



Short-term exposure to sulphur dioxide (SO₂) and all-cause and respiratory mortality: A systematic review and meta-analysis

Pablo Orellano^{a,*}, Julieta Reynoso^b, Nancy Quaranta^c

^a Centro de Investigaciones y Transferencia San Nicolás, Universidad Tecnológica Nacional (CONICET), San Nicolás, Argentina

^b Hospital General "San Felipe", San Nicolás, Argentina

^c Facultad Regional San Nicolás, Universidad Tecnológica Nacional, San Nicolás, Argentina, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, La Plata, Argentina

ARTICLE INFO

Handling Editor: Da Chen

Keywords:

Sulfur dioxide

Mortality

Time series studies

Observational study

Systematic review

Meta-analysis

ABSTRACT

Background: Many studies have assessed the harmful effects of ambient air pollution on human mortality, but the evidence needs further exploration, analysis, and refinement, given the large number of studies that have been published in recent years. The objective of this study was to evaluate all the available evidence of the effect of short-term exposure to ambient sulphur dioxide (SO₂) on all-cause and respiratory mortality.

Methods: Articles reporting observational epidemiological studies were included, comprising time-series and case-crossover designs. A broad search and wide inclusion criteria were considered, encompassing international and regional databases, with no geographical or language restrictions. A random effect meta-analysis was conducted, and pooled relative risk for an increment of 10 µg/m³ in SO₂ concentrations were calculated for each outcome. We analysed the risk of bias (RoB) in individual studies for specific domains using a new domain-based RoB assessment tool, and the certainty of evidence across studies with an adaptation of the Grading of Recommendations Assessment, Development and Evaluation approach. The certainty of evidence was judged separately for each exposure-outcome combination. A number of subgroup and sensitivity analyses were carried out, as well as assessments of heterogeneity and potential publication bias. The protocol for this review was registered with PROSPERO (CRD42019120738).

Results: Our search retrieved 1,128 articles, from which 67 were included in quantitative analysis. The RoB was low or moderate in the majority of articles and domains. An increment of 10 µg/m³ in SO₂ (24-hour average) was associated with all-cause mortality (RR: 1.0059; 95% CI: 1.0046–1.0071; p-value: <0.01), and respiratory mortality (RR: 1.0067; 95% CI: 1.0025–1.0109; p-value: <0.01), while the same increment in SO₂ (1-hour max.) was associated with respiratory mortality (RR: 1.0052; 95% CI: 1.0013–1.0091; p-value: 0.03). Similarly, the association was positive but non-significant for SO₂ (1-hour max.) and all-cause mortality (RR: 1.0016; 95% CI: 0.9930–1.0102; p-value: 0.60). These associations were still significant after the adjustment for particulate matter, but not for other pollutants, according to the results from 13 articles that evaluated co-pollutant models. In general, linear concentration–response functions with no thresholds were found for the two outcomes, although this was only evaluated in a small number of studies. We found signs of heterogeneity for SO₂ (24-hour average) – respiratory mortality and SO₂ (1-hour max.) – all-cause mortality, and funnel plot asymmetry for SO₂ (24-hour average) – all-cause mortality. The certainty of evidence was high in two combinations, i.e. SO₂ (24-hour average) – all-cause mortality and SO₂ (1-hour max.) – respiratory mortality, moderate in one combination, i.e. SO₂ (24-hour average) – respiratory mortality, and low in the remaining one combination.

Conclusions: Positive associations were found between short-term exposure to ambient SO₂ and all-cause and respiratory mortality. These associations were robust against several sensitivity analyses, and were judged to be of moderate or high certainty in three of the four exposure-outcome combinations.

* Corresponding author at: Colon 332, San Nicolas, Provincia de Buenos Aires ZIP 2900, Argentina.

E-mail address: porellano@frsn.utn.edu.ar (P. Orellano).

<https://doi.org/10.1016/j.envint.2021.106434>

Received 30 July 2020; Received in revised form 21 January 2021; Accepted 30 January 2021

Available online 15 February 2021

0160-4120/© 2021 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Sulphur dioxide (SO₂) is a common atmospheric pollutant naturally generated by geothermal activities, usually derived from volcanoes, or produced by human activity, e.g. the combustion of coal and petroleum (Cullis and Hirschler, 1980). In Europe, the emission of SO₂ increased since the beginning of 1900s, with a higher increment after the Second World War, and a substantial reduction after 1980 s, but only in some regions (Mylona, 1996). A similar decrease of anthropogenic SO₂ emissions was observed for the United States during the first decade of the 21st century, with the exception of specific seasons and regions, in which these concentrations increased (Hand et al., 2012). The deleterious effect of the short-term exposure to SO₂ on human health, particularly on the mortality risk, was investigated in many epidemiological studies, and summarized in a number of systematic reviews and meta-analyses (Atkinson et al., 2012; Shah et al., 2015; 2013; Stieb et al., 2002; Yang et al., 2014). According to the last Integrated Science Assessment for SO₂ prepared by the Environmental Protection Agency of the United States (US EPA, 2015), there is consistent epidemiologic evidence indicating a positive association between short-term SO₂ exposure and all-cause and respiratory mortality. However, this evidence is suggestive of, but not sufficient to infer, a causal relationship between the exposure and the outcome, mainly due to uncertainties related to the presence of confounders and effect modifiers, to seasonal differences, and to inconclusiveness regarding the shape of the concentration–response function (CRF) (US EPA, 2015).

The evidence of the effects of SO₂ and other common air pollutants on human health has been compiled to inform the air quality guidelines (AQGs) published by the World Health Organization (WHO), a primary reference for air pollution standards and policies worldwide. The last version of this document was released in 2006 (WHO, 2006), and involved analyses of the evidence for several air pollutants including SO₂, with insightful information regarding risk values, CRFs, and potential thresholds. However, numerous studies have been published since that year, and the exact shape of the CRFs is insufficiently defined for many pollutants and outcomes (Landrigan et al., 2018). As a result of the vast amount of evidence published in recent years, WHO has convened a Guideline Development Group to update the guidelines. This systematic review was commissioned by the WHO Regional Office for Europe's European Centre for Environment and Health in order to generate evidence to support the new update of the AQGs, with the aim of providing updated evidence-based numerical concentration levels (i.e. guidelines) and, where possible, an indication of the shape of the CRF for a number of ambient air pollutants, for relevant averaging times (i.e. long- and short-term exposure duration) and in relation to selected health outcomes. In particular, the objective of this systematic review and meta-analysis was to synthesize the worldwide evidence of the effects of short-term exposure to SO₂ on all-cause and respiratory mortality. This review complements the Orellano et al.'s paper (2020), published in this issue.

2. Methods

2.1. Protocol, registration, and reporting standards

The reporting of this systematic review complies with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards (Moher et al., 2009), with slight adaptations, since these were originally intended for the evaluation of health care interventions. The PRISMA checklist for this study can be found in [Table A.1 \(Supplementary data 1, Appendix A\)](#). The protocol for this study was registered with PROSPERO (<http://www.crd.york.ac.uk/PROSPERO/>) under registration number CRD42019120738, before the formal screening of search results ([Supplementary data 2, Appendix A](#)).

3. Research question

A summary of the Population, Exposure, Comparator, and Outcomes (PECO) question (Morgan et al., 2018) is presented below:

Among human population, what is the incremental effect of 10 µg/m³ increase in the short-term exposure to ambient SO₂ on all-cause and respiratory mortality?

3.1. Eligibility criteria

In [Table 1](#) we list the eligibility criteria for the articles, in relation to the population, exposure, comparator, outcome and study type elements.

When total or partial overlapping data was observed in two or more articles, the article for inclusion was selected according to the following criteria, and in the subsequent order: 1) wider geographical distribution; 2) longer duration of the study period; and 3) more recent publishing date. When analysing overlaps, multicity studies were always preferred over single-city studies. Regarding lag times, when multiple lag-estimates were reported in papers, the framework proposed by Atkinson et al. (Atkinson et al., 2014) was followed: if only one lag estimate was reported, it was included in the systematic review. If multiple lag-estimates were reported, the selection algorithm was as follows: 1) the most frequently used lag in all selected studies (0 and 1 days in this systematic review); and 2) single lags, but not cumulative/distributed lags.

3.2. Studies search and selection

The following bibliographic databases and citation indexes were searched comprehensively: Medical Literature Analysis and Retrieval System Online (MEDLINE) via PubMed, and Scopus via Elsevier. Moreover, regional databases in English and other languages as Literatura Latinoamericana y del Caribe en Ciencias de la Salud (LILACS), Western Pacific Region Index Medicus (WPRIM), Index Medicus for South-East Asia Region (IMSEAR), Index Medicus for the Eastern Mediterranean Region (IMEMR), and African Index Medicus (AIM) were

Table 1
Eligibility criteria.

