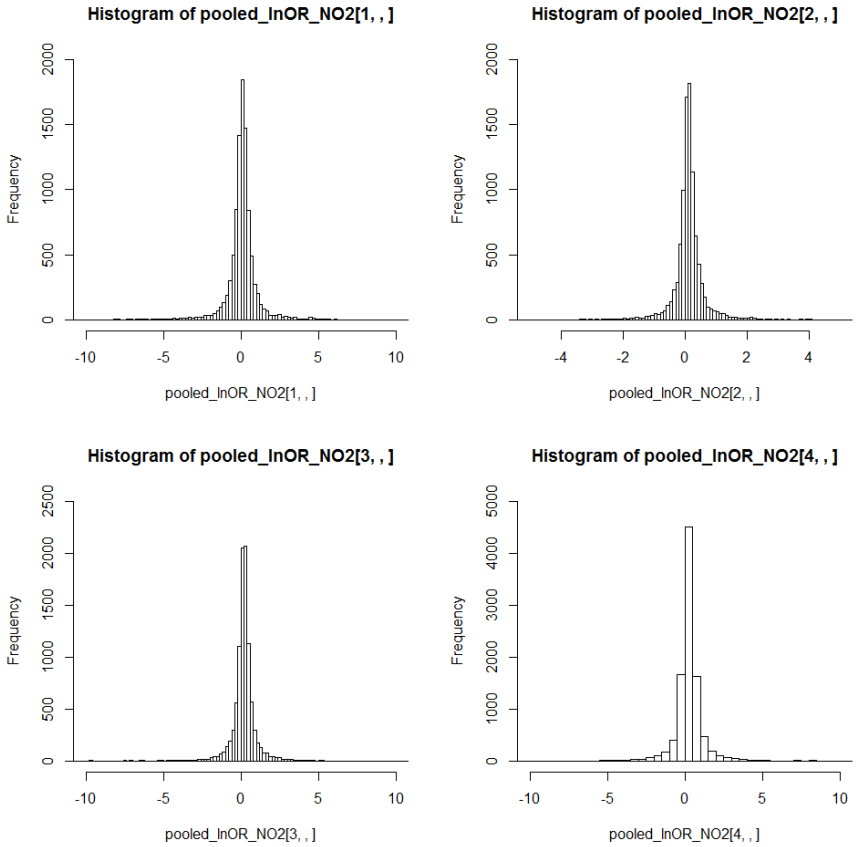

Figure 3.4 Distribution of lnOR of NO_2 on each outcome generated from fixed effects model.

Point estimate and confidence interval

	Mean of OR	95% CI of OR
ASD	0.982	[0.024, 43.146]
VSD	1.512	[0.185, 7.746]
CA	1.155	[0.043, 21.234]
TF	1.809	[0.056, 33.957]

Table 3.4 Estimated effect size of NO_2 for each disease and corresponding 95% confidence interval generated from fixed effects model

Homogeneity test

Homogeneity test is carried out using the summarized effect size synthesized from fixed effects model. With 10,000 ρs sampled from uniform [-1,1], 10,000 homogeneity tests are performed. For each homogeneity test, the total number of effect sizes provided by the including studies is 16 and the number of effect sizes estimated is 4, thus the degree of freedom (df) is 12. The critical Q statistics is 21.026 with df equals to 12 and α is 0.05. Comparing each Q statistics generated from 10,000 homogeneity tests with the critical value 21.026, 5098 of the datasets led to rejection. The proportion of rejection is then 50.98%. Hence random effects model should be employed.

Random effects model

The between-study variance for each disease is calculated. However, none of them has positive between-study variance. This suggests that the variation is not from random effects. So the concentration of NO2 is incorporated into the model to test if the variability could be explained by this covariate.

Weighted regression model

The mean concentration of NO2 from each study is added to the design matrix of fixed effects model as covariate. Using this new design matrix, weighted linear regression model is created. The parameters this model estimated are the coefficients for covariate and intercept in the linear regression model.

Point estimate and confidence interval

	β_0	95% CI of β_0	β_1	95% CI of β_1
ASD	0.782	[-2.018, 2.510]	-0.719	[-3.283, 3.248]
VSD	0.038	[-1.045, 1.294]	0.233	[-1.673, 1.787]
CA	-0.218	[-1.798, 2.075]	0.029	[-2.413, 2.256]
TF	-0.163	[-2.516, 1.658]	-0.275	[-2.101, 3.165]

Table 3.5 Estimated coefficients and their 95% confidence intervals obtained from weighted regression model

THE EFFECTS OF O3

The effects of O3 on 3 diseases from 5 studies are summarized here. The examining diseases are ASD, VSD and Conotruncal defects (CD) and are renamed as 1, 2 and 3 for simplicity discussed later.

