Aus dem Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie (IBE)

Institut der Universität München

Vorstand/Direktor: Prof. Dr. Ulrich Mansmann

Systematic Reviews and Evidence Synthesis of Complex Interventions: Traditional methods and potential future directions

Dissertation

zum Erwerb des Doktorgrades der Humanbiologie

an der Medizinischen Fakultät der

Ludwig-Maximilians-Universität zu München

vorgelegt von

Jacob Burns

aus

Beebe, AR, USA

Jahr

2022

Mit Genehmigung der Medizinischen Fakultät der

Universität München

Berichterstatter:	Prof. Eva Rehfuess
Mitberichterstatter:	Prof. Dr. Jochen Gensichen
	Prof. Dr. Eva Grill
	PD Dr. Michaela Schunk
	Prof. Dr. Volker Heinemann
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	02.08.2022

Affidavit

Eidesstattliche Versicherung

Burns, Jacob

Name, Vorname

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel:

Systematic Reviews and Evidence Synthesis of Complex Interventions: Traditional methods and potential future directions

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 03. August 2022 — Jacob Burns

Ort, Datum — Unterschrift Doktorandin bzw. Doktorand

Table of Contents

Affidavit **3**

Table of Contents **4**

List of abbreviations **6**

List of Publications **7**

Candidate's contribution to publications **8**

1.1 Paper I: Interventions to reduce ambient particulate matter air pollution and their effects on health 8

1.2 Paper II: Looking beyond the forest: combining harvest plots, gap analysis and expert consultations to assess effectiveness and inform policy 8

1.3 Paper III: COVID-19 mitigation measures and nitrogen dioxide – A quasi-experimental study of air quality in Munich, Germany 8

2. Introduction **9**

2.1 Evidence generation and synthesis for informing decisions 9

2.1.1 The role of the systematic review in health research 9

2.1.2 Complexity and systematic reviews 10

2.1.3 Rationale 12

2.2 Objectives 12

2.3 Overview of PhD thesis 12

2.4 Improving the evidence for complex interventions: an overview of the PhD thesis. 13

2.5 Publication I: Conducting a systematic review of a complex intervention 13

2.5.1 Background 13

2.5.2 Methods 14

2.5.3 Results 14

2.5.4 Discussion 15

2.5.5 Implications for thesis 15

2.6 Publication II: Making systematic reviews of complex interventions more relevant for users 16

2.6.1 Background 16

2.6.2 Methods 16

2.6.3 Results 16

2.6.4 Discussion 17

2.6.5 Implications for thesis 17

2.7 Publication III: Conducting an exemplary primary evaluation of a complex intervention 18

2.7.1 Background 18

2.7.2 Methods 18

2.7.3 Results 18

2.7.4 Discussion 19

2.7.5 Implications for thesis 19

2.8 Further applications of the work in this thesis 19

2.9 Conclusions 20

3. Zusammenfassung 21

4. Abstract 23

5. Paper I 24

6. Paper II 92

7. Paper III 105

8. References 119

Acknowledgements 124

List of abbreviations

CBA: controlled before-after

cITS: controlled interrupted time series

GRADE: Grading of Recommendations, Assessment, Development and Evaluation

ITS: Interrupted time series

LEZ: low emission zone

LMIC: lower and middle-income countries

MRC: Medical Research Council

NO_2: nitrogen dioxide

PM2.5: fine particulate matter

RCT: randomized controlled trial

SC: synthetic control

UBA: uncontrolled before-after

WHO: World Health Organization

List of Publications

Paper I: Interventions to reduce ambient particulate matter air pollution and their effect on health [1]
Journal: Cochrane Database of Systematic Reviews
Impact factor (2020): 9.266

Paper II: Looking beyond the forest: combining harvest plots, gap analysis and expert consultations to assess effectiveness and inform policy [2]
Journal: Research Synthesis Methods
Impact factor (2020): 5.263

Paper III: COVID-19 mitigation measures and nitrogen dioxide – A quasi-experimental study of air quality in Munich, Germany [3]
Journal: Atmospheric Environment
Impact factor (2020): 4.012

Candidate's contribution to publications

1.1 Paper I: Interventions to reduce ambient particulate matter air pollution and their effects on health

The candidate led in the conceptualization, planning, conduct and writing of this systematic review. This began with leading on the development and publication of the protocol. Subsequently, in conducting the systematic review, the candidate assumed the primary role at each step, from searching electronic databases and the screening of titles, abstracts and full texts to data extraction, risk of bias assessment and evidence synthesis. The candidate coordinated the remaining work – primarily the duplication of screening, data extraction and risk of bias assessment – among all coauthors. The candidate drafted the manuscript, and subsequently incorporated feedback from the coauthors, peer reviewers and editors in finalizing the manuscript.

1.2 Paper II: Looking beyond the forest: combining harvest plots, gap analysis and expert consultations to assess effectiveness and inform policy

The work described in this paper builds from a systematic review of effectiveness, for which the candidate led and coordinated each step, from protocol development to searching electronic databases, the screening of titles, abstracts and full texts, data extraction, risk of bias assessment and evidence synthesis. Subsequently, for the supplemental methods described in this paper, the candidate led in the conceptualization, planning, conduct and writing. This involved developing the methods, comprising a gap analysis and expert consultations, and then leading and coordinating the implementation of these methods. Specifically, the candidate led the author team in a structured gap analysis procedure, and then used the identified gaps to inform and conduct expert consultations. In a final step, the candidate led the author team in examining and analyzing the content of the expert consultations. The candidate drafted the manuscript, and subsequently incorporated feedback from the coauthors, peer reviewers and editors in finalizing the manuscript.

1.3 Paper III: COVID-19 mitigation measures and nitrogen dioxide – A quasi-experimental study of air quality in Munich, Germany

The candidate led in the conceptualization, planning, conduct and writing of this quasi-experimental study. This began with the planning of all methods, with input from the coauthors, and the registration of the study protocol. The candidate identified and processed all data from multiple sources, and conducted several main analyses and sensitivity analyses, with input from the coauthors. The candidate drafted the manuscript, and subsequently incorporated feedback from the coauthors, peer reviewers and editors in finalizing the manuscript.

2. Introduction

2.1 Evidence generation and synthesis for informing decisions

2.1.1 The role of the systematic review in health research

As early as the 1700's, with an examination of the effect of inoculation on smallpox in England, the benefit of looking beyond an individual study to a collection of studies asking the same question was recognized [4]. Then from the 1970's forward, evidence-based medicine and the associated systematic review increasingly gained widespread acceptance as critical tools for informing decision-making in medicine and beyond. By identifying, appraising and synthesizing all studies assessing a particular question, the systematic review can avoid "cherry picking" of individual studies and provide robust estimates of intervention effects. This aggregate 'pooled' effect protects against the over-interpretation of individual studies that are often under-powered or may produce spurious findings. Additionally, the a priori-defined, systematic and transparent methods ensure that bias is minimized and that uncertainty is clearly communicated [5].

It is, however, also broadly recognized that simply synthesizing multiple studies does not ensure rigor and accuracy, and thus the systematic reviews generally include only those study designs that can reliably answer the question of interest [6]. For systematic reviews of interventions related to health, this has led to a tradition of primarily including and synthesizing randomized controlled trials (RCTs). The RCT has been a cornerstone of comparative research for the better part of a century. The study design was conceptualized and adapted by early innovators of research methodology – R.A. Fisher, in studying agricultural fertilization for increasing crop yield, and Sir Austin Bradford Hill, in studying streptomycin for treating tuberculosis [7-9]. By randomizing the allocation of units into two or more groups, the RCT aims to balance all observed and unobserved confounders across groups, resulting in groups that are, on average, except for group assignment, similar to one another [10]. This ensures that between-group differences observed after an intervention are likely due to the intervention, not some other confounding factor. Because of this ability to produce a unbiased estimate of effectiveness, the RCT remains the gold standard for evaluating interventions in the field of medicine, as well as social sciences [11], education [12] and international development [13], among others.

Following these strands, a systematic review of RCTs represents a powerful tool for assessing the effectiveness of an intervention. Each included RCT, in principle, enables the estimation of an unbiased effect of the assessed intervention, with the systematic review providing a pooled estimate of intervention effectiveness across the entire evidence base. This rigorously generated evidence can subsequently inform decisions and practice. Take for example, the question of whether antihypertensive medications are effective in preventing cardiovascular morbidity and mortality. A rich evidence base exists for this broad question, with multiple systematic reviews comprising an interrelated network of 230 RCTs addressing various aspects of the question. Such an evidence base is well suited to informing decisions and practice, as reflected in multiple recent national and international guidelines [14-16].

2.1.2 Complexity and systematic reviews

The history of the systematic review, as described above, is closely related to clinical medicine; the example of hypertensive medications illustrates how such systematic reviews of RCTs can neatly feed into decisions in clinical practice. As researchers in other fields have begun asking questions of effectiveness, fields where the objects under study may be more complex, a natural first step was to apply this same proven systematic review frame – i.e. a synthesis of all available RCTs. Questions have accumulated over time, however, as to whether this frame is equally well suited to such complex interventions [17-19].

There is ongoing discussion about what complexity is with regard to interventions and what implications it may have for developing, implementing and, importantly, for evaluating interventions. Although these discussions have not yet settled, one can safely say that the importance of considering complexity in evaluating interventions has become increasingly accepted. This discussion was spurred by the publication of the UK Medical Research Council's (MRC) first guidance on complex interventions in 2000 [20], which was updated in 2006 [21], and is currently being updated again. Initially, conceptualizations of complexity focused on complexity in the intervention itself (e.g. multiple components, required skills of those delivering or receiving an intervention). Over the last several years, however, there has been an increasing focus on interventions as "events in complex systems" [22], and on embracing a complex systems approach [23]. Here, populations, settings and contexts represent components of a system, and an intervention may influence one or multiple of these other components within that system. Sources of complexity thus include interactions between the various system components, adaptivity of the system, and complexities along the causal pathway [24]. Seen through this lens, it is not sensible or perhaps even possible to apply a binary classification to interventions as being either simple or complex; most interventions and the systems in which they are implemented will be characterized by some degree of complexity. At some point along this complexity spectrum, however, challenges begin to emerge in applying the traditional systematic review frame as described above, informed primarily by RCTs.

Drawing on an example from the fields of environmental and public health, consider a low emission zone (LEZ). An LEZ defines a boundary around an urban area that only certain relatively low-polluting vehicles may enter [25]. An LEZ is implemented across a broad geographical area and targets the entire population of individuals living within this area – individuals with varying risk profiles and socioeconomic backgrounds. The causal chain from the implementation of an LEZ to changes in environmental outcomes (e.g. fine particulate matter (PM2.5) concentrations) and, subsequently, population-level health outcomes (e.g. cardiovascular/cardio-metabolic and respiratory mortality and hospitalizations) is long and influenced by many other individual and population-level aspects [26]. An illustrative schematic of this complex causal chain is provided in *Figure 1*. It is unsurprising that all published evaluations of LEZs are not RCTs – the logistical implications for conducting an experiment at the population-level are considerable, and most policies of this nature are implemented within a social and political context not well-sensitized or well-equipped to prospectively evaluate their effectiveness [27]. Thus for a systematic review setting out to assess the effectiveness of the LEZ, a search for RCTs will likely be in vain. Observational studies can assess the association between an LEZ and changes in air quality and/or health, however such studies will be at high risk of bias due to potential confounding. As highlighted in *Figure 1*, those confounding aspects could be: how the LEZ Is implemented and enforced, other measures implemented in the area to address air pollution, weather and climactic changes, advancements in healthcare and changes to the demographics of the population, among others. A systematic review of observational studies, therefore, is at risk of aggregating and synthesizing findings that are biased and thus potentially inappropriate for informing decisions. The LEZ represents just one example of a complex intervention, yet there are countless examples of others across health-related fields, such as public health.

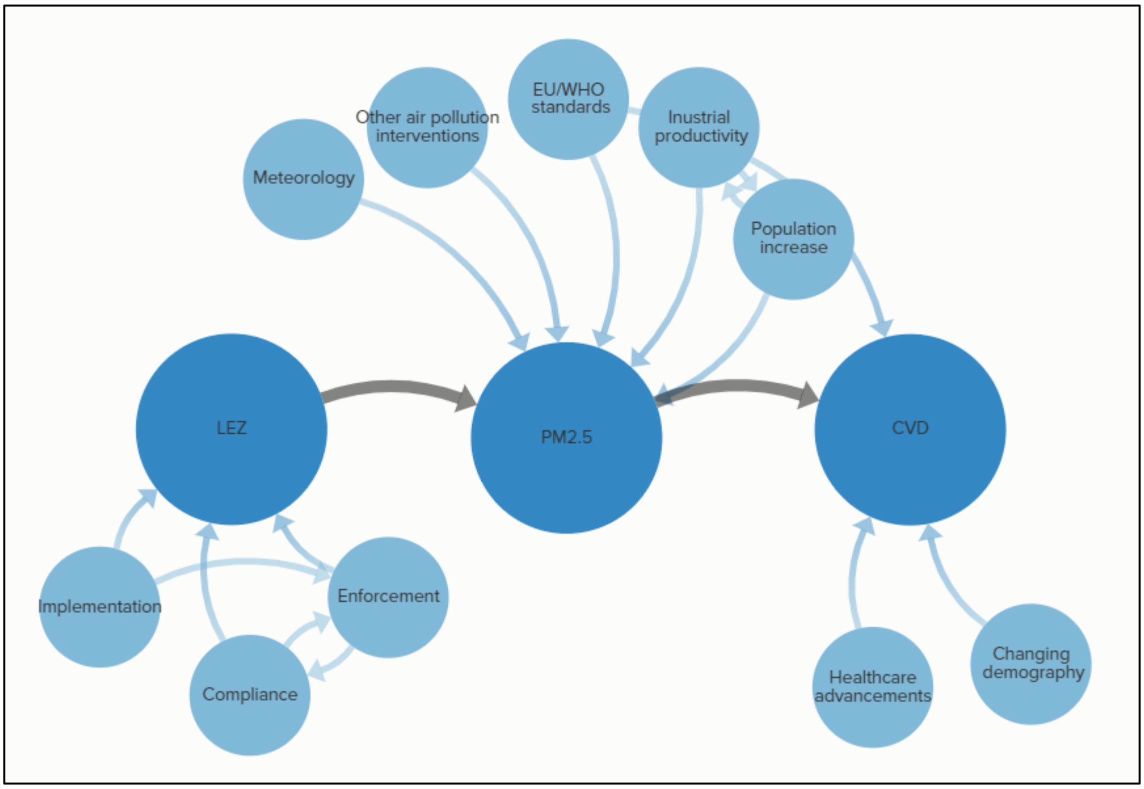


Figure 1: Dark blue circles lie upon the causal pathway from a low emission zone (LEZ) to air quality (e.g. PM2.5) and subsequently health (e.g. cardiovascular disease, CVD). Light blue circles represent other factors that may influence the causal pathway.

There have been multiple descriptions, discussions and examinations of the implications that complexity may have for systematic reviews of complex interventions. These discuss, for example, the potential appropriateness of including study designs other than RCTs and the implications that will have for the applied methods, including, for example, challenges associated with defining the question and conducting the review, as well as how complexity impacts the overall systematic review process [28-34]. Others have outlined particular challenges related specifically to synthesizing and reporting the results [35, 36], highlighting the challenge of producing results appropriate for informing decisions. As a result of these and many other works, the recently updated Cochrane Handbook – sixth edition, generally considered the gold standard guidance for systematic reviews, devotes much more space to discussions about the implications of complexity and the inclusion of non-randomized study designs [37, 38].

Similar aspects have been explored in relation to primary research. Various works have approached complexity and the evaluation of interventions from different perspectives; for example these have focused on public health [39, 40], health systems research [27], 'real-world' effectiveness of interventions [41], and complex interventions in general [42]. Despite the slightly varying focus of these works, each posits that in evaluating the effectiveness of complex interventions, researchers may have to consider study designs other than the traditional RCT; these other designs could include, for example, pragmatic RCTs, quasi-experimental studies and natural experiment studies [27, 39, 40, 42]. These works describe that such study designs represent a tradeoff compared to the traditional RCT: the internal validity may be lower, but they can be applied more flexibly and often offer a higher external validity [43].

2.1.3 Rationale

As outlined above, at the outset of this doctoral thesis, some research was published and much was underway related to evaluating complex interventions, both at the level of the systematic review and primary research. At the same time, however, there were numerous open questions and a clear lack of good practice examples. It was generally accepted, for example, that in many cases systematic reviews of complex interventions should include study designs other than RCTs, and broadly speaking that this would have implications for the methods. However, which specific designs were most appropriate and the extent to which the standard systematic review methods and tools were suited to these designs was much less clear. Additionally, from the few examples of systematic reviews of complex interventions that existed, it was evident that the evidence synthesis, for clinical questions usually comprising statistical pooling of data through a meta-analysis, was often not appropriate. However, it was unclear which alternative methods for evidence synthesis existed and to what extent these methods provided relevant and informative findings for decision-makers. The starting point for this thesis was that the traditional methods for systematic reviews, as well as the underlying primary evidence, could be improved for assessing the effectiveness of complex interventions, and this thesis thus set out to explore this question.

2.2 Objectives

In aiming to address several of the relevant gaps and move towards improved methods for evaluating complex interventions, this thesis had the following objectives:

i. Conduct a systematic review of a complex intervention, and explore the extent to which traditional methods are suitable;
ii. Demonstrate a method for making the results of systematic reviews of complex interventions more relevant to users, including decision-makers and practitioners;
iii. Conduct an exemplary good-practice quasi-experimental study, thus illustrating and promoting the improved primary evaluation of complex interventions.

2.3 Overview of PhD thesis

This PhD thesis comprises three articles that were published in peer-reviewed journals.

I. **Burns J**, Boogard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess EA. Interventions to reduce ambient particulate matter air pollution and their effect on health. 2019. *Cochrane Database of Systematic Reviews*, Issue 5. [1]

II. **Burns J**, Polus S, Brereton L, Chilcott J, Ward SE, Pfadenhauer LM, Rehfuess EA. Looking beyond the forest: Using harvest plots, gap analysis, and expert consultations to assess effectiveness, engage stakeholders, and inform policy. 2020. *Research Synthesis Methods*, 9(1): 132-140. [2]

III. **Burns J**, Hoffmann S, Kurz C, Laxy M, Polus S, Rehfuess EA. COVID-19 mitigation measures and nitrogen dioxide – A quasi-experimental study of air quality in Munich, Germany. *Atmospheric Environment*, 246: 118089. [3]

2.4 Improving the evidence for complex interventions: an overview of the PhD thesis

Although there is some overlap, broadly speaking publications I, II and III aim to address objectives i, ii and iii, respectively.

As described above, there has been much discussion related to the extent that traditional systematic review methods extend to public health and other fields where interventions tend to be complex [17-19]. Publication I was important in providing a concrete application of current best-practice systematic methods in evaluating the effectiveness of a complex public health intervention – specifically, interventions to reduce ambient air pollution. This application allowed for the assessment of how well these methods perform, and of where alternative or additional methods could help improve the planning, conduct and reporting of such systematic reviews. Publication I also allowed for the critical appraisal of the primary evidence base for a complex public health intervention, and subsequently the identification and description of shortcomings, as well as the exploration of opportunities for improvement.

Publication II also involved the application of best-practice methods for systematic reviews of complex interventions, in this case related to the effectiveness of palliative care interventions with an additional component focusing on the informal caregiver. Building from the knowledge and experiences of Publication I, we subsequently demonstrated the use of additional methods, with the aim of making the results from such reviews more relevant to users, i.e. decision-makers and practitioners. These supplemental methods, including a gap analysis and expert consultations, allowed us to use the expertise and experiences of researchers and practitioners working in palliative care to further explore and contextualize the results of the systematic review, and thus potentially provide more useful information for users relying on the systematic review to inform decisions.

Publication III dealt with the limitations of the primary evidence base for a complex public health intervention. Specifically, in line with the scope of the systematic review of interventions to reduce ambient air pollution, we assessed how the COVID-19 mitigation measures implemented abruptly in early spring 2020 impacted air pollution in Munich. We applied several of the recommendations for improving research that we made in Publication I, and thus illustrated how, in contrast to most of the studies we had identified in the Cochrane review, routine data and quasi-experimental methods can be utilized to conduct rigorous evaluations of complex interventions.

The three publications and their contributions to addressing the thesis objectives are described in detail below.

2.5 Publication I: Conducting a systematic review of a complex intervention

2.5.1 Background

Evidence on the association between various ambient air pollutants and a range of indicators for ill health, particularly respiratory and cardiovascular mortality and morbidity, has amassed over the past decades. Recently, a series of systematic reviews, conducted as part of the World Health Organization's (WHO) update of the Air Quality Guidelines, provided an up-to-date look at these associations. They showed consistently, from evidence from across the world, that higher concentrations of air pollutants, including ozone, carbon monoxide, nitrogen dioxide, and sulfur dioxide, are associated with poorer health outcomes [44-49]. The Global Burden of Disease study has highlighted that outdoor air pollution is among the top ten risk factors for health and the number one environmental risk factor globally. The study estimated that exposure to outdoor air pollution led to over 4 million excess deaths in 2015 [50]. This well-developed evidence base provides a strong argument for protecting public health through implementing interventions for improving ambient air quality. In 2013, however, no systematic review addressing the effectiveness of such public health interventions existed; thus we set out to fill this gap. Specifically, we set out to conduct the first systematic review to identify, appraise and synthesize the evidence on all interventions aiming to reduce ambient air pollution.

2.5.2 Methods

We employed rigorous systematic review methods in line with guidance and standards from the Cochrane Handbook – fifth edition [51]; we defined all methods a priori and published the protocol [52]. This involved searching a broad range of databases – including those with biomedical, social sciences, urban planning and environment and multidisciplinary focuses, as well as those pertaining specifically to low and middle-income countries (LMICs). We also searched additional sources, including grey literature databases and trial registries, the reference lists of included studies, and highly relevant journals. We included studies that assessed four main categories of interventions, i.e. those aiming to reduce ambient air pollution from industrial, residential, vehicular and multiple sources. We included studies assessing health outcomes, specifically all-cause mortality and cardiovascular and respiratory mortality and morbidity, as well as air quality outcomes, specifically particulate matter, ozone, carbon monoxide, nitrogen oxides, nitrogen dioxide, nitric oxide and sulfur dioxide concentrations.

We considered cluster randomized controlled trials, as well as multiple non-randomized study designs commonly used to evaluate population-level public health interventions, including controlled interrupted time series (cITS) studies, interrupted time series (ITS) studies and controlled before-after (CBA) studies. Additionally, we considered uncontrolled before-after (UBA) studies as supporting studies; however these were not used in generating conclusions.

Two authors independently conducted key steps of the review, i.e. title/abstract and full text screening, data extraction and risk of bias assessment. We synthesized the evidence narratively and graphically, and assessed the certainty of the evidence using the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system [53].

2.5.3 Results

We included 42 studies assessing 38 unique interventions in the main analyses. Included studies were heterogeneous across a range of aspects, including the study setting and context, the interventions themselves and assessed outcomes. With respect to the intervention types, five studies aimed at addressing industrial sources of pollution (e.g. factory closures), seven aimed at residential sources (e.g. coal bans), twenty-two aimed at vehicular sources (e.g. low emission zones), and four aimed at multiple sources (e.g. mix of measures targeting traffic and industrial polluters).

With regard to study design and analysis methods, included studies were also very heterogeneous. All included studies were non-randomized studies. Among studies assessing health outcomes, nine, five and three studies applied cITS, ITS and CBA study designs, respectively. Among studies assessing air quality outcomes, three, ten and seventeen applied cITS, ITS and

CBA study designs, respectively. The analysis methods employed also varied substantially among studies, even among the same study design.

We observed mixed evidence regarding effectiveness: most studies observed either a significant effect favoring the intervention, or no clear effect in either direction, while only few studies observed harmful effects of interventions. For example, for studies assessing the effect of interventions to reduce ambient air pollution from vehicular sources, across health outcomes, 5/9 studies observed a significant effect favoring the intervention and 4/9 observed no clear effect in either direction, while no studies observed a significant effect favoring the control. A similar pattern was observed across air quality outcomes, with 15/46 studies observing a significant effect favoring the intervention and 23/46 no clear effect in either direction, while 8/46 observed a significant effect favoring the control.

2.5.4 Discussion

As the first systematic review of this evidence base, this work addressed an important gap in the evidence by identifying and synthesizing up-to-date evidence on interventions aiming to reduce ambient particulate matter air pollution. Due to the substantial heterogeneity across included studies, deriving overall conclusions on the effectiveness of interventions was challenging. Overall, we identified some evidence suggesting that interventions have successfully improved health and air quality outcomes, with comparatively little indication that interventions cause harm. Included studies highlight several challenges in conducting this type of study, as well as several avenues for improving research in this field. Future studies, for example, should focus on better addressing confounding in studies; opportunities for doing so include the use of appropriate comparison populations or comparison outcomes unaffected by the intervention, and accounting for underlying time trends in outcomes. Other potential improvements could involve the prospective planning of evaluations, a priori definition and documentation of methods, and improved reporting.

2.5.5 Implications for thesis

Apart from being an important contribution to the evidence base on air pollution interventions, this systematic review was important from a methodological standpoint. We were able to explore the extent to which traditional best-practice systematic review methods are well suited for assessing the effectiveness of complex interventions, such as those in public health; and although each complex intervention is different to a certain extent, we learned or underscored valuable lessons that can be generalized.

Firstly, this systematic review highlighted that questions of complex, broad public health interventions will likely yield an extremely heterogeneous evidence base, making the synthesis and interpretation of the evidence and the generation of concise conclusions difficult. Without pooled estimates, even when using recently developed methods of graphical summary, users, i.e. decision-makers and practitioners, are likely left uncertain of how to correctly interpret and apply the results to inform a decision. This suggests that it may be helpful to supplement traditional systematic review methods with other methods, which could help make the heterogeneous results from this type of systematic review more informative and easier-to-use for users.

