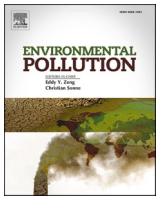
Environmental Pollution 336 (2023) 122465

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol



Health impact assessment for air pollution in the presence of regional variation in effect sizes: The implications of using different meta-analytic approaches☆

Duncan Lee [a,*], Heather Walton [b,c], Dimitris Evangelopoulos [b,c,d], Klea Katsouyanni [b,c,e], Alison M. Gowers [f], Gavin Shaddick [g], Christina Mitsakou [f]

[a] School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8SQ, UK
[b] Environmental Research Group, School of Public Health, Imperial College London, UK
[c] National Institute of Health Research Health Protection Research Unit in Environmental Exposures and Health, Imperial College London, UK
[d] MRC Centre for Environment and Health, Imperial College London, UK
[e] University of Athens, Greece
[f] Air Quality and Public Health Group, UK Health Security Agency, UK
[g] Royal Holloway University of London, UK

ARTICLE INFO

Keywords:
Air pollution
Health impact assessment
Meta analysis
Regional heterogeneity

ABSTRACT

The estimated health effects of air pollution vary between studies, and this variation is caused by factors associated with the study location, hereafter termed regional heterogeneity. This heterogeneity raises a methodological question as to which studies should be used to estimate risks in a specific region in a health impact assessment. Should one use all studies across the world, or only those in the region of interest? The current study provides novel insight into this question in two ways. Firstly, it presents an up-to-date analysis examining the magnitude of continent-level regional heterogeneity in the short-term health effects of air pollution, using a database of studies collected by Orellano et al. (2020). Secondly, it provides in-depth simulation analyses examining whether existing meta-analyses are likely to be underpowered to identify statistically significant regional heterogeneity, as well as evaluating which meta-analytic technique is best for estimating region-specific estimates. The techniques considered include global and continent-specific (sub-group) random effects meta-analysis and meta-regression, with omnibus statistical tests used to quantify regional heterogeneity. We find statistically significant regional heterogeneity for 4 of the 8 pollutant-outcome pairs considered, comprising NO_2, O_3 and $PM_{2.5}$ with all-cause mortality, and $PM_{2.5}$ with cardiovascular mortality. From the simulation analysis statistically significant regional heterogeneity is more likely to be identified as the number of studies increases (between 3 and 30 in each region were considered), between region heterogeneity increases and within region heterogeneity decreases. Finally, while a sub-group analysis using Cochran's Q test has a higher median power (0.71) than a test based on the moderators' coefficients from meta-regression (0.59) to identify regional heterogeneity, it also has an inflated type-1 error leading to more false positives (median errors of 0.15 compared to 0.09).

1. Introduction

Air pollution continues to be a global health problem, with the World Health Organization (WHO) linking seven million deaths to fine particulate matter each year (WHO, 2016), while more recent comprehensive assessments of its impacts are given by the United States Environmental Protection Agency's Integrated Science Assessments (https://www.epa.gov/isa). There is a wealth of epidemiological studies quantifying the risks associated with both short-term and long-term exposures to a range of pollutants (Brunekreef and Holgate, 2002), including particulate matter $\leq$10 μm in aerodynamic diameter (PM_{10}, Lu et al., 2015), particulate matter $\leq$2.5 μm ($PM_{2.5}$, Ban et al., 2021),

☆ This paper has been recommended for acceptance by Admir Créso Targino.

* Corresponding author.
E-mail address: Duncan.Lee@glasgow.ac.uk (D. Lee).

https://doi.org/10.1016/j.envpol.2023.122465
Received 31 January 2023; Received in revised form 24 August 2023; Accepted 25 August 2023
Available online 26 August 2023

nitrogen dioxide (NO_2, Williams et al., 2014) and ozone (O_3, e.g. Zheng et al., 2021). Associations have been found with general classifications such as all-cause mortality (e.g. Orellano et al., 2020), and cause-specific classifications including respiratory diseases (e.g. Horne et al., 2018; Duan et al., 2020) and stroke (Shah et al., 2015).

However, the magnitude of these associations can vary considerably between studies, and as larger associations lead to increases in the estimated health and economic costs of air pollution from health impact assessments, their accurate quantification is vital. This heterogeneity in effect size could be caused by numerous factors including: (i) the study design including the definitions of the health outcomes and choice of exposure measurements/models (e.g., Sheppard et al., 2012); and (ii) the populations under study and the components within the pollutant mix (e.g., Hajat et al., 2021). A review of the possible determinants of this heterogeneity is given by Samet (2008). As a result of this heterogeneity the quantification of the health impacts associated with air pollution for use in policy formulation is typically based on meta-analyses of multiple studies, with examples including Atkinson et al. (2018), Chen and Hoek (2020) and Pope et al. (2020) for the long-term health effects, and Walton et al. (2014) and Orellano et al. (2020) for the short-term health effects.

Meta-analytic methods have also been used to examine between study heterogeneity, such as in the Air Pollution and Health: a European Approach (APHEA) study that focused on short-term exposures and mortality. Katsouyanni et al. (1997) identified heterogeneity in the effects of short-term exposures to black smoke and sulphur dioxide between cities in Western and Central-Eastern Europe, with larger effects in the former where pollution was more traffic related. Furthermore, Katsouyanni et al. (2009) investigated effect modification across Europe and North America, and for PM_{10} found that a higher percentage of older people and a higher unemployment rate were both associated with a greater effect of PM_{10} on all-cause mortality.

While there are many important methodological considerations governing health impact assessment, this paper focuses on between study heterogeneity due to differences in study location, hereafter termed regional heterogeneity. Such heterogeneity could be due to regional differences in factors affecting the pollution-health association, including mean concentrations, sources and composition of pollutants, the magnitudes of other mediating factors, and underlying population susceptibilities. This heterogeneity raises an important question as to which epidemiological studies policymakers should use when estimating a summary risk for a specific country or region of interest. Should risk estimates be based on the results from all studies, or should the evidence base be restricted to studies from the country or surrounding region? The UK Committee on the Medical Effects of Air Pollutants (COMEAP) use estimates from all studies of sufficient quality regardless of their location when quantifying effects in the UK (e.g. COMEAP, 2018; COMEAP, 2022). In other cases more localised approaches have been used, with the concentration-response functions recommended by WHO's HRAPIE project (WHO, 2013) for quantification of effects in Europe including only European studies, while French assessments often use concentration-response functions estimated from French populations (e.g. Pascal et al., 2016).

Meta-analyses such as Chen and Hoek (2020), Huangfu and Atkinson (2020) and Orellano et al. (2020) have examined regional heterogeneity, although often this is a supplementary analysis applied only to pollutant-outcome pairs that exhibit a high I^2 heterogeneity statistic (I^2 measures the percentage of the variation in the effect estimates due to real heterogeneity rather than random chance). In addition, their results are mainly inconclusive due to the regional heterogeneity not being statistically significant (see for example Figure B8 of Huangfu and Atkinson, 2020). This paper expands the understanding of regional heterogeneity by addressing the following questions.