Population	Inclusion: General human population, including subgroups at risk, i.e. children, pregnant women, the elderly, and patients with particular conditions, of all ages, developed and developing areas, both urban and rural settings. We have not imposed geographical restrictions. Exclusion: Studies analysing populations with partial or complete geographical and temporal overlap were excluded, in order to avoid double-counting of participants. The procedure for selecting the articles for inclusion in case of overlapping data is described in the text.
Exposure	Inclusion: We considered only short-term exposures, defined in the order of one hour to 7 days (Shah et al., 2013), to outdoor ambient air SO ₂ , from any source, expressed in a concentration unit (e.g. µg/m ³ , ppb). Exposure to daily 24-hour averages and to daily 1-hour maximum were considered separately in our analyses. Articles reporting exposures measured through ground-based air monitoring stations or estimated using models were included. Exclusion: Long-term exposures, exposure in occupational settings, or as a result of indoor exposure exclusively.
Comparator	Inclusion: The comparison was with the same population exposed to a lower (lowest) level of air pollution, considering a given difference in a standardized concentration (10 µg/m ³).
Outcome	Inclusion: The outcomes were classified using the International Statistical Classification of Diseases and Related Health Problems (ICD), and encompassed all-cause natural mortality (ICD-10: A00 to R99), and respiratory mortality (ICD-10: J00 to J99) (WHO, 2015). Exclusion: Other causes of death.
Study type	Inclusion: We included human epidemiological studies, i.e. ecological time-series (ETS), case-crossover (CCO), cohort, and panel studies. Exclusion: Qualitative studies, reviews, methodological papers, and non-human studies (i.e. in vivo, in vitro) were not included.

searched to retrieve additional articles and grey literature. WHO Regional Databases are open access resources that comprise journal indexes of locally produced information, especially some low-and middle-income country health journals and other reports, and complement the internationally known bibliographic databases. Data search included studies from January 1980 up to December 2018, with an update in July 2020, in order to incorporate relevant studies that might have been published shortly before the finalization of the review. We developed a literature search strategy for each database, using free text and also controlled vocabulary terms, and considering all the eligibility criteria. An example of the search strategy applied to PubMed can be seen in [Table A.2 \(Supplementary data 1, Appendix A\)](#). In addition, a manual search in reference lists from other systematic reviews was performed to find additional relevant studies. The strategy was developed by PO, with input from the systematic review team.

PO and JR independently screened titles and abstracts, and potentially eligible studies were assessed again by the same reviewers based on the full-text, to ensure that those met all the eligibility criteria. Any disagreement on inclusion was resolved by discussion and, if no consensus was reached, a third reviewer (NQ) was consulted. The reasons for excluding articles at this stage were recorded.

3.3. Data extraction and process

Data from selected studies was extracted by one reviewer and checked by a second. The data was then transferred to electronic extraction forms developed in Microsoft Excel®. The bibliography was managed using the software Zotero® ([Ahmed and Al Dhubaib, 2011](#)). Extracted data included study details, exposures, outcomes, and data analysis. Association measures were relative risks (RRs), odds ratios (ORs), and percentage excess (increment) or change in mortality (Perc-Incr). The data on ambient concentrations of SO₂ in the areas of study included the mean or median, standard deviation (SD), range, interquartile range (IQR), and percentiles. Due to the relevance of the 5th percentile as a proxy for the lowest level of exposure in a given study, these values were registered, when available. If a given study did not report the 5th percentile, but the SD or the 10th percentile were available, we estimated the 5th percentile using a normal approximation. If only the range was available, we estimated the SD as the range divided by four ([Hozo et al., 2005](#)).

Where a study reported multiple relevant exposure-outcome combinations, each relevant effect-size was extracted. If a given study reported more than one effect size for the same exposure-outcome combination, e.g. one effect size per city, all these effect sizes were extracted, and subsequently treated as independent samples.

3.4. Risk of bias

The risk of bias (RoB) evaluation of included studies was conducted using a new domain-based RoB assessment tool, developed by a group of experts convened by the WHO. A detailed description of this tool is available in the WHO Regional Office for Europe's website ([Working Group on Risk of Bias Assessment, 2020](#)). The instrument allows rating sections according to 13 items grouped in six domains: confounding, selection bias, exposure assessment, outcome measurement, missing data, and selective reporting ([Morgan et al., 2019](#)). Each item can be judged as having low, moderate, or high RoB. For the item of potential confounders that were accounted for in the analysis, four critical potential confounders (temperature, seasonality, day-of-the-week, long-term trends) and two additional potential confounders (holidays, influenza epidemics) were considered. If all these confounders were included, the item was classified as having low RoB; if one or two of the additional confounders were not included, but the four critical confounders were incorporated, the item was classified as having moderate RoB; otherwise, a high RoB rating was assigned. In the articles following an ETS design, the models were considered controlled for seasonality,

long-term trends, and day-of-the-week, if they included a smooth function of time ([Bhaskaran et al., 2013](#); [Peng et al., 2006](#)). Conversely, in CCO designs, we considered seasonality, long-term trends, and day-of-the-week controlled by design ([Bateson and Schwartz, 1999](#)), and thus only temperature was considered as critical. In [Table A.3 \(Supplementary data 1, Appendix A\)](#), the criteria for assigning low, moderate or high RoB in all the domains are reported. The results of each domain were analysed separately, without considering a single result for the whole article/dataset. If only one item of the same domain was judged as having high RoB, the entire domain was classified as having high RoB. The same logic was applied to moderate vs. low RoB. The assessment of RoB across domains of studies was performed as one of the sensitivity analyses, as will be detailed later.

3.5. Data synthesis and analysis

3.5.1. Main analysis

We used RRs as the common association measure in pooled analyses. Meta-analyses input data were RRs for a standardized increment in SO₂ concentration (10 µg/m³), assuming a linear exposure-outcome relationship ([Shah et al., 2013](#)), and taking into account the original increment of the pollutant, as for the following equation ([Li et al., 2017](#); [Yang et al., 2014](#)):

$$RR_{(standardized)} = e^{Ln(RR_{(original)}) \times 10 / \text{Increment}_{(original)}}$$

The effects expressed as interquartile (or quintile, or percentile differences) were converted into effects per concentration unit increase with the previous equation. When ORs were the association measures reported in a study, they were supposed to approach the RRs, under the "rare disease assumption" ([Greenland and Thomas, 1982](#)), given the fact that a cumulative incidence lower than 10% was demonstrated or assumed in all articles. Effects expressed as Perc-Incr were also recalculated to reflect a RR for a concentration unit increase in the pollutant, assuming a linear relationship, according to the following equation ([Yang et al., 2014](#)):

$$RR = \frac{\text{Perc} - \text{Incr}}{100} + 1$$

For the summary measure (pooled RRs), a random-effects (RE) model was employed, assuming that the included studies were a random selection of all possible results. The DerSimonian-Laird estimator was used ([DerSimonian and Laird, 1986](#)), a straightforward method that allows the incorporation of heterogeneity in the analysis. When the pooled effect size was calculated from 20 or less effect sizes, the Hartung and Knapp adjustment was employed ([Hartung and Knapp, 2001](#)).

3.5.2. Heterogeneity analysis

Heterogeneity between studies was assessed using 80% prediction intervals (PI), as a means to estimate the 80% interval in which the true RR in a new air pollution study will lie ([Chiolero et al., 2012](#)). We have chosen not to measure heterogeneity using the I² parameter, because this statistic can be artificially inflated when increasing the sample size (number of included studies), or when increasing the precision of the estimates from primary studies ([Rücker et al., 2008](#)). Heterogeneity was assumed when PI included the null effect ([Int'Hout et al., 2016](#)), and concurrently this 80% PI was larger than the 95% CI around the pooled RR, by a factor of two.

3.5.3. Subgroup and sensitivity analyses

Subgroup analyses were carried out with the objective of explaining the source of heterogeneity, and for these analyses we used readily available study information (age group, sex, and continent). Differences between subgroups were assessed using the χ^2 test. Additionally, the associations in each subgroup were analysed separately to assess clinically relevant findings within these subgroups.

Sensitivity analysis was carried out in order to assess the extent to

which model assumptions could have influenced the association measures. This assessment included four analyses, i.e. the inclusion of effect sizes that considered only lags of 0, 1 and 0–1 days, an analysis comparing articles showing high RoB in some of the 6 domains, versus articles with low or moderate RoB, and an analysis comparing multicity versus single-city studies. The fourth sensitivity analysis was the evaluation of potential unmeasured confounders through the calculation of the E-value (VanderWeele and Ding, 2017). This type of sensitivity analysis considers a potential unmeasured confounder, which is associated with the exposure and with the disease. The E-value is the minimum strength of association, measured as RR, that an unmeasured confounder would need to have with both the exposure and the outcome, conditional on the measured covariates, to fully explain away the observed association. It takes into account the association of the unmeasured confounder with the exposure (RR_{EU}) and with the disease (RR_{UD}). We have chosen temperature as an example of unmeasured confounder, and focused on the association between temperature and mortality (RR_U). We selected RR_U values of temperature for all-cause and respiratory mortality from a systematic review on temperature and mortality (Song et al., 2017). In order to be conservative, the higher values of different central estimates reported in that paper were selected. Then a comparison was made between the RR_U and the RRs calculated by means of meta-analysis in this study ($RR_{observed}$), but considering a wider range of pollutants increase ($50 \mu\text{g}/\text{m}^3$). The E-value was then calculated using the following equation:

$$E\text{-value} = RR_{observed} + \sqrt{RR_{observed} \times (RR_{observed} - 1)}$$

The rule was that when the RR_U was higher than the lower confidence limit of the E-value, comparatively weaker confounder associations could explain away the observed association, i.e. the presence of unmeasured confounders is plausible.

3.5.4. Concentration-response functions

The shape of the CRFs was analysed for all-cause and respiratory mortality, to assess the suitability of linear assumptions, and the existence of thresholds. The CRF can be displayed as a graph that shows the relationship between levels of adverse health responses in exposed populations (vertical axis) and levels of ambient concentrations of a pollutant (horizontal axis), and is widely used to predict the public health impacts of proposed reductions in air pollutants (Cox, 2017).

3.6. Publication bias

The potential for publication bias was assessed through two methods, i.e. the visual examination (Sterne et al., 2011) and the numerical evaluation of the funnel plot asymmetry by means of the Egger's regression test (Egger et al., 1997).

3.7. Co-pollutant models

The interaction between pollutants was analysed by means of the assessment of co-pollutant models, considering studies that addressed the effect of SO_2 controlled by the inclusion of one or more co-pollutants in the regression model, e.g. particulate matter (PM), nitrogen dioxide (NO_2), ozone (O_3), or carbon monoxide (CO).