Fixed effects model

Distribution of lnOR

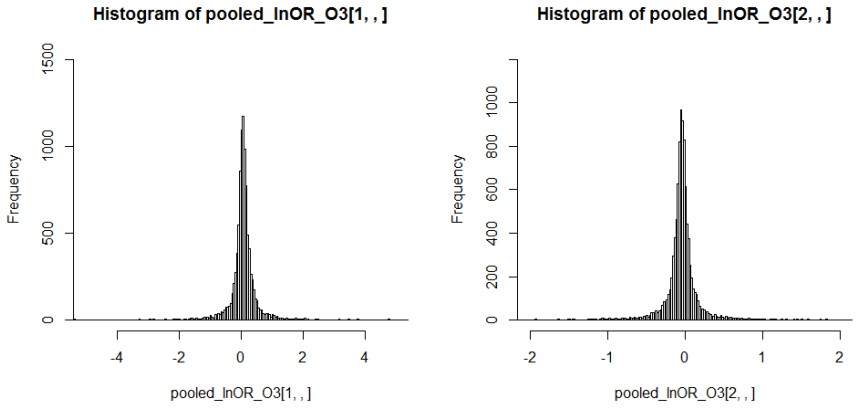

Figure 3.6 Distribution of lnOR of O_3 on each outcome generated from fixed effects model.

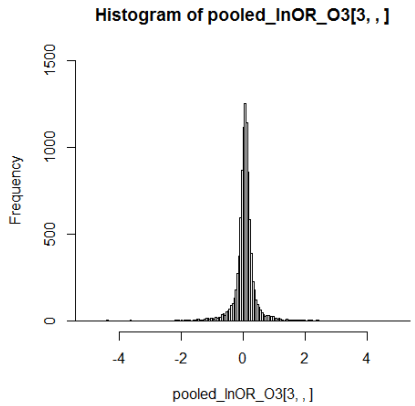

Figure 3.6 continued.

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	1.083	[0.285, 4.316]
VSD	0.975	[0.428, 1.990]
CD	1.106	[0.313, 3.370]

Table 3.6 Estimated effect size of O3 for each disease and its 95% confidence interval obtained from fixed effects model

Homogeneity test

A Homogeneity test is conducted using the summarized effect size synthesized from fixed effects model. With 10,000 ρs sampled from uniform [-1,1], 10,000 homogeneity tests are carried out. For each homogeneity test, the total number of effect sizes under investigation is 14 and the number of effect sizes estimated is 3, thus the degree of freedom (df) is 11. The critical Q statistics is 19.695 with df equals to 11 and α is 0.05. Comparing each Q statistics from 10,000 homogeneity tests with the critical value1 9.695, 6270 of the datasets led to rejection of the null hypothesis. The proportion of rejection is 62.7%. Thus the random effect model should be applied.

Random effects model

The between-study variance for each disease is calculated. Only disease 2 has positive between-study variance. This suggests that the between-study variation is from disease 2. So new covariance matrix is constructed as described above and summarized effect sizes are estimated by random effects model using the new covariance matrix