Secondly, we identified an evidence base that was substantially limited with regards to the study design and analysis methods, especially related to accounting for confounding and complete reporting. Additionally, a wide range of differences in study design and analysis methods precludes the possibility of statistically synthesizing the data. If systematic reviews are to meaningfully and robustly inform decisions, improvements in the underlying evidence base will be critical. Important improvements include more rigorous and standardized methods for study design and analysis, as well as prospective planning and the a priori definition of methods, where possible.

2.6 Publication II: Making systematic reviews of complex interventions more relevant for users

2.6.1 Background

A critical methodological decision in the conduct of every systematic review is whether or not the identified studies are sufficiently homogeneous to statistically pool them, i.e. to conduct a meta-analysis. For systematic reviews of complex interventions, a broad research question requiring intricate, multidisciplinary searches often yields an evidence base characterized by substantial heterogeneity; in such cases, authors often forego the conduct of a meta-analysis, deciding instead to synthesize the findings narratively. Such a synthesis, however, may result in a long, unwieldy summary text that may not be informative and accessible to users, such as decision-makers and practitioners. Indeed, often this summary text does not even undertake a synthesis across studies but simply provides a descriptive summary of each individual study. Using home-based palliative care interventions as an application case, we employed novel methods supplemental to more traditional systematic review methods, with the aim of increasing the relevance of the findings of systematic reviews for informing specific decisions.

2.6.2 Methods

Our application case involved updating a Cochrane systematic review on the effectiveness of home-based palliative care interventions with an explicit component targeting the informal caregiver [54]. We applied methods largely in line with the original review and guidance from the Cochrane Handbook, fifth edition [51], only diverging at the stage of the evidence synthesis, for which we applied a narrative synthesis and a graphical summary through the creation of harvest plots.

Recognizing that these results may be limited for users, we applied supplemental methods, including a gap analysis and expert consultations. The gap analysis was an iterative approach through which the author team examined the review findings in detail to identify ‘emerging aspects’; these could include, for example, open questions, inconsistencies or potential patterns in the evidence. These emerging aspects subsequently served as a flexible structure for consultations with palliative care practitioners and researchers. Specifically, we consulted these individuals regarding how their expertise and experiences could help to further explore the emerging aspects, by, for example, providing additional detail and context or explaining how such interventions are perceived and implemented in daily practice.

2.6.3 Results

The thirteen included studies evaluating home-based palliative care interventions with an explicit component targeting the informal caregiver exhibited extensive heterogeneity in the study setting, population, intervention, comparison, outcomes and applied study design and analysis methods. Although results varied across studies, overall there was little indication that these interventions led to better outcomes in caregivers or patients than home-based palliative care interventions without an explicit caregiver component.

Through the gap analysis we identified four aspects that may have influenced the effectiveness of the evaluated interventions, or the assessment of the effectiveness. Emerging aspects related to (1) the heterogeneity and ambiguous nature of the comparator, i.e. home-based palliative care without an explicit caregiver component; (2) the potential lack of tailoring care to patient and caregiver needs in structured interventions; (3) the questionable appropriateness of outcomes assessed (e.g. standardized health outcomes related to quality of life and psychological accounts likely not sensitive to the palliative population); and (4) the questionable appropriateness of the study designs used in included studies (e.g. non-randomized studies of low internal validity, which are additionally poorly suited to the flexibility and tailored nature of home-based palliative care interventions). Through expert consultations structured around these emerging aspects, we identified several important clinical aspects, like the need to embrace tailored, adaptive care and responsive outcomes in practice, as well as methodological aspects, like the need for improved study designs and improved reporting in primary research.

2.6.4 Discussion

Through the conduct of an application case on home-based palliative care with an explicit caregiver component, we supplemented rigorous traditional systematic review methods with further methods, including gap analysis and expert consultations, to further explore and contextualize the findings. This allowed us to highlight aspects not explicitly found in the included studies, yet that could be very helpful in helping users, including decision-makers and/or practitioners, to better interpret the systematic review findings, and more importantly, move towards a more informed decision.

2.6.5 Implications for thesis

Similar to Publication I, the application case for Publication II, home-based palliative care with an explicit informal caregiver component, highlighted an extremely diverse and challenging evidence base. Here, the traditional systematic review methods provided an overview of the effectiveness, but this information on its own would likely prove too broad and superficial for users attempting to make a context-specific and nuanced decision.

It is increasingly considered best-practice to engage with stakeholders, such as decision-makers, practitioners and researchers, during the conduct of a systematic review, especially one of a complex intervention. The supplemental methods we applied, a gap analysis and expert consultations, allow such stakeholders to be involved in the process, and to use their expertise and experiences to help explain and contextualize findings. Such information could be critical in helping users of the review better understand, not only the overall effectiveness of a complex intervention, but what aspects, including those described in the included studies and those external to the evidence base, may influence its effectiveness.

This is just one potential set of supplemental methods, and a range of other methods exist for incorporating additional quantitative and/or qualitative data with the findings of systematic reviews [30, 55]. Other methods include, for example, using Bayesian meta-analysis to combine quantitative systematic review findings with qualitative interview data [56], and using program theory or logic models to triangulate multiple streams of quantitative and/or qualitative data to explore causal pathways and assess aspects such as context and implementation [57]. Our demonstrated methods, as well as these others, provide an opportunity to expand traditional systematic review methods and thus produce more informative and relevant findings for informing decisions.

2.7 Publication III: Conducting an exemplary primary evaluation of a complex intervention

2.7.1 Background

There is a rich evidence base on the health effects air pollution, which is well summarized in the recent systematic reviews conducted as part of the update of the WHO Air Quality Guidelines [44-49]. Our systematic review of interventions to reduce outdoor air pollution, however, illustrated that the effect air pollution policies and interventions, as well as other external shocks that likely influence air quality, has been much less studied. Such evidence is key to help decision-makers determine how air quality can be improved in both the short and long-term to promote and improve public health.

The policy and behavioral response to the COVID-19 pandemic, because of the sudden and drastic change in human behavior, automobile traffic and industrial activity, provided a unique opportunity to study how short-term pollutant concentrations changed in response to an abrupt curtailment of such activities. Specifically, we assessed how the COVID-19 mitigation measures implemented in Munich, Germany, beginning on 16 March 2020, influenced concentrations of nitrogen dioxide (NO_2), a key indicator for automobile traffic, over the subsequent four weeks.

2.7.2 Methods

We applied two quasi-experimental designs, including a controlled interrupted time series (cITS) approach and a synthetic control (SC) approach. Both approaches used underlying time trends in NO_2 concentrations in 2020, the intervention year, as well as in 2014-2019, the control years, to determine how observed concentrations after the implementation of the COVID-19 mitigation measures differed from what would have been expected in the absence of the mitigation measures. The two approaches differ in how the control years are treated; the cITS approach uses the data from all years equally, while the SC approach uses a data-driven process to weight the control years so that the most similar years, with regard to the outcome and key confounders, contribute more weight to the analysis.

We assessed changes in NO_2 measurements available from five different air quality monitors in Munich: two traffic sites, one urban background site, and two background sites situated in comparatively rural areas. We hypothesized a priori that the largest changes would occur at traffic sites, with moderate changes at urban background sites and small or no changes at background sites. In addition to main analyses, we conducted additional analyses, as well as post-hoc analyses to further explore changes in NO_2 and to test our methodological assumptions. All hypotheses, main and additional analyses were defined a priori, and the study protocol was registered with the Open Science Framework (https://osf.io/dmt74/).

2.7.3 Results

Observed changes in NO_2 after the introduction of the COVID-19 mitigation measures largely supported our hypotheses. At the two traffic sites we observed reductions in concentrations of 9.34 µg/m^3 (95% confidence interval: -23.58; 4.90) and 10.02 µg/m^3 (-19.25; -0.79) using the cITS approach, and of 15.65 µg/m^3 (-27.58; -4.09) and 15.1 µg/m^3 (-24.82; -9.83) using the SC approach. At the urban background and background sites, we observed smaller changes or no change.

Through our additional analyses, we observed that the reduction in concentrations was largest in the two weeks immediately following the introduction of the measures, as compared to three and four weeks after; we also observed that there was a lag of approximately three days after March 16 before the reduction occurred. With our post-hoc analyses, we observed that at least some of the observed effect may have been explained by changes in NO_2 concentrations that were occurring before the measures were implemented, and by uncharacteristically high NO_2 concentrations in January 2020.

2.7.4 Discussion

We applied two quasi-experimental study designs to assess the impact of the COVID-19 mitigation measures in March 2020 on NO_2 concentrations in Munich, Germany. Across analyses, we observed a consistent pattern: that concentrations were reduced by about 15-25% and 24-36% at the two traffic sites. This suggests that substantially reducing automobile traffic in Munich may be an effective option for improving air quality in the city. The results were similar to studies examining the impact of the COVID-19 lockdown in other contexts, including in China [58-60], Brazil [61], India [62, 63], as well as in other European cities [60, 64, 65]. In the context of a policy environment very much active in discussing how future air quality improvements can be accomplished, this has meaningful implications: reducing automobile traffic should likely play a role in any air quality management program, however single interventions addressing this source will likely need to be combined with a coordinated range of other interventions.

2.7.5 Implications for thesis

Researchers in the methodological literature have highlighted non-randomized, quasi-experimental study designs as an attractive option for assessing the effectiveness of interventions where randomization is impossible, infeasible or inappropriate [27, 39, 41]. However, systematic reviews of complex public health interventions show that there is a lack of such rigorous evaluations in many fields, as Publication I did for interventions to reduce outdoor air pollution.

With this application, we aimed to implement several of the recommendations for research that we had made in that systematic review. Specifically, we defined all hypotheses and planned the methods a priori, and registered the study protocol; we applied quasi-experimental methods which utilized the serial nature of the underlying data, and controlled for confounding through design; and we conducted and clearly reported a range of additional and sensitivity analyses that allowed us and the reader to judge the robustness of our methods and conclusions. This study can serve as an example of how evaluations of public health evaluations can be planned, conducted and reported in the future; this would ensure that systematic reviews could synthesize more standardized and internally valid evidence, and that users have access to a reliable evidence base for informing decisions.

2.8 Further applications of the work in this thesis

The work described in this thesis was not conducted in a vacuum, and the knowledge and expertise gained through this work were applied in several other projects and activities throughout the period of the thesis.

Through Publications I and III the candidate gained an extensive knowledge of the effectiveness of interventions to reduce ambient air pollution, as well as the methods with which the effectiveness is assessed. Because of this, the candidate was invited to be involved with the update of the

WHO Air Quality Guidelines as an external advisor. As part of this role, the candidate presented the findings from Publication I and discussed with the guideline committee what role air pollution interventions could play in the Air Quality Guidelines [66]. Additionally, the candidate served as a guest editor for the journal Environment International, which involved editing a special issue comprising the series of systematic reviews of the health effects of several air pollutants; this special issue served as the evidence base for updating the Air Quality Guidelines (https://www.sciencedirect.com/journal/environment-international/special-issue/10MTC4W8FXJ).

Regarding the methodological expertise related to systematic reviews gained through Publications I and II, particularly related to the handling of diverse non-randomized study designs and methods for evidence synthesis, the candidate has been heavily involved in planning and conducting several further systematic reviews of complex interventions. Several of these reviews, focusing on international travel-related control measures to contain the COVID-19 pandemic [67, 68], environmental interventions to reduce the consumption of sugar-sweetened beverages [69], interventions to reduce exposure to lead through consumer products and drinking water [70], and workflow disruptions in the surgical operating room [71] have been published in peer reviewed journals. Several others are currently underway, which focus on population-level interventions to increase physical activity and mitigate hypertension and diabetes [72], interventions to reduce road traffic injuries among pedestrians [73], and measures implemented in schools [74] and in long-term care facilities [75] to contain the COVID-19 pandemic. The candidate was also involved in multiple methods-based papers describing methods to improve systematic reviews of complex interventions [70, 76, 77].

The methodological expertise related to the appraisal and conduct of non-randomized evaluations of complex interventions gained through Publications I, II and III, has been applied in multiple primary research activities. This includes the quasi-experimental evaluation of the effect of the Bavarian smoking ban legislation on pregnancy outcomes using an interrupted time series design [78]. It also includes two forthcoming publications, one evaluating the effect of hypertension screening on long-term cardiovascular outcomes [79] and the evaluation of the effect of the German Disease Management Programs on cardiovascular mortality [80]. The candidate was also involved in a methodological review, which highlighted the partially inconsistent and inappropriate use of non-randomized studies in evaluating interventions [81].

2.9 Conclusions

This thesis involved the conduct of a systematic review of a complex intervention, the demonstration of novel methods for making the results of systematic reviews of complex interventions more relevant to users, and the conduct of a good-practice quasi-experimental study. These activities and publications not only highlighted several gaps in the status quo of systematic reviews and primary evaluations of complex interventions, but also described potential steps forward towards improved methods. Due to the nature of complex interventions and the system in which they are implemented, there is no one-size-fits-all set of methods for evaluating their effectiveness, either at the level of systematic review or primary evaluation. Nevertheless, the knowledge generated through these publications provides a valuable and useful contribution to a larger methodological toolbox for evaluating complex interventions. This knowledge can be used both by researchers and practitioners seeking appropriate methods for evaluation, and in the development and further refinement of methods.

3. Zusammenfassung

Seit mehreren Jahrzehnten stellt die systematische Übersichtsarbeit ein wertvolles Instrument für Entscheidungen zu gesundheitsbezogenen Interventionen dar. A priori definierte, Methoden zur systematischen Suche, Identifizierung, Bewertung und Synthese aller Evidenz zu einer bestimmten Intervention bereiten die beste verfügbare Evidenz für eine Nutzung durch Entscheidungsträger, Praktiker und andere Stakeholder auf. Für systematische Übersichtsarbeiten zu Interventionen ist die randomisierte kontrollierte Studie traditionell das wichtigste Studiendesign, da sie das Risiko von Bias und Confounding minimiert und somit einen unverzerrten Schätzer zur Wirksamkeit einer Intervention liefert. Zunehmend stellen Forscher außerhalb der klinischen Medizin, in Bereichen wie Public Health, Fragen zur Wirksamkeit. Public Health Interventionen und das System, in dem sie umgesetzt werden, sind komplexer. So wird diskutiert, ob die traditionellen Methoden der systematischen Übersichtsarbeit auch auf komplexe Interventionen in Public Health und anderen Bereichen anwendbar sind.

Der Ausgangspunkt für diese Arbeit war, dass traditionelle Methoden für systematische Übersichtsarbeiten zu komplexen Interventionen sowie die zugrundeliegenden Primärstudien verbessert werden könnten. Um mehrere der relevanten Lücken zu schließen und verbesserte Methoden für die Evaluation komplexer Interventionen zu erzielen, waren die Ziele dieser Arbeit, i) eine systematische Übersichtsarbeit zu einer komplexen Intervention durchzuführen und zu untersuchen, inwieweit traditionelle Methoden zur Bewertung ihrer Wirksamkeit geeignet sind; ii) eine Methode zu entwickeln und beispielhaft umzusetzen, mithilfe derer die Ergebnisse systematischer Übersichtsarbeiten zu komplexen Interventionen für Entscheidungsträger und Praktiker relevanter und informativer gemacht werden können; und iii) eine qualitativ hochwertige („Good-Practice“) quasi-experimentelle Studie durchzuführen und so exemplarisch eine verbesserte Evaluation komplexer Interventionen darzustellen.

Publikation I, eine Cochrane systematische Übersichtsarbeit zur Wirksamkeit von Maßnahmen zur Verringerung der Luftverschmutzung, stellt eine konkrete Anwendung aktueller Best-Practice Methoden zur Evaluation der Wirksamkeit einer komplexen Public Health Intervention dar. Im Rahmen dieser Anwendung wurde die Eignung der Methodik geprüft und eruiert , wo alternative oder zusätzliche Methoden den Nutzen systematischer Übersichtsarbeiten unterstützen könnten. Publikation I ermöglichte auch die kritische Bewertung der zugrundeliegenden Primärstudien für eine komplexe Public Health Intervention und damit die Identifizierung und Beschreibung von Schwachstellen und Verbesserungsmöglichkeiten.

Publikation II baute auf den Erkenntnissen und Erfahrungen von Publikation I auf. Eingebettet in eine systematische Übersichtsarbeit zur Wirksamkeit häuslicher Palliativversorgung wurde die Anwendung zusätzlicher Methoden – eine Gap-Analyse und Expertenkonsultationen – demonstriert. Diese hatte das Ziel, die Ergebnisse solcher Übersichtsarbeiten für die Nutzer relevanter und aussagekräfiger zu machen.

Publikation III befasste sich mit den Limitationen von Primärstudien zur Evaluation einer komplexen Public Health Intervention. Entsprechend dem Umfang der systematischen Übersichtsarbeit von Publikation I wurde untersucht, wie sich die COVID-19-Beschränkungen, die im Frühjahr 2020 abrupt umgesetzt wurden, auf die Luftverschmutzung in München auswirkten. So wurden mehrere der Empfehlungen aus Publikation I zur Verbesserung der Evaluationsforschung angewandt. Insbesondere wurde beispielhaft gezeigt, wie Routinedaten und quasi-experimentelle Methoden genutzt werden können, um rigorose Evaluationen von komplexen Interventionen durchzuführen.

Diese Publikationen haben nicht nur einige Lücken im Status quo der systematischen Übersichtsarbeiten und zugrundeliegenden Primärstudien zur Evaluation komplexer Interventionen gezeigt; sie haben auch beschrieben wie die Methodik noch weiter verbessert werden könnte. Aufgrund der Komplexität einer Interventionen und des Systems, in dem sie umgesetzt wird, gibt es kein einheitliches, immer gültiges Set von Methoden für die Evaluation ihrer Wirksamkeit. Nichtsdestotrotz liefert das Wissen, das durch diese Publikationen generiert wurde, einen wertvollen und nützlichen Beitrag zu einem größeren methodischen Werkzeugkasten für die Evaluation komplexer Interventionen. Dieses Wissen kann sowohl von Forschern und Praktikern genutzt werden, die nach geeigneten Methoden für die Evaluation ihrer Maßnahmen suchen, als auch bei der Entwicklung und weiteren Verfeinerung von Methoden unterstützen.

4. Abstract

For decades, the systematic review has represented an invaluable tool for informing decisions on health-related interventions. A priori-defined, systematic methods for searching, identifying, appraising and synthesizing all evidence related to a specific intervention ensure that users of the evidence, such as decision-makers and practitioners, have the best-available evidence at hand to inform decisions. Within systematic reviews, the randomized controlled trial has traditionally been the main study design of interest, due to its ability to minimize the risk of bias and confounding and thus deliver an unbiased estimate of intervention effectiveness. Researchers beyond the bounds of clinical medicine, in fields such as public health, have increasingly asked questions of effectiveness. In such fields, interventions and the system in which they are implemented become more complex; as a result, discussions arose as to whether traditional systematic review methods extend to such fields.

The starting point for this thesis was that the traditional methods for systematic reviews of complex interventions, as well as the underlying primary studies, could be improved. Thus, to address several of the relevant gaps and move towards improved methods for evaluating complex interventions, this thesis aimed to i) conduct a systematic review of a complex intervention, and explore the extent to which traditional methods are suitable; ii) demonstrate a method for making the results of systematic reviews of complex interventions more informative for users; and iii) conduct an exemplary good-practice quasi-experimental study, thus illustrating and promoting the improved primary evaluation of complex interventions.

Publication I, a Cochrane systematic review assessing the effectiveness of interventions to reduce ambient air pollution, provided a concrete application of current best-practice systematic review methods for evaluating the effectiveness of a complex public health intervention. This application allowed for the assessment of the suitability of these methods, and of where alternative or additional methods could help improve such systematic reviews. Publication I also allowed for the critical appraisal of the primary study base for a complex public health intervention, and consequently the identification and description of shortcomings and opportunities for improvement.

Publication II built upon the knowledge and experiences of Publication I. Embedded in a systematic review of the effectiveness of home-based palliative care interventions, we demonstrated the use of additional methods – a gap analysis and expert consultations – with the aim of making the results from such reviews more relevant to users.

Publication III dealt with the limitations of the primary evidence base for a complex public health intervention. Specifically, in line with the scope of the systematic review of Publication I, we assessed how the COVID-19 mitigation measures implemented abruptly in early spring 2020 impacted air pollution in Munich. We applied several of the recommendations for improving research made in Publication I, and thus illustrated how routine data and quasi-experimental methods can be utilized to conduct rigorous evaluations of complex interventions.

These publications not only highlighted several gaps in the status quo of systematic reviews and primary evaluations, but also described potential steps forward towards improved methods. Due to the nature of complex interventions, there is no one-size-fits-all set of methods for evaluating their effectiveness. Nevertheless, the knowledge generated through these publications provides a valuable and useful contribution to a larger methodological toolbox for evaluating complex interventions. This knowledge can be used both by researchers and practitioners seeking appropriate methods for evaluation, and in the development and further refinement of methods.

5. Paper I

Cochrane
Library

Cochrane Database of Systematic Reviews

Interventions to reduce ambient particulate matter air pollution and their effect on health (Review)

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess E

Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess E.
Interventions to reduce ambient particulate matter air pollution and their effect on health.
Cochrane Database of Systematic Reviews 2019, Issue 5. Art. No.: CD010919.
DOI: 10.1002/14651858.CD010919.pub2.

www.cochranelibrary.com

TABLE OF CONTENTS

HEADER 1
ABSTRACT 1
PLAIN LANGUAGE SUMMARY 2
SUMMARY OF FINDINGS 4
BACKGROUND 10
Figure 1. 11
Figure 2. 12
OBJECTIVES 13
METHODS 13
RESULTS 18
Figure 3. 19
Figure 4. 20
Figure 5. 27
Figure 6. 28
Figure 7. 30
Figure 8. 31
Figure 9. 32
Figure 10. 33
Figure 11. 35
Figure 12. 36
Figure 13. 38
Figure 14. 39
DISCUSSION 40
AUTHORS' CONCLUSIONS 45
ACKNOWLEDGEMENTS 46
REFERENCES 47
CHARACTERISTICS OF STUDIES 65
ADDITIONAL TABLES 95
APPENDICES 111
Figure 15. 126
Figure 16. 141
Figure 17. 142
CONTRIBUTIONS OF AUTHORS 251
DECLARATIONS OF INTEREST 251
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 251
INDEX TERMS 252

[Intervention Review]

Interventions to reduce ambient particulate matter air pollution and their effect on health

Jacob Burns[1], Hanna Boogaard[2], Stephanie Polus[1], Lisa M Pfadenhauer[1], Anke C Rohwer[3], Annemoon M van Erp[2], Ruth Turley[4], Eva Rehfuess[1]

[1]Institute for Medical Informatics, Biometry and Epidemiology, Pettenkofer School of Public Health, Ludwig-Maximilians-University Munich, Munich, Germany. [2]Health Effects Institute, Boston, MA, USA. [3]Centre for Evidence-based Health Care, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa. [4]Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), Cardiff University, Cardiff, UK

Contact address: Eva Rehfuess, Institute for Medical Informatics, Biometry and Epidemiology, Pettenkofer School of Public Health, Ludwig-Maximilians-University Munich, Marchioninistr. 15, Munich, Bavaria, Germany. rehfuess@ibe.med.uni-muenchen.de.

Editorial group: Cochrane Public Health Group.
Publication status and date: New, published in Issue 5, 2019.

Citation: Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, Turley R, Rehfuess E. Interventions to reduce ambient particulate matter air pollution and their effect on health. *Cochrane Database of Systematic Reviews* 2019, Issue 5. Art. No.: CD010919. DOI: 10.1002/14651858.CD010919.pub2.

ABSTRACT

Background

Ambient air pollution is associated with a large burden of disease in both high-income countries (HICs) and low- and middle-income countries (LMICs). To date, no systematic review has assessed the effectiveness of interventions aiming to reduce ambient air pollution.

Objectives

To assess the effectiveness of interventions to reduce ambient particulate matter air pollution in reducing pollutant concentrations and improving associated health outcomes.

Search methods

We searched a range of electronic databases with diverse focuses, including health and biomedical research (CENTRAL, Cochrane Public Health Group Specialised Register, MEDLINE, Embase, PsycINFO), multidisciplinary research (Scopus, Science Citation Index), social sciences (Social Science Citation Index), urban planning and environment (Greenfile), and LMICs (Global Health Library regional indexes, WHOLIS). Additionally, we searched grey literature databases, multiple online trial registries, references of included studies and the contents of relevant journals in an attempt to identify unpublished and ongoing studies, and studies not identified by our search strategy. The final search date for all databases was 31 August 2016.

Selection criteria

Eligible for inclusion were randomized and cluster randomized controlled trials, as well as several non-randomized study designs, including controlled interrupted time-series studies (cITS-EPOC), interrupted time-series studies adhering to EPOC standards (ITS-EPOC), interrupted time-series studies not adhering to EPOC standards (ITS), controlled before-after studies adhering to EPOC standards (CBA-EPOC), and controlled before-after studies not adhering to EPOC standards (CBA); these were classified as main studies. Additionally, we included uncontrolled before-after studies (UBA) as supporting studies. We included studies that evaluated interventions to reduce ambient air pollution from industrial, residential, vehicular and multiple sources, with respect to their effect on mortality, morbidity and several air pollutant concentrations. We did not restrict studies based on the population, setting or comparison.

Data collection and analysis

After a calibration exercise among the author team, two authors independently assessed studies for inclusion, extracted data and assessed risk of bias. We conducted data extraction, risk of bias assessment and evidence synthesis only for main studies; we mapped supporting studies with regard to the types of intervention and setting. To assess risk of bias, we used the Graphic Appraisal Tool for Epidemiological studies (GATE) for correlation studies, as modified and employed by the Centre for Public Health Excellence at the UK National Institute for Health and Care Excellence (NICE). For each intervention category, i.e. those targeting industrial, residential, vehicular and multiple sources, we synthesized evidence narratively, as well as graphically using harvest plots.