(i) To what extent is there regional heterogeneity in the estimated short-term health effects of air pollution, and how do these results vary by pollutant-outcome pair?
(ii) To what extent could a lack of statistically significant regional heterogeneity be caused by a lack of statistical power?
(iii) What is the most appropriate meta-analytic approach for estimating a regional effect of air pollution on health in the presence of regional heterogeneity?

The first question is addressed by re-analysis of a recent database of studies compiled by Orellano et al. (2020), while the second and third questions are answered by extensive simulation studies.

2. Materials and methods

This section describes the database of studies used in our analysis, the statistical methods we adopt and the simulation study design. Following the Population, Exposure, Comparator, and Outcomes classification (PECO, Morgan et al., 2018): studies included were for both sexes and all ages in general populations; we consider exposures to NO_2, O_3, $PM_{2.5}$, and PM_{10}; relative risks are calculated for differences of 10 μg m^{-3} in exposures; and the disease endpoints are all-cause, cardiovascular and respiratory mortality.

2.1. Database of studies

The database is taken from the supplementary material of Orellano et al. (2020). It contains 795 pollutant-outcome pair relative risks (for a 10 μg m^{-3} increase in concentrations) from 196 studies published between January 1990 and November 2017, of which 50 are case crossover (CCO) studies and 146 are ecological time series (ETS) studies. Due to the equivalence of CCOs and ETSs (Lu and Zeger, 2007) we combine the estimates across these studies. The database was restricted to studies focusing on: (i) all ages and both sexes; (ii) all-cause, cardiovascular and respiratory mortality outcomes; (iii) NO_2, O_3, $PM_{2.5}$ and PM_{10} exposures; and (iv) single pollutant analyses. For NO_2, $PM_{2.5}$ and PM_{10} the exposure metric was restricted to a 24-hr mean, while for O_3 either maximum 8-hr or maximum 24-hr means were considered. No distinction was made between the two different metrics for O_3 in the main analysis, but a sensitivity analysis assessing the impact of the averaging period is presented in Table S2 of the supplementary material. Additionally, no distinction was made between the different lags of exposures (lags ranged between 0 and 7) to ensure there were enough studies to reliably estimate regional effects.

Regional heterogeneity is assessed at the continent rather than the country level, which although less desirable was the only practical approach given the very small number of studies available in some countries. The database contains 59 pollutant-outcome pair estimates in America (North, Central and South), 84 in Asia and 166 in Europe, with the remaining continents excluded due to their small number of studies. A summary of the number of studies for each pollutant-outcome pair by continent is shown in Table S1 of the supplementary material, which shows that studies examining. NO_2 and O_3 only relate to all-cause mortality, while studies examining $PM_{2.5}$ and PM_{10} additionally relate to cardiovascular and respiratory mortality.

2.2. Statistical methods

We compare the following approaches for estimating global and continent specific effects.

- **Global:** a random effects meta-analysis is applied to studies across all 3 continents, which estimates a single global effect.
- **Sub-group (SG):** a random effects meta-analysis is applied to studies in each continent separately to yield continent specific effects. The statistical significance of the regional heterogeneity is assessed using

an omnibus test based on Cochran's Q (Cochran, 1954), whose null hypothesis is that there are no differences in the effects between any pair of continents. Further details are given in Harrer et al. (2022).

- **Meta-regression (MR):** a random effects meta-regression is applied to all studies, where the single moderator (covariate) is a categorical variable defining which continent each study relates to. This approach yields continent specific effects, and the statistical significance of the regional heterogeneity is assessed by an omnibus test whose null hypothesis is that there are no differences in the effects between any pair of continents. This test is based on the estimated coefficients of the moderators, and further details are given in Viechtbauer (2010). Note, although not a problem here, if a meta-regression is run with other covariates in addition to continent, then this omnibus test will relate to all covariates, not just those relating to continent. In this case an alternative method to assess regional heterogeneity would be using the p-values of the continent-level regression coefficients.

All analyses are conducted in the statistical programme **R** (R Core Team, 2023), using the **meta** (Balduzzi et al., 2019) and **metafor** (Viechtbauer, 2010) packages. In all cases the Knapp and Hartung (2003) adjustment is used as recommended by Viechtbauer (2010), which accounts for the uncertainty in the random effects variance. This variance is estimated by the DerSimonian-Laird estimator (DerSimonian and Laird, 1986). Non-continent between study heterogeneity is estimated by the I^2 statistic, which measures the percentage of the variation in the individual study effect estimates due to heterogeneity in the true effect size rather than sampling error. Finally, the percentage of the between study heterogeneity that is explained by continent is computed as described in Viechtbauer (2010), as the percentage reduction in the random effects variance from including continent in the meta-regression. R code provided to re-produce our analyses is described in Section 2 of the supplementary material.

2.3. Simulation study design

We use simulated data to address the second and third motivating questions, and generate data under multiple scenarios to assess the sensitivity of the results to input parameters. We apply the three meta-analytic methods outlined above to 1000 simulated data sets generated under each scenario, and each simulated data set consists of log relative risks and standard errors from N studies from 2 different regions, denoted (A,B). Thus, N_A and N_B studies come from regions A and B respectively, with $N=N_A+N_B$. The aim is to estimate the relative risk in region A. The true relative risks for regions A and B are denoted by $RR_A = \exp(\beta_A)$ and $RR_B = \exp(\beta_B)$, where (β_A, β_B) are the true log relative risks. Here N log relative risks from the two regions are generated from

$$\begin{aligned}\widehat{\beta}_{A_i} &\sim N\left(\beta_A, T_A^2\right) \quad for \quad i=1,\ldots,N_A \quad and \quad \widehat{\beta}_{B_i} \\ &\sim N\left(\beta_B, T_B^2\right) \quad for \quad i=N_A+1,\ldots,N, \end{aligned} \tag{1}$$

Where $N(P, Q)$ denotes a normal distribution with mean P and variance Q *and* i denotes the ith study out of N. The corresponding relative risks are obtained by exponentiating, e.g., $\widehat{RR}_{A_i} = \exp\left(\widehat{\beta}_{A_i}\right)$. The between study variances (T_A^2, T_B^2) control the extent of the within region heterogeneity between the effect estimates from the individual studies. Then the individual study estimated standard errors are generated in a similar manner from

$$\begin{aligned}\widehat{\sigma}_{A_i} &\sim N\left(\sigma_A, V_A^2\right) \quad for \quad i=1,\ldots,N_A \quad and \quad \widehat{\sigma}_{B_i} \\ &\sim N\left(\sigma_B, V_B^2\right) \quad for \quad i=N_A+1,\ldots,N, \end{aligned} \tag{2}$$

where (σ_A, σ_B) denote the mean standard errors from studies in each region, while (V_A^2, V_B^2) are the between study variances for each region, which control how much within region heterogeneity there is between the estimated standard errors from the individual studies. Note, any negative estimated standard errors generated are replaced by 0.0001.