Distribution of lnOR

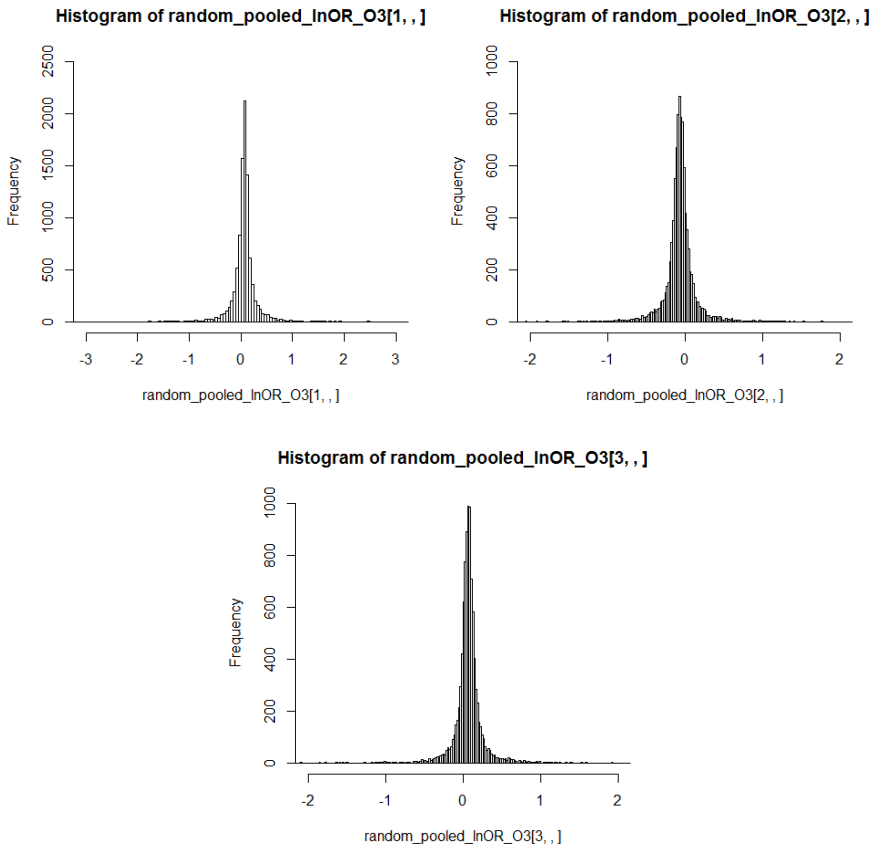

Figure 3.7 Distribution of lnOR of O_3 on each outcome generated from random effects model.

Point estimate and confidence interval

	Mean OR	95%CI of OR
ASD	1.071	[0.417, 2.681]
VSD	0.933	[0.424, 1.980]
CD	1.107	[0.584, 2.104]

Table 3.7 Estimated effect size of O3 for each disease and its 95% confidence interval obtained from random effects model

Homogeneity test

Homogeneity test is also carried out using the summarized effect size provided by random effects model. Out of 10000 tests, 5685 are rejected. Suggests the variation cannot only be explained by sampling error and random effects. Therefore, in order to finding the source of variability, the mean concentration of O3 from each study is added to the random effects model as covariate to create mixed effects model.

Mixed effects model

Mixed effects models are constructed for this multivariate meta-analysis. In the model, 2 coefficients for each outcome are estimated. One of the estimated coefficients is the intercept and the other one is coefficient for covariate which is the mean concentration of O3 in this case. From equation 16, the coefficients can be obtained.

Point estimate and confidence interval

	β_0	95% CI of β_0	β_1	95% CI of β_1
ASD	-0.015	[-3.305, 2.001]	0.024	[-0.640, 1.150]
VSD	-1.274	[-2.592, 2.165]	0.403	[-0.710, 0.807]
CD	1.528	[-3.155, 3.890]	-0.526	[-1.339, 1.106]

Table 3.8 Estimated coefficients and their 95% confidence intervals generated by mixed effects model

THE EFFECTS OF PM_{10}

The effects of PM10 on 4 diseases from 5 studies are summarized. These four diseases are ASD, Ventricular, Coarctation of the aorta (CA) and Tetralogy of Fallot (TF). As shown above, the four diseases are renamed again as 1, 2, 3 and 4.