Main results

We included 42 main studies assessing 38 unique interventions. These were heterogeneous with respect to setting; interventions were implemented in countries across the world, but most (79%) were implemented in HICs, with the remaining scattered across LMICs. Most interventions (76%) were implemented in urban or community settings.

We identified a heterogeneous mix of interventions, including those aiming to address industrial (n = 5), residential (n = 7), vehicular (n = 22), and multiple sources (n = 4). Some specific interventions, such as low emission zones and stove exchanges, were assessed by several studies, whereas others, such as a wood burning ban, were only assessed by a single study.

Most studies assessing health and air quality outcomes used routine monitoring data. Studies assessing health outcomes mostly investigated effects in the general population, while few studies assessed specific subgroups such as infants, children and the elderly. No identified studies assessed unintended or adverse effects.

The judgements regarding the risk of bias of studies were mixed. Regarding health outcomes, we appraised eight studies (47%) as having no substantial risk of bias concerns, five studies (29%) as having some risk of bias concerns, and four studies (24%) as having serious risk of bias concerns. Regarding air quality outcomes, we judged 11 studies (31%) as having no substantial risk of bias concerns, 16 studies (46%) as having some risk of bias concerns, and eight studies (23%) as having serious risk of bias concerns.

The evidence base, comprising non-randomized studies only, was of low or very low certainty for all intervention categories and primary outcomes. The narrative and graphical synthesis showed that evidence for effectiveness was mixed across the four intervention categories. For interventions targeting industrial, residential and multiple sources, a similar pattern emerged for both health and air quality outcomes, with essentially all studies observing either no clear association in either direction or a significant association favouring the intervention. The evidence base for interventions targeting vehicular sources was more heterogeneous, as a small number of studies did observe a significant association favouring the control. Overall, however, the evidence suggests that the assessed interventions do not worsen air quality or health.

Authors' conclusions

Given the heterogeneity across interventions, outcomes, and methods, it was difficult to derive overall conclusions regarding the effectiveness of interventions in terms of improved air quality or health. Most included studies observed either no significant association in either direction or an association favouring the intervention, with little evidence that the assessed interventions might be harmful. The evidence base highlights the challenges related to establishing a causal relationship between specific air pollution interventions and outcomes. In light of these challenges, the results on effectiveness should be interpreted with caution; it is important to emphasize that lack of evidence of an association is not equivalent to evidence of no association.

We identified limited evidence for several world regions, notably Africa, the Middle East, Eastern Europe, Central Asia and Southeast Asia; decision-makers should prioritize the development and implementation of interventions in these settings. In the future, as new policies are introduced, decision-makers should consider a built-in evaluation component, which could facilitate more systematic and comprehensive evaluations. These could assess effectiveness, but also aspects of feasibility, fidelity and acceptability.

The production of higher quality and more uniform evidence would be helpful in informing decisions. Researchers should strive to sufficiently account for confounding, assess the impact of methodological decisions through the conduct and communication of sensitivity analyses, and improve the reporting of methods, and other aspects of the study, most importantly the description of the intervention and the context in which it is implemented.

PLAIN LANGUAGE SUMMARY

Ambient air quality – what works to reduce pollution and improve health?

Why did we conduct this review?

Globally, outdoor air pollution is a serious public health problem. In 2016, approximately 4 million deaths were attributable to air pollution, mostly from cardiovascular and respiratory diseases. Air pollution has also been linked to other health problems, like asthma. It is of much concern both in low- and middle-income countries, where air quality may still be worsening, as well as in high-income countries, where pollution levels have decreased over several decades.

Many different policies and programmes have been put into place to reduce air pollution; examples include vehicle restrictions to reduce traffic, fuel standards for cars, buses and other motorized transport, industrial regulations to limit pollution from factories, and the replacement of inefficient heating stoves with more efficient, cleaner burning stoves. So far, no review has investigated systematically whether these measures have impacted air pollution and health as intended.

What is the aim of this review?

We investigated whether measures put into place to reduce outdoor air pollution have actually reduced air pollution and improved health.

What were the main results of this review?

We found 42 studies evaluating a broad range of measures to reduce air pollution in different countries around the world, although most were from high-income countries. Most aimed to reduce air pollution from cars and other vehicles. However, we also identified measures addressing heating and cooking, industry, or a combination of different sources.

We wanted to know whether these measures led to a reduction in the overall number of deaths, and in the number of deaths from cardiovascular and respiratory causes. We also investigated whether the measures led to fewer people going to hospitals for cardiovascular and respiratory problems. We also examined whether there were any changes in outdoor air quality, looking at different pollutants, such as particulate matter, fine particulate matter and other criteria pollutants.

Studies were very diverse with respect to the policies or programmes they assessed, the settings and contexts in which they were implemented, and the methods used to evaluate them.

The evidence we identified was of low and very low certainty, which means we cannot be very confident in the overall findings. Questions around certainty arose because of how studies were designed, conducted and analyzed. While some studies applied rigorous methods, others did not.

Overall, we observed mixed results across studies. Many studies observed no clear changes in health or air quality associated with the measures, while others did observe clear improvements. We identified very few studies that reported worsened health or air quality associated with the measures.

How do we interpret these results?

Differences in the studies make it difficult to draw general conclusions about whether the measures worked. Detecting changes in population health and air pollution levels is challenging, and assessing whether changes that occur are due to a specific measure is complex. Air pollution levels are changing constantly and often unpredictably due to weather and other factors, and other changes happening at the same time could also impact population health and air pollution. When regulations to limit industrial pollution are introduced, one must keep in mind that several other changes may be occurring in the background: an increase in traffic and an upgrade of residential heating systems, for example, or an economic downturn that leads to reduced pollution. It can sometimes take a long time before improvements in health become apparent. In interpreting the review's findings it is important to remember that just because a study did not detect an improvement does not mean that there really was no improvement.

Further evaluations of measures to reduce outdoor air pollution in different countries, in particular in low- and middle-income countries, are needed. Wherever possible, future evaluations should apply more reliable and standardized methods to analyze the data. This should help improve the quality of individual studies as well as our confidence in the findings across studies.

How up to date is this review?

This review includes studies up to 31 August 2016; any studies that were published after that date are not included in this review.

SUMMARY OF FINDINGS

Summary of findings for the main comparison. Interventions targeting vehicular sources compared to practice as usual for improving health and air quality

Interventions targeting vehicular sources compared to practice as usual for improving health and air quality

Population: General population

Setting: Urban and rural areas in high-, middle-, and low-income countries

Intervention: Vehicle charging scheme; speed limit change; low emission zone; road closure; alternating vehicle restriction based on licence plate number; infrastructure changes; fuel requirements; vehicle ban; compulsory vehicle standards

Comparison: Practice as usual

Outcomes	№ of studies	Certainty of the evidence (GRADE)†*	Impact
All-cause mortality Assessed with: routine mortality data Follow-up: 12 years	**1 study**: 1 cITS-EPOC	⊕⊕⊝⊝ LOW	1 cITS-EPOC study showed a significant 2.1% decrease in all-cause mortality associated with the intervention (Yorifuji 2016).
Cardiovascular mortality assessed with: routine mortality data follow-up: 12 years	**1 study**: 1 cITS-EPOC	⊕⊕⊝⊝ LOW	1 cITS-EPOC study showed a significant 5.9% decrease in cardiovascular mortality associated with the intervention (Yorifuji 2016).
Respiratory mortality Assessed with: routine mortality data Follow-up: 12 years	**1 study**: 1 cITS-EPOC	⊕⊕⊝⊝ LOW	1 cITS-EPOC study showed a significant 10% decrease in respiratory mortality associated with the intervention (Yorifuji 2016).
Particulate matter (PM_{10}) Assessed with: routine and study-specific air quality monitors Follow-up: range 4 months to 10 years	**10 studies**: 2 cITS-EPOC 3 ITS-EPOC 2 CBA-EPOC 3 CBA	⊕⊝⊝⊝ VERY LOW [1 2]	4 studies, including 2 ITS-EPOC (Bel 2013b, Viard 2015**) and 2 CBAs (Dijkema 2008, Fensterer 2014), showed significant decreases of 14.7%, 31% , 7.4% and 13%, respectively, in PM_{10} concentrations associated with the intervention. 5 studies, including 1 cITS-EPOC (Cowie 2012), 1 ITS-EPOC (Peel 2010), 1 CBA-EPOC (Boogaard 2012) and 1 CBA (Ruprecht 2009**) observed no effect associated with the intervention. 2 studies, including 1 cITS-EPOC (Bel 2013a) and 1 CBA-EPOC (Kim 2011**) showed significant 5.4% and 14.7% increases, respectively, in concentrations associated with the intervention.
Fine particulate matter ($PM_{2.5}$) Assessed with: routine and study-specific air quality monitors Follow-up: range 2 years to 3 years	**2 studies**: 1 cITS-EPOC 1 CBA-EPOC	⊕⊕⊝⊝ LOW	1 CBA-EPOC study showed a significant 30% decrease in PM2.5 concentrations associated with the intervention (Boogaard 2012). 1 cITS-EPOC study observed no effect associated with the intervention (Cowie 2012).
Coarse particulate matter	0 studies	-	No studies assessed the effect of interventions to reduce ambient air pollution from vehicular sources on coarse particle concentrations.
Combustion-related particulate matter	**4 studies**:	⊕⊕⊝⊝ LOW	2 studies, including 2 CBAs (Titos 2015a**; Titos 2015b**), showed significant decreases in black carbon of 72% and 37%

Assessed with: routine and study-specific air quality monitors Follow-up: range 2 months to 2 years	1 CBA-EPOC 3 CBA		associated with the intervention. 2 studies, including 1 CBA-EPOC (Boogaard 2012) and 1 CBA (Dijkema 2008) observed no effect associated with the intervention.

† All studies included for this comparison were non-randomized; thus each body of evidence started the GRADE assessment with a rating of 'Low quality'.

***** The certainty of evidence ratings from GRADE should not be confused with those from the NICE modified GATE Risk of Bias tool, which uses a (++); (+); (-) rating system for individual study risk of bias.

**Denotes that effectiveness was determined in parallel analyses for intervention and control sites before and after the intervention. The separate effect estimates obtained through the parallel analyses were then compared in order to draw indirect conclusions about intervention effectiveness, e.g. if a statistically significant improvement was observed at intervention sites, while no change was observed at control sites, this was assigned an "effect favouring the intervention".

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[1] Rated −1 for risk of bias, due to the selection of intervention and control sites and pollution monitors, and methods of statistical analysis.
[2] Rated −1 for inconsistency, as effects from the studies range from positive to negative effects. Some of this is likely to be due to differences in the intervention and/or context, however this inconsistency is nevertheless a concern.

Summary of findings 2. Interventions targeting industrial sources compared to practice as usual for improving health and air quality

Interventions targeting industrial sources compared to practice as usual for improving health and air quality

Population: General population, as well as age-specific subgroups (< 1 year; < 14 years; > 65 years)

Setting: Urban and rural areas in high- and middle-income countries

Intervention: Cap and trade programme; factory closure; compulsory power plant standards; power plant fuel conversion

Comparison: Practice as usual

Outcomes	№ of studies	Certainty of the evidence (GRADE)†*	Impact
All-cause mortality Assessed with: routine mortality data Follow-up: range 5 years to 10 years	**3 studies**: 2 cITS-EPOC 1 CBA-EPOC	⊕⊕⊝⊝ LOW	1 cITS-EPOC study found a statistically significant 2.5% decrease in all-cause mortality at intervention sites compared to control sites (Pope 2007). 2 studies, 1 cITS-EPOC (Deschênes 2012) and 1 CBA-EPOC (Tanaka 2015), observed no effect associated with the intervention.
Cardiovascular mortality Assessed with: routine mortality data Follow-up: 10 years	**1 study**: 1 cITS-EPOC	⊕⊕⊝⊝ LOW	1 cITS-EPOC study observed no effect associated with the intervention (Deschênes 2012).

Respiratory mortality	0 studies	-	No studies assessed the effect of interventions to reduce ambient air pollution from industrial sources on coarse particle concentrations.
Particulate matter (PM_{10}) Assessed with: routine and study-specific air quality monitors Follow-up: range 2 years to 10 years	**3 studies**: 1 cITS-EPOC 1 ITS-EPOC 1 CBA	⊕⊝⊝⊝ VERY LOW [1 2]	1 CBA study showed a statistically significant 14% decrease in PM_{10} concentrations associated with the intervention (Saaroni 2010). 1 cITS-EPOC study observed no effect associated with the intervention (Deschênes 2012). 1 ITS-EPOC study showed a significant 13.2% increase in PM_{10} concentrations associated with the intervention (Sajjadi 2012).
Fine particulate matter ($PM_{2.5}$) Assessed with: routine and study-specific air quality monitors Follow-up: 10 years	**1 study**: 1 cITS-EPOC	⊕⊕⊝⊝ LOW	1 cITS-EPOC study observed no effect associated with the intervention (Deschênes 2012).
Coarse particulate matter	0 studies	-	No studies assessed the effect of interventions to reduce ambient air pollution from industrial sources on coarse particle concentrations.
Combustion-related particulate matter	0 studies	-	No studies assessed the effect of interventions to reduce ambient air pollution from industrial sources on concentrations of combustion-related particulate matter concentrations.

† All studies included for this comparison were non-randomized; thus each body of evidence started the GRADE assessment with a rating of 'Low quality'.

***** The certainty of evidence ratings from GRADE should not be confused with those from the NICE modified GATE Risk of Bias tool, which uses a (++); (+); (-) rating system for individual study risk of bias.

**Denotes that effectiveness was determined in parallel analyses for intervention and control sites before and after the intervention. The separate effect estimates obtained through the parallel analyses were then compared in order to draw indirect conclusions about intervention effectiveness, e.g. if a statistically significant improvement was observed at intervention sites, while no change was observed at control sites, this was assigned an "effect favouring the intervention".

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[1] Rated −1 for risk of bias, due to potential selection bias and the lack of adjustment for potentially important confounders.
[2] Rated −1 for inconsistency, as effects from the studies range from positive to negative effects. Some of this is likely explainable due to differences in the intervention and /or context, however this inconsistency is nevertheless a concern.

Summary of findings 3. Interventions targeting residential sources compared to practice as usual for improving health and air quality

Interventions targeting residential sources compared to practice as usual for improving health and air quality
Population: General population

Setting: Urban and rural areas in high- and low-income countries

Intervention: Stove exchange; ban on wood burning; ban on sale, distribution and burning of coal

Comparison: Practice as usual

Outcomes	№ of studies	Certainty of the evidence (GRADE)†*	Impact
All-cause mortality Assessed with: routine mortality data Follow-up: range 13 years to 23 years	**4 studies:** 4 cITS-EPOC	⊕⊝⊝⊝ VERY LOW [1]	4 cITS-EPOC studies observed no effect associated with the intervention (Dockery 2013a**; Dockery 2013b**; Dockery 2013c**; Johnston 2013**).
Cardiovascular mortality Assessed with: routine mortality data Follow-up: range 13 years to 23 years	**4 studies:** 4 cITS-EPOC	⊕⊕⊝⊝ LOW	4 cITS-EPOC studies observed no effect associated with the intervention (Dockery 2013a**; Dockery 2013b**; Dockery 2013c**; Johnston 2013**).
Respiratory mortality Assessed with: routine mortality data Follow-up: range 13 years to 23 years	**4 studies:** 4 cITS-EPOC	⊕⊝⊝⊝ VERY LOW [1]	1 cITS-EPOC study showed a significant 16.8% decrease in respiratory mortality associated with the intervention (Dockery 2013a**). 3 cITS-EPOC studies observed no effect associated with the intervention (Dockery 2013b**; Dockery 2013c**; Johnston 2013**).
Particulate matter (PM_{10})	0 studies	-	No studies assessed the effect of interventions to reduce ambient air pollution from residential sources on PM_{10} concentrations.
Fine particulate matter ($PM_{2.5}$) Assessed with: routine and study-specific air quality monitors Follow up: range 3 months to 6 years	**3 studies:** 1 ITS-EPOC 2 CBA	⊕⊝⊝⊝ VERY LOW [1 2]	1 ITS-EPOC showed a significant 12.3% decrease in PM2.5 concentrations associated with the intervention (Yap 2015). 2 CBAs observed no effect associated with the intervention (Allen 2009**; Aung 2016**).
Coarse particulate matter Assessed with: routine air quality monitors Follow-up: 6 years	**1 study:** 1 ITS-EPOC	⊕⊝⊝⊝ VERY LOW [3]	1 ITS-EPOC showed a significant 8.5% decrease in coarse particle concentrations associated with the intervention (Yap 2015).
Combustion-related particulate matter Assessed with: study-specific air quality monitors Follow-up: 3 months	**1 study:** 1 CBA	⊕⊝⊝⊝ VERY LOW [1 2]	1 CBA observed no effect associated with the intervention (Aung 2016**).

† All studies included for this comparison were non-randomized; thus each body of evidence started the GRADE assessment with a rating of 'Low quality'.

***** The certainty of evidence ratings from GRADE should not be confused with those from the NICE modified GATE Risk of Bias tool, which uses a (++); (+); (-) rating system for individual study risk of bias.

**Denotes that effectiveness was determined in parallel analyses for intervention and control sites before and after the intervention. The separate effect estimates obtained through the parallel analyses were then compared in order to draw indirect conclusions about intervention effectiveness, e.g. if a statistically significant improvement was observed at intervention sites, while no change was observed at control sites, this was assigned an "effect favouring the intervention".

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

Low certainty: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[1] Rated −1 for imprecision, due to very wide confidence intervals spanning from a meaningful effect to a potential harmful effect .
[2] Rated −2 for risk of bias, due to the risk of contamination between intervention and control sites, an inappropriately short follow-up time, and the lack of consideration of potentially important confounders.
[3] Rated −1 for risk of bias, due to the timing of the intervention introduction, and the lack of consideration of potentially important confounders.

Summary of findings 4. Interventions targeting multiple sources compared to practice as usual for improving health and air quality

Interventions targeting multiple sources compared to practice as usual for improving health and air quality

Population: General population

Setting: Urban and rural areas in high countries

Intervention: Coordinated vehicular and industrial measures during periods of heavy pollution; definition of attainment/non-attainment status and tailored measures for reaching attainment status

Comparison: Practice as usual

Outcomes	№ of studies	Certainty of the evidence (GRADE)†*	Impact
All-cause mortality Assessed with: routine mortality data Follow-up: range 11 years to 19 years	**2 studies**: 1 ITS-EPOC 1 CBA-EPOC	⊕⊝⊝⊝ VERY LOW [1,2]	2 studies, 1 CBA-EPOC (Zigler 2016) and 1 ITS-EPOC (Mullins 2014), observed no effect associated with the intervention.
Cardiovascular mortality	0 studies	-	No studies assessed the impact of interventions to reduce ambient air pollution from multiple sources on cardiovascular mortality.
Respiratory mortality	0 studies	-	No studies assessed the impact of interventions to reduce ambient air pollution from multiple sources on respiratory mortality.
Particulate matter (PM_{10}) Assessed with: routine and study-specific air quality monitors Follow-up: range 11 years to 19 years	**2 studies**: 1 ITS-EPOC 1 CBA-EPOC	⊕⊝⊝⊝ VERY LOW [2]	1 ITS-EPOC study showed a significant 5.6% decrease in PM_{10} concentrations associated with the intervention (Mullins 2014). 1 CBA-EPOC observed no effect associated with the intervention (Zigler 2016).
Fine particulate matter ($PM_{2.5}$)	0 studies	-	No studies assessed the impact of interventions to reduce ambient air pollution from multiple sources on $PM_{2.5}$ concentrations.
Coarse particulate matter	0 studies	-	No studies assessed the impact of interventions to reduce ambient air pollution from multiple sources on coarse particle concentrations.

Combustion-related particulate matter	0 studies	-	No studies assessed the impact of interventions to reduce ambient air pollution from multiple sources on concentrations of combustion-related particulate matter concentrations.

† All studies included for this comparison were non-randomized; thus each body of evidence started the GRADE assessment with a rating of 'Low quality'.

***** The certainty of evidence ratings from GRADE should not be confused with those from the NICE modified GATE Risk of Bias tool, which uses a (++); (+); (-) rating system for individual study risk of bias.

**Denotes that effectiveness was determined in parallel analyses for intervention and control sites before and after the intervention. The separate effect estimates obtained through the parallel analyses were then compared in order to draw indirect conclusions about intervention effectiveness, e.g. if a statistically significant improvement was observed at intervention sites, while no change was observed at control sites, this was assigned an "effect favouring the intervention".

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect

[1] Rated −1 for risk of bias, due to potential contamination in the aggregate outcome data, and the use of potentially non-appropriate covariates in the analysis.
[2] Rated −1 for imprecision, due to concerns regarding whether there is sufficient precision to detect the presence of an effect.

BACKGROUND

Description of the condition

Ambient air pollution is a complex mixture of particles and gases. Their concentrations and composition vary from place to place, depending on what sources are present, weather conditions, and how they mix in the atmosphere. Particulate matter (PM) is one of the most widely monitored and studied components of air pollution, namely PM_{10} (particles smaller than 10 micrometres in aerodynamic diameter, and particularly $PM_{2.5}$ (particles with an average aerodynamic diameter smaller than 2.5 micrometres). Both PM_{10} and $PM_{2.5}$ can be readily inhaled, and $PM_{2.5}$ is considered especially harmful because of its ability to penetrate deep into the lungs (Chow 1995).

Exposure to PM and other pollutants is associated with numerous health outcomes in adults, including premature deaths from all causes, and cardiovascular and respiratory diseases (Pope 2006). In addition to mortality, ambient PM air pollution has been associated with respiratory morbidity, including asthma attacks, pneumonia, decreased lung function and hospital admissions due to respiratory events, as well as with cardiovascular morbidity, including heart attack and hospital admissions due to cardiovascular events (Pope 2006; Rückerl 2011).

Description of the intervention

In order to improve air quality and reduce particulate matter and other air pollutant concentrations, a variety of interventions have been implemented. These range from national and regional regulations to local actions, and may involve either single or multiple governmental sectors (van Erp 2012). They range from those that influence air quality over a long period of time to those with short-term goals. Interventions that improve air quality may be implemented for a range of reasons, including meeting air quality standards, reducing emissions, reducing contamination of water bodies or improving visibility. An improvement in air quality could also occur as a side effect of an intervention with different goals, for example reducing congestion or improving traffic flow (van Erp 2012).

Interventions can be categorized with regard to the target source of air pollution directly or indirectly affected by the intervention. Globally, on top of the 18% stemming from natural and 22% from unspecified sources, approximately 15% of urban ambient pollution stems from industrial sources, 20% from residential sources and 25% from vehicular sources (Campbell-Lendrum 2019). In line with this, the categories of interventions considered in this review, along with some examples of each, are as follows.

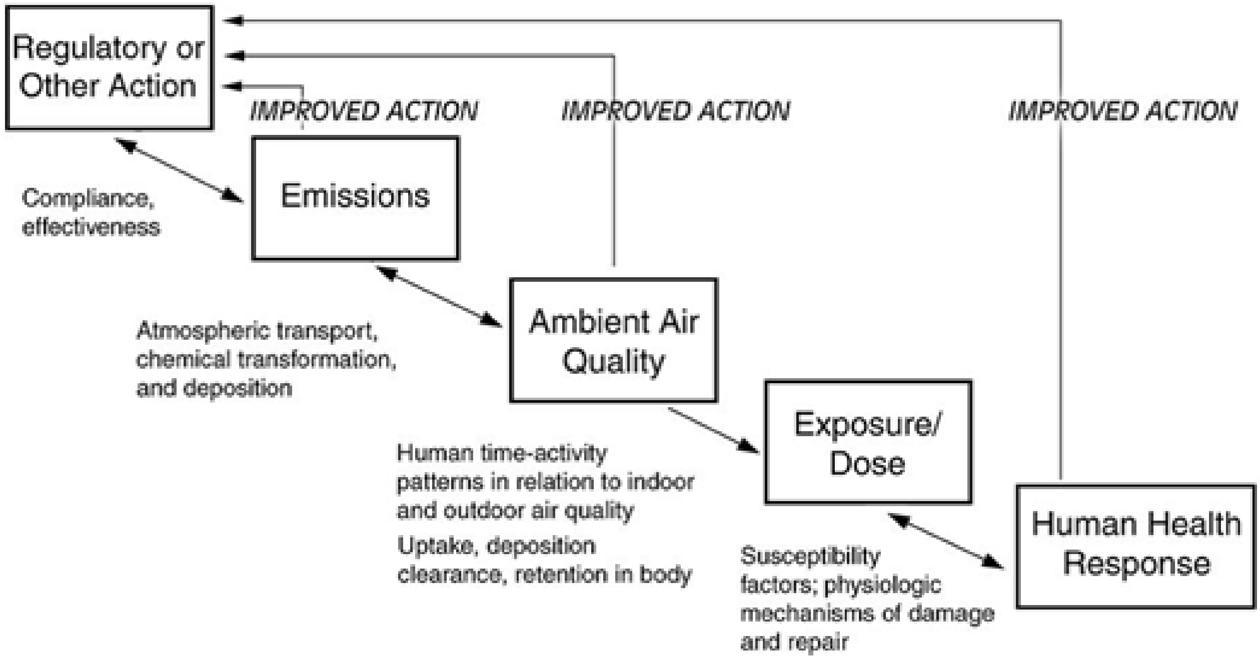
- Industrial: emission standards and regulations for power plants and other industrial sources, fuel changes.
- Residential: stove changeout programmes, banning the sale and use of coal.
- Vehicular: low emission zones, vehicle charging schemes, public transportation expansion; fuel and technology changes; these could apply to the road-based fleet, but also to air and marine fleets.
- Multiple: coordinated policies such as the European National Emission Ceilings Directive, measures during international sporting events, such as the 2008 Beijing Olympic Games.