3.8. Software

All analyses and graphs were performed using the "meta" package (version 4.9–2) (Schwarzer et al., 2015) in the statistical software R, version 3.4.4 (<https://www.r-project.org/>). The script used for the analysis was based on a previously published systematic review, and can be obtained from that publication in this issue (Orellano et al., 2020).

3.9. Certainty of evidence across studies

The certainty of evidence (CoE) was judged using an adaptation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Schunemann et al., 2013), developed by a group of experts convened by the WHO. The approach is briefly described here, with its full version available as [Supplementary data 3 \(Appendix A\)](#). The CoE was assessed across eight domains: limitations in studies, indirectness, inconsistency, imprecision, publication bias, magnitude of the effect, the occurrence of all possible confounding factors shifting towards the null effect, and the evidence of a concentration–response gradient (equivalent to CRFs). In short, the procedure was as follows: the evidence related to an exposure–outcome combination, e.g. SO_2 exposure and all-cause mortality, was initially judged as being of moderate CoE, and then was potentially downgraded or upgraded according to these domains. The approach implies that there is always a risk of unmeasured confounding in observational studies, and therefore it starts at moderate certainty. The method involves the possible downgrade of the evidence using the first five domains, and a potential upgrade using the last three domains. After applying this tool, the overall certainty was rated as: high, meaning that further research is very unlikely to change the confidence in the estimate of the effect; moderate, meaning that further research is likely to have an important impact on the confidence in the estimate of the effect; low, meaning that further research is very likely to have an important impact on the confidence in the estimate of the effect; or very low, meaning that the estimate of the effect is very uncertain.

Limitations in studies: the level of evidence was downgraded if there were statistical differences detected between studies showing high versus moderate or low RoB in the sensitivity analysis. However, we also analysed the number of studies and the impact they had in the meta-analysis. For example, the presence of small studies with high RoB but limited influence on the meta-analysis was not a reason to downgrade. If the sensitivity analysis for RoB showed a considerable influence on the pooled effect-size, the conclusions were based on the high-quality studies only, and the evidence was not downgraded. This was a judgement and there were no clear pre-set cut-off points.

Indirectness: the evidence was not downgraded based on this domain, as the research question in the included studies always reflected the PECO question.

Inconsistency: the evidence was downgraded if heterogeneity was detected, i.e. the PI included unity, and was more than twice the random effects confidence interval (see Section 2.7.2).

Imprecision: We did not use the cut-off value proposed in the adapted GRADE approach ([Supplementary data 3, Appendix A](#)), because that value was computed for rate ratios in long-term studies. Therefore, we decided not to use a cut-off point for short-term studies, as the procedure of finding an appropriate value still deserves further analysis. Instead, we adopted the following criterion: when at least one multicity study found a clinically meaningful association for a given combination, the evidence was not downgraded. The idea behind this reasoning is that if the number of events is sufficient for a given multicity study to derive significant effect sizes, the same number will be adequate for meta-analysis.

Publication bias: the evidence was downgraded if publication bias was detected by visual inspection of the funnel plot, or through the Egger's test. However, if analyses considering only multicity studies were still statistically significant, publication bias was dismissed, and the evidence was not downgraded. The reasoning behind this decision is that the probability of not publishing multicity studies due to negative results is assumed to be minimal.

Large effect size: unmeasured confounders and the results of the E-value were used, i.e. when the RR_U between air temperature and mortality was higher than the lower confidence limit of the E-value (the presence of unmeasured confounders is plausible), the evidence was not upgraded. Otherwise, the evidence was upgraded.

Confounding domain: the evidence was not upgraded using this domain, as several potential confounders could shift the RR in both directions.

Concentration- response gradient domain: the evidence was upgraded when the associations were statistically significant, and thus a positive linear or nonlinear relationship could be assumed.

4. Results

4.1. Studies included

Database searches retrieved 1,026 studies, with the addition of 93 studies in the update, and 9 additional articles that we identified through reference lists of selected reviews (Atkinson et al., 2012; Stieb et al., 2002; US EPA, 2015; WHO, 2013). After removing duplicates, 843 records were screened by title and abstract, and 154 articles were selected for full-text eligibility assessment. Finally, we included 67 articles for quantitative analysis, showing at least one specific combination of exposure (SO₂, 24-hour average or 1-hour maximum) and outcome (Fig. 1). These 67 articles represented 132 effect sizes. Conversely, in the full-text selection stage, 87 articles were excluded due to different reasons, i.e. a total or partial geographical or temporal data overlap (32), missing data (11), full-text articles not found (9), pollutants or outcomes other than selected (6), specific subgroups out of the

objective of this review (6), predictive models (6), no association measures (4), other than observational studies (4), analyses based on restricted data (4), no data on pollutant increase (3), study protocols (1), and a longer lag-time (1) (see Supplementary data 4, Appendix A). A table was added in the Supplementary data 4 (Appendix A) where the details on replacements due to overlap between studies can be examined. Within the search, we identified 16 articles that can be considered grey literature, e.g. book sections, conference papers, special reports from health agencies. However, all these articles were excluded from the quantitative analysis due to any of the aforementioned reasons.

Extracted information from included articles can be seen in the Supplementary data 5 (Appendix A). These articles comprised a total study period of 45 years, between 1972 and 2017, while the publication period ranged from 1986 to 2019. The mean duration of the studies was 5.6 years (SD: 4.2, range: 1–19). The majority of studies were carried out in Europe (36) and Asia (25), while the Americas were less represented (6). We have not found any studies from Africa or Oceania. Among selected articles, 15 reported results from multicenter collaborative studies, i.e. APHEA, CAPES, EMECAM, MISA, MISA-2, NMMAPS, and PAPA. The most frequently used study designs were ETS studies (53), followed by CCO (12) and cohort (2) studies. However, the studies evaluating a cohort of people were analysed as time-series studies. There were no studies self-defined as panel studies. Regarding the individual characteristics of the cases considered in the exposure-outcome

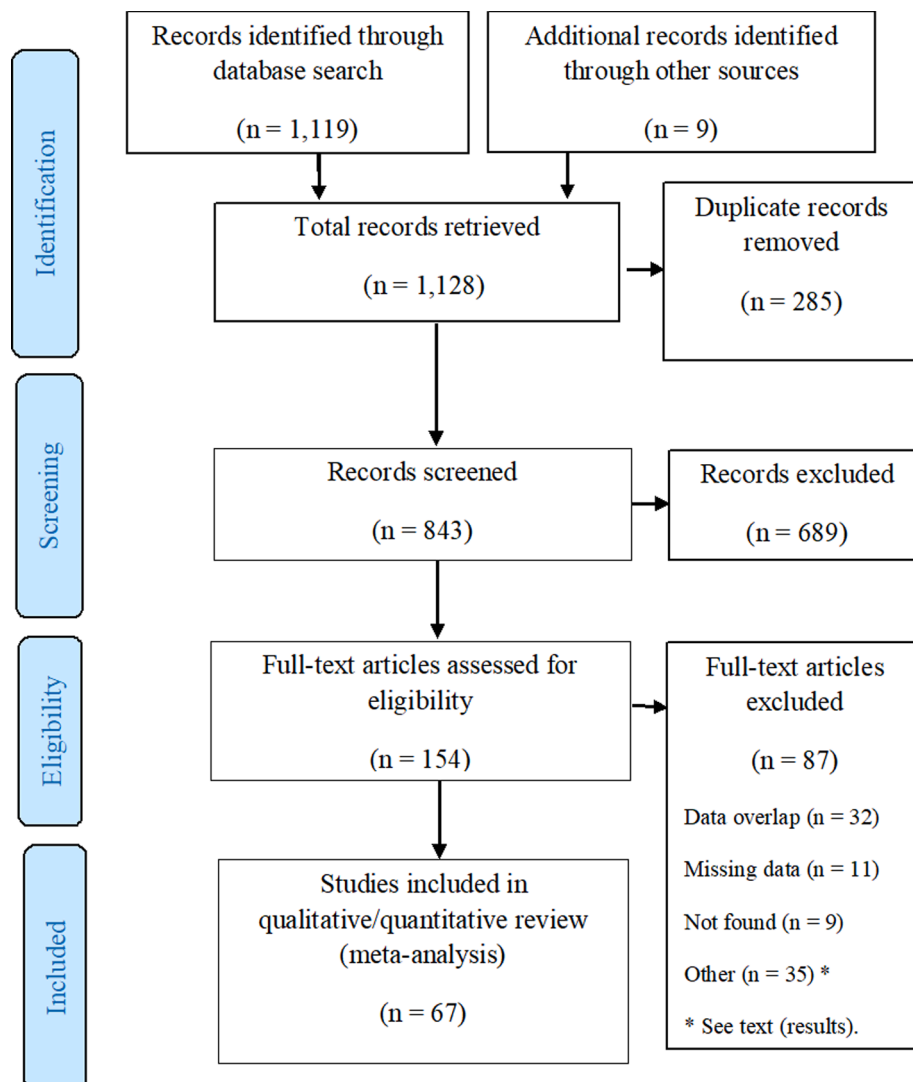


Fig. 1. Flowchart of assessment of eligible studies.

combinations, 88 effect sizes were retrieved for all-ages, 39 for elderly people, and five for children, among which four have dealt with neonatal and postneonatal deaths. Only 22 effect sizes were sex-specific, while 110 effect sizes considered both sexes in different age groups.