Fixed effects model

Distribution of lnOR

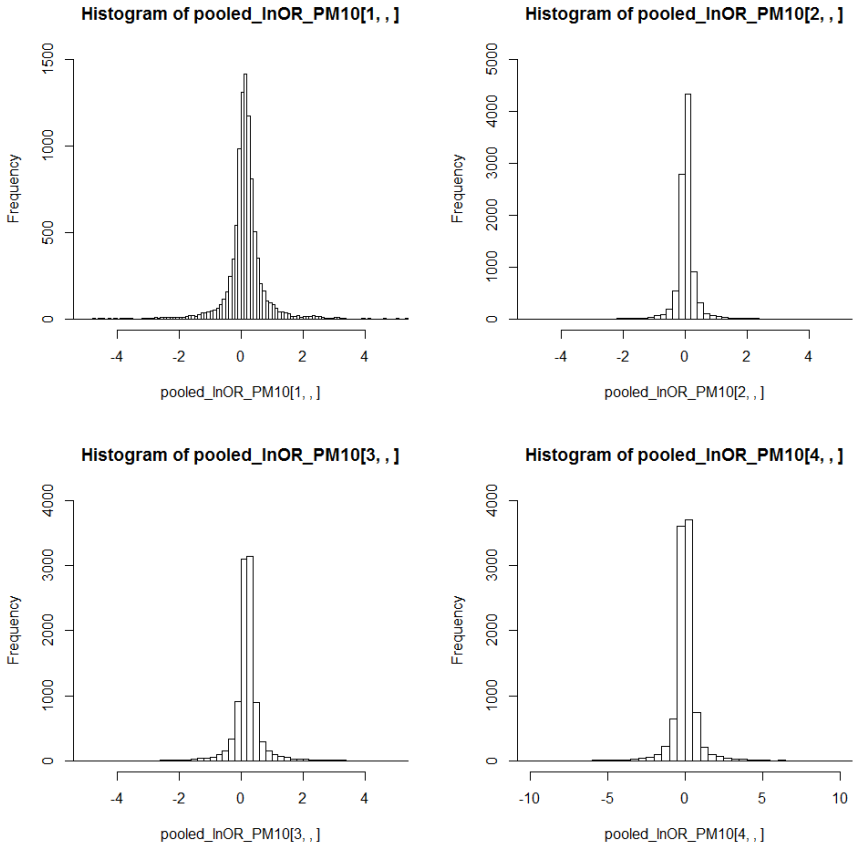

Figure 3.9 Distribution of lnOR of PM_{10} on each outcome generated from fixed effects model.

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	1.089	[0.118, 12.692]
VSD	1.176	[0.373, 3.347]
CA	1.367	[0.245, 6.406]
TF	1.273	[0.076, 16.054]

Table 3.9 Estimated coefficients and their 95% confidence intervals generated by fixed effects model

Homogeneity test

Homogeneity test is performed using the summarized effect size generated by fixed effects model. With 10,000 ρ sampled from uniform [-1,1], 10,000 homogeneity tests are performed. For each homogeneity test, the number of effect sizes provided is 16 and the number of effect sizes estimated is 4, thus the degree of freedom (df) is 11. The critical Q statistics is 21.026 with df equals to 12 and α is 0.05. Comparing each Q statistics from 10000 homogeneity tests with the critical value 21.026, 5870 of the datasets led to rejection of the null hypothesis. The proportion of rejection is 58.7%. Thus random effect model should be applied.

Random effects model

The between-study variance for each disease is calculated. Only disease 2 has positive between-study variance. This suggests that the between-study variation is from disease 2. So new covariance matrix is constructed as described above and summarized effect sizes are estimated by random effects model using the new covariance matrix.

Distribution of lnOR

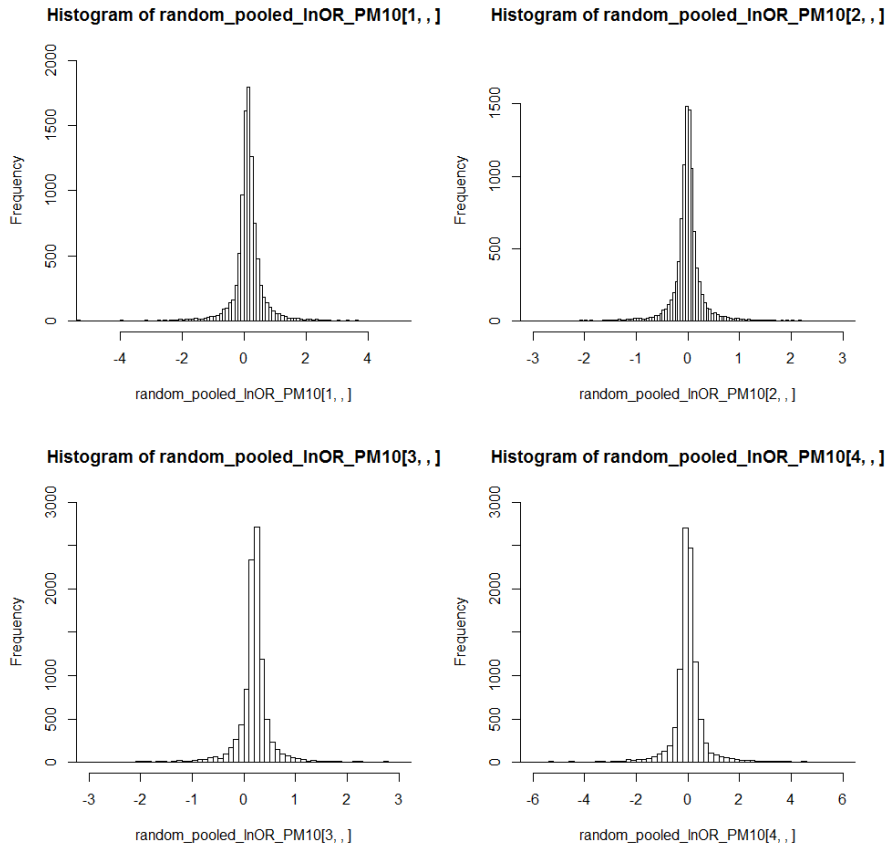

Figure 3.10 Distribution of lnOR of PM_{10} on each outcome generated from random effects model.