We choose (RR_A, RR_B, σ_A, σ_B, T_A, T_B, V_A, V_B, N_A, N_B) to be representative of the real data results summarised in Tables 1, 2 and Table S1. In all scenarios we fix the relative risk from region A, the region of interest, at $RR_A = 1.005$ based on the *$PM_{2.5}$ – all-cause mortality* estimate for Europe. We then consider scenarios covering all possible combinations of the following three characteristics.

- **Between region heterogeneity:** $RR_B = 1.005, 1.009, 1.013$, which correspond to none, moderate and large regional heterogeneity ($RR_A = 1.005$).
- **Within region heterogeneity (T):** is the same for both regions, i.e., $T = T_A = T_B$. We consider $T = 0.0025, 0.005$ or 0.01, which correspond to small, moderate and large within region heterogeneity.
- **Number of studies (N):** all pairwise combinations of $N_A = 3, 5, 10, 20, 30$ and $N_B = 3, 5, 10, 20, 30$, allowing us to examine situations where studies from both regions are rare or relatively common.

The within study variation in the estimates (σ_A, σ_B) is fixed at 0.005, while (V_A, V_B) are fixed at $\sigma_A/3$ to ensure that the simulated values of $\{\widehat{\sigma}_{A_i}, \widehat{\sigma}_{B_i}\}$ will be positive. The following metrics are used to investigate which of the methods best estimate regional relative risks (first 3) and identify statistically significant regional heterogeneity (last 2).

- **Percentage Bias:** The bias in the estimated effect from region A as a percentage of its true value. A bias of zero means the estimate is the correct overall size on average.
- **Percentage root mean square error (RMSE):** Measures how close the estimated effect is to the true value as a percentage of the true value. Smaller values of the RMSE equate to a better estimate.
- **Coverage percentage:** the percentage of the 95% confidence intervals for the estimated effect from region A that contain the true region A effect. The coverage percentage should be 95%.
- **Power:** the probability that a statistical test identifies statistically significant regional heterogeneity when it is present. Large values indicate better tests.
- **Type 1 error:** the probability that a statistical test identifies statistically significant regional heterogeneity when it is not present. The optimum value is 0.05 for a 5% test.

R code provided to re-produce the simulation study is described in Section 2 of the supplementary material.

3. Results

3.1. Real data: regional heterogeneity in short-term air pollution and health studies

These results extend the original Orellano et al. (2020) study, which only considered continent-level heterogeneity via sub-group analyses for the *$PM_{2.5}$ – respiratory mortality* and *NO_2 – all-cause mortality* associations. Here, Tables 1 and 2 present the following results for America, Asia and Europe: (i) global estimates for all studies (**Global**); (ii) continent specific estimates from sub-group analyses (**SG**); and (iii) continent specific estimates from meta-regression analyses (**MR**). All estimates are presented as relative risks and 95% confidence intervals for a 10 μg/m^3 increase in each pollutant. Additionally, the tables present the I^2 statistic as a general measure of heterogeneity, p-values quantifying whether there is statistically significant between continent heterogeneity, and the percentage of the between study heterogeneity due to continent.

The global estimates are similar to those presented in Table 1 of Orellano et al. (2020), but are not identical because we only use studies from America, Asia and Europe. All global effects are statistically significant at the 5% level, showing clear evidence that short-term

Table 1
Summary of the regional meta-analyses for the short-term effects of each pollutant (NO_2, O_3, $PM_{2.5}$, PM_{10}) on all-cause mortality (per 10 μg/m^3).The number of studies N is presented in brackets. I^2 represents the percentage of the variation in the effect estimates due to heterogeneity in the true effect size. The **Heterogeneity** column contains the p-values from the omnibus tests of regional heterogeneity discussed in Section 2.2.

(A) - NO_2 and all-cause mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (51)	1.0072 (1.0054, 1.0090)	77.5	–
Sub-group	America (7)	1.0053 (1.0040, 1.0065)	19.7	p-value – 0.0008
	Asia (16)	1.0110 (1.0078, 1.0141)	79.6	
	Europe (28)	1.0047 (1.0023, 1.0071)	59.4	
Meta-regression	America (7)	1.0062 (1.0025, 1.0100)	NA	p-value - 0.0041
	Asia (16)	1.0107 (1.0081, 1.0132)	NA	
	Europe (28)	1.0047 (1.0024, 1.0070)	NA	
Continental heterogeneity	*23.2%*			
(B) – O_3 and all-cause mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (42)	1.0043 (1.0030, 1.0055)	81.0	–
Sub-group	America (10)	1.0022 (1.0010, 1.0033)	61.7	p-value – 0.0008
	Asia (9)	1.0043 (1.0002, 1.0083)	83.9	
	Europe (23)	1.0058 (1.0040, 1.0075)	77.8	
Meta-regression	America (10)	1.0020 (0.9999, 1.0042)	NA	p-value - 0.0361
	Asia (9)	1.0041 (1.0016, 1.0066)	NA	
	Europe (23)	1.0056 (1.0040, 1.0073)	NA	
Continental heterogeneity	*5.5%*			
(C) – $PM_{2.5}$ and all-cause mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (28)	1.0065 (1.0037, 1.0093)	84.0	–
Sub-group	America (8)	1.0116 (1.0070, 1.0163)	38.1	p-value – 0.0008
	Asia (6)	1.0037 (0.9995, 1.0079)	49.7	
	Europe (14)	1.0042 (0.9998, 1.0087)	86.2	
Meta-regression	America (8)	1.0116 (1.0078, 1.0155)	NA	p-value - 0.0086
	Asia (6)	1.0038 (1.0008, 1.0068)	NA	
	Europe (14)	1.0049 (1.0009, 1.0089)	NA	
Continental heterogeneity	*63.4%*			
(D) – PM_{10} and all-cause mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (64)	1.0041 (1.0031, 1.0051)	76.0	–
Sub-group	America (14)	1.0058 (1.0025, 1.0092)	86.5	p-value – 0.0890
	Asia (19)	1.0030 (1.0021, 1.0038)	59.9	
	Europe (31)	1.0044 (1.0025, 1.0092)	61.0	
Meta-regression	America (14)	1.0052 (1.0032, 1.0072)	NA	p-value - 0.2483
	Asia (19)	1.0031 (1.0015, 1.0047)	NA	
	Europe (31)	1.0045 (1.0028, 1.0061)	NA	
Continental heterogeneity	*2.2%*			

exposure to each of the pollutants is harmful to each of the health outcomes considered. The associations between $PM_{2.5}$/PM_{10} and cardiovascular/respiratory mortality are all higher than the corresponding associations with all-cause mortality, with the latter being relative risks of 1.0065 for $PM_{2.5}$ and 1.0041 for PM_{10}.

Forest plots for the 8 global associations are displayed in Figs. S1–S8 in Section 1 of the supplementary material.