The mean/median concentration of SO₂ ranged from 1.5 to 79.4 µg/m³ (24-hour average), and from 2.8 to 113.2 µg/m³ (1-hour maximum). Fifth (5th) percentiles reported in the articles or estimated in this study ranged from 0.0 to 21.0 µg/m³ (24-hour average), and from 0.0 to 24.9 µg/m³ (1-hour max.). All these data regarding included studies can be seen in the [Supplementary data 5 \(Appendix A\)](#).

4.2. Risk of bias

In three out of 6 domains, i.e. selection bias, exposure assessment, and selective responding, the RoB was found to be low or moderate only. In two other domains, a small proportion of articles were judged as having high RoB. For example, only three articles showed high RoB in the confounding domain, and only one article in the outcome domain. A different situation was observed for the missing data domain, in which more than half of the articles was judged as having high RoB. The classification of high RoB in the confounding domain was related to a lack of adjustment for the confounders identified as critical. In these articles, the models were controlled for meteorological variables, but not for seasonality, long-term trends, or day-of-the-week. In the only article with high RoB in the outcome domain, the authors apparently did not use the ICD for the identification of respiratory mortality cases. The reasons for the high RoB in the missing data domain were related to the lack of information on the number of missing values in the exposure, or to the absence of information regarding imputation methods. The same judgment was applied when the number of missing data was higher than 5%. A summary of the results of the RoB analysis can be seen in [Figure A.1 \(Supplementary data 1, Appendix A\)](#). The description of the RoB analysis per item and domain in individual studies, together with the rationale to justify each judgment, are presented as [Supplementary data 6 \(Appendix A\)](#).

4.3. Meta-analysis

4.3.1. Main analysis

The detailed results regarding pooled effect sizes (RRs) for a 10 µg/m³ increase in SO₂, p-values, heterogeneity, and funnel plot asymmetry can be seen in [Table 2](#). A 10 µg/m³ increase in the average 24-hour concentrations of SO₂ was associated with a higher risk of all-cause mortality in the short-term (RR: 1.0059; 95% CI: 1.0046–1.0071). We also found positive associations between SO₂ (24-hour average) and respiratory mortality (RR: 1.0067; 95% CI: 1.0025–1.0109), and SO₂ (1-hour max.) and respiratory mortality (RR: 1.0052; 95% CI: 1.0013–1.0091). Similarly, the association was positive but non-significant for SO₂ (1-hour max.) and all-cause mortality (RR: 1.0016; 95% CI: 0.9930–1.0102). The forest plots for these analyses are shown in [Figs. 2 to 5](#).

4.3.2. Heterogeneity analysis

In the associations between SO₂ (24-hour average) and all-cause mortality, and between SO₂ (1-hour average) and respiratory

mortality, the 80% PIs excluded the null effect, i.e. the heterogeneity between studies was not substantial. Conversely, in the associations between SO₂ (24-hour average) and respiratory mortality, and between SO₂ (1-hour average) and all-cause mortality, the 80% PIs included the null effect, and was more than twice the random effects meta-analysis confidence interval. This suggests heterogeneity to some extent, sufficient enough to find some populations in which the effects of these pollutants on the outcome could be null.

4.3.3. Subgroup and sensitivity analyses

The only difference between subgroups was found for SO₂ (24-hour) and all-cause mortality, when analysing the age of cases. In this analysis, the associations were still significant for all ages and children, but turned to be non-significant for the elderly. In this sense, this appears to be the only evident source of heterogeneity, even though this specific association did not show signs of substantial heterogeneity. Interestingly, in the association between SO₂ (24-hour) and respiratory mortality, there were no statistical differences between subgroups, but the association was significant in Asia, and non-significant in Europe. Nevertheless, with a few exceptions, the number of studies/effect sizes included in these subgroup analyses were low, which prevents an adequate statistical power for the tests. These association values are shown in the [Table A.4 \(Supplementary data 1, Appendix A\)](#).

The sensitivity analysis by lag showed positive associations for all-cause and respiratory mortality ([Table A.5 in the Supplementary data 1, Appendix A](#)). When discriminated by study design, the associations were significant and comparable with those found in the main analysis, for both ETS and CCO designs ([Table A.6 in the Supplementary data 1, Appendix A](#)). When considering only multicity studies, the associations were also positive and significant ([Table A.7 in the Supplementary data 1, Appendix A](#)). In the analysis comparing studies with low or moderate against studies with high RoB, the subgroups did not show statistical differences among them ([Table A.8 in the Supplementary data 1, Appendix A](#)). Sensitivity analyses were applied in general for 24-hour average exposures, because the number of articles considering 1-hour maximum exposures was, in most cases, not sufficient enough to ensure a minimum number of effect sizes.

E-values with 95% CIs and RR_{US} can be seen in [Table A.9 \(Supplementary data 1, Appendix A\)](#). For the association between SO₂ (24-hour average) and all-cause mortality, the RR_{US} was below the lower limit of the E-value, meaning that potential unmeasured confounders are not supposed to have a major influence on the association. The other combinations showed RR_{US} above the lower limit of the E-value, and thus the presence of unmeasured confounders that might explain away the exposure-outcome associations cannot be ruled out.

4.3.4. Concentration-response functions

The linearity assumptions were investigated in five of the included articles, the five articles evaluating all-cause mortality, and two of them also evaluating respiratory mortality, for 24-hour average exposures. In relation to all-cause mortality, the behaviour of the curve was almost linear, with no evidence of apparent thresholds. However, the curve seemed to level off at higher SO₂ concentrations, i.e. above approximately 40–50 µg/m³ (Lu et al., 2015; Moolgavkar et al., 2013; Wong et al., 2010), or above 20 µg/m³ in one location (Yorifuji et al., 2019). In

Table 2
Exposures, outcomes and pooled effect sizes.

Pollutant	Outcome	Number of effect sizes	RR (95% CI)	p-value	PI	Egger's test (p-value)
SO ₂ (24-hour average)	All-cause mortality	36	1.0059 (1.0046–1.0071)	<0.0001	1.0025–1.0092	0.016
SO ₂ (24-hour average)	Respiratory mortality	23	1.0067 (1.0025–1.0109)	0.0018	0.9976–1.0159	0.082
SO ₂ (1-hour max.)	All-cause mortality	4	1.0016 (0.9930–1.0102)	0.6045	0.9952–1.0080	N/A
SO ₂ (1-hour max.)	Respiratory mortality	3	1.0052 (1.0013–1.0091)	0.0287	1.0013–1.0091	N/A

RR, pooled relative risks; 95% CI, 95% confidence interval; p-value, significance of the association or statistical tests; PI, 80% prediction interval; N/A, not applicable (<10 studies); statistically significant results in bold.

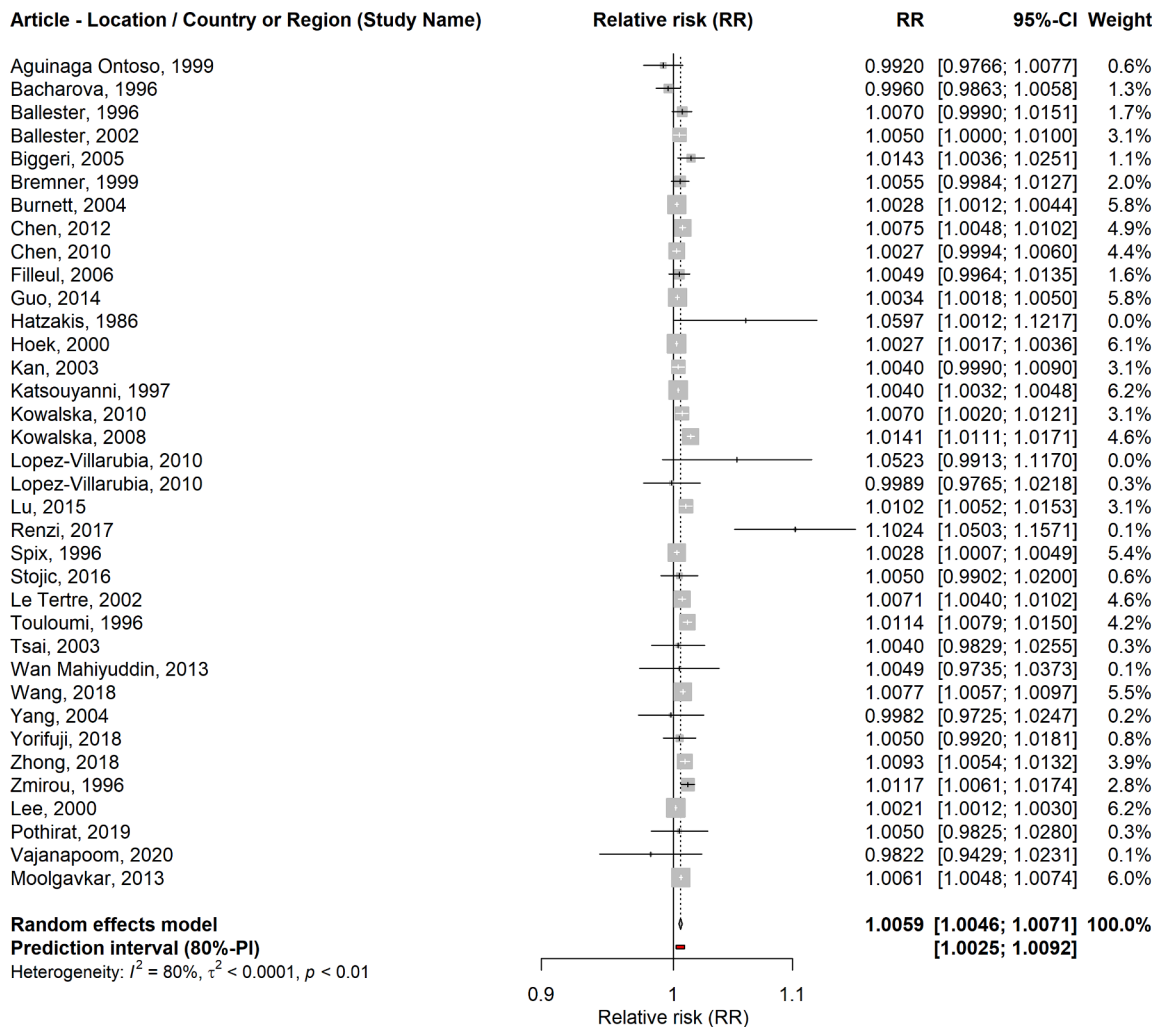


Fig. 2. Forest plot: SO₂ (24-h average) and all-cause mortality.

another article, this levelling off was not apparent (Wang et al., 2018). As for respiratory mortality, the nature of the curve was similar in one multicity study from China (Wang et al., 2018), with a linear behaviour, and no threshold. However, in contrast to the above-mentioned studies, in one study conducted in Hong Kong (Wong et al., 2010) the nature of the CRF clearly demonstrated a non-linearity behaviour for all-cause and respiratory mortality, with a potential threshold at low concentrations (around 10 µg/m³), and the curve levelling off at the same concentrations as mentioned before for all-cause mortality.