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	1.123	[0.190, 6.971]
VSD	0.972	[0.337, 3.186]
CA	1.278	[0.389, 3.856]
TF	1.357	[0.140, 7.326]

Table 3.10 Estimated effect size of PM10 for each disease and its 95% confidence interval given by random effects model

Homogeneity test

Homogeneity test is also carried out using the summarized effect size generated from random effects model. Out of 10000 tests, 6708 of them are rejected. Suggesting the variation cannot only be explained by random error and systematic difference between studies. Therefore, covariates need to be added to the model in order to interpret the variation. Mean concentration of PM10 from each study is added to the random effects model as covariate to create mixed effects model.

Mixed effects model

Mixed effects model is constructed for this multivariate meta-analysis. In the model, 2 coefficients for each outcome are estimated. One of the estimated coefficients is the intercept and the other one is coefficient for covariate which is the mean concentration of PM_{10} in this case.

Point estimate and confidence interval

	β_0	95% CI of β_0	β_1	95% CI of β_1
ASD	-0.039	[-1.066 0.880]	0.113	[-0.544 0.845]
VSD	-0.341	[-0.866 0.406]	0.171	[-0.367 0.582]
CA	-0.220	[-1.461 0.937]	0.186	[-0.483 0.945]
TF	-0.22428336	[-1.671, 1.006]	0.115	[-0.740, 1.099]

Table 3.11 Estimated coefficients and their 95% confidence intervals from mixed effects model

THE EFFECTS OF SO2

The effects of SO2 on 4 diseases from 6 studies are summarized. The four diseases are ASD, VSD, CA and TF. As before, they are renamed as 1, 2, 3 and 4.

Fixed effects model

Distribution of lnOR

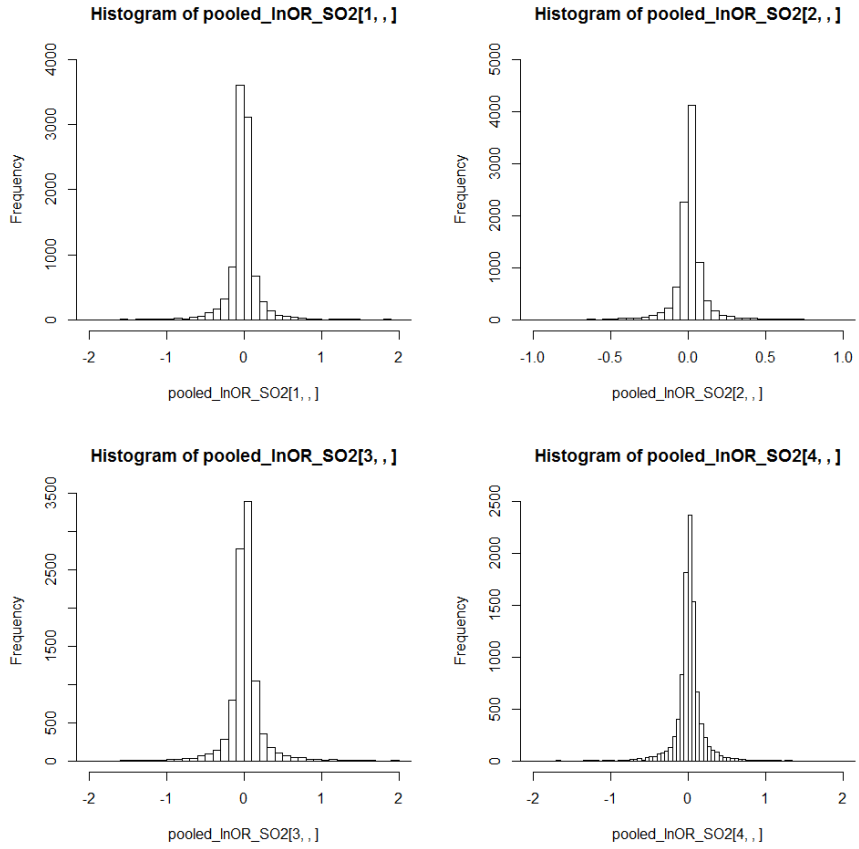

Figure 3.12 Distribution of lnOR of SO_2 on each outcome generated from fixed effects model.