All pollutant-outcome pairs show between continent heterogeneity in their relative risks, with the smallest heterogeneity occurring for *PM_{10} – all-cause mortality* which exhibits continent specific relative risks ranging between 1.0031 (Asia) and 1.0052 (America) from the meta-regression. However, despite these differences statistically significant continental heterogeneity is only observed for 4 out of the 8 pollutant-outcome pairs at a 5% significance level, which includes NO_2, O_3 and $PM_{2.5}$ with all-cause mortality, and $PM_{2.5}$ with cardiovascular mortality. In contrast, the effect estimates for *PM_{10} – all-cause mortality, PM_{10} – cardiovascular mortality* and *$PM_{2.5}$ – respiratory mortality* show no statistically significant regional heterogeneity, while *PM_{10} – respiratory mortality* shows inconclusive results with one of the two p-values being significant at the 5% level. Tables 1 and 2 also show that the percentage of the between study heterogeneity explained by continent is largest for $PM_{2.5}$, ranging between 61.4% and 79.4% across the 3 health outcomes. The smallest percentages are generally observed for O_3 and PM_{10}, which typically account for less than 10% of this variation (except for *PM_{10} – respiratory mortality*).

The continent specific effect estimates using sub-group analysis and meta-regression are mostly similar, and where they differ there is no clear pattern as to which estimate is larger. The 95% confidence intervals from the sub-group analyses are typically wider than those from the meta-regressions (also Tables 1 and 2), being wider in 13 cases, narrower in 7 cases and similar in 4 cases. For the four pollutant outcome pairs that were identified as having statistically significant regional heterogeneity, there are clear differences between the relative risks for at least 1 pair of continents. For *NO_2 – all-cause mortality* the relative risk for Asia (1.0107) is significantly higher than that for Europe (1.0047), as the corresponding 95% confidence intervals do not overlap. For *O_3 – all-cause mortality* the relative risk for Europe (1.0056) is borderline significantly higher than that for America (1.0020), as the 95% confidence intervals do not overlap from the sub-group analysis and only marginally overlap from the meta-regression. Similar borderline statistically significant differences are observed for the *$PM_{2.5}$ – all-cause mortality* and *$PM_{2.5}$ – cardiovascular mortality* associations. For the former the relative risk for America (1.0116) is higher than that for Asia (1.0038), while for the latter the relative risk for America (1.0121)

Table 2
Summary of the regional meta-analyses for the short-term effects of each pollutant ($PM_{2.5}$, PM_{10}) on cardiovascular and respiratory mortality (per 10 μg/m^3). The number of studies N is presented in brackets. I^2 represents the percentage of the variation in the effect estimates due to heterogeneity in the true effect size. The **Heterogeneity** column contains the p-values from the omnibus tests of regional heterogeneity discussed in Section 2.2.

(A) – $PM_{2.5}$ and cardiovascular mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (27)	1.0093 (1.0059, 1.0126)	91.7	–
Sub-group	America (7)	1.0136 (1.0066, 1.0206)	63.4	p-value – 0.0056
	Asia (7)	1.0044 (1.0018, 1.0070)	59.8	
	Europe (13)	1.0091 (1.0026, 1.0157)	17.7	
Meta-regression	America (7)	1.0121 (1.0082, 1.0159)	NA	p-value - 0.0157
	Asia (7)	1.0048 (1.0018, 1.0078)	NA	
	Europe (13)	1.0096 (1.0035, 1.0157)	NA	
Continental heterogeneity	*79.4%*			
(B) – PM_{10} and cardiovascular mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (40)	1.0059 (1.0039, 1.0078)	62.8	–
Sub-group	America (4)	1.0066 (0.9978, 1.0154)	90.5	p-value – 0.1755
	Asia (11)	1.0036 (1.0006, 1.0066)	83.9	
	Europe (25)	1.0071 (1.0042, 1.0101)	45.4	
Meta-regression	America (4)	1.0063 (1.0018, 1.0108)	NA	p-value - 0.3564
	Asia (11)	1.0040 (1.0006, 1.0073)	NA	
	Europe (25)	1.0071 (1.0043, 1.0100)	NA	
Continental heterogeneity	*0.2%*			
I – $PM_{2.5}$ and respiratory mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (19)	1.0073 (1.0029, 1.0118)	86.8	–
Sub-group	America (4)	1.0109 (0.9982, 1.0166)	66.5	p-value – 0.2252
	Asia (4)	1.0038 (1.0002, 1.0073)	23.0	
	Europe (11)	1.0056 (0.9947, 1.0166)	27.9	
Meta-regression	America (4)	1.0110 (1.0051, 1.0170)	NA	p-value - 0.2639
	Asia (4)	1.0048 (0.9994, 1.0102)	NA	
	Europe (11)	1.0057 (0.9971, 1.0143)	NA	
Continental heterogeneity	*61.4%*			
(D) – PM_{10} and respiratory mortality				
Model	Continent (N)	RR (95% CI)	I^2 (%)	Heterogeneity
Global	All (38)	1.0094 (1.0061, 1.0127)	62.3	–
Sub-group	America (5)	1.0121 (1.0018, 1.0225)	79.6	p-value – 0.0084
	Asia (12)	1.0039 (1.0014, 1.0065)	4.7	
	Europe (21)	1.0115 (1.0057, 1.0173)	41.2	
Meta-regression	America (5)	1.0116 (1.0058, 1.0174)	NA	p-value - 0.1136
	Asia (12)	1.0053 (1.0006, 1.0100)	NA	
	Europe (21)	1.0118 (1.0069, 1.0168)	NA	
Continental heterogeneity	*38.3%*			

is again higher than that for Asia (1.0048).

3.2. Simulation study: identification of statistically significant regional heterogeneity

We address the second motivating question by first comparing the sub-group and meta-regression methods in the simulation study scenarios with regional heterogeneity, i.e. where $RR_B = 1.009$ or 1.013 ($RR_A = 1.005$ in all cases). The results are displayed in the bottom portion of Fig. 1 in panels (D) – (I), which display the statistical power of the two tests outlined in Section 2.2 for the 6 pairwise combinations of $RR_B = 1.009$, 1.013 and $T = 0.0025$, 0.005 or 0.01. The results within a single panel correspond to all pairwise combinations of $N_A = 3$, 5, 10, 20, 30 and $N_B = 3$, 5, 10, 20, 30 for that scenario. Note, all tests are conducted at the 5% level.

Fig. 1 shows that just because there is between region heterogeneity in the true relative risks, it does not mean that either test will detect it. Over all scenarios considered the power ranges between 0.07 and 1, with a median value of 0.63. For example, if the between region heterogeneity is small and the within region heterogeneity is large (panel F), then the maximum power is only 0.32 even with 30 studies in each of the two regions. In contrast, if the between region heterogeneity is large and the within region heterogeneity is moderate (panel H), then having only 10 studies in each region gives a power of 0.93. Thus, power increases when there are: (a) more studies available in each region; (b) larger between region heterogeneity; and (c) smaller within region heterogeneity.