How the intervention might work

Air quality interventions may comprise multiple components, are often carried out over an extended period of time and may involve multiple governmental sectors including environment, transport, energy, energy generation and health. Also, such interventions may not lead to immediate changes in human exposure or health outcomes. This complexity, as well as multiple, interacting environmental and biological pathways leading to a health response, greatly complicate the assessment of these effects (HEI 2003).

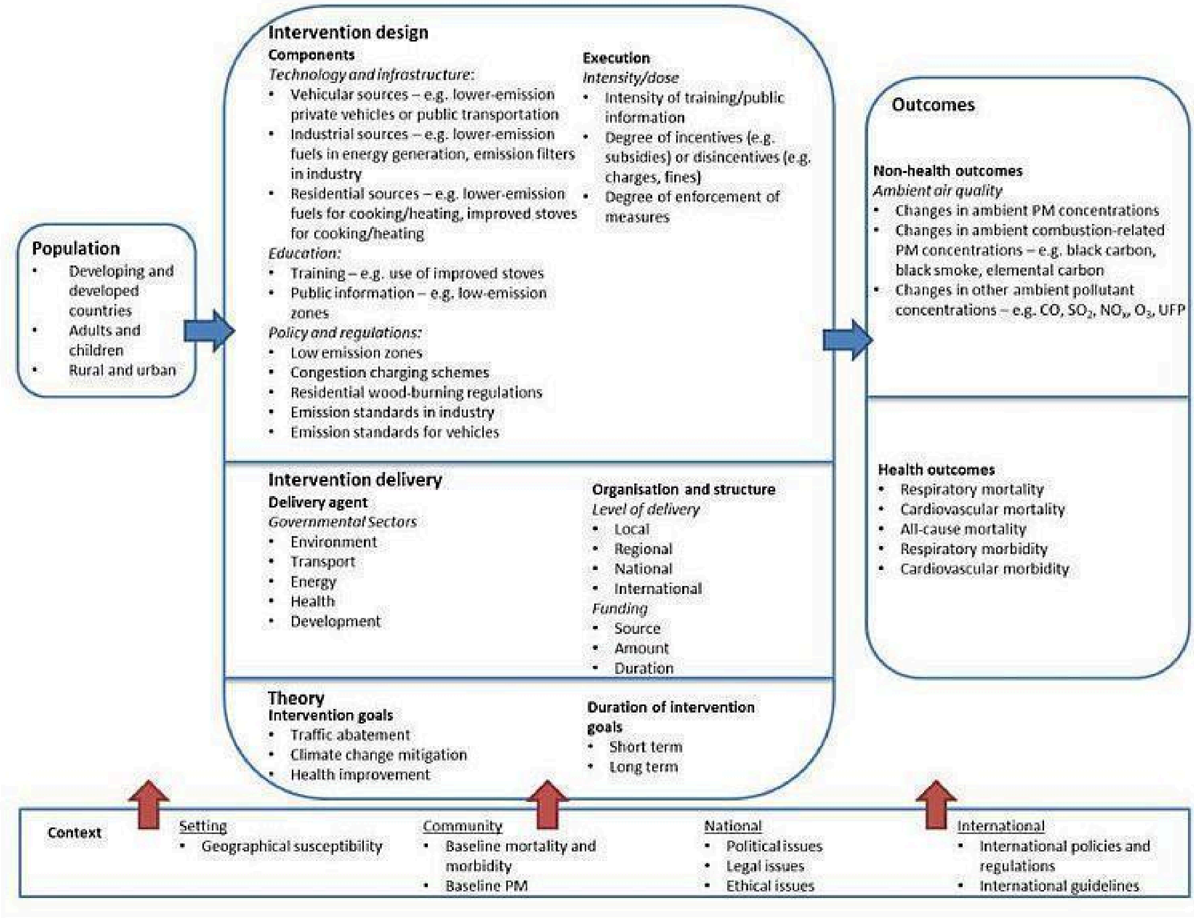
The US National Research Council's Committee on Research Priorities for Airborne Particulate Matter set out a conceptual framework for linking air pollution sources to adverse health effects (NRC 2002). This 'chain of accountability' has been adapted by the Health Effects Institute, as shown in Figure 1, with each stage affording its own opportunities to evaluate how interventions affect emissions, ambient air quality, human exposures and doses, and ultimately health effects (HEI 2003). Each stage provides a checkpoint at which one can assess whether an intervention has been effective; studies may include evaluations of one or several of the stages. This 'cycle' is often used in studies investigating the health effects of interventions.

Figure 1.



At the protocol stage we developed a system-based logic model to visualize and communicate the relationship between various ambient pollutants and interventions in their broader societal and environmental context, as well as to structure and guide the review process (Figure 2) (Rehfuess 2017; Rohwer 2017).

Figure 2. System-based logic model depicting the relationship between various interventions, air pollutants and health in their broader societal and environmental context



Why it is important to do this review

Air quality has improved substantially over recent years in most HICs, with downward trends in concentrations of several major regulatory pollutants such as PM, ozone (O_3), carbon monoxide (CO), nitrogen dioxide (NO_2), and sulphur dioxide (SO_2). In large part, these air quality improvements have been achieved through air quality regulations and effective control of emissions from both stationary and mobile air pollution sources. However, new research has strengthened the evidence for adverse health effects of air pollution at low ambient concentrations, even those below current ambient air quality standards, supporting the case for further regulatory action (Di 2017; Pinault 2017). Additionally, outdoor air pollution exposures and trends differ widely across different parts of the globe, with many LMICs experiencing very high average annual concentrations and increasing trends (Cohen 2017; van Donkelaar 2015).

The contrasting situations (i.e. improvement versus deterioration of air quality) around the globe present challenges in evaluating air-pollution-related health effects and the impact of air quality interventions. In the HICs, interest in assessing the health effects of air quality interventions has grown in response to questions about the benefit of further tightening air pollution regulations. The cost of the air-pollution-control technologies and mechanisms needed to implement and enforce regulations can be substantial (WHO 2016). For example, the US Environmental Protection Agency (US EPA) estimated the cost of air pollution control in 2000 at approximately USD 20 billion, USD 53 billion in 2010, and USD 65 billion has been projected for 2020. Estimated benefits, however, in terms of fewer deaths and hospital admissions, as well as reduced absence at school or work due to illness, exceed those costs by a factor of 30 to 1 (US EPA 2011). In contrast, there is interest in many LMICs to generate local scientific documentation of associations between air pollution and health as well as the impact of air quality interventions. For these settings, there is uncertainty as to whether the concentration-response functions from existing epidemiologic studies primarily conducted in HICs are directly applicable to the differing pollution mixtures and concentrations, as well as the differing demographic compositions, found in many LMICs (Tonne 2017).

Typically, assessments of the benefits of air quality regulations have relied on concentration-response functions from existing epidemiologic studies, which are then used to predict health outcomes that might be avoided under alternative air pollution policy scenarios. Such assessments can be done either retrospectively, by calculating health benefits based on actual observed or modelled air quality improvements (Tonne 2008), or prospectively, by calculating benefits based on improvements predicted in advance of a new policy (Schmitt 2016). To date, however, such estimates have not been extensively validated by comparison with results of 'real world' studies of regulatory programmes using actual health outcome data. Accountability studies (sometimes referred to as intervention studies), which refer to empirical studies assessing the effects of regulatory actions, interventions, or natural experiments (e.g. the sudden closure of a factory or a public transportation strike) on air pollution and health, have emerged to fulfil that role.

Accountability studies typically compare air quality or population health (or both) before and after implementation of a policy intervention, although they often defy a clear study design classification. Accountability studies are appealing since they are the closest epidemiologic equivalent to controlled experimental studies in the field of air pollution research, and thus may provide evidence for causal relationships.

Several recent reviews have summarized the evidence to assess the effectiveness of air quality interventions to improve air quality and health (Bell 2011; Boogaard 2017; Henneman 2017; Henschel 2012; Rich 2017); however, no review has been performed to date with standardized and transparent and systematic review methods.

A protocol including 'a priori defined' methods for this review has been published (Burns 2014).

OBJECTIVES

To assess the effectiveness of interventions to reduce ambient particulate matter air pollution in reducing pollutant concentrations and improving associated health outcomes.

METHODS

Criteria for considering studies for this review

Types of studies

The randomized evaluation of large-scale public health interventions is often not feasible or practical (Craig 2017; Higgins 2012), thus non-randomised studies (NRS) of interventions comprise the main source of evidence to assess the effectiveness of ambient air quality interventions. The following study designs were therefore eligible for inclusion.

- Individually randomized trials.
- Cluster-randomized trials.
- Controlled before-after studies adhering to EPOC standards (CBA-EPOC) – assessed pre- and post-intervention data for at least two intervention sites and two control sites (Cochrane EPOC 2017).
- Interrupted time series studies adhering to EPOC standards (ITS-EPOC) – with at least three data points before and after a clearly defined intervention (in terms of content and timing) (Cochrane EPOC 2017).
- Controlled before-after studies not adhering to EPOC standards (CBA) – assessed pre- and post-intervention data at fewer than two intervention and/or control sites.
- Uncontrolled before-after studies (UBA) – assessed pre- and post-intervention data only at one or multiple intervention sites.
- Interrupted time series studies not adhering to EPOC standards (ITS) – with fewer than three data points before and after a clearly defined intervention (in terms of content and timing).
- Controlled ITS studies (cITS-EPOC) – After publication of the protocol, we identified several publications that applied an ITS-EPOC study design, and also included data from one or more control sites. These, for example, conducted separate, parallel ITS analyses at intervention and control sites, or conducted an ITS analysis at intervention sites that was adjusted for contemporaneous changes at control sites. Although these studies meet the study design inclusion criteria, none of the 'a priori defined' study designs appropriately captured the design and analysis features. We decided post hoc to classify these studies as cITS-EPOC.

As we expected inconsistencies in the terminology and naming of study designs, we were cautious not to exclude studies based on study design labels. For example, a study labelled a cohort study, which was clearly linked to an intervention and where effect data were collected both pre- and post-intervention at an intervention site, but without a control site, was considered an uncontrolled before-and-after study according to our definition, and was thus included.

Types of participants

Interventions to reduce ambient PM air pollution are usually intended for the general population and are of global relevance. As discussed above, concentrations at which ambient PM air pollution has been shown to affect health are experienced by both children and adults in urban and rural settings in both developed and developing countries (Dadvand 2013; Gakidou 2017; WHO Europe 2013). For this reason, we made no exclusions with regard to age group or any other individual, population or setting-related characteristics.

Types of interventions

We categorized interventions with regard to the target PM source, and thus included interventions belonging to the following categories.

- Industrial interventions: those interventions aimed at reducing ambient PM stemming from industrial and power-generating sources.
- Residential interventions: those interventions aimed at reducing ambient PM stemming from residential heating and cooking, or those aimed at reducing indoor PM from these sources, but resulting in changes in ambient PM concentrations.
- Vehicular interventions: those interventions aimed at reducing ambient PM originating from any vehicular source, including automobiles, but also other forms of transportation such as public transportation, aeroplanes or ships. We also included interventions aimed at reducing traffic and/or congestion that also resulted in changes in ambient PM concentrations.
- Multiple interventions: those interventions aimed at reducing ambient PM originating from multiple sources, which could include any of the above-listed sources.

Certain interventions, for example forms of personal protection including masks and filtration systems, were not included. Additionally, we did not include studies assessing changes to agricultural practices.

The comparison was expected to be no intervention or practice as usual in most cases; we did not exclude studies based on the comparison.

Types of outcome measures

Effects of interventions can be assessed with regard to the impact on air quality or impact on the health of individuals or populations, or both. For this review, studies that measured any primary or secondary outcome were eligible for inclusion.

Primary outcomes

Health

An association between health and exposure to ambient air pollution, and in particular to PM, has been observed for several health outcomes, including cardiovascular, respiratory and all-cause mortality, as well as acute cardiovascular and respiratory events. As approximately 4 million deaths worldwide were attributed to air pollution in 2016 (Gakidou 2017), and given that mortality data is often collected on a routine basis, the primary health outcomes we considered for this review were the following mortality-related outcomes.

- All-cause mortality
- Cardiovascular mortality
- Respiratory mortality

Ambient air quality

Ambient air pollution is a complex mixture of particles and gases, such as PM, carbon monoxide (CO), sulphur dioxide (SO_2), nitrogen oxides (NOx) (including nitric oxide (NO) and nitrogen dioxide (NO_2)), and Ozone (O_3) (Hoek 2013; Rückerl 2011; WHO Europe 2013). PM is the indicator pollutant used most broadly for monitoring, with one of the most stringent standards, and has been shown to be associated with numerous health outcomes. It was therefore the primary outcome used to assess ambient air quality for this review. As other pollutants are also monitored and associated with health effects, we considered these as secondary outcomes.

PM is measured using various sampling methods, most often gravimetrically on filters, and is often classified using size ranges, such as PM_{10}, $PM_{2.5}$ and coarse particles (i.e. particles with an average aerodynamic diameter between 2.5 and 10 micrometres). Additionally, since there is some evidence that combustion-related PM may be more harmful to health than PM generated from other sources (Janssen 2011; Lippmann 2013), we also considered studies that focused on combustion-related indicators of PM. Thus the PM-related primary outcomes included:

- PM_{10};
- $PM_{2.5}$;
- coarse PM;
- soot;
- black carbon (BC);
- black smoke (BS);
- elemental carbon (EC);
- absorption of PM (a measure of soot).

For these PM-related outcomes, studies were eligible for inclusion if ambient PM concentrations were measured over 24 hours or over multiples of 24 hours (e.g. 48-hour, weekly, monthly or annual averages).

As the focus of this review is on the effectiveness of interventions to reduce ambient PM concentrations, we did not include those studies measuring only indoor air pollution. While studies that use biomarkers as proxies of exposure are becoming more common, this field is still in its infancy, and uncertainties remain with respect to the reliability of these biomarkers (Turner 2017). We therefore did not consider such studies.

Secondary outcomes

This review also assessed the following secondary outcomes, where available.

Health

- Respiratory effects
  * Lung function
  * Respiratory events, including symptoms
  * Hospital admissions due to respiratory events
- Cardiovascular effects
  * Cardiovascular events, including symptoms
  * Hospital admissions due to cardiovascular events

Ambient air quality

Concentrations of:

- CO;
- SO_2;
- NOx;
- O_3;
- ultrafine particles (UFP) – particles with an average aerodynamic diameter smaller than 0.1 micrometres, or 100 nanometres (measured as particle number concentration);
- personal PM exposure.

Unintended adverse outcomes

As PM interventions may also generate unintended adverse effects, which would be of relevance to decision makers, we attempted to document these where reported in primary studies. Examples could include:

- reductions in physical activity;
- loss of employment;
- economic losses;
- safety.

Search methods for identification of studies

We performed searches within the following electronic databases:

- Health/biomedical
  * CENTRAL
  * Cochrane Public Health Group Specialised Register
  * MEDLINE (1947 to date)
  * MEDLINE (In-Process)
  * Embase (1947 to date)
  * PsycINFO (1806 to date)
- Multidisciplinary
  * Scopus (1960 to date)
  * Science Citation Index (1960 to date)
- Social sciences
  * Social Science Citation Index (1956 to date)
- Urban planning/environment
  * Greenfile
- Lower/middle-income country-relevant
  * Global Health Library sources
    - Regional indexes: AIM (AFRO), LILACS (AMRO/PAHO), IMEMR (EMRO), IMSEAR (SEARO), WPRIM (WPRO)
  * WHOLIS (World Health Organization (WHO) Library)
- Grey literature/unpublished/in press
  * HMIC (1979 to date)
  * WHO ICTRP (inception to date)
  * ClinicalTrials.gov (inception to date)
  * IDEAS (inception to date)
  * JOLIS (inception to date)
  * 3ie impact database (inception to date)
  * PubMed (all-topic search for e-publications ahead of print in title and abstract)

We first designed the search strategy in MEDLINE, and combines four search concepts: 1) the phenomenon of interest (ambient PM air pollution, ambient air quality); 2) ambient air quality and health outcomes of interest; 3) interventions expected to reduce ambient PM concentrations from vehicular, industrial or residential sources; and 4) eligible study designs (this search filter returns those study designs used in epidemiological research, i.e. no toxicological, pharmaceutical or animal studies). The search strategy was then adapted for each remaining database, as shown in Appendix 1. The electronic searches were conducted in two rounds, first during January to February 2014, followed by a search update in August 2016.

In addition to the electronic search, we handsearched the references of included studies, and the tables of contents of Environmental Health Perspectives and Atmospheric Environment for the 12 months preceding the last search date.

Searches were conducted in English but we endeavoured not to exclude any studies on the basis of language, with the team being able to assess papers published in English, Dutch, German, French, Italian and Afrikaans. For papers not published in any of these languages, we explored options for translation and assessment for inclusion. All search results were stored in EndNote.

Data collection and analysis

Selection of studies

Following removal of duplicate studies, we performed a multistage screening process. In the first stage, JB and LP screened all titles, removing those clearly not relevant with regard to population, intervention, outcomes or study design (e.g. animal studies, chamber studies, letters to the editor). In a subsequent calibration exercise, all review authors independently screened 100 randomly selected titles and abstracts and discussed any disagreements to ensure a standardized screening process. In the protocol, we had planned a single-reviewer title- and abstract-screening round at this stage, to further remove any clearly irrelevant evidence. Given that only very few studies appeared to be clearly irrelevant we did not perform this step, and continued with duplicate title and abstract screening, as described below.

In the second stage, two review authors (from JB, HB, SP, LP, AR, ER) independently screened all remaining titles and abstracts. An inclusive approach was taken, and studies for which we could not ascertain certain key criteria for inclusion from the abstract were kept for full-text screening. Review authors resolved disagreements through discussion; or invited a third review author to arbitrate when necessary.

In the final screening stage, two review authors (from JB, HB, SP, LP, AR, ER) independently examined the full text of all potentially relevant studies, assessing each against a checklist of inclusion criteria. Review authors resolved disagreements through discussion; or invited a third review author to arbitrate when necessary. Review authors documented the reasons for exclusion at the full-text screening stage.

We conducted all stages of the screening process using Endnote.

We made the post hoc decision to further divide the included studies into main studies that contributed intervention effects to the evidence synthesis, and supporting studies that contributed descriptive data to the review results. Supporting studies included two different types of study: those conducting non-analytical descriptive comparisons; and those applying a UBA study design. We made this decision completely independent of the results of included studies.

With regard to the first type of supporting study, although the study design technically met the a priori inclusion criteria, no analytical comparison providing a quantitative effect estimate relevant for our review was conducted. Such studies, for example, might have collected air quality and/or health data at intervention and control sites before and after an intervention, but presented only descriptive data at these sites, without any further statistical analysis.

With regard to the second type of supporting study, after extracting data and assessing the risk of bias of approximately half of the included UBA studies, we realized that these would only provide a very weak argument for a causal link between the intervention and the air quality and/or health, and very low confidence that the estimated effect indeed represented intervention effectiveness. Problems with UBA studies were compounded by 1) poor internal validity due to data collection, study and intervention timing, selection of sites, statistical analysis, and 2) weak reporting with respect to the intervention, the intervention timing, the expected intervention effect, as well as study design and statistical analysis.

Thus, as described above, we included as supporting studies the studies with a descriptive comparison and the studies applying a UBA study design. These studies represent a record of the types of interventions and settings covered but did not undergo full data extraction or risk of bias assessment and did not contribute to the evidence synthesis to examine intervention effectiveness. Consequently, the description of data extraction and management and data synthesis in the following section only refers to main studies.

Data extraction and management

As considerable heterogeneity was expected with respect to the interventions, outcomes, study designs and analyses of included main studies, we extracted extensive data on these aspects. Additionally, over the past years the importance of the setting, context and implementation on the effectiveness of public health interventions has also been emphasized (Wells 2012). We therefore aimed to extract potentially relevant data using the Context and Implementation of Complex Interventions (CICI) framework (Pfadenhauer 2017). We used a standardized form adapted from the Data Extraction and Assessment Template provided by Cochrane Public Health (see Appendix 2).

After developing the data extraction form, we performed a calibration exercise in which all review authors extracted data from the same two studies; we then discussed and clarified any differences in extraction between review authors before continuing. For all included main studies, two review authors (from JB, HB, SP, LP, AR, ER) independently extracted data using the standardized data extraction form. The two review authors resolved inconsistencies or disagreements through discussion, or consulted a third review author where necessary.

Assessment of risk of bias in included studies

We assessed the risk of bias of all primary and secondary outcomes. To do so, we used the Graphic Appraisal Tool for Epidemiological studies (GATE) for correlation studies, as modified and employed by the Centre for Public Health Excellence at the UK National Institute for Health and Care Excellence (NICE) (Jackson 2006; NICE 2012). This modified GATE tool is well suited to the assessment of non-randomized intervention studies, and is therefore practical in a review such as this (NICE 2012; Voss 2013). The GATE appraisal checklist is divided into five sections consisting of 18 criteria, and allows for a systematic assessment of aspects related to the external validity (section 1: population) and internal validity or risk of bias (sections 2 to 4: method of selection of exposure or comparison group; outcomes; analyses) of a study (see Appendix 3). Although external validity is not relevant for assessing the risk of bias, we assessed and reported external validity in this review given that it was included in the modified GATE tool.

We rated the individual criteria within sections 1 to 4 as follows (NICE 2012).

- ++ Indicates that for that particular aspect of study design, the study has been designed or conducted in such a way as to minimize the risk of bias.
- + Indicates that either the answer to the checklist question is not clear from the way the study is reported, or that the study may not have addressed all potential sources of bias for that particular aspect of study design.
- - Reserved for those aspects of study design in which significant sources of bias may persist.
- Not reported (NR): Reserved for those study design aspects in which the study under review fails to report how they have (or might have) been considered.
- Not applicable (NA): Reserved for those study design aspects that are not applicable given the study design under review.

A fifth section then allows the review authors to give each study an overall rating for both external and internal validity. In section 5 we used the following rating system.

- ++ All or most of the checklist criteria have been fulfilled; where they have not been fulfilled the conclusions are very unlikely to alter.
- + Some of the checklist criteria have been fulfilled; where they have not been fulfilled, or are not adequately described, the conclusions are unlikely to alter.
- - Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

The individual checklist criteria can be found in Appendix 3.

Some studies applied different study design and analysis methods to assess health and air quality outcomes. Where applicable, we therefore conducted two separate assessments for these outcome categories.

After a pilot exercise to calibrate the assessment, two authors (from JB, HB, SP, LP, AR, ER) independently appraised all included main studies. The review authors resolved disagreements through discussion; or asked a third review author to arbitrate when necessary.

Measures of treatment effect

We had initially aimed to convert effects from all main studies into common measures of treatment effect: mean differences (MDs) for continuous outcomes and risk ratios (RRs) for dichotomous outcomes. However the observed effects reported by included main studies were so heterogeneous, due to varying analytical methods and reporting practices, that this undertaking was deemed infeasible. Thus we extracted any measure of intervention effectiveness reported in the included main studies which reported an association between included interventions and outcomes.

Where multiple relevant analyses were conducted in a study, review authors discussed and agreed upon which were most relevant for the review. For example, where unadjusted and adjusted estimates were provided, we considered the adjusted estimates more appropriate. Where multiple studies assessed the same outcome for a given intervention, we included the effect estimate from the study with the lowest risk of bias in the evidence synthesis and in the summary of findings. Where the same risk of bias rating was given to multiple studies assessing the same intervention, we chose the effect estimate from the study with the most recent follow-up.

Dealing with missing data

In the case that missing information on study features (e.g. number of time points, selection of intervention and control sites), intervention characteristics (e.g. timing or duration) or outcome data (e.g. missing values, variance measure) prevented or limited use of a study, we contacted the investigators via email for more information. Where authors were initially non-responsive, we contacted them a second time.

Assessment of heterogeneity

At the protocol stage we had planned to assess statistical heterogeneity graphically, using a forest plot; and statistically, using I^2 statistic calculations. Given the heterogeneity of the identified evidence base, and the narrative nature of our evidence synthesis (see below), such an assessment was not feasible. Instead, and as laid out in our protocol, we carefully documented and described methodological and population, intervention, comparator and outcome (PICO)-related heterogeneity for both main and supporting studies through the narrative synthesis and the creation of tables.

Assessment of reporting biases

At the protocol stage, we had planned to examine funnel plot asymmetry to investigate the risk of publication bias by intervention type and outcome measure. Given the heterogeneity of the identified evidence base, and the narrative nature of our evidence synthesis (see below), such an assessment was not feasible. For all included studies, we checked whether a study protocol or analysis plan was cited; where a protocol or analysis plan was available we checked whether all described outcomes were also assessed in the published study.

Data synthesis

We described the characteristics and methods of all included studies, including main and supporting studies, by creating summary tables.

For reasons described above, we only considered main studies in the evidence synthesis regarding intervention effectiveness. For each intervention category (interventions targeting vehicular, industrial, residential and multiple sources), where two or more studies reported on the same primary outcome and for which sufficient methodological and PICO-related homogeneity existed, we had planned to conduct a random effects meta-analysis.

As the evidence proved too heterogeneous to conduct meta-analyses, in line with the review protocol we synthesized evidence narratively as well as graphically using harvest plots. Harvest plots have been shown to be an effective, clear and transparent way to summarize evidence of effectiveness for complex interventions (Ogilvie 2008; Turley 2013). We created eight separate harvest plots, one for health outcomes and one for air quality outcomes for each intervention category. We arranged studies, represented by bars, in rows according to outcomes, and columns according to the direction of effect: effect favours control; unclear effect due to lack of statistical significance; effect favours intervention. Please note that this distinction relies on statistical significance but acknowledges that 'unclear effects' may include effects favouring the intervention or favouring the control, as well as true null effects. In the narrative synthesis we refer to this mixed category as either “no change” or “no significant effect in either direction”. The risk of bias of the study is illustrated by the height of the bar, with the height of the bar corresponding to the rating from the GATE tool (++, +, -).