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	1.157	[0.517, 1.989]
VSD	1.019	[0.669, 1.443]
CA	0.881	[0.456, 2.281]
TF	1.071	[0.525, 1.915]

Table 3.12 Estimated effect size of SO_2 for each disease and its 95% confidence interval given by fixed effects model

Homogeneity test

As described above, 10,000 homogeneity tests are carried out. For each homogeneity test, the number of effect sizes provided is 20 and the number of effect sizes estimated is 4, thus the degree of freedom (df) is 16. The critical Q statistics is 26.296 with df equals to 16 and α is 0.05. Comparing each Q statistics generated from 10,000 homogeneity tests with the critical value 26.296, 6536 of the datasets led to rejection of the null hypothesis. The proportion of rejection is 65.36%. Therefore, random effect model should be applied.

Random effects model

The between-study variance for each disease is calculated. Only disease 3 has positive between-study variance. This suggests that the between-study variation is from disease 3. So new covariance matrix is constructed as described above and summarized effect sizes are estimated by applying random effects model using the new covariance matrix.

Distribution of lnOR

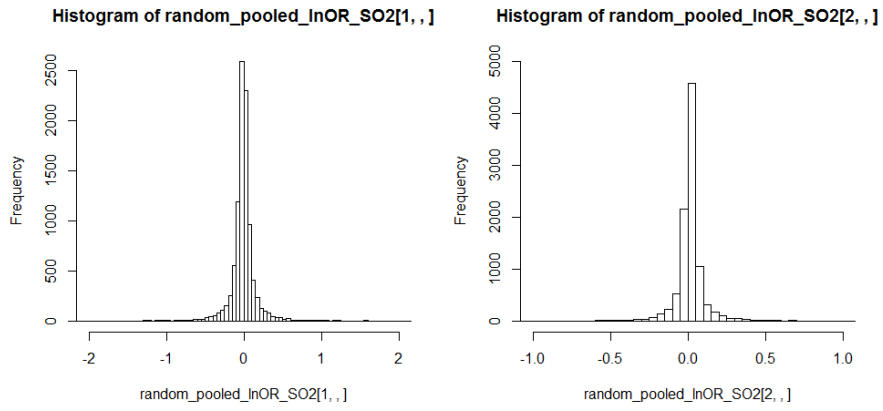

Figure 3.13 Distribution of lnOR of SO2 on each outcome generated from random effects model.

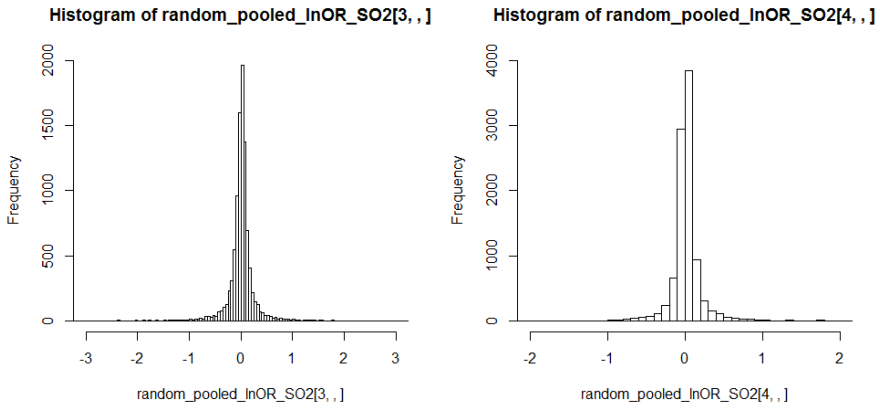

Figure 3.13 continued.

Point estimate and confidence interval

	Mean OR	95% CI of OR
ASD	0.951	[0.581, 1.766]
VSD	1.051	[0.751, 1.402]
CA	0.941	[0.438, 2.112]
TF	0.997	[0.567, 1.871]

Table 3.13 Estimated effect size of SO_2 for each disease and its 95% confidence interval given by random effects model

Homogeneity test

Homogeneity test is also carried out using the summarized effect size generated from random effects model. Out of 10000 tests, 6626 of them are rejected. Suggesting the variation cannot only be explained by random error and systematic difference between studies. Therefore, covariates need to be added to the model in order to interpret the variation. The mean concentration of SO2 from each study is added to the random effects model as covariate to create mixed effects model.

Mixed effects model

The mixed effects model is constructed for this multivariate meta-analysis. In this model, 2 coefficients for each outcome are estimated. One of the estimated coefficients is

the intercept and the other one is the coefficient for covariate which is the mean concentration of SO_2 in this case.