The heterogeneity test used in the sub-group analysis exhibits higher power than the heterogeneity test from the meta-regression analysis in all cases, with median powers of 0.71 (SG) and 0.59 (MR) respectively. However, this increased power comes at the cost of a higher type 1 error for the sub-group test compared with the meta-regression test in all scenarios when $RR_A = RR_B = 1.005$ (no regional heterogeneity). These type 1 errors are displayed in the top portion of Fig. 1 in panels (A)–(C). The theoretical type 1 error is 0.05 but in practice ranges between 0.05 and 0.32 for the sub-group test and between 0.04 and 0.32 for the meta-regression test, with median values of 0.15 (sub-group) and 0.09 (meta-regression). Thus, as a test of whether regional heterogeneity is statistically significant, the meta-regression test is more conservative.

3.3. Simulation study: estimation of region-specific effects

The accuracy of the three methods for estimating the effects in region A is displayed in Fig. 2 (percentage bias), Fig. 3 (coverage percentages)

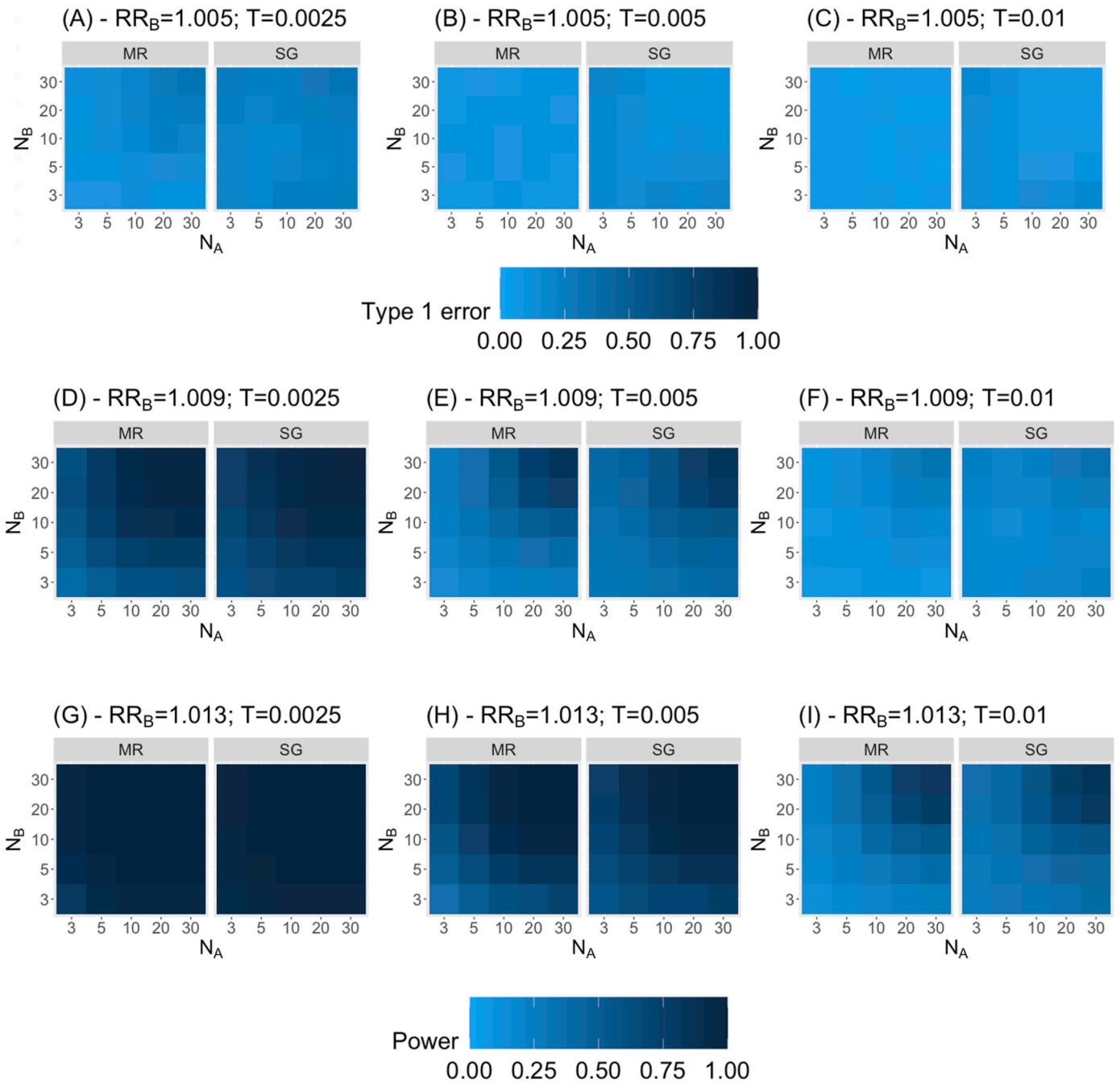


Fig. 1. Summary of the type 1 error in panels A–C where there is no difference between the two region's relative risks, and the statistical power in panels D–F where there is a difference between the two region's relative risks for the sub-group (SG) and meta-regression (MR) analyses for each simulation scenario. All results are based on 1000 simulated data sets and a significance level of 5%. Note, in all scenarios the true relative risk in region A is $RR_A = 1.005$, and the within region standard deviations are the same in both regions, i.e., $T = T_A = T_B$.

and Fig. S9 (percentage RMSE), the latter being in Section 3 of the supplementary material. Firstly, as expected, if there is no regional heterogeneity (panels (A)–- (C)) then no bias is observed and the coverage percentages are always close to their desired 95% levels for all three methods. In this setting the RMSEs are smallest when doing a simple random effects meta-analysis of all studies, because these estimates are based on the largest amount of data.

Secondly, if there is regional heterogeneity (panels (D) – (I)), then performing a random effects meta-analysis of all studies regardless of region leads to biased estimates of the relative risk in the region of interest. This bias ranges between 8% and 147% of the true relative risk, with a median value of 60% across the scenarios considered. The bias increases as both the magnitude of the between region heterogeneity increases and the proportion of studies from the region not of interest (Region B) increases. In contrast, sub-group and meta-regression analyses always lead to negligible bias, with percentage biases ranging between −7% and 3%. The bias from using a global meta-analysis also deleteriously affects the coverage percentages of its estimates, with the coverages ranging between 0% and 95% across all scenarios, with a median value of only 54%. These poor coverage percentages are not observed for the sub-group and meta-regression methods, whose values range between 70% and 96% with a median value of 90%.