We made the post hoc decision to also include information on the nature of the statistical comparison through the colour of the bar. Black bars represent studies with standard comparisons based on a statistical comparison of intervention and control sites before and after the intervention. White bars represent studies for which the effectiveness was determined in parallel analyses for intervention and control sites before and after the intervention. Specifically, these studies conducted two parallel and separate before-after statistical analyses for intervention and control sites, without comparing these sites directly. Effects from these studies were interpreted and portrayed in the harvest plots so that if a statistically significant improvement in the outcome was observed at intervention sites, while no change was observed at control sites, this was classified as an “effect favouring the intervention”; and if significant improvements were seen both at intervention and control sites, this was classified as “no change”, etc. We created harvest plots in Microsoft Excel.

Subgroup analysis and investigation of heterogeneity

In order to assess the impact of potentially important sources of heterogeneity, we performed a subgroup analysis focusing on the temporal aim of the intervention, i.e. whether the intervention aimed to temporarily or permanently affect air quality. To accomplish this, we stratified the evidence into temporary and permanent interventions, and assessed the effectiveness of each narratively, as well as using harvest plots.

Other subgroup analyses were planned — based on, for example population characteristics, intervention goal, delivery characteristics and inequality characteristics — but these were not conducted. For many of these aspects, suitable data were not reported in included studies; additionally, we felt that further fragmenting the very heterogeneous evidence base was not appropriate.

Sensitivity analysis

As NRS designs were important for this review, we had originally planned to conduct a sensitivity analysis assessing whether the effectiveness evidence from randomized study designs (RCT, cRCT), EPOC-recognised NRS designs (cITS-EPOC, ITS-EPOC, CBA-EPOC) and non-EPOC NRS designs (CBA, UBA, ITS) differed. Given the absence of randomized evidence and the incorporation of very few main studies from the non-EPOC study designs category in the evidence synthesis, we did not conduct this sensitivity analysis.

Certainty of evidence

In order to assess the certainty of the body of evidence used in the data syntheses for primary outcomes, we applied the GRADE system for grading evidence (Guyatt 2008). GRADE allows for the systematic and transparent grading of the certainty of the body of evidence for each outcome based on the following factors.

- Factors decreasing certainty of evidence
  - Limitations in study design or execution (risk of bias)
  - Inconsistency of results
  - Indirectness of evidence
  - Imprecision
  - Publication bias

- Factors increasing certainty of evidence
  * Large magnitude of effect
  * Plausible confounding, which would reduce a demonstrated effect
  * Dose-response gradient.

Based on these criteria, we graded each the evidence base for each intervention category and primary outcome as one of the following.

- High certainty – we are very confident that the true effect lies close to that of the estimate of the effect.
- Moderate certainty – we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
- Low certainty – our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
- Very low certainty – we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

According to the recommendation from the GRADE working group, all non-randomized studies started the GRADE assessment rated as 'low certainty'. We created a ‘Summary of findings’ table for each of the four intervention categories to summarize our evidence synthesis and the results of the GRADE assessment. The initial GRADE assessment was undertaken by one review author (JB), and was then discussed in detail and finalized with a second review author (ER).

Review Advisory Group

A draft protocol draft was sent to a Review Advisory Group (RAG). The RAG comprised air pollution and health experts as well as potential end users of the review from a wide range of countries and contexts, who all provided feedback to ensure the review will meet its intended goal of assessing the effectiveness of ambient PM interventions in a systematic and comprehensive way and that the review will appropriately inform policy.

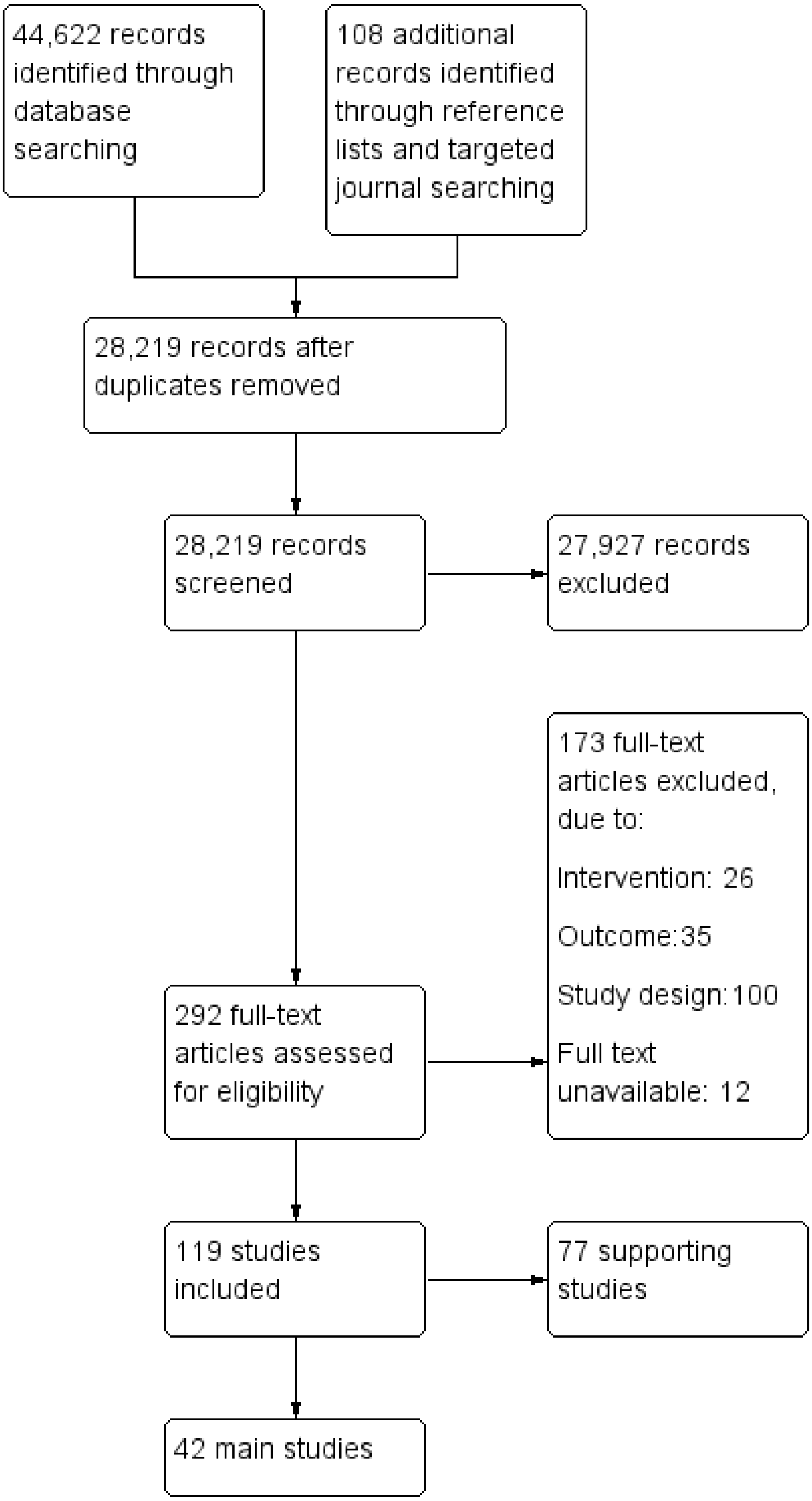
RESULTS

Description of studies

Results of the search

The results of the selection of studies are shown in Figure 3. From a total of 28,219 unique records, 292 full texts were deemed potentially relevant, and 119 met the a priori eligibility criteria and were included in the review. Reasons for exclusion at the full-text screening stage are documented in Figure 3 and in the Characteristics of excluded studies; most studies (n = 100; 58%) were excluded due to the study design.

Figure 3. Study flow diagram.



Of the 119 included studies, 42 were included as main studies, and 77 as supporting studies. The characteristics of the 42 main studies are described in detail in the Characteristics of included studies table and in the following text, while the characteristics of the 77 supporting studies are described in Appendix 4 and Appendix 5.

Of the 42 main studies, 23 were identified during the first round of searching, 9 during the second round of searching, and 10 during handsearching. One study was published in German and one study in Italian, while all others were published in English. These 42 included studies evaluated 38 unique interventions.

Given that some unique interventions were evaluated by multiple studies, which could not be considered individual parts of a single evaluation, and that some studies evaluated multiple distinct interventions, we describe the evaluated 'interventions' rather than individual 'studies' in the following detailed description of the evidence base.

The main studies are described in the following sections according to the setting, population, intervention and comparison, outcomes, study design and risk of bias. This descriptive section is followed by a section presenting the effects of these interventions using harvest plots and narrative synthesis.

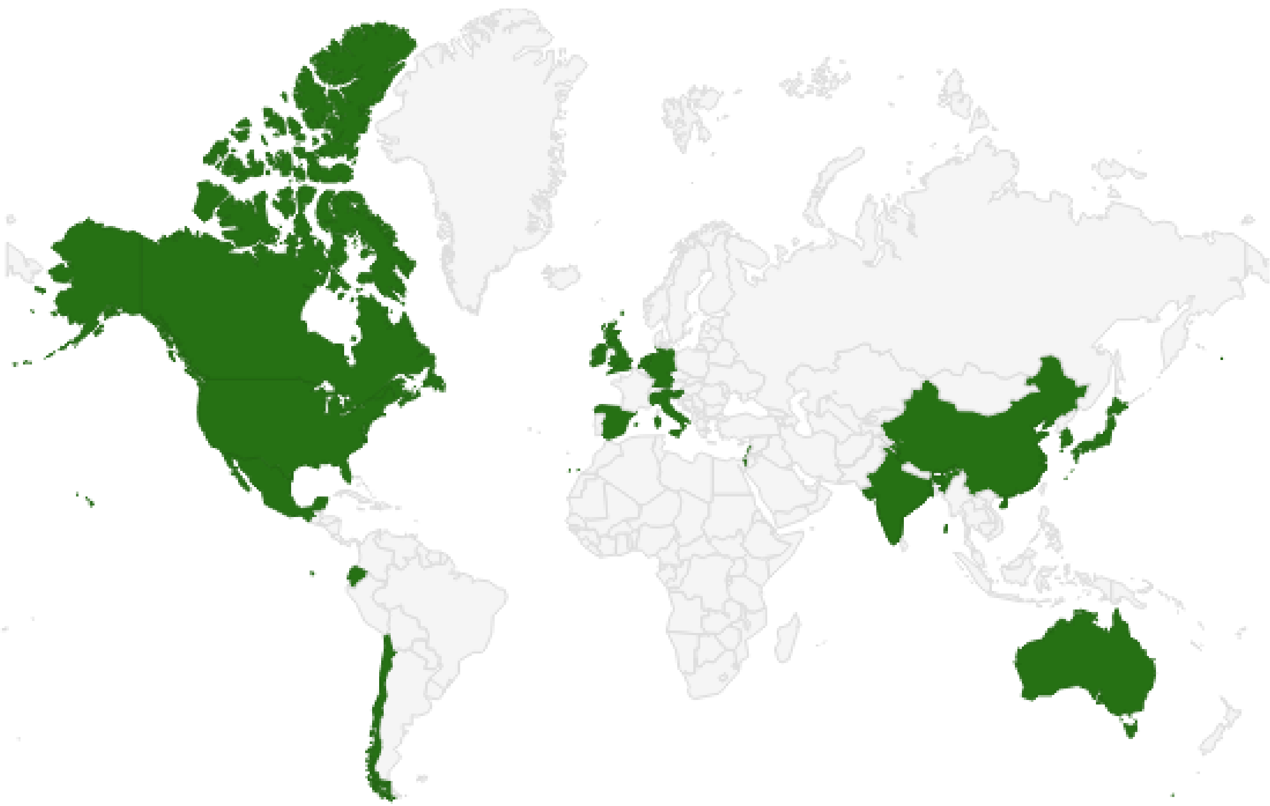
Included studies

The characteristics of each of the 42 main studies are summarized below and described in detail in Table 1 and in the Characteristics of included studies table.

Setting

Included main studies assessed interventions from 19 different countries (Figure 4). Although there was a wide geographical distribution of included studies, using the Global Burden of Disease (GBD) super-region classification (Gakidou 2017), most of the assessed interventions were from HICs (n = 30) (Allen 2009; Atkinson 2009; Bel 2013a; Bel 2013b; Boogaard 2012; Burr 2004; Cowie 2012; Deschênes 2012; Dijkema 2008; Dockery 2013a; Dockery 2013b; Dockery 2013c; Dolislager 1997; Fensterer 2014; Gallego 2013b; Giovanis 2015; Hasunuma 2014; Johnston 2013; Kim 2011; Morfeld 2014; Mullins 2014; Peel 2010; Pope 2007; Ruprecht 2009; Saaroni 2010; Sajjadi 2012; Titos 2015b; Yap 2015; Yorifuji 2016; Zigler 2016). Interventions in LMICs were also included, but most of the non-HIC super-regions were poorly represented; three interventions were assessed in the Southeast Asia, East Asia and Oceania region (Li 2011; Tanaka 2015; Viard 2015); two interventions in the Latin America and the Caribbean region (Carrillo 2016; Davis 2008); one intervention in Central Europe, Eastern Europe and Central Asia (Titos 2015a); one intervention in the North Africa and Middle East region (El-Zein 2007); and one intervention in the South Asia region (Aung 2016). Notably, we did not identify any interventions in the sub-Saharan Africa region.

Figure 4. Geographic location of the 38 interventions evaluated in the main studies.



Most interventions (n = 29) evaluated in the main studies were implemented in an urban or community setting (Atkinson 2009; Bel 2013a; Bel 2013b; Boogaard 2012; Burr 2004; Carrillo 2016; Cowie 2012; Davis 2008; Dijkema 2008; Dockery 2013a; Dockery

2013b; Dockery 2013c; Dolislager 1997; El-Zein 2007; Fensterer 2014; Gallego 2013b; Johnston 2013; Kim 2011; Li 2011; Morfeld 2014; Mullins 2014; Peel 2010; Ruprecht 2009; Saaroni 2010; Tanaka 2015; Titos 2015a; Titos 2015b; Viard 2015; Yorifuji 2016). Two studies examined interventions in rural settings (Allen 2009; Aung 2016); and a further seven examined interventions in mixed urban/rural settings (Deschênes 2012; Giovanis 2015; Hasunuma 2014; Pope 2007; Sajjadi 2012; Yap 2015; Zigler 2016).

Population

This review comprises both studies that measure air quality only and studies that measure health, either alone or in combination with air quality. In studies assessing air quality only, most used routinely monitored data collected for regulatory purposes, although some collected data from study-specific pollutant monitors. In studies assessing only health or health and air quality combined, the population of interest tended to be the general population. Due to the ecological nature as well as the use of routine data of the included studies, exact demographic characteristics were often not provided. Selected studies, however, did assess specific subsets of the population.

Main studies assessing a subset of the population assessed children under the age of 1 year (Tanaka 2015), under the age of 3 years (Hasunuma 2014), under the age of 14 years (Sajjadi 2011), and under the age of 17 years (El-Zein 2007). One study specifically assessed individuals over the age of 65 years (Sajjadi 2011).

Interventions and comparisons

Among the 38 unique interventions included in the main studies, five aimed to reduce ambient air pollution from industrial sources, seven from residential sources, 22 from vehicular sources, and four from multiple sources. Each of these broad intervention categories, however, consists of a wide range of intervention types. Thus in an attempt to provide a more meaningful and precise categorization, we further classified interventions post hoc into intervention subcategories, such as “cap and trade program”, “temporary infrastructure changes”, “low emission zone” and “wood burning ban”.

In all studies, the comparison against which the intervention was compared can be considered no intervention or practice as usual.

A description of each of the interventions from the main studies is included in the following table.

Description of the interventions evaluated in the included main studies

Study ID	Intervention subcategory	Intervention description	Level of implementation	Introduction and duration of intervention
Industrial sources				
Butler 2011/ Deschênes 2012/ Lin 2013	Cap and trade programme	Cap and trade programme regulating large combustion sources (EGUs, industrial boilers, etc.). NOx emissions are monitored by and reported to the EPA. To meet the cap sources may utilized control technologies, switch fuels or buy and sell allowances at a free market price.	Region	2003 to 2008 (ozone season only)
Pope 2007	Factory closure	National copper smelter strike that was especially relevant in the Southwest US where much copper smelting took place	Region	15 July 1967 to April 1968
Saaroni 2010	Power plant conversion	Converting the Tel Aviv power station from oil to gas	Factory	2005 – permanent (specific timing unclear)
Sajjadi 2011/ Sajjadi 2012	Factory closure	Closure of the local steel works industry	Factory	October 1999 – permanent
Tanaka 2015	Required industry requirements	Two Control Zone policy which designated areas exceeding acid rain or SO2 thresholds as TCZ status. These areas were then subject to more stringent regulations with regard to coal mining and burning.	Country	January 1998 – permanent
Residential sources				

Allen 2009	Stove exchange	Stove exchanges, along with financial incentives for purchasing new stoves	Community	2012 – permanent (specific timing unclear)
Aung 2016	Stove exchange	Removal of traditional stoves from intervention homes, installation of new stoves, assistance with stove operation and maintenance	Community	2007 or 2008 – permanent (specific timing unclear)
Dockery 2013a/ Clancy 2002	Coal ban	Ban on marketing, sale and distribution of coal used for heating	City	1990 – permanent
Dockery 2013b	Coal ban	Ban on marketing, sale and distribution of coal used for heating	City	1995 – permanent
Dockery 2013c	Coal ban	Ban on marketing, sale and distribution of coal used for heating	City	1998 – permanent
Johnston 2013	Stove exchange	Wood Heater Replacement Program; education campaign; monitoring	City	July 2001 to June 2004
Yap 2015	Wood burning ban	Mandatory ban on residential wood burning when poor air quality was forecast, and strict regulations regarding fireplaces and wood stoves when a home is to be sold.	Region	November 2003 – permanent
Vehicular sources				
Atkinson 2009	Charging scheme	Congestion charging scheme applied to four-wheeled vehicles entering the charging zone on workdays	City centre	February 2003 – permanent
Bel 2013a	Speed limit change	80 km/h speed limit on motorways;	City	1 January 2008 to 31 December 2010 (80 km/h speed limit)
Bel 2013b	Speed limit change	Variable speed limit (minimum 40, maximum 80 km/h) based on traffic density and specific conditions, such as accidents, construction, air pollution, poor weather.	City	1 January 2009 to 31 December 2010 (variable speed limit)
Boogaard 2012	Low emission zone	Low emission zones limiting the types of trucks allowed to enter the city centres of the assessed cities. Limits became more stringent over time.	City centre	July 2007 – permanent
Burr 2004	Infrastructure changes	Opening of bypass around an area subject to heavy traffic congestion	Street	1997 or 1998 – permanent (specific timing unclear)
Carrillo 2016	Even-odd restriction	Restriction of the city centre during weekday peak traffic hours based on the last digit of a vehicle's license plate number. Establishment of free parking areas on the periphery of the restriction zone, allowing drivers to utilize public transportation	City centre	3 May 2010 – permanent (subject to annual reassessment)
Cowie 2012	Tunnel construction; Road restructuring	3.6 km tunnel linking two major roadways, along with concomitant road changes to a nearby main road to reduce traffic, including lane number reduction and a dedicated bus lane	Community	25 March 2007 – permanent (tunnel opening);

				March 2008 – permanent (road changes)
Davis 2008/ Gallego 2013a	Even-odd restriction	Even-odd driving ban: Banning of drivers from using their vehicles one day per week based on the last digit of the license plate	City	20 November 1989 – permanent
Dijkema 2008	Speed limit change	Speed limit reduction on urban traffic ring	Street	November 2009 – permanent
Dolislager 1997	Fuel requirements	Requiring gasoline sold during months prone to high CO concentrations to have a low oxygen content	Regional	November 1991 – permanent (winter only)
El-Zein 2007	Vehicle ban	Ban on the import of all light – and medium duty diesel engines	Country	June 2002 – permanent
Gallego 2013b/ Gramsch 2013	Public transport restructuring	Restructuring of the entire public transport system, including changes to the subway system and bus network	City	10 February 2007 – permanent
Hasunuma 2014	Required vehicle standards	Ban on automobiles not conforming to the Automobile NOx/PM Law, in areas designated enforcement areas	Country	June 2001 – permanent
Kim 2011	Clean fuel use	Natural Gas Vehicle Supply program led to the replacement of the entire fleet of diesel-powered city buses with natural gas buses in large cities	Country	1 June 2000 – permanent
Morfeld 2013/ Fensterer 2014	Low emission zone	Low emission zone in line with EURO regulations, becoming gradually more stringent	City centre	October 2008 – permanent
Morfeld 2014	Low emission zone	Low emission zone, restricting entrance of diesel cars below Euro II and gasoline cars Euro I standards	City centre	Approximately 2008 – permanent (start date differs for individual cities)
Peel 2010/ Friedman 2001	Comprehensive traffic reduction strategy	Various traffic-reduction strategies including increased availability of public transportation, comprehensive traveller information and updates, encouraging businesses to provide telecommuting and alternative work hours for employees	City centre	19 July 1996 to 4 August 1996
Ruprecht 2009	Charging scheme	Ecopass congestion charging scheme, requiring payment during the week for entering the city centre	City centre	8 January 2008 – permanent
Titos 2015a	Road restructuring	Partial closure and reconstruction of 400 m of a major street. Only public buses and taxis were allowed after implementation	Street	22 September 2013 – permanent
Titos 2015b	Public transport	Redesign of the bus transportation system, including the reduction in overlap between bus lines, and new buses with higher passenger capacities and meeting EURO V requirements	City	29 June 2014 – permanent

	restructuring			
Viard 2015	Even-odd restriction	Even-odd driving restriction policy, restricting cars to drive only every-other-day, applying seven days a week from 3 a.m. to 12 a.m.; This was then relaxed to a policy restricting cars to drive one day per week	City	20 July 2008 to 20 September 2008 11 October 2008 – permanent
Yorifuji 2016/ Yorifuji 2011	Required vehicle standards	Standards for diesel vehicles, which represented stricter controls than the nationally mandated standards. Diesel vehicles not meeting the standards were required to be replaced or be retrofitted to reduce emissions; These standards were then further tightened in some regions.	Region	October 2003 – permanent; April 2006 – permanent
Multiple sources				
Giovanis 2015	Repeated coordinated measures	Coordinated measures for reducing pollution on days where high levels of pollution were expected. These include postponing high-emitting activities, changes in business operations, alternative scheduling, public education, and the promotion of alternative modes of transportation	Region	March 2006 – permanent (intermittent operation: implemented on days where especially high levels are expected, then relaxed when levels drop)
Li 2011	Even-odd restriction; Vehicle restriction; Power plant restriction	Alternative transportation strategy banning trucks not meeting emission standards, even-odd ban on private vehicles every other day, and strict restrictions on polluting industries in Beijing and the surrounding provinces	City	1 July 2008 to 7 August 2008
Mullins 2014	Repeated coordinated measures	Identification of high pollution days, which triggered mandatory restrictions on driving, the shutdown of certain major stationary emitters, street sweeping, traffic enforcement activities, restriction on the use of biomass combustion for residential heating	City	1997 – permanent (Intermittent operation: implemented on specific high pollution days)
Zigler 2016	Tailored selection of measures	As part of the US Clean Air Act, areas in the Western United States were classified as either attainment or non-attainment of the 1987 National Ambient Air Quality Standards for PM_{10}. Non-attainment areas were required to develop a strategy for further reducing PM_{10} below the standard	Region	1990 – permanent

Interventions targeting industrial sources

Among the main studies of interventions aiming to reduce ambient air pollution from industrial sources, we included the US NOx Budget Trading Program, a nationally coordinated and monitored cap and trade programme (Butler 2011; Deschênes 2012; Lin 2013); the Chinese Two Control Zone policy, a set of nationally coordinated and monitored compulsory industrial standards (Tanaka 2015); a power plant conversion from oil to gas in Tel Aviv, Israel (Saaroni 2010); as well as two natural experiments, including a temporary short-term copper smelter strike in the Southwest US (Pope 2007), and a permanent steel works closure in New South Wales, Australia (Sajjadi 2012).

Interventions targeting residential sources

Among the main studies of interventions aiming to reduce ambient air pollution from residential sources, we included a ban on the marketing, sale and distribution of coal for heating purposes, implemented originally in Dublin, Ireland (Clancy 2002; Dockery 2013a) and subsequently expanded to several other Irish cities (Dockery 2013b; Dockery 2013c); wood stove exchange programmes in British Columbia, Canada (Allen 2009), in rural southern India (Aung 2016) and in Tasmania, Australia (Johnston 2013); and an air-quality-dependent wood burning ban in California, USA (Yap 2015).