Point estimate and confidence interval

	β_0	95% CI of β_0	β_1	95% CI of β_1
ASD	-0.118	[-1.669, 1.628]	0.012	[-0.176, 0.185]
VSD	0.063	[-0.392, 0.512]	-0.011	[-0.068, 0.050]
CA	0.111	[-1.151, 1.442]	-0.020	[-0.199, 0.159]
TF	0.077	[-0.951, 1.070]	-0.009	[-0.125, 0.112]

Table 3.14 Estimated coefficient and its 95% confidence interval given by mixed effects model

CONCLUSION

The effects of air pollutants CO, NO_2, O_3, PM_{10} and SO_2 on congenital heart anomalies are represented by the odds ratio of each disease per unit increase in the concentration of each pollutant. The effect was summarized by fixed effects model for NO2 and random effects model for the rest of pollutants using multivariate GLS approach with correlation sampled from uniform [-1,1]. Moreover, the coefficients in mixed models which measure the linear relationship between the concentration of each pollutant and the risk of getting a disease were also reported in this analysis. Based on the result pasted above, none of the odds ratios significantly differs from 1. Therefore, my conclusion is that when correlation between outcomes is considered, the summarized effect sizes synthesized by GLS approach suggested that air pollutants didn't have effects on the occurrence of congenital heart anomalies.

When talking about coefficients in mixed models, all of them contained 0 in their 95% CI, which may suggest that there is no linear relationship between the concentration of air pollutants and the natural logarithm of the odds ratio of congenital heart anomalies.

DISCUSSION

The summarized effect sizes obtained by multivariate GLS method in my study are different from those provided by Vrijheid etc 2011 in their meta-analysis. They suggested that NO_2 and SO_2 exposures were related to increase in risk of Coarctaion of the aorta and Tetralogy of Fallot. However, from the meta-analysis I conducted, none of the exposure to any air pollutants seems to be related to the risk of any congenital heart anomalies. The biggest difference between the meta-analysis Vrijheid etc conducted and mine is that they didn't consider the correlation between different outcomes within a study and used univariate meta-analysis method in analyzing their data. In my study, the correlations between outcomes were considered and multivariate GLS approach was applied. The difference in conclusions from two meta-analysis suggests that the incorporation of correlations between outcomes within a study can make a big difference when summary effect sizes are synthesized. Therefore, before any suggestion is made, thorough consideration about whether the correlation should be incorporated should be given.

The 95% CI of all of the reported coefficients from the mixed models I constructed contains 0. However, I won't say that there is no linear relationship between the concentration of air pollutant and the risk of heart anomalies. Simply because only 4 or 5 studies are used in estimating the parameters from the mixed linear regression model, thus the power is not high enough to give a meaningful conclusion. To make accurate estimations on the coefficients, more studies would be included in the mixed model.

REFERENCES

1. Tsukue N, Toda N, Tsubone H, et al. Diesel exhaust (DE) affects the regulation of testicular function in male Fischer 344 rats. J Toxicol Environ Health A 2001; 63:115-26.

2. Singh J. Nitrogen dioxide exposure alters neonatal development. Neurotoxicology 1988; 9:545-50.

3. Somers CM, Yauk CL, White PA, et al. Air pollution induced heritable DNA mutations. Proc Natl Acad Sci USA 2002; 99:15904-7.

4. Somers CM, McCarry BE, Malek F, et al. Reduction of particulate air pollution lowers the risk of heritable mutations in mice. Science 2004;304:1008-10.

5.Samet JM, DeMarini DM, Malling HV. DO airborne particles induce heritable mutations? Science 2004;304:971-2.

6. Ginianaia SV, Rankin J, Bell R, Pless-Mulloli T, Howel D. Particulate air pollution and fetal health: a systematic review of the epidemiologic evidence. Epidemiology.2004b;15:36–45.

7.Maisonet M, Correa A, Misra D, Jaakkola JJ. A review of the literature on the effects of ambient air pollution on fetal growth. Environ Res. 2004;95:106–115.

8. Sr ám RJ, Binkov á B, Dejmek J, Bobak M. Ambient air pollution and pregnancy outcomes: a review of the literature. Environ Health Perspect. 2005;113:375–382.

9. Vrijheid M, Martinez D, Manzanares S, Dadvand P, Schembari A, Rankin J, Nieuwenhuijsen M. Ambient air pollution and risk of congenital anomalies: a systematic review and meta-analysis. Environ Health Perspect. 2011;119:598-606.

10. Ritz B. Air pollution and congenital anomalies. Occup Environ Med. 2010;

2010;67:221–222.

11. Ritz B, Yu F, Fruin S, Chapa G, Shaw GM, Harris JA. Ambient air pollution and risk of birth defects in Southern California. Am J Epidemiol. 2002;155:17–25.