Finally, the RMSEs presented in Fig. S9 show that across panels (D) to (I) where regional heterogeneity is present a global meta-analysis has a

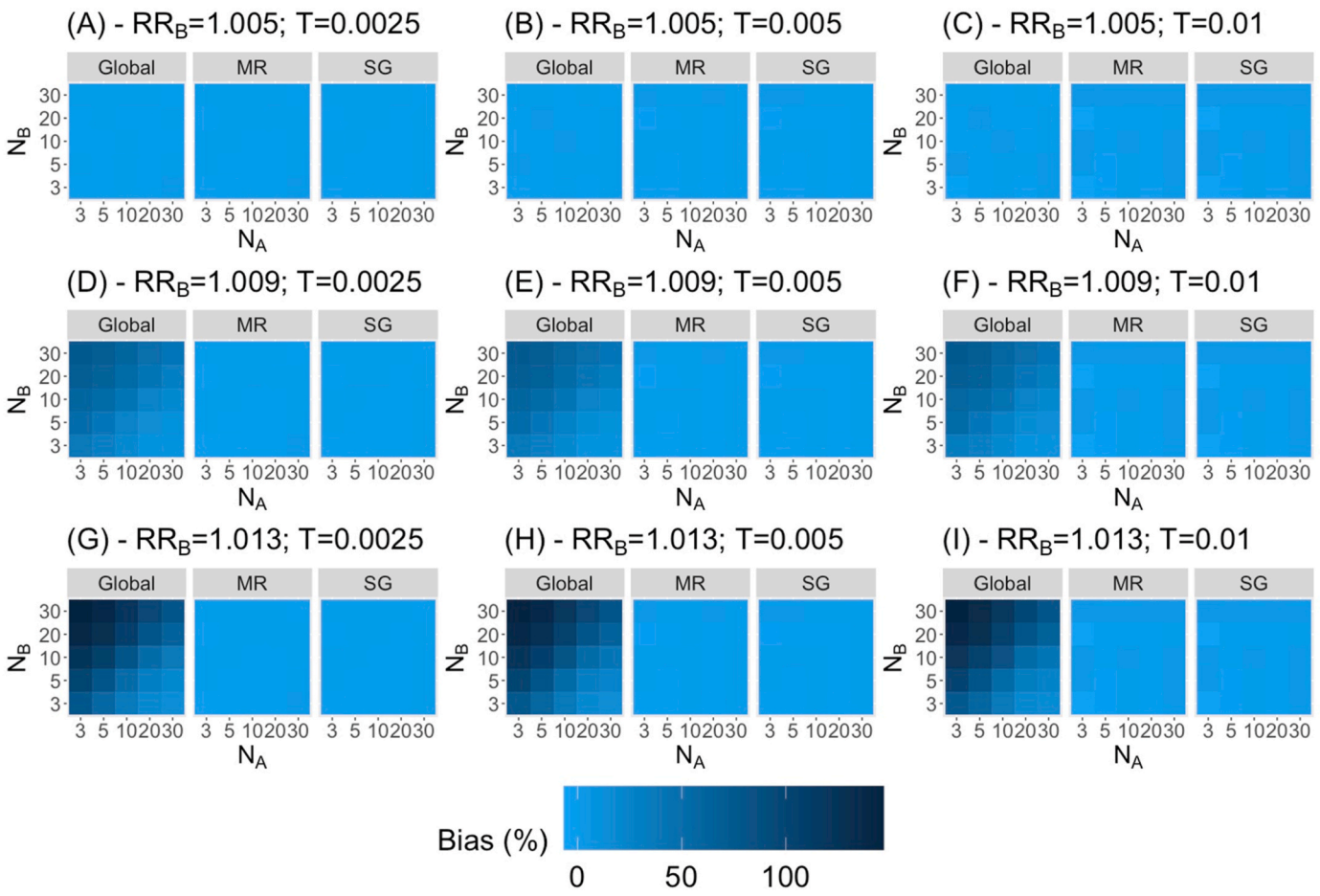


Fig. 2. Summary of the percentage biases from the 3 methods in the different scenarios in the simulation study. Note, in all scenarios the true relative risk in region A is $RR_A = 1.005$, and the within region standard deviations are the same in both regions, i.e., $T = T_A = T_B$.

higher RMSE in 85% of cases than either the sub-group or meta-regression analyses, with the latter two having very similar values and neither being preferable in general. In contrast, the only cases where a global meta-analysis outperforms a sub-group or meta-regression analysis in the presence of regional heterogeneity is when there is small between region heterogeneity, large within region heterogeneity, and a small number of studies in the region of interest (region A, see panel (F)).

4. Discussion

The main finding from our re-analysis is that regional heterogeneity is statistically significant at the 5% level for four of the eight pollutant-outcome pairs, namely for NO_2, O_3 and $PM_{2.5}$ with all-cause mortality, and $PM_{2.5}$ with cardiovascular mortality. Regional heterogeneity was not clearly statistically significant for the four remaining pollutant-outcome pairs, despite most of them having substantial heterogeneity in their continental-level relative risks in epidemiological terms. The main finding from the simulation study is that in the presence of regional heterogeneity, a sub-group analysis or a meta-regression analysis is more likely to lead to accurate estimation of region-specific effects than a global meta-analysis. This is evidenced by these methods almost always having lower absolute biases, lower RMSEs and higher coverage percentages than those from a global meta-analysis of all studies. The only counterpoint is when there is small between region heterogeneity, large within region heterogeneity and a small number of studies in the region of interest, in which case a global meta-analysis that borrows strength across all studies is likely to be superior. For example, in the scenario where there is low between region heterogeneity and large within region heterogeneity ($RR_B = 1.009$ and $T = 0.01$), then a global estimate generally outperforms a regional estimate in terms of RMSE if the number of studies in the region of interest is 10 or less. Conversely if between region heterogeneity is large ($RR_B = 1.013$), then a global estimate performs worse than a regional estimate in all scenarios except for the specific case where there are 3 studies in each of the two regions and within region heterogeneity is large.

The effect estimates from sub-group analyses and meta-regression are generally very similar, with neither being more accurate in general. Furthermore, in the presence of regional heterogeneity the median power across all scenarios was only 0.71 for sub-group analyses and 0.59 for meta-regression, and these powers decrease with decreasing sample size, decreasing between region heterogeneity and increasing within region heterogeneity. However, the increased power for the sub-group analyses compared to meta-regression comes at the cost of an increased type 1 error when there is no regional heterogeneity, which have median values of 0.15 (sub-group) and 0.09 (meta-regression) over all scenarios considered. This means that meta-regression yields improved specificity at the cost of reduced sensitivity compared to a sub-group analysis, and thus arguably should be preferred in practice as it is more conservative. Collectively, these findings imply that some instances of real regional heterogeneity will go undetected by these meta-analytic approaches, meaning that more than 4 of the 8 pollutant-outcome pairs from this study are likely to exhibit real heterogeneity. This may be the reason why despite having large differences in effect sizes, the $PM_{2.5}$ – respiratory mortality association based on only 19 studies (America – 4, Asia – 4, Europe – 11) did not exhibit statistically significant heterogeneity.