Interventions targeting vehicular sources

Among the main studies of interventions aiming to reduce ambient air pollution from vehicular sources, we identified compulsory standards for fuel composition in California, USA (Dolislager 1997); and for vehicles in Tokyo (Yorifuji 2016) and several other urban areas in Japan (Hasunuma 2014). We included schemes that restrict the frequency with which individuals can use vehicles (e.g. by limiting use on certain days to those with an even or odd number plate, from here on referred to as 'even-odd ban') in several cities across the world, including Quito (Ecuador), Mexico City (Mexico), and Beijing (PRC) (Carrillo 2016; Davis 2008, Gallego 2013a; Viard 2015). The Natural Gas Vehicle Supply (NGVS) programme led to the replacement of the diesel-powered bus fleet with natural gas buses in urban areas of South Korea (Kim 2011). One intervention consisted of a comprehensive traffic reduction strategy during the 1996 Olympic Games in Atlanta (Friedman 2001; Peel 2010). Other interventions comprised permanent infrastructure changes, including the construction of a bypass around a heavily congested area in Northern Wales (UK) (Burr 2004); the construction of a tunnel for congestion relief in Sydney (Australia) (Cowie 2012); the restructuring of the public transportation systems in Santiago (Chile) (Gallego 2013b; Gramsch 2013), and Granada (Spain) (Titos 2015b); and the redesign of a major street allowing access only to public buses and taxis in Ljubljana (Slovenia) (Titos 2015a). We identified low emission zones across the Netherlands and Germany (Boogaard 2012; Fensterer 2014; Morfeld 2014). Other interventions included a reduction of the speed limit in Barcelona (Spain) and Amsterdam (the Netherlands) (Bel 2013a; Dijkema 2008), as well as an adaptive speed limit system in Barcelona (Spain) (Bel 2013b). One study assessed a nationwide ban on diesel vehicles in Beirut (Lebanon) (El-Zein 2007); and two studies assessed vehicle charging schemes in London (UK) (Atkinson 2009), and in Milan (Italy) (Ruprecht 2009).

Interventions targeting multiple sources

Among the main studies of interventions aiming to reduce ambient air pollution from multiple sources, we included broad, nationwide policies such as the US National Ambient Air Quality Standards attainment status designation, part of the US Clean Air Act amendments of 1990 (Zigler 2016), combined measures to reduce vehicular traffic and industrial pollution during the Beijing Olympic Games of 2008 (Li 2011), and repeated, tailored measures at the city level on high-pollution days in Charlotte (North Carolina in the USA) (Giovanis 2015) and in Santiago (Chile) (Mullins 2014).

Level of implementation of interventions

The level of intervention implementation varied substantially across included main studies, from national level (El-Zein 2007; Hasunuma 2014; Kim 2011; Tanaka 2015), to regional level (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; Dolislager 1997; Pope 2007; Sajjadi 2012; Yap 2015; Zigler 2016), city/community level (Allen 2009; Atkinson 2009; Aung 2016; Bel 2013a; Bel 2013b; Boogaard 2012; Carrillo 2016; Cowie 2012; Davis 2008; Gallego 2013b; Giovanis 2015; Johnston 2013; Li 2011; Morfeld 2013; Morfeld 2014; Mullins 2014; Peel 2010; Ruprecht 2009; Saaroni 2010; Titos 2015b; Viard 2015; Yorifuji 2016), and street level (Burr 2004; Dijkema 2008; Titos 2015a).

Timing and duration of interventions

The timing and duration of the interventions is another important aspect to consider, as some measures, e.g. the construction of a tunnel (Cowie 2012) or a permanent even-odd vehicle ban (Davis 2008), aimed to permanently improve air quality, while more temporary measures, e.g. traffic reduction strategies during the 1996 Atlanta Olympic Games (Peel 2010) or measures to reduce vehicle traffic and industrial pollution during the 2008 Beijing Olympic Games (Li 2011), had a much more time-limited impact on air quality and health. Other interventions also had an intermittent effect, as they were only active during certain times, for example when pollution levels were predicted to be above a certain threshold (Mullins 2014). Another important aspect of timing involves seasonal implementation: most interventions remained in place regardless of season, while others were implemented or only expected to impact air quality during the higher pollution winter season. Such examples include California's winter-time oxygenated fuels programme (Dolislager 1997); and those targeting heating practices (Allen 2009; Dockery 2013a; Dockery 2013b; Dockery 2013c; Johnston 2013; Yap 2015).

Outcomes

Health outcomes

Of the 38 unique interventions, only 18 were evaluated with respect to their effect on health outcomes (Table 1). With regard to the primary health outcomes of the review, the effects of 10 interventions were assessed in relation to all-cause mortality (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; Giovanis 2015; Johnston 2013; Pope 2007; Tanaka 2015; Yorifuji 2016; Zigler 2016); of six interventions in relation to cardiovascular mortality (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; Johnston 2013; Yorifuji 2016); and of six interventions in relation to respiratory mortality (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; Johnston 2013; Yorifuji 2016).

The effects of a further 12 interventions were evaluated in relation to secondary health outcomes of the review, i.e. cardiovascular hospitalizations, respiratory hospitalizations, or both for 10 interventions (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; El-Zein 2007; Li 2011; Peel 2010; Sajjadi 2012; Yap 2015; Zigler 2016), and lung function and/or measures of respiratory symptoms for two interventions (Burr 2004; Hasunuma 2014).

Air quality outcomes

Of the 38 unique interventions, 27 were assessed with respect to their effect on air quality outcomes (Table 1). With regard to the primary AQ outcomes of the review, the effects of 16 interventions were assessed with respect to PM_{10} (Atkinson 2009; Bel 2013a; Bel 2013b; Boogaard 2012; Burr 2004; Cowie 2012; Dijkema 2008; Fensterer 2014; Kim 2011; Li 2011; Mullins 2014; Ruprecht 2009; Saaroni 2010; Sajjadi 2012; Viard 2015; Zigler 2016), 9 interventions with respect to $PM_{2.5}$ (Allen 2009; Aung 2016; Boogaard 2012; Burr 2004; Cowie 2012; Li 2011; Sajjadi 2012; Yap 2015; Yorifuji 2016), 1 intervention with respect to coarse PM (Yap 2015), and 6 interventions with respect to combustion-related PM (Aung 2016; Boogaard 2012; Dijkema 2008; Gallego 2013b; Titos 2015a; Titos 2015b).

The effects of a further 21 interventions were evaluated in relation to secondary outcomes of the review, including 14 interventions with respect to NO, NO_2 and/or NOx (Atkinson 2009; Bel 2013a; Bel 2013b; Boogaard 2012; Cowie 2012; Davis 2008; Dijkema 2008; Hasunuma 2014; Kim 2011; Morfeld 2014; Peel 2010; Saaroni 2010; Sajjadi 2012; Yorifuji 2016), 4 with respect to SO_2 (Saaroni 2010, Sajjadi 2012, Davis 2008, Peel 2010), 5 with respect to O_3 (Davis 2008; Deschênes 2012; Giovanis 2015; Li 2011; Peel 2010), and 5 with respect to CO (Carrillo 2016; Davis 2008; Dolislager 1997; Gallego 2013b; Peel 2010). No main studies assessed effectiveness of interventions with respect to UFP concentrations.

Unintended outcomes

No identified studies assessed unintended or adverse effects.

Study designs

It should be noted that many included studies did not define or report an exact study design, meaning that a study design label was assigned by review authors. Additionally, in several included studies there was a stark discrepancy between the data collection and the analysis, also rendering the definition of study design more complicated. Two review authors extensively discussed study design classification both at the full-text screening and the data extraction stage, and discussed any unclear cases with other members of the review team. We included cITS-EPOC, ITS-EPOC, CBA-EPOC, and CBA studies in the evidence synthesis; we identified no RCTs, cRCTs or ITS studies not adhering to EPOC criteria. The study designs are listed in Table 1, and a more in-depth description of the study methodology, including aspects of the design and analysis can be found in Table 2 and Table 3 for studies assessing health and air quality outcomes, respectively. As some studies applied different study designs to assess the health and air quality outcomes, we have described these separately in the following.

Studies assessing health outcomes

Among the main studies, nine studies assessing health outcomes applied a cITS-EPOC study design (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; Johnston 2013; Pope 2007; Sajjadi 2012; Tanaka 2015; Yorifuji 2016), five studies applied an ITS-EPOC design (El-Zein 2007; Li 2011; Mullins 2014; Peel 2010; Yap 2015), two studies applied a CBA-EPOC study design (Hasunuma 2014; Zigler 2016), and one study applied a CBA study design not adhering to the EPOC criteria (Burr 2004).

Studies assessing air quality outcomes

Among the main studies, four studies assessing air quality outcomes applied a cITS-EPOC study design (Bel 2013a; Cowie 2012; Deschênes 2012), ten studies applied an ITS-EPOC study design (Bel 2013b; Butler 2011; Davis 2008; Dolislager 1997; Gallego 2013a; Gallego 2013b; Mullins 2014; Sajjadi 2012; Viard 2015; Yap 2015), eight studies applied a CBA-EPOC study design (Boogaard 2012; Carrillo 2016; Giovanis 2015; Hasunuma 2014; Kim 2011; Morfeld 2014; Peel 2010; Zigler 2016), and 11 applied a CBA study design not adhering to the EPOC criteria (Allen 2009; Aung 2016; Burr 2004; Dijkema 2008; Fensterer 2014; Gramsch 2013; Ruprecht 2009; Saaroni 2010; Titos 2015a; Titos 2015b; Yorifuji 2016).

Excluded studies

We excluded 174 studies at the full-text screening stage, as they did not meet our review inclusion criteria with respect to study design (n = 100), intervention (n = 26), or outcome (n = 35). The full texts of an additional 12 records were not available; four of these were conference presentations with no associated full publication and one appeared to be a non-quantitative report. A further five evaluated interventions evaluated by other included studies, including the Beijing Olympic Games, the switch to natural gas for heating in Urumqi (PRC) and a range of coordinated measures in Taiwan. For a further two studies we simply were unable to identify any further record. A full list of these excluded studies, along with reason for exclusion, can be found in Characteristics of excluded studies.

Risk of bias in included studies

Using the NICE-modified GATE tool, we assessed the risk of bias (i.e. internal validity) and external validity of all included main studies; as specified above, we do not report on the risk of bias or external validity assessment of supporting studies. The overall judgements for internal validity, external validity and our additional criterion addressing causality for included main studies can be found in Figure 5 and Figure 6 for studies assessing health and air quality outcomes, respectively. These judgements consist of one of the following.

Figure 5. Overall judgements for risk of bias, external validity and our additional criterion addressing causality for included main studies assessing health outcomes. Symbols should be interpreted as follows: (++) All or most of the checklist criteria have been fulfilled; where they have not been fulfilled the conclusions are very unlikely to alter; (+) Some of the checklist criteria have been fulfilled; where they have not been fulfilled, or are not adequately described, the conclusions are unlikely to alter; (-) Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter

	Internal validity	External validity
Industrial interventions		
Deschenes 2012	(++)	(++)
Lin 2013	(+)	(++)
Pope 2007	(++)	(++)
Sajjadi 2011	(-)	(+)
Tanaka 2015	(++)	(++)
Residential interventions		
Dockery 2013a	(++)	(++)
Dockery 2013b	(++)	(++)
Dockery 2013c	(++)	(++)
Johnston 2013	(+)	(++)
Yap 2015	(-)	(+)
Vehicular interventions		
Burr 2004	(-)	(+)
El-Zein 2007	(-)	(+)
Hasunuma 2014	(+)	(++)
Peel 2010	(++)	(++)
Yorifuji 2016	(++)	(++)
Multiple interventions		
Giovanis 2015	(+)	(+)
Li 2011	(+)	(++)
Mullins 2014	(++)	(++)
Zigler 2016	(++)	(++)

Figure 6. Overall judgements for risk of bias, external validity and our additional criterion addressing causality for included main studies assessing AQ outcomes. Symbols should be interpreted as follows: (++) All or most of the checklist criteria have been fulfilled; where they have not been fulfilled the conclusions are very unlikely to alter; (+) Some of the checklist criteria have been fulfilled; where they have not been fulfilled, or are not adequately described, the conclusions are unlikely to alter; (-) Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter

	Internal validity	External validity
Industrial interventions		
Butler 2011	(+)	(++)
Deschenes 2012	(++)	(++)
Lin 2013	(+)	(++)
Saaroni 2010	(-)	(+)
Sajjadi 2012	(-)	(+)
Residential interventions		
Allen 2009	(-)	(++)
Aung 2016	(-)	(+)
Yap 2015	(+)	(+)
Vehicular interventions		
Atkinson 2009	(+)	(+)
Bel 2013a	(+)	(+)
Bel 2013b	(+)	(+)
Boogaard 2012	(++)	(+)
Burr 2004	(-)	(+)
Carillo 2013	(++)	(++)
Cowie 2012	(++)	(++)
Davis 2008	(+)	(++)
Gallego 2013a	(+)	(++)
Dijkema 2008	(++)	(++)
Dollslager 1997	(-)	(++)
Gallego 2013b	(+)	(++)
Gramsch 2013	(+)	(++)
Hasunuma 2014	(+)	(++)
Kim 2011	(+)	(+)
Morfeld 2013	(+)	(+)
Fensterer 2014	(++)	(++)
Morfeld 2014	(++)	(++)
Peel 2010	(+)	(++)
Ruprecht 2009	(-)	(+)
Titos 2015a	(+)	(++)
Titos 2015b	(+)	(++)
Viard 2015	(++)	(++)
Yorifuji 2016	(-)	(++)
Multiple interventions		
Giovanis 2015	(+)	(+)
Mullins 2014	(++)	(+)
Zigler 2016	(++)	(++)

- ++ All or most of the checklist criteria have been fulfilled; where they have not been fulfilled the conclusions are very unlikely to alter.
- + Some of the checklist criteria have been fulfilled; where they have not been fulfilled, or are not adequately described, the conclusions are unlikely to alter.

- \- Few or no checklist criteria have been fulfilled and the conclusions are likely or very likely to alter.

Judgements for the individual criteria for each included main study are summarized in Appendix 6 and Appendix 7, and described in detail in Appendix 8 for studies assessing health and air quality outcomes, respectively.

Studies assessing health outcomes

The judgements regarding the internal validity of main studies assessing health outcomes were mixed. We appraised 11 studies (58%) as (++), four studies (21%) as (+), and four studies (21%) as (-). The judgements across the individual studies varied widely (Appendix 6). Several studies inappropriately selected and justified the selection of covariates (criterion 2.2), which likely introduced bias into study results (Deschênes 2012; Dockery 2013a; Dockery 2013b; Dockery 2013c; El-Zein 2007; Sajjadi 2011; Yap 2015; Yorifuji 2016). The analysis methods (criteria 4.1 to 4.4) of several studies, especially those assessing vehicular interventions, likely also introduced bias into individual study results where, for example, models were not adjusted or poorly adjusted, analyses were under-powered, or effect estimates or measures of precision (or both) were reported insufficiently (Burr 2004; El-Zein 2007; Hasunuma 2014; Johnston 2013; Sajjadi 2011; Yap 2015).

The external validity of these studies was high overall. We rated 14 studies (74%) as (++) and five studies (26%) as (+), meaning that in most cases, the selected and analyzed populations represented the eligible and source populations well. We did not rate the external validity of any studies as (-).

Studies assessing air quality outcomes

With respect to the internal validity of studies assessing air quality outcomes, we judged 10 studies (29%) as (++), 17 studies (49%) as (+), and eight studies (23%) as (-), indicating high variability (Appendix 7). Several studies likely introduced bias through the selection of intervention and control sites (criterion 2.1) (Aung 2016; Bel 2013a; Bel 2013b; Kim 2011; Quiros 2013; Saaroni 2010). Similar to the studies assessing health outcomes, the selection of and justification for explanatory variables (criterion 2.2) was poorly described and likely biased the results of several included studies (Aung 2016; Cowie 2012; Davis 2008; Deschênes 2012; Gallego 2013a; Gallego 2013b; Gramsch 2013; Ruprecht 2009; Sajjadi 2012; Saaroni 2010; Yorifuji 2016). Many studies, especially those assessing vehicular interventions, did not report the completeness of outcome data, or were missing a meaningful proportion of outcome data (criterion 3.2) (Aung 2016; Bel 2013a; Bel 2013b; Burr 2004; Cowie 2012; Kim 2011; Ruprecht 2009; Sajjadi 2012). There were concerns with the analysis methods (criteria 4.1 to 4.4) of several studies, with regard to the choice of statistical test, model selection, model adjustment, study power, and the overall poor reporting of effect estimates and precision (Allen 2009; Aung 2016; Bel 2013a; Bel 2013b; Burr 2004; Gramsch 2013; Hasunuma 2014; Kim 2011; Ruprecht 2009; Saaroni 2010; Titos 2015a; Titos 2015b; Yorifuji 2016).

We rated the external validity of 21 studies (60%) as (++), 14 studies (40%) as (+), and no studies as (-). Thus a lack of representativeness of selected and analyzed intervention and control areas with respect to the eligible and source populations was of no significant concern.

Effects of interventions

See: **Summary of findings for the main comparison** Interventions targeting vehicular sources compared to practice as usual for improving health and air quality; **Summary of findings 2** Interventions targeting industrial sources compared to practice as usual for improving health and air quality; **Summary of findings 3** Interventions targeting residential sources compared to practice as usual for improving health and air quality; **Summary of findings 4** Interventions targeting multiple sources compared to practice as usual for improving health and air quality

We summarized the observed associations between included interventions and outcomes compared to practice as usual using harvest plots. In the following, we provide a more detailed narrative summary of the observed associations between each of the four intervention categories and health and air quality outcomes based on main studies (corresponding to the evidence synthesized in the harvest plots). Appendix 9 provides details on the measured data and associations reported in the individual studies that correspond to the data portrayed in the harvest plots and described below.

Industrial interventions versus practice as usual

As illustrated in Figure 7 and Figure 8, observed associations between interventions to reduce ambient air pollution from industrial sources and both health and air quality outcomes were mixed, with the majority of studies observing either no clear association in either direction or a significant association in favour of the intervention. Summary of findings 2 outlines details regarding the effectiveness of interventions for each primary outcome, as well as a description of the certainty of evidence drawn from our application of GRADE.

Figure 7. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from industrial sources on health outcomes

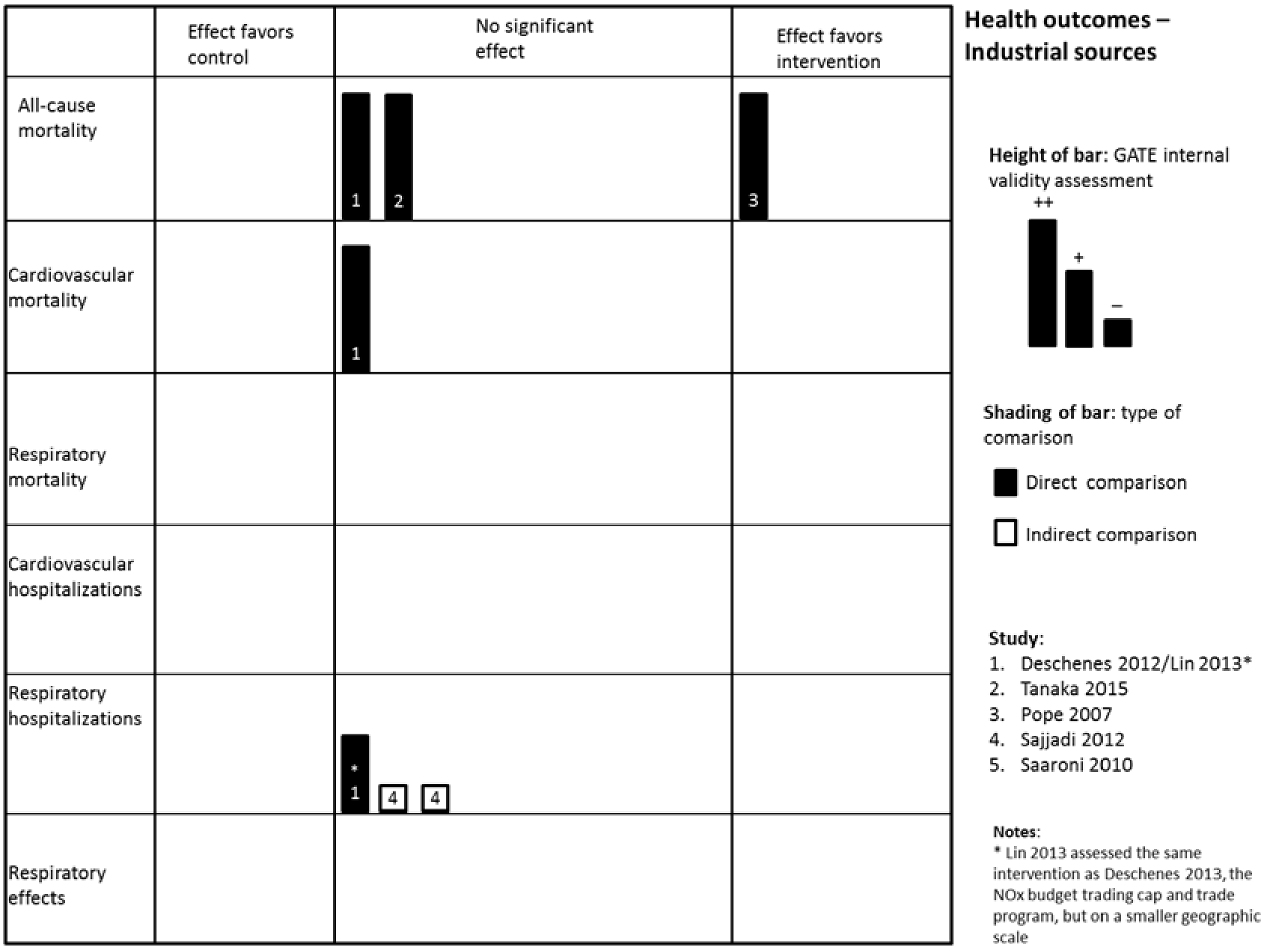
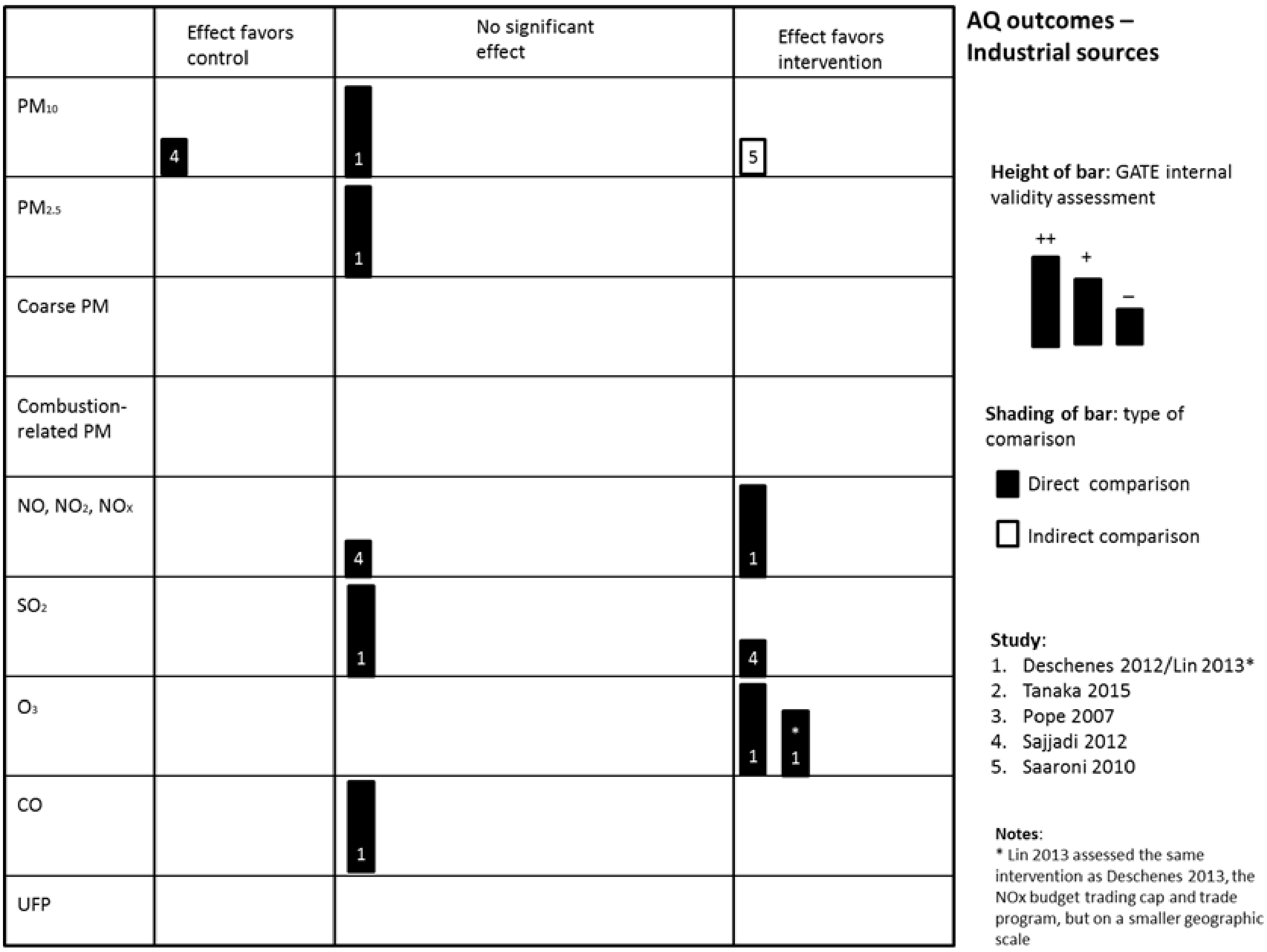


Figure 8. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from industrial sources on AQ outcomes



Health outcomes

Five studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from industrial sources on health outcomes, with three studies reporting all-cause mortality, one study reporting cardiovascular mortality, one study reporting respiratory hospitalizations and one study cardiovascular hospitalizations. No studies reported on respiratory mortality or respiratory effects. Most studies reported no clear associations in either direction, while one study observed a significant association favouring the intervention. No study observed a significant association favouring the control.