4.4. Publication bias

The visual inspection of funnel plots and Egger's tests were only performed in two combinations, particularly daily (24-hour average) SO₂ exposures, as for 1-hour exposures the number of articles was not sufficient for an adequate analysis of publication bias. Funnel plot asymmetry was observed for SO₂ and all-cause mortality, both in the visual inspection and in the results of the Egger's test. For respiratory mortality, the asymmetry of the funnel plot was not evident, as can be observed in Figure A.2 (Supplementary data 1, Appendix A).

4.5. Co-pollutant models

Co-pollutant models were evaluated in only 13 articles. Two-pollutant models showed significant associations for SO₂ (24-hour average) and all-cause and respiratory mortality, when adjusting for PM.

The same fact did not occur when the adjustments were made considering other gases, as NO₂ or O₃ (Table A.10 in the Supplementary data 1, Appendix A). The pooled effect sizes were in general lower in two-pollutant models, as compared to single-pollutant models, when considering studies comparing both models (Table A.10 in the Supplementary data 1, Appendix A). The exception was all-cause mortality when adjusted by O₃, with similar association values in single and co-pollutant models. Two-pollutant models showed more precise estimates when adjusting by PM, but wider confidence intervals when adjusting by O₃ or NO₂. Three-pollutant models were not evaluated, because we have not found more than two articles analysing associations adjusted for the same three pollutants. On the other hand, in a high proportion of articles the correlations between pollutants were moderate to high (correlation coefficient > 0.4) (Dai and Zhou, 2017), as can be seen in the Supplementary data 5 (Appendix A).

4.6. Certainty of evidence

The only reason for downgrading the CoE was the high heterogeneity, particularly in the association between SO₂ (24-hour average) and respiratory mortality, and in the association between SO₂ (1-hour max.) and all-cause mortality. On the other hand, the evidence was upgraded due to large effect size in one combination, and to the assumption of concentration-response gradients in three combinations. As for the final judgment regarding the CoE, it was high in 2 combinations, moderate in one combination, and low in another. In particular, the evidence was

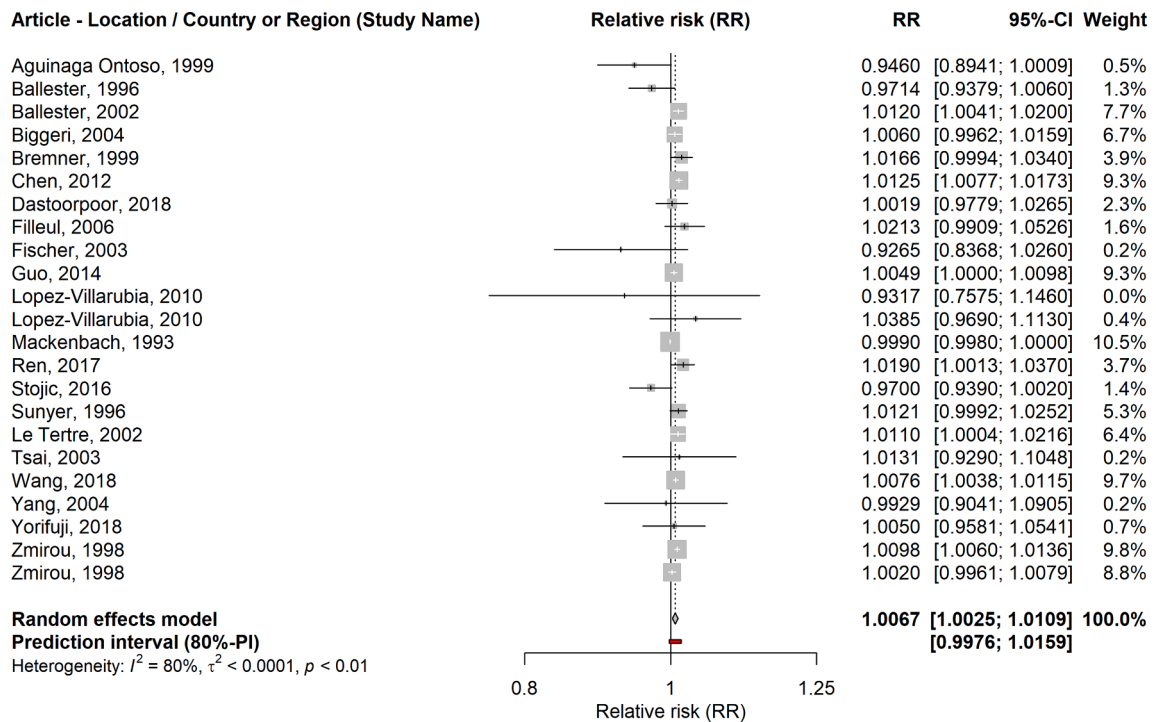


Fig. 3. Forest plot: SO₂ (24-h average) and respiratory mortality.

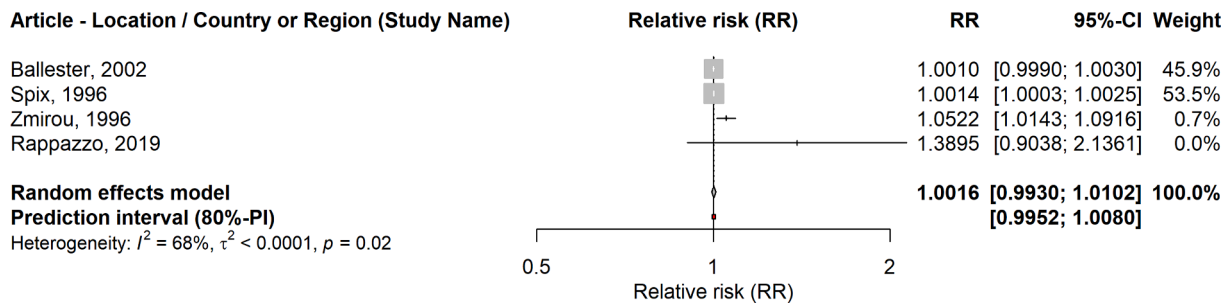


Fig. 4. Forest plot: SO₂ (1-h max.) and all-cause mortality.

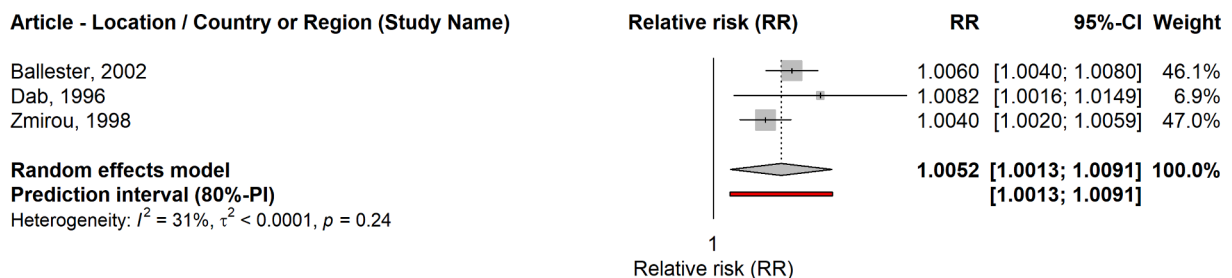


Fig. 5. Forest plot: SO₂ (1-h max.) and respiratory mortality.

low in the association between SO₂ (1-hour max.) and all-cause mortality, being also the only combination in which the association was not significant. The descriptions associated with this analysis, together with the explanations of the rationale behind the judgements made, can be found in Table 3.

5. Discussion

5.1. Summary of evidence

We found evidence of a positive association between SO₂ exposure and all-cause and respiratory mortality. The only exception was between SO₂ and all-cause mortality, but when measured as 1-hour maximum concentration, which showed a positive but non-significant relationship. The systematic reviews and meta-analyses evaluating these associations are scarce, as compared to other air pollutants. Three reviews evaluated

Table 3
Certainty of evidence profile.