Our study does have a number of limitations that provide natural areas for future work. Firstly, we were only able to assess between

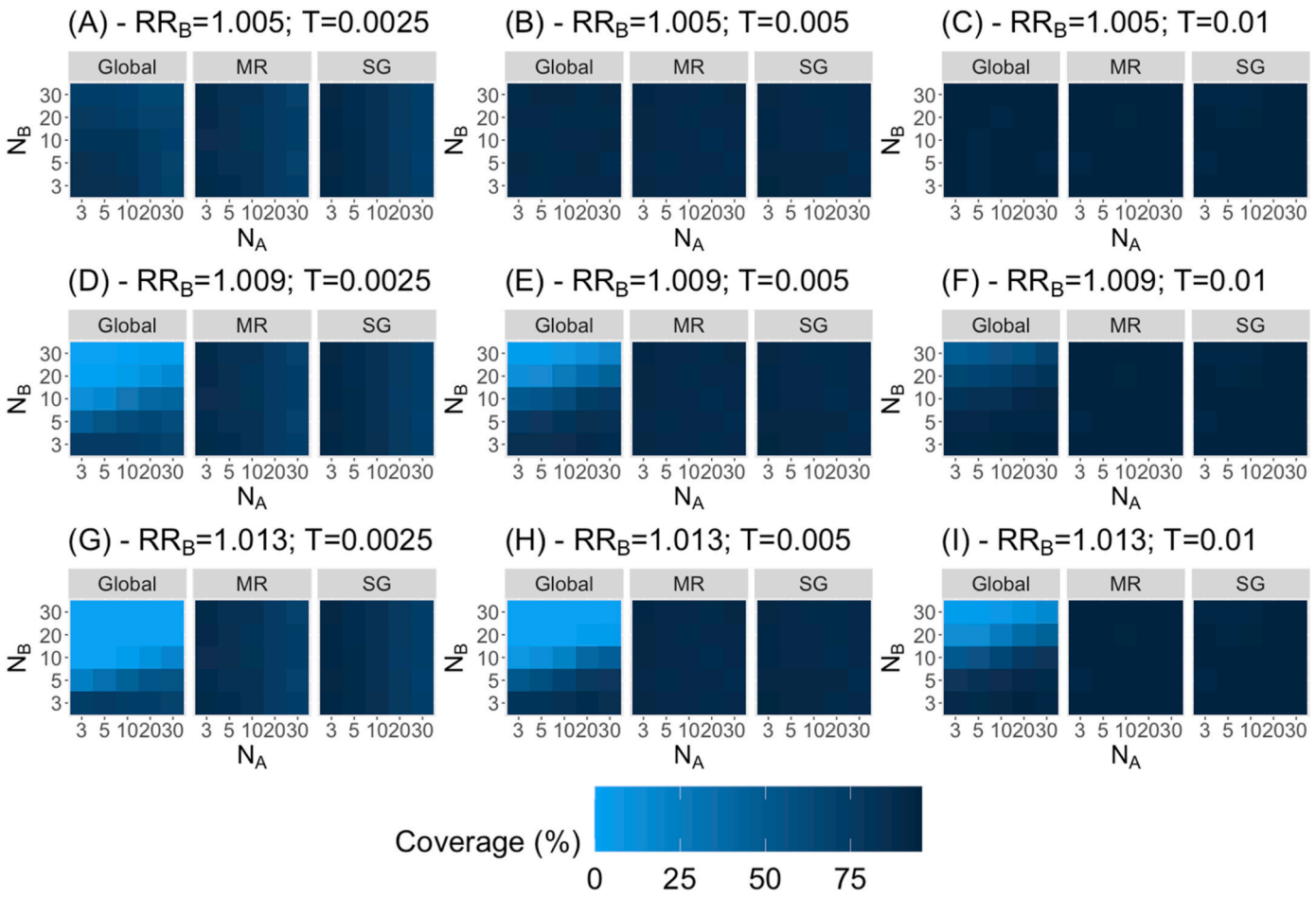


Fig. 3. Summary of the coverage percentages of the 95% confidence intervals from the 3 methods in the different scenarios in the simulation study. Note, in all scenarios the true relative risk in region A is $RR_A = 1.005$, and the within region standard deviations are the same in both regions, i.e., $T = T_A = T_B$.

continent heterogeneity due to data limitations, and it is likely that relative risks will be heterogeneous between different countries within the same continent, as evidenced, for example, by Katsouyanni et al. (1997) in the APHEA study who reported heterogeneous effects between Eastern and Western Europe. Secondly, we have quantified the magnitude of between continent heterogeneity but have not attempted to identify its drivers, which will include factors such as differences in pollution compositions, differences in confounders such as meteorology, spatial variation in population characteristics and underlying susceptibilities, and the exposure assessment and statistical methods used in the individual studies. As suggested by COMEAP (2018), if information were available on these drivers then it could be taken into account when considering the transferability of evidence between regions and which studies should be used to produce summary concentration response functions for quantification purposes. Finally, the analyses presented here are based on studies examining the effects of short-term exposure to air pollution, because a comprehensive database of existing studies was available from Orellano et al. (2020). However, given the importance of understanding the risks associated with long-term exposures to air pollution for quantifying mortality burdens, an important avenue of future work is to perform a similar analysis of regional heterogeneity for long-term air pollution studies.

5. Conclusion

Our meta-analysis found that continent-level heterogeneity appears to be present in most of the pollutant outcome pairs considered, although some of it was not statistically significant. Our simulation studies show that this latter point may be because meta-analyses are likely to be underpowered to detect such statistically significant regional heterogeneity, as a result of the small numbers of studies and within region heterogeneity in the effect estimates. These results suggest that regional heterogeneity is an important consideration for policymakers constructing concentration response functions and estimating the mortality burden of air pollution for specific regions. In this setting we recommend that unless the number of studies in the region of interest is small and their estimated relative risks are heterogeneous, then regional concentration response functions should be estimated using meta-regression analyses, rather than relying on global effect estimates.

Author statement

Duncan Lee - Methodology, Formal analysis, Writing – original draft and Review and Editing. Heather Walton – Conceptualization, Writing – original draft and Review and Editing. Dimitris Evangelopoulos - Methodology, Writing – original draft and Review and Editing. Klea Katsouyanni - Writing – original draft and Review and Editing. Alison M Gowers – Conceptualization, Project administration, Writing – review & editing. Gavin Shaddick - Methodology, Formal analysis, Writing – review & editing. Christina Mitsakou – Visualisation, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Data availability

The research data are publicly available

Acknowledgements

The authors gratefully acknowledge comments from the editor and 3 anonymous reviewers that have improved the motivation for and content of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2023.122465.