Deschênes 2012, a cITS-EPOC study with no substantial risk of bias concerns, observed no clear change in either all-cause mortality (1.57 fewer deaths per 100,000 population) or cardiovascular mortality (0.547 fewer deaths per 100,000 population) associated with the NOx Budget Trading Program, a US cap-and-trade initiative. Lin 2013, an ITS-EPOC with some risk of bias concerns, also assessed the NOx Budget Trading Program, but only for New York State, and observed no clear change in respiratory hospitalizations (0.15% reduction, 95% confidence interval (CI) −9.83 to 10.55) associated with the intervention. Tanaka 2015, a CBA-EPOC study with no substantial risk of bias concerns, observed no clear change in all-cause infant mortality (3.3 fewer deaths per 1000 live births) associated with the Chinese Two Zone Control policy. Pope 2007, a cITS-EPOC study with no substantial risk of bias concerns that evaluated the closure of copper smelters in the US Southwest due to a strike, observed a significant decrease (2.5% reduction, 95% CI −4.0 to −1.1) in all-cause mortality associated with the intervention. Sajjadi 2011, a cITS-EPOC study with serious risk of bias concerns, in parallel analyses observed similar changes at both intervention and control sites in COPD hospitalizations in the elderly (aged 65+) (36.9% increase at intervention sites; 31.5% increase at control sites) and asthma in children (aged < 15) (34.1% reduction at intervention sites; 36.6% reduction at control sites) associated with the closure of a local steel works in Australia.

Ambient air quality outcomes

Four studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from industrial sources on air quality outcomes, with studies reporting PM_{10}, $PM_{2.5}$, NO_2, SO_2, O_3 and CO. No studies reported on coarse PM, combustion-related PM, or UFP. Observed associations between interventions and different air quality outcomes were mostly spread between significant associations favouring the intervention and no clear association in either direction, although one study observed a significant association favouring the control.

Sajjadi 2012, an ITS-EPOC study with serious risk of bias concerns, observed a significant increase in PM_{10} (13.2% increase), no clear change in NO_2 (3.3% reduction), and a significant decrease in SO_2 (40.5% reduction) associated with the closure of a local steel works in Australia. Deschênes 2012, a cITS-EPOC study with no substantial risk of bias concerns, observed no clear change in either PM_{10} (3.0% decrease), $PM_{2.5}$ (2.3% reduction), SO_2 (2.1% increase) or CO (8.1% reduction), and a significant decrease in NO_2 (7.2% reduction) and O_3 (5.8% reduction) associated with the US NOx Budget Trading Program. Lin 2013, an ITS-EPOC with some risk of bias concerns, also assessed the US NOx Budget Trading Program, but only for New York State, and observed a significant decrease in O_3 associated with the intervention (2.5% reduction, 95% CI −3.22 to −1.72). Saaroni 2010, a CBA study with serious risk of bias concerns, in parallel analyses at intervention and control sites, observed a significant decrease in PM_{10} concentrations (14% reduction at intervention sites; 31% increase at control sites) associated with the conversion of a Tel Aviv power station from oil to gas.

Residential interventions versus practice as usual

As illustrated in Figure 9 and Figure 10, observed associations between interventions to reduce ambient air pollution from residential sources and both health and air quality outcomes were mixed, with all studies observing either a significant association favouring the intervention or no clear association in either direction. Summary of findings 3 outlines details regarding the effectiveness of interventions for each primary outcome, as well as a description of the quality of evidence drawn from our application of GRADE.

Figure 9. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from residential sources on health outcomes

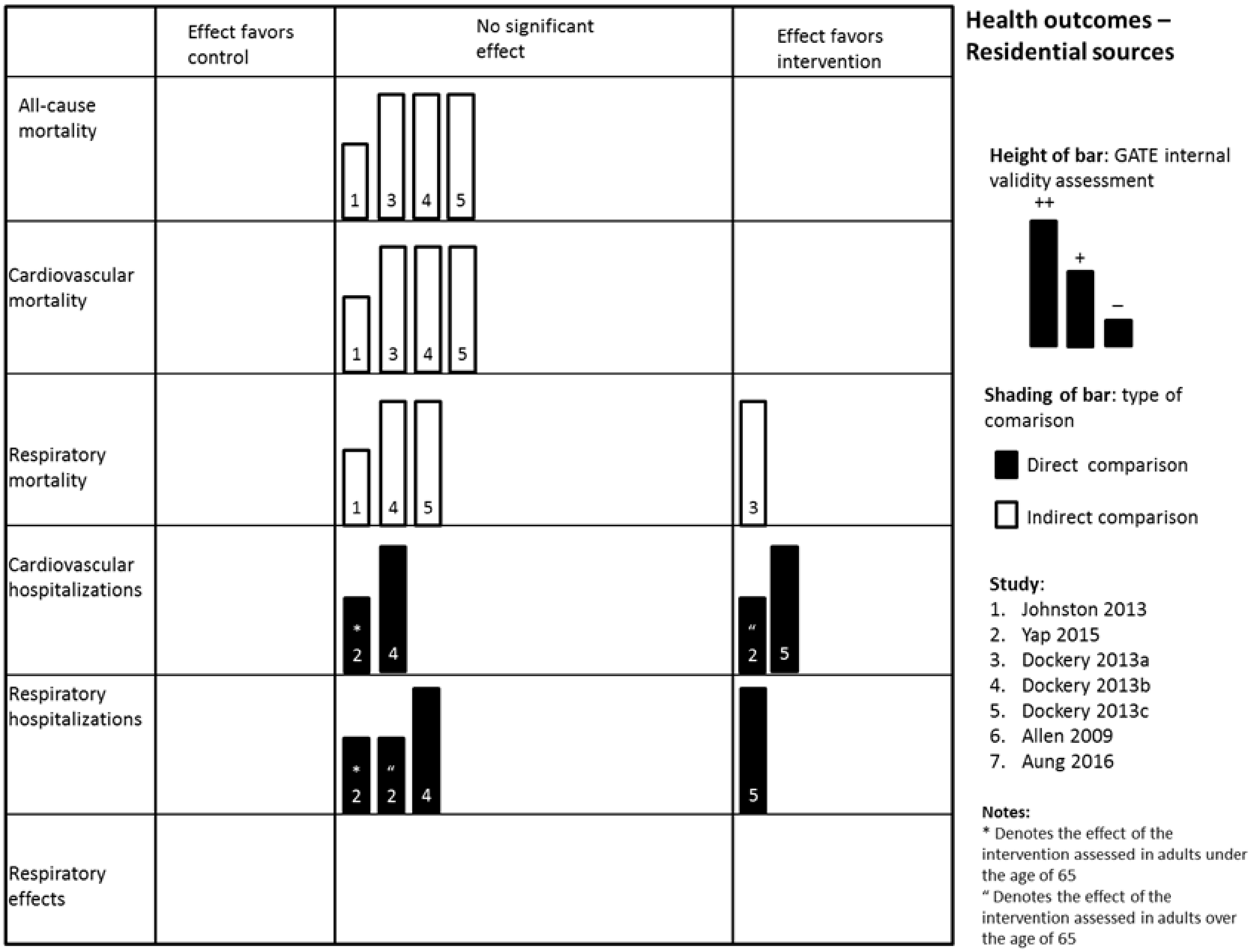
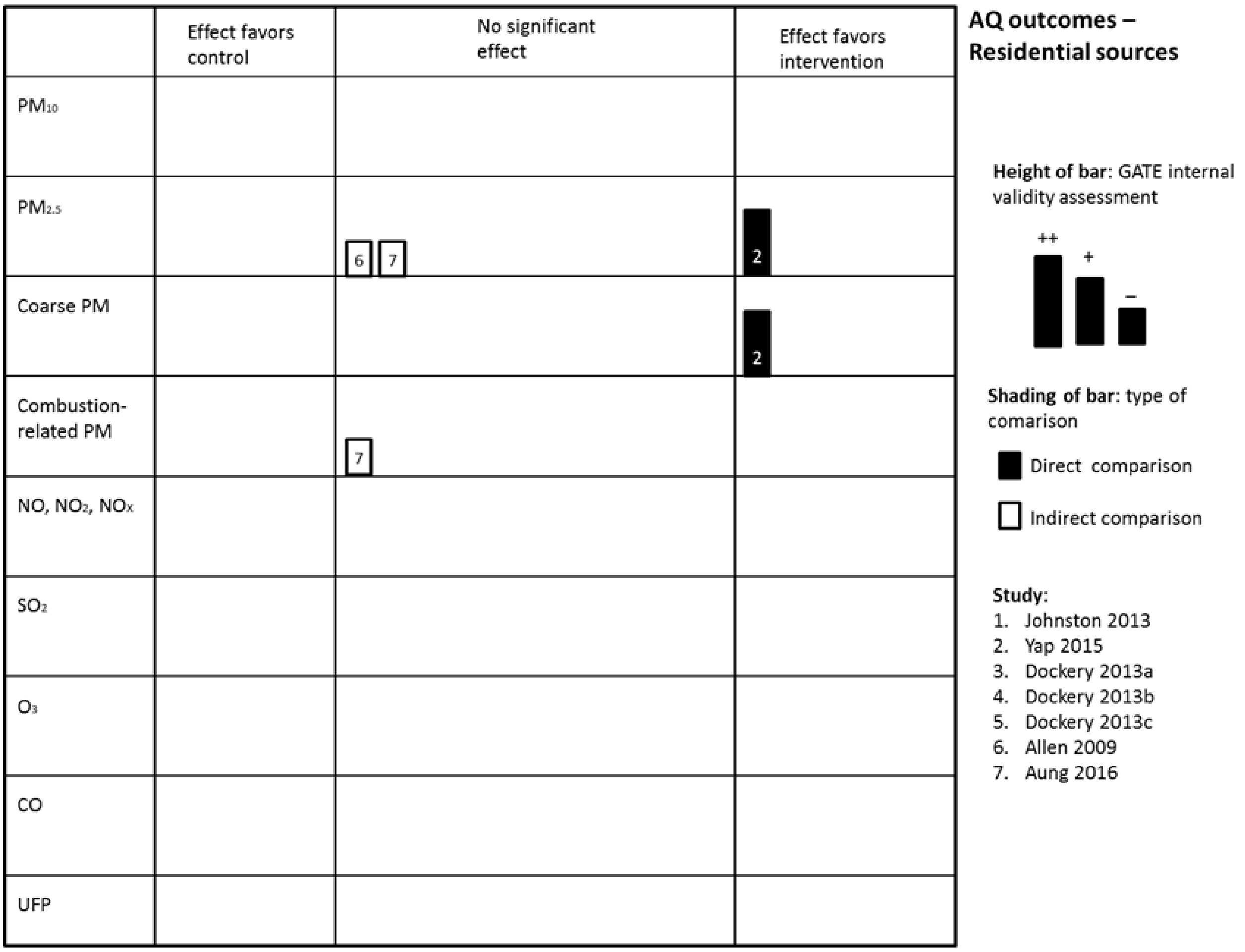


Figure 10. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from residential sources on AQ outcomes



Health outcomes

Five studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from residential sources on health outcomes; studies evaluated all-cause, cardiovascular and respiratory mortality, as well as cardiovascular and respiratory hospitalizations. No studies reported on respiratory effects. Studies showed a mix of significant associations favouring the intervention and no clear association in either direction. No study observed a significant association favouring the control.

Johnston 2013, a cITS-EPOC study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed no clear change in all-cause mortality (2.7% reduction at intervention sites, 95% CI −8.7 to 3.7; 1.4% increase at control sites, 95% CI −3.0 to 6.0), cardiovascular mortality (4.9% reduction at intervention sites, 95% CI −15.5 to 7.0; 0.9% increase at control sites, 95% CI −7.1 to 9.6) or respiratory mortality (8.5% reduction at intervention sites, 95% CI −23.2 to 9.0; 4.8% increase at control sites, 95% CI −7.4 to 18.6) associated with a stove exchange programme in Tasmania (Australia). Three studies with no substantial risk of bias concerns, assessed the effectiveness of coal ban interventions in Dublin (Dockery 2013a), in Cork (Dockery 2013b) and in five smaller Irish cities (Dockery 2013c); these studies applied a cITS-EPOC study design for mortality outcomes and an ITS-EPOC study design for hospitalization outcomes. The 1990 coal ban in Dublin, in parallel analyses at intervention and control sites, was associated with a significant reduction in respiratory mortality (16.8% reduction at intervention sites, 95% CI −24.4 to −8.4; 2.3% reduction at control sites, 95% CI −11.5 to 7.9), but no clear change was observed for all-cause mortality (1.0% reduction at intervention sites, 95% CI −6.0 to 4.4; 2.7% reduction at control sites, 95% CI −7.7 to 2.7) or cardiovascular mortality (0.1% increase at intervention sites, 95% CI −8.5 to 9.5; −1.8% reduction at control sites, 95% CI −10.0 to 7.2). In Cork, in parallel analyses at intervention and control sites, no clear changes were observed in all-cause mortality (4.4% reduction at intervention sites, 95% CI −9.6 to 1.1; 3.6% reduction at control sites, 95% CI −8.8 to 2.0), cardiovascular mortality (3.7% reduction at intervention sites, 95% CI −12.2 to 5.6; 3.4% reduction at control sites, 95% CI −12.0 to 6.1), respiratory mortality (9.3% reduction at intervention sites, 95% CI −18.2 to 0.7; 1.4% reduction at control sites, 95% CI −10.9 to 9.1), cardiovascular hospitalizations (3.6% reduction, 95% CI −9.8 to 2.9) or respiratory hospitalizations (3.6% increase, 95% CI −2.5 to 10) associated with the coal ban. In the five smaller Irish cities, in parallel analyses at intervention and control sites, no clear changes were observed for all-cause mortality (0.2% increase at intervention sites, 95% CI −3.1 to 3.6; 0.2% decrease at control sites, 95% CI −6.7 to 6.8), cardiovascular mortality (1.1% reduction at intervention sites, 95% CI −6.1 to 4.1; 3.1% reduction at control sites, 95% CI −12.6 to 7.3) or respiratory mortality (2.6% reduction at intervention sites, 95% CI −8.1 to 3.4; 1.4% increase at control sites, 95% CI −10.2 to 14.5) associated with the coal ban. This coal ban, however, was associated with a significant decrease in cardiovascular hospitalizations (3.2% decrease, 95%, CI −5.7 to −0.6) and a significant decrease in respiratory hospitalizations (8.5% decrease, 95% CI −10.5 to −6.2). Yap 2015, an ITS study with some risk of bias concerns, observed a significant decrease in cardiovascular hospitalizations in the population over 65 years of age (7% decrease, 95% CI −11 to −3), yet no clear change in the population under 65 years of age (3% decrease, 95% CI −10 to 15) associated with an intermittent, air-quality-dependent wood burning ban in the San Joaquin Valley of California. The study also observed no clear change in respiratory hospitalizations in either the population over 65 years of age (7% reduction, 95% CI −17 to 4.0) or the population under 65 years of age (10% reduction, 95% CI −22 to 5.0) associated with the wood burning ban.

Ambient air quality outcomes

Three studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from residential sources on air quality outcomes; these evaluated $PM_{2.5}$, coarse PM and combustion-related PM. No studies reported on PM_{10}, NO, NO_2, NOx, SO_2, O_3, CO or UFP. The few observed associations were mixed, with all studies observing either no clear association in either direction or a significant association in favour of the intervention.

Allen 2009, a CBA study with serious risk of bias concerns, in parallel analyses at intervention and control sites, observed no clear change in $PM_{2.5}$ concentrations (−2.7 ug/m^3 median change at intervention sites; −3.4 ug/m^3 median change at control sites) associated with a stove exchange programme in British Columbia (Canada). Aung 2016, a CBA study with serious risk of bias concerns, in parallel analyses at intervention and control sites, observed no clear change in $PM_{2.5}$ or BC concentrations associated with a stove exchange programme in southern India. Yap 2015, an ITS study with some risk of bias concerns, observed a significant decrease in $PM_{2.5}$ concentrations (−12.3% reduction, 95% CI −14.6 to −7.3) and coarse PM (−8.5% reduction, 95% CI −11.8 to −6.6) associated with an intermittent, air-quality-dependent wood burning ban in the San Joaquin Valley of California.

Vehicular interventions versus practice as usual

As illustrated in Figure 11 and Figure 12, observed associations between interventions to reduce ambient air pollution from vehicular sources and both health and air quality outcomes were mixed, with most studies observing either no clear association in either direction or a significant association in favour of the intervention. A small number of studies observed a significant association favouring the control. Summary of findings for the main comparison outlines details regarding the effectiveness of interventions for each primary outcome, as well as a description of the certainty of evidence drawn from our application of GRADE.

Figure 11. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from vehicular sources on health outcomes

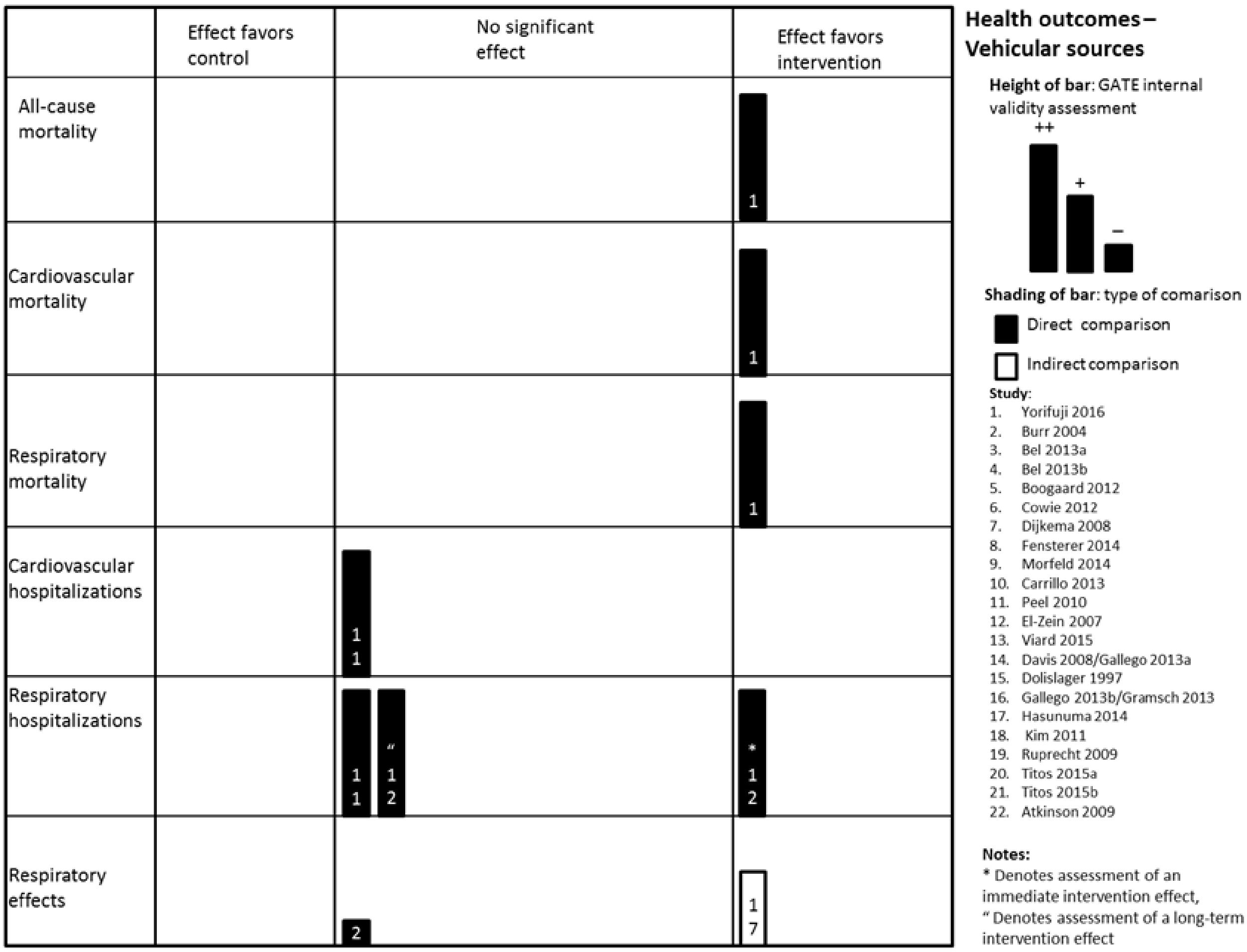
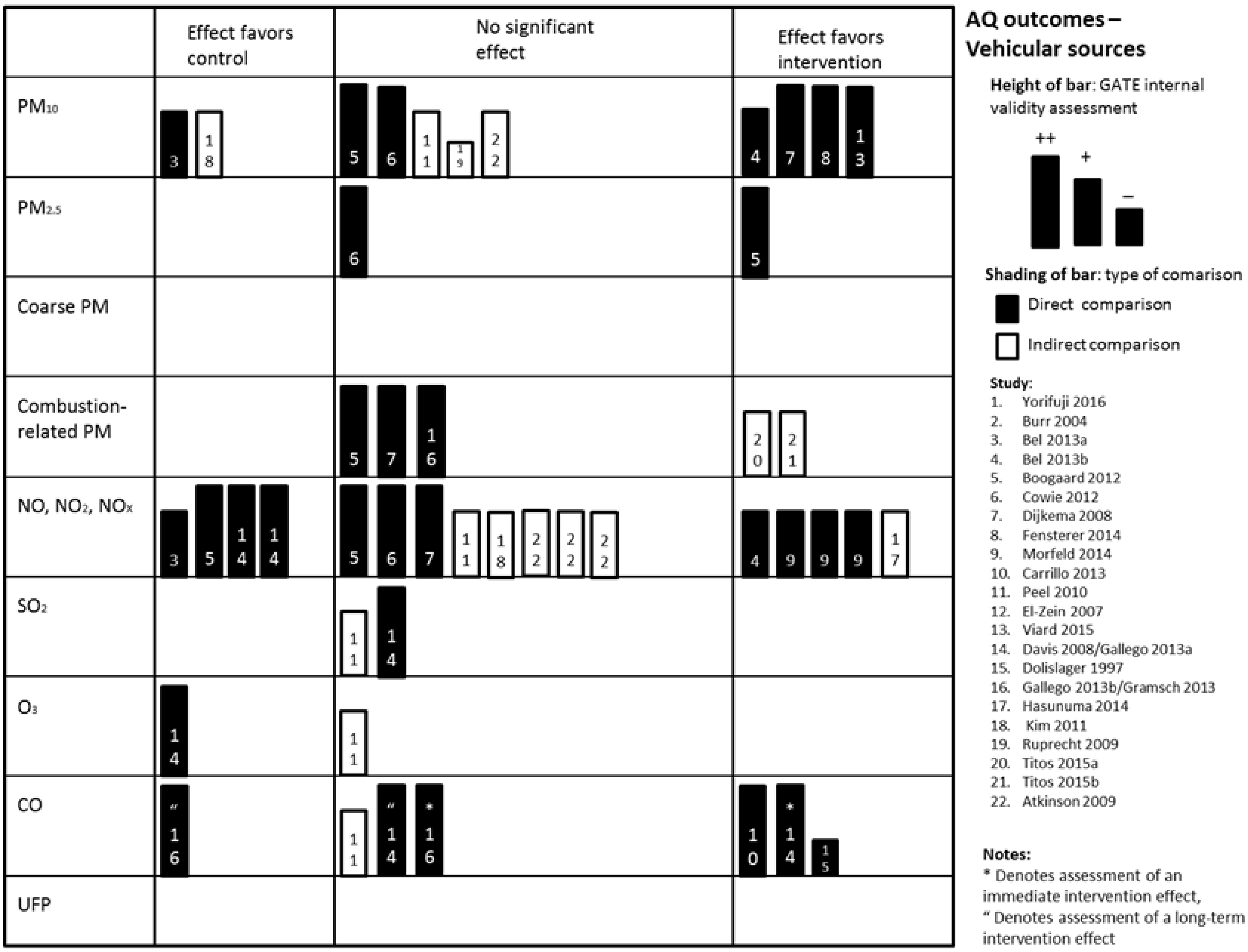


Figure 12. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from vehicular sources on AQ outcomes



Health outcomes

Five studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from vehicular sources on health outcomes; at least one study assessed each health outcome. Studies showed a mix of significant associations favouring the intervention and no clear association in either direction. No study observed a significant association favouring the control.

Yorifuji 2016, a cITS-EPOC study with no substantial risk of bias concerns, observed a significant decrease in all-cause mortality (2.1% reduction, 95% CI −2.8 to −1.4), cardiovascular mortality (5.9% reduction, 95% CI −7.2 to −4.6) and respiratory mortality (10% reduction, 95% CI −12 to −8.1) associated with mandatory standards for diesel vehicles entering the Tokyo metropolitan area. Peel 2010, an ITS-EPOC study with no substantial risk of bias concerns, observed no clear change in cardiovascular hospitalizations (Risk ratio (RR) 0.996, 95% CI 0.83 to 1.20) or respiratory hospitalizations (RR 1.01, 95% CI 0.92 to 1.11) associated with the coordinated measures aimed at reducing traffic during the 1996 Atlanta Olympic Games. El-Zein 2007, an ITS-EPOC study with serious risk of bias concerns, observed an immediate yet significant slight reduction, yet no longer-term change in respiratory hospitalizations in children under 14 associated with a ban on diesel automobiles in Beirut (Lebanon). Burr 2004, a CBA study with severe risk of bias concerns, observed no clear change in asthma symptoms associated with the opening of a bypass to reduce traffic congestion in northern Wales. Hasunuma 2014, a CBA-EPOC study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed a significant decrease in respiratory symptoms in children three years old or younger (17.4% reduction at intervention sites, 95% CI −25.9 to −9.1; 3.5% reduction at control sites, 95% CI −12.5 to 5.4) associated with standards required by the NOx/PM Law in Japan.

Ambient air quality outcomes

Nineteen studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from vehicular sources on air quality outcomes. Most studies assessed PM_{10}, NO, NO_2, NOx, and CO; very few studies assessed $PM_{2.5}$, SO_2 and O_3; while no studies reported on coarse PM or UFP. Studies showed a mix of significant associations favouring the intervention, significant associations favouring the control, and no clear association in either direction.