Exposure - Outcome	Limitations in studies	Indirectness	Inconsistency	Imprecision	Publication bias	Large effect size	Confounding	Concentration - response gradient	Certainty of evidence
SO ₂ (24-hour average) – all-cause mortality	(0) No differences between studies with low/moderate versus high RoB	(0) The research question in the studies reflects the PECO question.	(0) 80% prediction interval did not include unity.	(0) At least one multicentric study showed a clinically meaningful association.	(0) Publication bias possibly detected, but a positive significant association was estimated when using multicentric studies only.	(+1) Unmeasured confounding would not suffice to explain away the effect estimate.	(0) Several potential confounders that would shift the RR in both directions.	(+1) Significant association detected in the main analysis.	High ⊗ ⊗ ⊗ ⊗
SO ₂ (24-hour average) – respiratory mortality	(0) No differences between studies with low/moderate versus high RoB	(0) The research question in the studies reflects the PECO question.	(-1) 80% prediction interval included unity, and was twice the size of the 95% CI.	(0) At least one multicentric study showed a clinically meaningful association.	(0) Publication bias was not detected.	(0) According to the analysis of the E-value, the presence of unmeasured confounders is plausible.	(0) Several potential confounders that would shift the RR in both directions.	(+1) Significant association detected in the main analysis.	Moderate ⊗ ⊗ ⊗ ○

(continued on next page)

Table 3 (continued)

Exposure - Outcome	Limitations in studies	Indirectness	Inconsistency	Imprecision	Publication bias	Large effect size	Confounding	Concentration - response gradient	Certainty of evidence
SO ₂ (1-hour max.) – all-cause mortality	(0) Not enough studies to detect differences in the RoB.	(0) The research question in the studies reflects the PECO question.	(-1) 80% prediction interval included unity, and was twice the size of the 95% CI.	(0) At least one multicentric study showed a clinically meaningful association.	(0) Not enough studies to analyse publication bias.	(0) Unmeasured confounding was not analysed, because the association was not significant.	(0) Several potential confounders that would shift the RR in both directions.	(0) The association was not significant.	Low ⊗⊗⊗⊗○
SO ₂ (1-hour max.) – respiratory mortality	(0) Not enough studies to detect differences in the RoB.	(0) The research question in the studies reflects the PECO question.	(0) 80% prediction interval did not include unity.	(0) At least one multicentric study showed a clinically meaningful association.	(0) Not enough studies to analyse publication bias.	(0) According to the analysis of the E-value, the presence of unmeasured confounders is plausible.	(0) Several potential confounders that would shift the RR in both directions.	(+1) Significant association detected in the main analysis.	High ⊗⊗⊗⊗⊗

Certainty of evidence, starting from moderate certainty (⊗⊗⊗⊗○); (0), between brackets is the downgrading of levels in that domain; RoB, risk of bias in individual studies; RR, relative risk; CI, 95% confidence interval; PECO, population, exposure, comparator, and outcomes.

the influence of SO₂ on mortality, but considering other specific mortality causes (Shah et al., 2015; 2013; Yang et al., 2014). In these articles, the exposure to higher levels of ambient SO₂ was associated with a higher risk of cardio- or cerebrovascular mortality. Two systematic reviews analysed the effect of SO₂ on all-cause and respiratory mortality.

The review by Atkinson et al. (Atkinson et al., 2012) found higher pooled RRs, and wider confidence intervals. The reasons for these differences might be attributed to the smaller number of included articles, the progressive reduced emissions of this pollutant (Lawal et al., 2018), and the fact that the aforementioned systematic review focused on

studies from Asia. On the contrary, another systematic review and *meta*-analysis found lower associations for all-cause mortality (Stieb et al., 2002). The pooled RRs reported in that article, after applying a standardization like the one we used in the present study, leads to a central value of 1.0011 for single-city studies, and 1.0026 for multicity studies. The reasons for these lower values deserve further analysis; a possibility is the difference in specific aspects of the methodology. In the paper by Stieb et al. (Stieb et al., 2002), more studies were included in the estimation of the pooled effect size, without performing a procedure to exclude those studies which showed temporal or spatial overlaps. As a consequence, in the aforementioned review more primary studies were included in the calculation of the pooled effect size, and thus the RR values could have been diluted. In our study, we excluded studies with complete or even with partial overlapping data, in order to avoid double-counting of information. Furthermore, we prioritized the inclusion of multicity studies, as this type of evidence might be of higher quality, i.e. less subject to bias. The potential problems of double-counting data in *meta*-analyses is that it violates the assumption of independent observations (Tarp et al., 2020). On the other hand, some degree of overlap could be tolerated, if minor overlap can be assumed (Lee, 2018). Moreover, the decision about which article to include in an overlapping situation is subjective, and this might be another source of bias in the *meta*-analysis (Ku et al., 2020). All things considered, the appropriate decision regarding the inclusion of overlapping data in *meta*-analyses still warrants further discussion.

The Integrated Science Assessment for Sulfur Oxides (US EPA, 2015), a comprehensive review of the United States Environmental Protection Agency focused mainly on multicity studies, found consistent evidence of association between short-term exposure to ambient SO₂ and all-cause and respiratory morbidity, although the biological mechanism by which this pollutant might lead to these outcomes is still uncertain. In the same line, the Review of Evidence on Health Aspects of Air Pollution (REVIHAAP Project), a technical document developed by the WHO, reached a similar conclusion for all-cause and respiratory mortality (WHO, 2013), and gave a number of considerations regarding heterogeneity and publication bias.

In our study, signs of heterogeneity were found only in some combinations. The association between SO₂ (24-hour average) and all-cause mortality, the combination with the largest number of included studies, did not show signs of heterogeneity. Conversely, the *meta*-analysis by Atkinson et al. (Atkinson et al., 2012) reported heterogeneous results for all-cause mortality (although not for respiratory mortality). Stieb et al. (Stieb et al., 2002) found signs of heterogeneity as well, possibly due to the interaction with other pollutants and effect modifiers. Many are the variables potentially related to observed heterogeneity, from differences in the spatial and seasonal distribution of ambient SO₂, to factors associated with subpopulations susceptibility to air pollution. We made an attempt to explain the source of variation by subgroup analyses based on age, sex, and continent, and we found some differences in groups regarding the age of cases. Accordingly, we cannot rule out the influence of these variables on the associations. Taken together, this means that these and other unmeasured factors might be acting as modulators of these associations.

Regarding modelling decisions and potential biases, our study was robust to different variations appraised in the sensitivity analyses. The same combinations that were positive and significant in the main analysis, were further associated in a more restrictive analyses, including specific lags (zero and one days), and specific study designs. When comparing single-city and multicity studies, the results were also similar. There were no statistical differences between single- and multicity studies, and the latter group showed significant associations. This is particularly relevant because multicity studies are assumed to be more robust to potential biases. The RoB in individual studies was in general low or moderate, and quite similar across studies. The large number of studies analysing the association of air pollutants and mortality, the historical background of this research, and the high degree of

standardization, made the methodologies between studies comparable. This fact was evident for example in the control for potential confounders, with almost all studies using models that control for the same critical confounders. The exception was the missing data domain, with a high proportion of studies not reporting imputation methods for missing data, nor the number or proportion of missing data. It is worth noting that missing data might affect summary statistics, and could introduce bias and loss of power (Hadeed et al., 2020) due to systematic differences between observed and unobserved data. However, it is to be pointed out that we used a rather strict decision rule for this domain, as the articles needed to report imputation methods, or otherwise a maximum of 5% missing data were permitted, this last value requiring a further revision. The analysis of E-values highlighted the presence of potential unmeasured confounders for respiratory mortality. Unmeasured confounders are variables associated both with the exposure and the outcome, which might explain away this association (VanderWeele and Ding, 2017). Influenza and respiratory infections can be considered relevant unmeasured confounders, which have the effect of producing seasonal patterns in mortality (Peng et al., 2006). In this sense, it is expected that respiratory mortality will be more affected by this phenomenon, while all-cause mortality depends on these and many other variables.

Publication bias was analysed for SO₂ (24-hour average) and all-cause and respiratory mortality, and suspected in the former association, based on the visual inspection of the funnel plot and on the Egger's test. However, the analysis focused on multicity studies still showed statistically significant associations, and it is not likely that multicity studies could be subject to publication bias, due to the large number of institutions usually involved. Hence, the asymmetry in the funnel plot might be related to other factors, e.g. some heterogeneity between individual studies, although we did not detect substantial heterogeneity in this association. All things considered, publication bias could have inflated the size of the true effect, but it could not have affected the general conclusion, i.e. a positive effect of air pollutants on mortality.

The relationship between pollutant concentrations and RRs were found to be linear in general, with no obvious thresholds, and a levelling off after 40 to 50 µg/m³ was reported. This is consistent with the idea of the absence of a limit below which no adverse health effects occur, as recognized in the Integrated Science Assessment for Sulfur Oxides (US EPA, 2015). The exception was one article that analysed the effects of SO₂ on respiratory mortality in Hong Kong (Wong et al., 2010), in which a non-linear behaviour and an apparent threshold at unspecified low concentrations can be observed.

Co-pollutant models have been analysed in a restricted number of studies. For all-cause and respiratory mortality, the associations remain significant after the adjustment for particulate air pollution, with lower values of RRs but more precise estimates (i.e. narrower confidence intervals). In this sense, some of the heterogeneity observed for SO₂ can be explained by the simultaneous effect of PM. On the contrary, when adjusting by other gases (O₃ and NO₂), the estimates were less precise, probably due to problems related to collinearity in the predictors (Mason and Perreault, 1991). A more extensive discussion about the effect of co-pollutant exposures on the estimates can be found in our previous article analysing the effects of PM, NO₂, and O₃ on mortality in this issue (Orellano et al., 2020). More studies applying co-pollutant models should be carried out, in order to enhance the validity of these associations between pollutants and mortality.