12. KerenAgay-Shay a,n,1, MichaelFriger b, ShaiLinn a,c . Environmental Research. 2013; 124:28-34

13. Gilboa SM, Mendola P, Olshan AF, Langlois PH, Savitz DA, Loomis D, et al. Relation between ambient air quality and selected birth defects, seven county study, Texas, 1997–2000. Am J Epidemiol. 2005;162:238–252.

14. Rankin J, Chadwick T, Natarajan M, Howel D, Pearce MS, Pless-Mulloli T. Maternal exposure to ambient air pollutants and risk of congenital anomalies. Environ Res.2009;109:181–187.

15. Strickland MJ, Riehle-Colarusso TJ, Jacobs JP, Reller MD, Mahle WT, Botto LD, et al. The importance of nomenclature for congenital cardiac disease: implications for research and evaluation. Cardiol Young.2008;18(suppl 2):92–100.

16. Hansen CA, Barnett AG, Jalaludin BB, Morgan GG. Ambient air pollution and birth defects in Brisbane, Australia.PLoS One. 2009;4:e5408.

17. Dolk H, Armstrong B, Lachowycz K, Vrijheid M, Rankin J, Abramsky L, et al. Ambient air pollution and risk of congenital anomalies in England, 1991–1999. Occup Environ Med. 2010;67:223–227.

18. Dadvand P, Rankin J, Rushton S, Pless-Mulloli T. Association between maternal exposure to ambient air pollution and congenital heart disease: a register-based spatio-temporal analysis. Am J Epidemiol. 2011a;173:171–182.

19. Dadvand P, Rankin J, Rushton S, Pless-Mulloli T. Ambient air pollution and congenital heart disease, a registry-based study. Environ Res. 2011b;111:435–441.

20. Harris Cooper, Larry V. Hedges, Jeffrey C. Valentine. The Handbook of Research Synthesis and Meta-Analysis. 2nd edition.

21. Gleser J and Olkin I. Stochasstically Dependent Effect sizes. In the Handbook of Research Synthesis, edited by Harris Cooper and Larry Hedges. Nre York: Russekk Sage Foundation.

22. Raudenbush W, Betsy JB, Kalaian HA. Modeling multivariate effect sizes. Psychological Bulletin. 1988; 103(11):111-120

23. Timm NH. Testing Multivariate Effects Sizes in Multiple-Endpoint Studies. Multivariate Behavioral Research. 1999; 34(4):457-465

24. Dimitris M and Georgia S. A practical introduction to multivariate meta-analysis. Statistical Methods in Medical Research . 2011;0:0-26.

25. Hedges L. A random effects model for effect sizes. Psychological Bulletin, 1983;93:388-395

26. Hedges L. Meta-analysis. Journal of Education of Statistics, 1992; 17:279-296

27. Hedges L. Fixed- and Random-Effects models in meta-Analysis. Psychological Methods. 1998; 4:486-504

28. DerSimonian R, & Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials, 1986;7:177-188

29. Dankmar B, Malzahn U, Dietz E et al. Some General Points in Estimating Heterogeneity Variance with the DerSimonian-Laird Estimator. Biostatistics. 2002;3(4):445-457

30.Pierre J.B, Blanchard J, Panak E Considerations for the meta-analysis of randomized clinical trails. Controlled Clinical Trials. 1989; 10(3):254-281

31. Joseph F.L, Gross A.J Meta-Analysis in epidemiology, with special reference to studies f the association between exposure o environmental tobacco smoke and lung cancer: A critique. Journal of Clinical Epidemiology. 1991; 44(2):127-139

32. Berkey CS, Hoaglin DC, Antczak-Bouckoms A, Mosteller F and Colditz GA. Meta-analysis of multiple outcomes by regression with random effects. Stat Med 1998; 17:2537-2550.

33. Becker BJ , Wu MJ. The Synthesis of Regression Slopes in Meta-Analysis. Statistical Sciences. 2007: 22(3):414-429

34. Little RJ, Rubin DB. Statistical Analysis With Missing Data. New York:John Wiley & Sons. 1987.

35. Little RJ, Schenker N. In handbook for Statistical Modeling in the Social and Behavioral Sciences, edited by Gerhard Arminger, 1994

36. Becker BJ Coaching for the Scholastic Aptitude Test : Further synthesis and appraisal. Review of Educational Research. 1990;60(3):373-417

37. Powers DE, Rock DA. Effects of Coaching on SAT I: Reasonng scores. College Board Research Report. New York: College Board. 1998:98-106.