References

Atkinson, R.W., Butland, B.k., Anderson, H.R., Maynard, R.L., 2018. Long-term concentrations of nitrogen dioxide and mortality – a meta-analysis of cohort studies. Epidemiology 29, 460–472.
Balduzzi, S., Rücker, G., Schwarzer, G., 2019. How to perform a meta-analysis with R: a practical tutorial. Evid. Base Ment. Health 22, 153–160.
Brunekreef, B., Holgate, S., 2002. Air pollution and health. Lancet 360, 1233–1242.
Ban, J., Wang, Q., Ma, R., Zhang, Y., Shi, W., Zhang, Y., Chen, C., Sun, Q., Wang, Y., Guo, X., Li, T., 2021. Associations between short-term exposure to $PM_{2.5}$ and stroke incidence and mortality in China: a case-crossovere study and estimation of the burden. Environ. Pollut. 268, 115743.
Chen, J., Hoek, G., 2020. Long-term exposure to PM and all-cause and cause-specific mortality: a systematic review and meta-analysis. Environ. Int. 143, 105974.
Cochran, G., 1954. Some methods for strengthening the common chi-squared tests. Biometrics 10, 417–451.
COMEAP, 2018. Associations of long-term average concentrations of nitrogen dioxide with mortality. In: Public Health England Committee on the Medical Effects of Air Pollutants (COMEAP).
COMEAP, 2022. Statement on quantifying mortality associated with long-term exposure to fine particulate matter. Comm. Med. Effects of Air Pollut.
DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. Contr. Clin. Trials 7, 177–188.
Duan, R., Hao, K., Yan, T., 2020. Air pollution and chronic obstructive pulmonary disease. Chronic Dis. Transl. Med. 6, 260–269.
Hajat, A., MacLehose, R.F., Rosofsky, A., Walker, K.D., Clougherty, J.E., 2021. Confounding by socioeconomic status in epidemiological studies of air pollution and health: challenges and opportunities. Environ. Health Perspect. 129, 65001.
Harrer, M., Cuijpers, P., Furukawa, T., Ebert, D., 2022. Doing Meta-Analysis with R: A Hands-On Guide, first ed. CRC Press.
Horne, B.D., Joy, E.A., Hofmann, M.G., Gesteland, P.H., Cannon, J.B., Lefler, J.S., Blagev, D.P., Korgenski, E.K., Torosyan, N., Hansen, G.I., Kartchner, D., Pope, C.A., 2018. Short-term elevation of fine particulate matter air pollution and acute lower respiratory infection. Am. J. Respir. Crit. Care Med. 198, 759–766.
Huangfu, P., Atkinson, R.W., 2020. Long-term exposure to NO_2 and O_3 and all-cause and respiratory mortality: a systematic review and meta-analysis. Environ. Int. 144, 105998.
Katsouyanni, K., Touloumi, G., Spix, C., Schwartz, J., Balducci, F., Medina, S., Rossi, G., Wojtyniak, B., Sunyer, J., Bacharova, L., Schouten, J.P., Ponka, A., Anderson, H.R., 1997. Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities: results from time series data from the APHEA project. Br. Med. J. 314, 1658–1663.
Katsouyanni, K., Samet, J.M., Anderson, H.R., Atkinson, R.W., Le Tertre, A., Medina, S., Samoli, E., Touloumi, G., Burnett, R.T., Krewski, D., Ramsay, T., Dominici, F., Peng, R.D., Schwartz, J., Zanobetti, A., 2009. Air Pollution and Health: a European and North American Approach (APHENA), vol. 142. Research Reports Heatlh Effects Institute, pp. 5–90.
Knapp, G., Hartung, J., 2003. Improved tests for a random effects meta-regression with a single covariate. Stat. Med. 22, 2693–2710.
Lu, Y., Zeger, S., 2007. On the equivalence of case-crossover and time series methods in environmental epidemiology. Biostatistics 8, 337–344.
Lu, F., Xu, D., Cheng, Y., Dong, S., Guo, C., Jiang, X., Zheng, X., 2015. Systematic review and meta-analysis of the adverse health effects of ambient $PM_{2.5}$ and PM_{10} pollution in the Chinese population. Environ. Res. 136, 196–204.
Morgan, R.L., Whaley, P., Thayer, K.A., Schunemann, H.J., 2018. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. Environ. Int. 121, 1027–1031.
Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., Ciapponi, A., 2020. Short-term exposure to particulate matter (PM_{10} and $PM_{2.5}$), nitrogen dioxide (NO_2), and ozone (O_3) and all-cause and cause-specific mortality: systematic review and meta-analysis. Environ. Int. 142, 105876.
Pascal, M., Chanel, P.d.C., Wagner, V., Corso, M., Tillier, C., Bentayeb, M., Blanchard, M., Cochet, A., Pascal, L., Host, S., Goria, S., Le Tertre, A., Chatignoux, E., Ung, A., Beaudeau, P., Medina, S., 2016. The mortality impacts of fine particles in France. Sci. Total Environ. 571, 416–425.
Pope, C.A., Coleman, N., Pond, Z.A., Burnett, R.T., 2020. Fine particulate air pollution and human mortality: 25+ years of cohort Studies. Environ. Res. 183, 108924.
R Core Team, 2023. _R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Samet, J., 2008. Air pollution risk estimates: determinants of Heterogeneity. J. Toxicol. Environ. Health, Part A 71, 578–582.
Shah, A., Lee, K.K., McAllister, D.A., Hunter, A., Nair, H., Whiteley, W., Langrish, J.P., Newby, D.E., Mills, N.L., 2015. Short-term exposure to air pollution and stroke: systematic review and meta-analysis. Br. Med. J. 350.
Sheppard, L., Burnett, R.T., Szpiro, A.A., Kim, S.-Y., Jerrett, M., Pope III, C Arden, Brunekreef, B., 2012. Confounding and exposure measurement error in air pollution epidemiology. Air Qual., Atmos. Health 5, 203–216.
Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. J. Stat. Software 36, 3.
Walton, H.A., Kang, S., Atkinson, R.W., Mills, I.C., Anderson, H.R., 2014. Quantitative systematic review of the associations between short-term exposure to ambient ozone and mortality and hospital admissions. In: Atkinson, R.W., Mills, I.C., Walton, H.A., Kang, S., Anderson, H.R. (Eds.), Systematic Review and Quantitative Meta-Analysis of the Evidence for Associations between Chronic and Short-Term Exposure to Outdoor Air Pollutants and Health. Department of Health Policy Research Programme Project, 002/0037.
WHO, 2013. Health Risks of Air Pollution in Europe – HRAPIE Project. Recommendations for Concentration–Response Functions for Cost–Benefit Analysis of Particulate Matter, Ozone and Nitrogen Dioxide. World Health Organisation.
Williams, M.L., Atkinson, R.W., Anderson, H.R., Kelly, F.J., 2014. Associations between daily mortality in London and combined oxidant capacity, ozone and nitrogen dioxide. Air Quality Atmosphere and Health 7, 407–414.
WHO, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease. World Health Organisation.
Zheng, X.Y., Orellano, P., Lin, H.L., Jiang, M., Guan, W.J., 2021. Short-term exposure to ozone, nitrogen dioxide, and sulphur dioxide and emergency department visits and hospital admissions due to asthma: a systematic review and meta-analysis. Environ. Int. 150, 106435.