5.2. Strengths and limitations of the review

This study has a number of strengths that enhance the validity of our conclusions. First, the searches encompassed several international and regional databases, including articles usually considered as "grey literature", with no restrictions regarding language or place of publication. The inclusion criteria were also wide, comprising different study designs and different subpopulations. Second, our procedures included specific

instruments developed for this project, which were specially elaborated for systematic reviews of air pollution studies. In general, the common available instruments for analysing RoB in individual studies and the CoE across studies were developed for randomized controlled trials or particular observational designs, and are more generic. Moreover, the original GRADE tool for the evaluation of the CoE is also more suitable for interventional studies. In this review, we employed instruments developed in the contexts of this project, that allowed us a more specific assessment of the RoB domains for individual studies, and the CoE domains for associations and outcomes, without assigning a default level of evidence to any specific design. Third, our associations were consistent through several sensitivity analyses, which reinforces the confidence in the methodological decisions, models, and conclusions. Finally, the CoE was high for all-cause mortality and respiratory mortality, when the exposures to SO₂ were measured as 24-hour averages and 1-hour maximum, respectively. Moreover, the level of evidence was moderate for the combination SO₂ (24-hour average) and respiratory mortality.

Our study was subject to some limitations as well. First, even if time-series studies evaluating short-term exposures to air pollution are not supposed to be confounded by factors that do not change on a daily basis, as cigarette smoking or socioeconomic factors, they are still susceptible to residual or unobserved confounding (Pope and Burnett, 2007) other than ambient temperature. Second, the RoB and CoE tools are instruments that have a substantial degree of subjectivity, and might not capture all the relevant aspects to analyse the possibility of bias, and the certainty of the relationship. Many subdomains of the RoB tool were not sensitive enough to detect some differences between studies. Concurrently, we had to introduce a number of assumptions for the evaluation of the CoE, e.g. the lack of a clear cut-off point criterion for the imprecision domain. A further development of these tools is warranted, in order to achieve more objective judgements regarding the potential bias and the level of the evidence. Finally, there were differences in the geographical representation, with a large number of studies from Europe and Asia, fewer studies from North America, and a virtual absence of studies from South America, Africa or Oceania. It is worth noting that in these locations, the proportion of low-and middle-income countries is higher as compared to Europe or North America. It is not clear if some variables, characteristic of these areas, could have any influence on the associations between pollutants and mortality.

The results of this systematic review contribute to the evidence on the influence of selected air pollutants on total and specific mortality, meant to inform the update of WHO air quality guidelines. First, this information includes numerical values for the associations between air pollutants and specific risks, which might be used in studies on the economic and disease burden attributable to air pollution, risk assessments, and other analyses. Second, the differences between single- and multiple-pollutant exposures can be considered in the discussions about unifactorial and multifactorial causation. Third, the results related to CRFs and potential thresholds are useful for the determination of exposure levels in international recommendations, and national or local legislations. Fourth, subgroup analyses contribute to the evaluation of the differential risk associated with different subpopulations. Finally, the assessment of the CoE is relevant to understand the level of the evidence presented in this review and in future guidelines.

6. Conclusions

In this study, it was shown that a rise in SO₂ concentrations increases the risk of all-cause and respiratory mortality in humans. These associations were proven to be stable through a number of sensitivity analyses, which enhance the validity of the conclusions presented here. As for CRFs, these curves have shown in general a linear behaviour, without thresholds. This is consistent with the idea of harmful effects of SO₂,

even at low concentration exposures, although these results were based on a small number of studies which analysed these functions. The high consistency in the direction of the associations, and the high or moderate CoE in three of the four exposure-outcome combinations, reinforce the hypothesis of a positive association between air pollution and human mortality, while it also supports the evidence of a true causal relationship between ambient exposure to SO₂ and mortality. The reduced number of studies in some continents or economies is not a limitation of this particular study, but rather a shortcoming of research in this field. To overcome this deficiency, more resources should be allocated to research in these areas, preferably in the form of multi-collaborative projects.

CRedit authorship contribution statement

Pablo Orellano: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition. **Julieta Reynoso:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Nancy Quaranta:** Conceptualization, Methodology, Formal analysis, Investigation, 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.

Acknowledgements

The systematic review was funded by the World Health Organization Regional Office for Europe, supported by the European Commission (DG Environment), Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (Germany), Federal Ministry of Health (Germany), Government of the Republic of Korea, Federal Office for the Environment (Switzerland) and United States Environmental Protection Agency, and delivered as part of the evidence base that informs the ongoing development of WHO global air quality guidelines. All rights in the work, including ownership of the original work and copyright thereof, are vested in WHO. The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions or the stated policy of the World Health Organization. The authors would like to recognize the valuable contribution of Jos Verbeek, Román Pérez Velasco, Hanna Yang, Dorota Jarosińska, Rebecca Morgan, and experts from the Guideline Development Group (GDG). They all provided a consistent support during the full extension of the project, along with guidance, evaluation, and useful suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envint.2021.106434>.

References

- Ahmed, K.K.M., Al Dhubaib, B.E., 2011. Zotero: A bibliographic assistant to researcher. *J. Pharmacol. Pharmacother.* 2, 303–305. <https://doi.org/10.4103/0976-500X.85940>.
- Atkinson, R.W., Cohen, A., Mehta, S., Anderson, H.R., 2012. Systematic review and meta-analysis of epidemiological time-series studies on outdoor air pollution and health in Asia. *Air Qual. Atmosphere Health* 5 (4), 383–391. <https://doi.org/10.1007/s11869-010-0123-2>.

- Atkinson, R.W., Kang, S., Anderson, H.R., Mills, I.C., Walton, H.A., 2014. Epidemiological time series studies of PM_{2.5} and daily mortality and hospital admissions: a systematic review and meta-analysis. *Thorax* 69 (7), 660–665. <https://doi.org/10.1136/thoraxjnl-2013-204492>.
- Bateson, T.F., Schwartz, J., 1999. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiol. Camb. Mass* 10 (5), 539–544.
- Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L., Armstrong, B., 2013. Time series regression studies in environmental epidemiology. *Int. J. Epidemiol.* 42, 1187–1195. <https://doi.org/10.1093/ije/dyt092>.
- Chiolero, A., Santschi, V., Burnand, B., Platt, R.W., Paradis, G., 2012. Meta-analyses: with confidence or prediction intervals? *Eur. J. Epidemiol.* 27 (10), 823–825. <https://doi.org/10.1007/s10654-012-9738-y>.
- Cox, L.A.C., 2017. Do causal concentration-response functions exist? A critical review of associational and causal relations between fine particulate matter and mortality. *Crit. Rev. Toxicol.* 47 (7), 609–637. <https://doi.org/10.1080/10408444.2017.1311838>.
- Cullis, C.F., Hirschler, M.M., 1980. Atmospheric sulphur: Natural and man-made sources. *Atmospheric Environ.* 1967 (14), 1263–1278. [https://doi.org/10.1016/0004-6981\(80\)90228-0](https://doi.org/10.1016/0004-6981(80)90228-0).
- Dai, Y.-H., Zhou, W.-X., 2017. Temporal and spatial correlation patterns of air pollutants in Chinese cities. *PLoS One* 12, e0182724. <https://doi.org/10.1371/journal.pone.0182724>.
- DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. *Control. Clin. Trials* 7 (3), 177–188. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2).
- Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315 (7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>.
- Greenland, S., Thomas, D.C., 1982. On the need for the rare disease assumption in case-control studies. *Am. J. Epidemiol.* 116, 547–553. <https://doi.org/10.1093/oxfordjournals.aje.a113439>.
- Hadeed, S.J., O'Rourke, M.K., Burgess, J.L., Harris, R.B., Canales, R.A., 2020. Imputation methods for addressing missing data in short-term monitoring of air pollutants. *Sci. Total Environ.* 730, 139140. <https://doi.org/10.1016/j.scitotenv.2020.139140>.
- Hand, J.L., Schichtel, B.A., Malm, W.C., Pitchford, M.L., 2012. Particulate sulfate ion concentration and SO₂ emission trends in the United States from the early 1990s through 2010. *Atmospheric Chem. Phys.* 12, 10353–10365. <https://doi.org/10.5194/acp-12-10353-2012>.
- Hartung, J., Knapp, G., 2001. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat. Med.* 20 (24), 3875–3889. [https://doi.org/10.1002/\(ISSN\)1097-025810.1002/sim.v20:2410.1002/sim.1009](https://doi.org/10.1002/(ISSN)1097-025810.1002/sim.v20:2410.1002/sim.1009).
- Hozo, S.P., Djulbegovic, B., Hozo, I., 2005. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med. Res. Methodol.* 5, 13. <https://doi.org/10.1186/1471-2288-5-13>.
- Int'Hout, J., Ioannidis, J.P.A., Rovers, M.M., Goeman, J.J., 2016. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* 6 (7), e010247. <https://doi.org/10.1136/bmjopen-2015-010247>.
- Ku, P.-W., Hamer, M., Hsueh, M.-C., Chen, L.-J., 2020. Response to the letter titled “Double counting individuals in meta-analysis artificially inflates precision”. *Scand. J. Med. Sci. Sports* 30 (6), 1085–1086. <https://doi.org/10.1111/sms.v30.610.1111/sms.13692>.
- Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N.C., Baldé, A.B., Bertolotti, R., Bose-O'Reilly, S., Boufford, J.L., Breyse, P.N., Chiles, T., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter, D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K.V., McTeer, M.A., Murray, C.J.L., Ndashimana, J.D., Perera, F., Potočnik, J., Preker, A.S., Ramesh, J., Rockström, J., Salinas, C., Samson, L.D., Sandilya, K., Sly, P.D., Smith, K.R., Steiner, A., Stewart, R.B., Suk, W. A., van Schayck, O.C.P., Yadama, G.N., Yumkella, K., Zhong, M.a., 2018. The Lancet Commission on pollution and health. *Lancet Lond. Engl.* 391 (10119), 462–512. [https://doi.org/10.1016/S0140-6736\(17\)32345-0](https://doi.org/10.1016/S0140-6736(17)32345-0).
- Lawal, A.S., Guan, X., Liu, C., Henneman, L.R.F., Vasilakos, P., Bhogineni, V., Weber, R. J., Nenes, A., Russell, A.G., 2018. Linked Response of Aerosol Acidity and Ammonia to SO₂ and NO_x Emissions Reductions in the United States. *Environ. Sci. Technol.* 52 (17), 9861–9873. <https://doi.org/10.1021/acs.est.8b0071110.1021/acs.est.8b00711.s001>.
- Lee, P.N., 2018. Improving the conduct of meta-analyses of observational studies. *World J. Meta-Anal.* 6, 21–28. <https://doi.org/10.13105/wjma.v6.i3.21>.
- Li, J., Woodward, A., Hou, X.-Y., Zhu, T., Zhang, J., Brown, H., Yang, J., Qin, R., Gao, J., Gu, S., Li, J., Xu, L., Liu, X., Liu, Q., 2017. Modification of the effects of air pollutants on mortality by temperature: A systematic review and meta-analysis. *Sci. Total Environ.* 575, 1556–1570. <https://doi.org/10.1016/j.scitotenv.2016.10.070>.
- Lu, F., Zhou, L., Xu, Y., Zheng, T., Guo, Y., Wellenius, G.A., Bassig, B.A., Chen, X., Wang, H., Zheng, X., 2015. Short-term effects of air pollution on daily mortality and years of life lost in Nanjing, China. *Sci. Total Environ.* 536, 123–129. <https://doi.org/10.1016/j.scitotenv.2015.07.048>.
- Mason, C.H., Perreault, W.D., 1991. Collinearity, Power, and Interpretation of Multiple Regression Analysis. *J. Mark. Res.* 28 (3), 268. <https://doi.org/10.2307/3172863>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 339 (jul21 1), b2535. <https://doi.org/10.1136/bmj.b2535>.
- Moolgavkar, S.H., McClellan, R.O., Dewanji, A., Turim, J., Luebeck, E.G., Edwards, M., 2013. Time-series analyses of air pollution and mortality in the United States: a subsampling approach. *Environ. Health Perspect.* 121 (1), 73–78. <https://doi.org/10.1289/ehp.1104507>.
- Morgan, R.L., Thayer, K.A., Santesso, N., Holloway, A.C., Blain, R., Eftim, S.E., Goldstone, A.E., Ross, P., Ansari, M., Akl, E.A., Filippini, T., Hansell, A., Meerpohl, J. J., Mustafa, R.A., Verbeek, J., Vinceti, M., Whaley, P., Schünemann, H.J., 2019. A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE. *Environ. Int.* 122, 168–184. <https://doi.org/10.1016/j.envint.2018.11.004>.
- Morgan, R.L., Whaley, P., Thayer, K.A., Schünemann, 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. <https://doi.org/10.1016/j.envint.2018.07.015>.
- MYLONA, SOPHIA, 1996. Sulphur dioxide emissions in Europe 1880–1991 and their effect on sulphur concentrations and depositions. *Tellus B* 48 (5), 662–689. <https://doi.org/10.1034/j.1600-0889.1996.t01-2-00005.x>.
- Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., Ciapponi, A., 2020. Short-term exposure to particulate matter (PM₁₀ and PM_{2.5}), nitrogen dioxide (NO₂), and ozone (O₃) and all-cause and cause-specific mortality: Systematic review and meta-analysis. *Environ. Int.* 142, 105876. <https://doi.org/10.1016/j.envint.2020.105876> (this issue).
- Peng, R.D., Dominici, F., Louis, T.A., 2006. Model choice in time series studies of air pollution and mortality. *J. R. Stat. Soc. Ser. A Stat. Soc.* 169 (2), 179–203. <https://doi.org/10.1111/rssa.2006.169.issue-210.1111.j.1467-985X.2006.00410.x>.
- Pope, C.A., Burnett, R.T., 2007. Confounding in air pollution epidemiology: the broader context. *Epidemiol. Camb. Mass* 18 (4), 424–426.
- Rücker, G., Schwarzer, G., Carpenter, J.R., Schumacher, M., 2008. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med. Res. Methodol.* 8, 79. <https://doi.org/10.1186/1471-2288-8-79>.
- Schunemann, H., Brożek, J., Guyatt, G., Oxman, A., 2013. GRADE handbook [WWW Document]. URL <https://gdt.gradepro.org/app/handbook/handbook.html> (accessed 7.28.20).
- Schwarzer, G., Carpenter, J.R., Rücker, G., 2015. *Meta-Analysis with R*. Use R! Springer International Publishing. <https://doi.org/10.1007/978-3-319-21416-0>.
- Shah, A.S.V., Langrish, J.P., Nair, H., McAllister, D.A., Hunter, A.L., Donaldson, K., Newby, D.E., Mills, N.L., 2013. Global association of air pollution and heart failure: a systematic review and meta-analysis. *Lancet Lond. Engl.* 382 (9897), 1039–1048. [https://doi.org/10.1016/S0140-6736\(13\)60898-3](https://doi.org/10.1016/S0140-6736(13)60898-3).
- Shah, A.S.V., 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. *BMJ* 350, h1295. <https://doi.org/10.1136/bmj.h1295>.
- Song, X., Wang, S., Hu, Y., Yue, M., Zhang, T., Liu, Y., Tian, J., Shang, K., 2017. Impact of ambient temperature on morbidity and mortality: An overview of reviews. *Sci. Total Environ.* 586, 241–254. <https://doi.org/10.1016/j.scitotenv.2017.01.212>.
- Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., Terrin, N., Jones, D.R., Lau, J., Carpenter, J., Rucker, G., Harbord, R.M., Schmid, C.H., Tetzlaff, J., Deeks, J.J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D.G., Moher, D., Higgins, J.P.T., 2011. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 343 (jul22 1), d4002. <https://doi.org/10.1136/bmj.d4002>.
- Stieb, D.M., Judek, S., Burnett, R.T., 2002. Meta-analysis of time-series studies of air pollution and mortality: effects of gases and particles and the influence of cause of death, age, and season. *J. Air Waste Manag. Assoc.* 1995 (52), 470–484. <https://doi.org/10.1080/10473289.2002.10470794>.
- Tarp, J., Dalene, K.E., Hansen, B.H., Steene-Johannessen, J., Ekelund, U., 2020. Double counting individuals in meta-analysis artificially inflates precision. Comment on “Device-measured light-intensity physical activity and mortality: A meta-analysis”. *Scand. J. Med. Sci. Sports* 30 (6), 1083–1084. <https://doi.org/10.1111/sms.v30.610.1111/sms.13693>.
- US EPA, O., 2015. Integrated Science Assessment (ISA) for Sulfur Oxides - Health Criteria [WWW Document]. US EPA. URL <https://www.epa.gov/isa/integrated-science-assessment-isa-sulfur-oxides-health-criteria> (accessed 7.16.20).
- VanderWeele, T.J., Ding, P., 2017. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann. Intern. Med.* 167, 268–274. <https://doi.org/10.7326/M16-2607>.
- Wang, L., Liu, C., Meng, X., Niu, Y., Lin, Z., Liu, Y., Liu, J., Qi, J., You, J., Tse, L.A., Chen, J., Zhou, M., Chen, R., Yin, P., Kan, H., 2018. Associations between short-term exposure to ambient sulfur dioxide and increased cause-specific mortality in 272 Chinese cities. *Environ. Int.* 117, 33–39. <https://doi.org/10.1016/j.envint.2018.04.019>.
- WHO, 2015. *International statistical classification of diseases and related health problems*. World Health Organization.
- WHO, 2013. Review of evidence on health aspects of air pollution – REVIHAAP project: final technical report [WWW Document]. URL <https://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/2013/review-of-evidence-on-health-aspects-of-air-pollution-revihaap-project-final-technical-report> (accessed 7.16.20).
- WHO, 2006. WHO | Air quality guidelines - global update 2006 [WWW Document]. WHO. URL http://www.who.int/phe/health_topics/outdoorair/outdoorair_aq/en/ (accessed 7.16.20).
- Wong, C.-M., Thach, T.Q., Chau, P.Y.K., Chan, E.K.P., Chung, R.Y., Ou, C.-Q., Yang, L., Peiris, J.S.M., Thomas, G.N., Lam, T.-H., Wong, T.-W., Hedley, A.J., HEI Health Review Committee, 2010. Part 4. Interaction between air pollution and respiratory viruses: time-series study of daily mortality and hospital admissions in Hong Kong. *Res. Rep. Health Eff. Inst.* 283–362.
- Working Group on Risk of Bias Assessment, 2020. Risk of bias assessment instrument for systematic reviews informing WHO global air quality guidelines (2020) [WWW Document]. URL <https://www.euro.who.int/en/health-topics/environment-and-health>

- health/air-quality/publications/2020/risk-of-bias-assessment-instrument-for-systematic-reviews-informing-who-global-air-quality-guidelines-2020 (accessed 11.29.20).
- Yang, W.-S., Wang, X., Deng, Q., Fan, W.-Y., Wang, W.-Y., 2014. An evidence-based appraisal of global association between air pollution and risk of stroke. *Int. J. Cardiol.* 175 (2), 307–313. <https://doi.org/10.1016/j.ijcard.2014.05.044>.
- Yorifuji, T., Kashima, S., Suryadhi, M.A.H., Abudureyimu, K., 2019. Acute exposure to sulfur dioxide and mortality: Historical data from Yokkaichi, Japan. *Arch. Environ. Occup. Health* 74 (5), 271–278. <https://doi.org/10.1080/19338244.2018.1434474>.