Boogaard 2012, a CBA-EPOC study with no substantial risk of bias concerns, observed no clear change in PM_{10} (11% reduction at intervention sites; 14.7% reduction at control sites); soot (1.4% reduction at intervention sites; 7.4% reduction at control sites); or NOx (9.2% reduction at intervention sites; 15.9% reduction at control sites); a significant decrease in $PM_{2.5}$ (30% reduction at intervention sites; 19.6% at control sites); and a significant increase in NO_2 (3.2% reduction at intervention sites; 17.4% reduction at control sites) associated with multiple low emission zones in the Netherlands. Cowie 2012, a cITS-EPOC study with no substantial risk of bias concerns, observed no clear change in concentrations of PM_{10} (3.8% reduction, 95% CI −8.0 to 0.40), $PM_{2.5}$ (2.9% increase, 95% CI −4 to 9.7), NOx (8.1% reduction, 95% CI −18.7% to 2.4%) or NO_2 (2.9% reduction, 95% CI −7.2 to 1.5) associated with a tunnel meant to relieve traffic congestion in suburban Sydney (Australia). Dijkema 2008, a CBA study with no substantial risk of bias concerns, observed a significant decrease in PM_{10} concentrations (7.4% reduction at intervention sites, 95% CI −10 to −4.8; 3.9% reduction at control sites, 95% CI −6.7 to −1), but no clear change in concentrations of BS (15% reduction at intervention sites, 95% CI −23.7 to −6.2; 12% reduction at control sites, 95% CI −18.9 to 5.2) or NOx (2.4% reduction at intervention sites, 95% CI −8.1 to 3.3; 2.7% reduction at control sites, 95% CI −8.3 to 2.8) associated with a speed limit reduction on a heavily trafficked roadway in Amsterdam. Peel 2010, a CBA-EPOC study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed no clear change in concentrations of PM_{10} (17% reduction at intervention sites; 16.4% and 13.3% reduction at control sites), NO_2 (slight reduction at all intervention and control sites; see Appendix 9), O_3 (reductions at intervention and control sites; see Appendix 9), SO_2 (slight increase at intervention sites, mixed changes at control sites; see Appendix 9) or CO (reductions at intervention sites, mixed changes at control sites; see Appendix 9) associated with the coordinated measures aimed at reducing traffic during the 1996 Atlanta Olympic Games. Ruprecht 2009, a CBA study with serious risk of bias concerns, in parallel analyses at intervention and control sites, observed no clear change in concentrations of PM_{10} (4.8% reduction at intervention sites; 5.0% reduction at control sites) associated with the Ecopass congestion charging scheme in Milan (Italy). Atkinson 2009, a CBA study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed no clear change in concentrations of PM_{10} (5.6% increase at intervention sites; 2.5% increase at control sites), NOx (5% reduction at intervention sites; 4% reduction at control sites), NO_2 (2.1% increase at intervention sites; 3.7% increase at control sites) or NO (9.5% reduction at intervention sites; 9.4% reduction at control sites) at streetside sites associated with the London congestion charge scheme. Bel 2013b, an ITS-EPOC study with some risk of bias concerns, observed a significant decrease in concentrations of PM_{10} (14.7% reduction) and NOx (16% reduction) associated with an adaptive speed limit scheme in Barcelona (Spain). Fensterer 2014, a CBA study with no substantial risk of bias concerns, observed a significant decrease in PM_{10} concentrations associated with the low emission zone in Munich (Germany) both in summer (19.6% reduction, 95% CI −22.75 to −16.52) and winter (6.8% reduction, 95% CI −10.14 to −3.47). Viard 2015, an ITS-EPOC study with no substantial risk of bias concerns, observed a significant decrease in PM_{10} concentrations associated with an even-odd driving restriction policy (31% reduction), which was then relaxed to a one-day per vehicle (27% reduction) driving ban in Beijing. Bel 2013a, a cITS-EPOC study with some risk of bias concerns, observed a significant increase in concentrations of PM_{10} (5.4% increase) and NOx (1.7% increase) associated with a speed limit reduction in Barcelona (Spain). Kim 2011, a CBA-EPOC study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed a significant increase in PM_{10} concentrations (14.7% increase at intervention sites; 4.7% reduction at control sites), yet no clear change in NO_2 concentrations (1.1% reduction at intervention sites; 1.0% increase at control sites) associated with the Natural Gas Vehicle Supply programme that led to the introduction of natural-gas-powered buses in South Korean cities. Gramsch 2013, a CBA study with some risk of bias concerns, observed no clear change in BC (4.8% increase at intervention sites; 17.4% increase at control sites) associated with Transantiago, a restructuring of the public transportation system in Santiago (Chile). Gallego 2013b, an ITS-EPOC study with no substantial risk of bias concerns, also evaluated Transantiago in Santiago (Chile) and observed no clear immediate change (5.9% reduction), yet a significant long-term increase in CO concentrations (26.8% increase). Titos 2015a, a CBA study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed a significant decrease in BC concentrations (72% reduction at intervention sites; 6% increase at control sites) associated with a partial closure and reconstruction of a major street in Ljubljana (Slovenia). Titos 2015b, a CBA study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed a significant decrease in BC concentrations (37% reduction at intervention sites; 14% reduction at control sites) associated with the restructuring of the public bus system in Granada (Spain). Davis 2008, an ITS-EPOC study with some risk of bias concerns, observed a significant 17.3% increase in NOx concentrations , an 8.9% increase in NO_2 concentrations, and a 28% increase in O_3 concentrations, yet no clear change in SO_2 concentrations (9.2% decrease) associated with Hoy no Circula, an even-odd driving ban in Mexico City. Gallego 2013a, which also evaluated Hoy no Circula in Mexico City, observed an immediate significant decrease in CO concentrations (13% reduction), yet no clear long-term change in CO concentrations (11.3% increase) associated with the intervention. Morfeld 2014, a CBA-EPOC study with no substantial risk of bias concerns, observed a significant decrease in concentrations of NOx (3.5% reduction, 95% CI −4.7 to −2.3), NO_2 (2.2% reduction, 95% CI −2.3 to −2.0) and NO (2.3% reduction, 95% CI −3.1 to −1.4) associated with LEZs in 17 German cities. Hasunuma 2014, a CBA-EPOC study with some risk of bias concerns, in parallel analyses at intervention and control sites, observed a significant decrease in NO_2 concentrations (22.5% reduction at intervention sites, 95% CI −26.4 to −18.5; 21.6% reduction at control sites, 95% CI −30.0 to 13.4) associated with the NOx/PM Law which introduced the designation of "enforcement areas" and associated vehicle standards in Japan. Carrillo 2016, a CBA-EPOC study with no substantial risk of bias concerns, observed a significant decrease in CO concentrations (9% reduction) associated with an even-odd driving ban in Quito (Ecuador). Dolislager 1997, an ITS-EPOC study with serious risk of bias concerns, observed a significant decrease in CO concentrations (8.5% reduction) associated with fuel standards in California restricting the oxygen content of gasoline in winter months.

Multiple interventions versus practice as usual

As illustrated in Figure 13 and Figure 14, observed associations between interventions to reduce ambient air pollution from multiple sources and both health and air quality outcomes were mixed, with all studies showing either no clear association or a significant association in favour of the intervention. Summary of findings 4 outlines details regarding the effectiveness of interventions for each primary outcome, as well as a description of the certainty of evidence drawn from our application of GRADE.

Figure 13. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from multiple sources on health outcomes

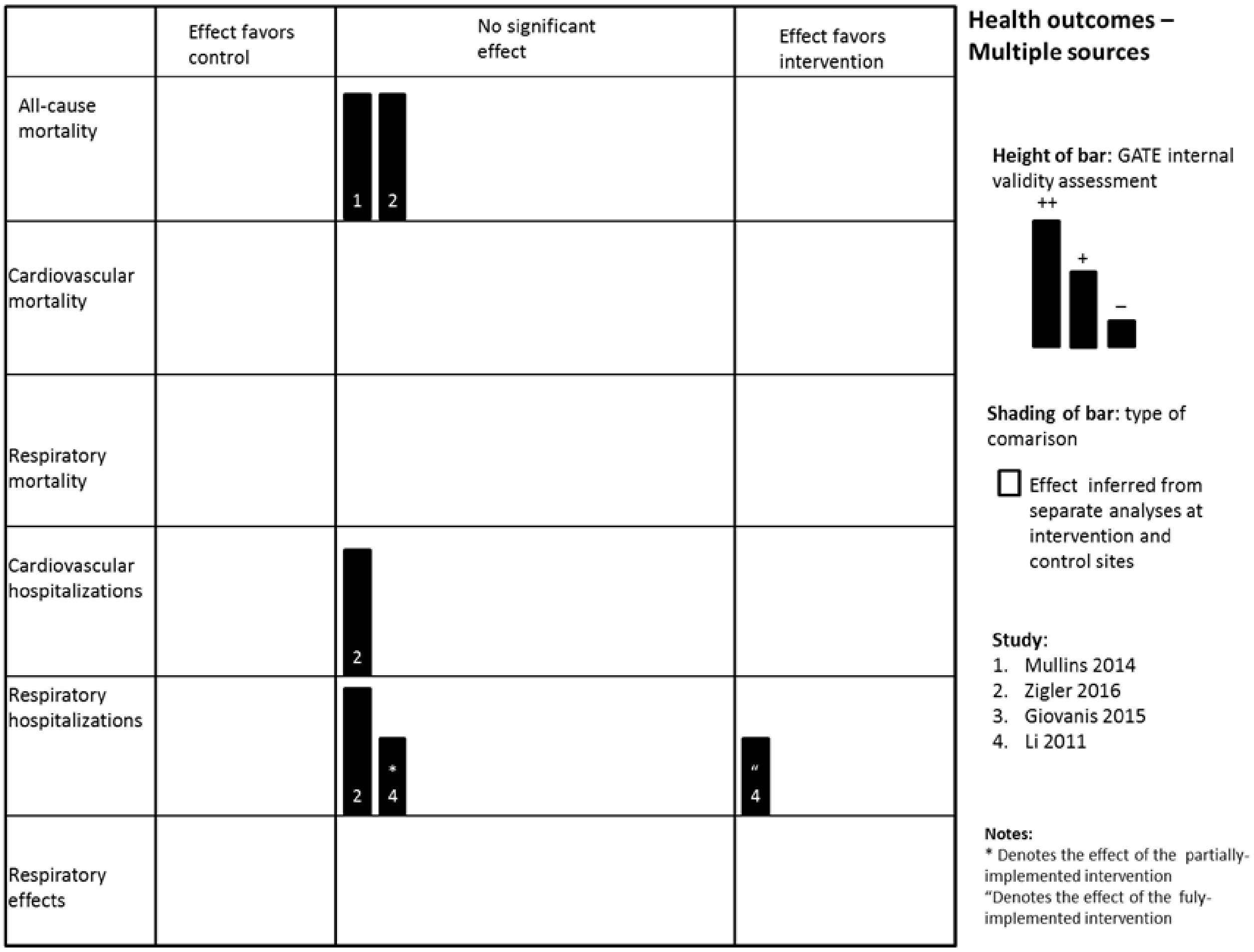
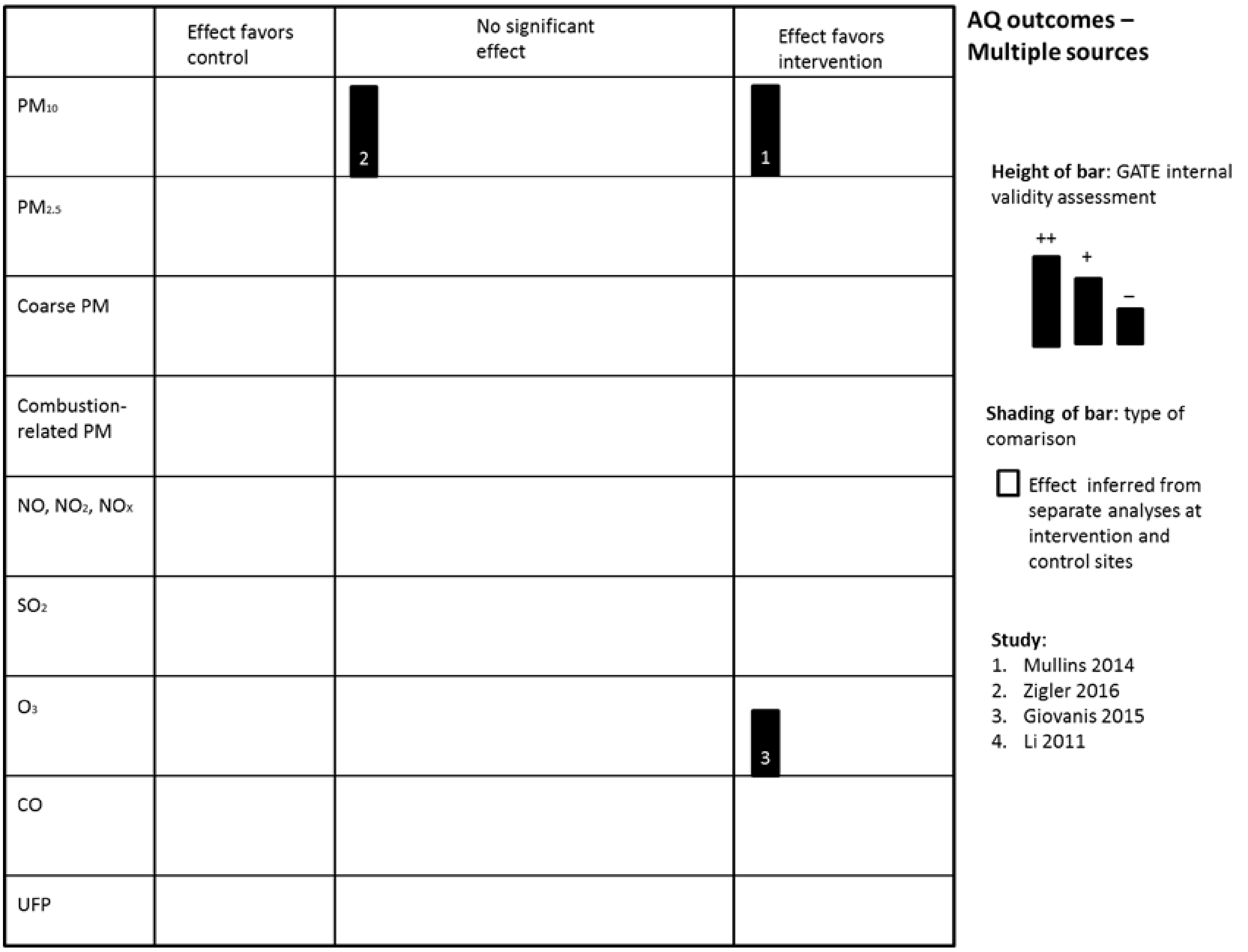


Figure 14. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from multiple sources on AQ outcomes



Health outcomes

Three studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from multiple sources on health outcomes, with studies measuring all-cause mortality or cardiovascular and respiratory hospitalizations, or mortality and hospitalizations. No studies reported on cardiovascular mortality, respiratory mortality or respiratory effects. All studies observed either a significant association favouring the intervention or no clear association in either direction. No study observed a significant association favouring the control.

Mullins 2014, an ITS-EPOC study with no substantial risk of bias concerns, observed no clear change in all-cause mortality (5.6% reduction) associated with coordinated measures to reduce vehicular and industrial pollution enacted in Santiago (Chile) on days for which poor air quality is forecast. Zigler 2016, a CBA-EPOC study with no substantial risk of bias concerns, observed no clear change in all-cause mortality (1.7% reduction, 95% CI −5.2 to 1.6), cardiovascular hospitalizations (1.6% increase, 95% CI −5.0 to 6.7) or respiratory hospitalizations (5.2% reduction, 95% CI −13.6 to 4.5) associated with the US National Ambient Air Quality Standards non-attainment designation, given as part of the US Clean Air Act to areas which did not meet the air quality standards. Li 2011, an ITS-EPOC study with some risk of bias concerns, observed no clear change in respiratory hospitalizations when the intervention was only partially implemented (adjusted risk ratio 1.24, 95% CI 0.93 to 1.76), then a significant decrease (adjusted risk ratio 0.50, 95% CI 0.47 to 0.55) associated with the full set of measures aiming to decrease vehicular and industrial pollution during the 2008 Beijing Olympic Games.

Ambient air quality outcomes

Three studies contributed data to the evidence synthesis of interventions to reduce ambient air pollution from multiple sources on air quality outcomes, with studies assessing PM_{10} and O_3. No studies assessed $PM_{2.5}$, coarse PM, combustion-related PM, NO, NO_2, NOx, SO_2, CO or UFP. All studies observed either a significant association favouring the intervention or no clear change in either direction. No studies observed effects favouring the control.

Mullins 2014, an ITS-EPOC study with no substantial risk of bias concerns, observed a significant decrease in PM_{10} concentrations (16.9% reduction) associated with coordinated measures to reduce vehicular and industrial pollution enacted in Santiago (Chile) on days for which poor air quality is forecast. Zigler 2016, a CBA-EPOC study with no substantial risk of bias concerns, observed no clear change in PM_{10} concentrations (2.9% reduction, 95% CI

−18.1 to 9.9) associated with non-attainment designation given as part of the US Clean Air Act to areas not meeting the National Ambient Air Quality Standards. Giovanis 2015, a CBA-EPOC study with some risk of bias concerns, observed a significant decrease on O_3 concentrations (2.3% reduction) associated with coordinated measures to reduce vehicular and industrial pollution enacted in Charlotte (North Carolina, USA) on days for which poor air quality is forecast.

Subgroup analysis of temporary interventions

One temporary intervention targeted industrial sources (Pope 2007); one temporary intervention targeted vehicular sources (Peel 2010); and one temporary intervention targeted multiple sources (Li 2011). No temporary interventions aimed to decrease air pollution from residential sources. The rest of the interventions aimed to affect air quality permanently. Potential differences were assessed graphically through the creation of harvest plots stratified for temporary and permanent interventions. Overall, it appears that the temporary and permanent interventions did not differ substantially with regard to effectiveness. Given the limited number of studies assessing temporary interventions, these harvest plots are not shown.

Supporting studies

The supporting studies, which are described narratively in Appendix 4 and summarized in table form in Appendix 5, were largely similar to main studies with regard to the assessed populations, interventions and outcomes. One notable difference is that a larger proportion of supporting studies were conducted in LMICs (56% vs 29%).

DISCUSSION

Summary of main results

This is the first systematic review to assess the effectiveness of interventions in reducing pollutant concentrations and improving associated health outcomes. Given the heterogeneity across interventions, outcomes, and study methods, it was difficult to derive any overall conclusions regarding the effectiveness of interventions in improving air quality or health.

Most interventions, whether aiming to reduce pollution from industrial, residential, vehicular or multiple sources, observed either no significant association in either direction or an association favouring the intervention. There is very little evidence suggesting that any of the assessed interventions were harmful.

In interpreting these results, however, it is important to consider several factors that may have impacted individual study results. Establishing a causal relationship between air pollution interventions, changes in air quality and health outcomes is challenging for a range of reasons. First, the nature of the causal pathway between air pollution interventions and changes in health, as illustrated by the Health Effects Institute (HEI) chain of accountability (HEI 2003), is long. The introduction of an intervention must first lead to reductions in source emissions, followed by reduced ambient pollutant concentrations, reduced exposure/dose for the individual, and finally improvements in health; all of these steps in the chain may also be influenced by the broader environmental and social context in which an intervention is embedded.

Second, these interventions do not exist in a vacuum, and often multiple interventions are implemented within the same time frame, and at multiple levels (e.g. local, regional, and national) in the context of a host of other long-term environmental and societal changes. Large-scale multi-year regulatory programmes are particularly challenging since they may not have immediate effects on either air quality or public health; they are typically implemented in multiple separate steps, often on different spatial scales, and over an extended period of time to address emissions from a variety of sources. Also, the biological processes that underlie adverse health effects of air pollution may take years to manifest, and are also associated with a complex array of genetic, biological, social, cultural and environmental factors (Dahlgren 1991; Graham 2016). This poses a challenge for epidemiologists since the longer the time between implementation of an intervention and its effects, the greater the possibility that other factors influencing air quality and health outcomes (e.g. an economic downturn, changes in medical practices, and the availability of health care) may come into play and interfere with demonstrating the effects of the intervention itself. In this context it is particularly noteworthy that all ambient air pollution interventions are evaluated against the backdrop of long-term trends of demographic change (i.e. population growth, increasing life expectancies and ageing), industrialization and economic development, which directly influence all sources of air pollution covered in this review, leading to increased motorized vehicle traffic, more potentially polluting industries and greater energy use for lighting, cooking, heating and various electric appliances in residences.

Third, as previously discussed, ambient air pollution represents a complex mix of pollutants, originating from a range of sources, with approximately 15% of urban ambient pollution stemming from industrial sources, 20% from residential sources and 25% from vehicular sources (Karagulian 2015). Thus, interventions aiming to reduce air pollution from a single source inherently only address part of the problem, and air pollution from other sources, including industrial, residential and vehicular sources, but also agricultural and other transport-related sources such as shipping and flight traffic may adversely affect health. Efforts to improve air quality and associated human health are therefore likely to require a systems approach that targets multiple sources through a combination of different measures in a context- and setting-specific manner (Rutter 2017).

All of these aspects contribute to the challenge of firstly, improving ambient air quality and population health outcomes through specific interventions, and secondly, detecting these changes through rigorous research methods. These aspects should, therefore, be considered when interpreting effects from individual studies, including those described in this review. It should be emphasized that *no evidence of an effect* is not equivalent to *evidence of no effect*; it is possible that some interventions assessed in this review may have improved air quality and the associated health outcomes, even where no improvement was observed in the primary studies.

Interventions targeting industrial sources

For interventions targeting industrial sources, the evidence base with respect to primary outcomes ranged from low certainty (for all-cause mortality, respiratory mortality, and $PM_{2.5}$) to very low certainty (for PM_{10}) (Summary of findings 2). The associations observed in these studies were mixed for both health and air quality outcomes, (Figure 6, Figure 7). The closure of a copper smelter in the US Southwest (Pope 2007) and the conversion of a power station from oil to gas in Tel Aviv, Israel (Saaroni 2010) were associated with improvements in all-cause mortality and PM_{10}, respectively. The US NOx Budget Trading Program (Deschênes 2012), whose impact on all-cause mortality, cardiovascular mortality, PM_{10} and $PM_{2.5}$ was assessed, and the Chinese Two Zone Control policy (Tanaka 2015), evaluated for its impact on all-cause mortality, were not associated with clear changes in these outcomes. The closure of a steel works in New South Wales (Australia) was associated with an increase in PM_{10}, no change in respiratory hospitalizations, or NO_2, and a decrease in SO_2 (Sajjadi 2012). Associations with regard to secondary outcomes were similarly mixed (Figure 6, Figure 7).

Interventions targeting residential sources

For interventions targeting residential sources, the evidence base with respect to primary outcomes ranged from low certainty for cardiovascular mortality to very low certainty for all-cause and respiratory mortality, $PM_{2.5}$, coarse PM and combustion-related PM (Summary of findings 3). The associations observed in these studies were mixed for both health and air quality outcomes, (Figure 8, Figure 9). A coal ban in Dublin was associated with a decrease in respiratory mortality, but no clear change in all-cause or cardiovascular mortality (Dockery 2013a). A stove exchange programme in Tasmania (Australia) (Johnston 2013) and a coal ban in Cork (Dockery 2013b), and in five smaller Irish cities (Dockery 2013c) showed no clear change in all-cause, cardiovascular or respiratory mortality. A stove exchange programme in British Columbia and another in southern India were not associated with clear changes in $PM_{2.5}$, while an intermittent wood burning ban in the San Joaquin Valley of California (USA) showed a decrease in $PM_{2.5}$ concentrations (Yap 2015). Associations with regard to secondary outcomes were similarly mixed (Figure 8, Figure 9).

Interventions targeting vehicular sources

For interventions targeting vehicular sources, the evidence base with respect to primary outcomes ranged from low certainty for all-cause mortality, cardiovascular mortality, respiratory mortality and $PM_{2.5}$ to very low certainty for PM_{10} and combustion-related PM (Summary of findings for the main comparison). The associations observed in these studies were mixed for both health and air quality outcomes (Figure 10, Figure 11). Mandatory standards for diesel vehicles entering the metropolitan area in Tokyo were associated with improvements in all-cause, cardiovascular and respiratory mortality. An adaptive speed limit scheme in Barcelona (Spain) (Bel 2013b), a low emission zone in Munich (Germany) (Fensterer 2014), and an even-odd driving restriction policy in Beijing (PRC) (Viard 2015) were all associated with decreased PM_{10} concentrations. Similarly, low emission zones in several Dutch cities showed a decrease in $PM_{2.5}$ concentrations (Boogaard 2012). The partial closure and reconstruction of a major street in Ljubljana (Slovenia) (Titos 2015a) and the restructuring of the public bus system in Granada (Spain) (Titos 2015b) were associated with decreases in combustion-related PM. Several interventions, including the low emission zones in Dutch cities (Boogaard 2012), the construction of a tunnel to relieve traffic congestion in Sydney (Australia) (Cowie 2012), a speed limit reduction in Amsterdam (the Netherlands) (Dijkema 2008), the 1996 Olympic Games in Atlanta (USA) (Peel 2010), the Ecopass congestion charging scheme in Milan (Italy) (Ruprecht 2009), and the London congestion charging scheme (Atkinson 2009) did not show clear changes in PM_{10}. The construction of a tunnel for relieving congestion was not associated with a clear change in $PM_{2.5}$ (Cowie 2012). Low emission zones in several Dutch cities (Boogaard 2012), a speed limit reduction in Amsterdam (the Netherlands) (Dijkema 2008), and a restructuring of the public transportation system in Santiago (Chile) (Gallego 2013b; Gramsch 2013) reported no clear changes in combustion-related PM. A speed limit reduction in Barcelona (Spain) (Bel 2013a), and the Natural Gas Vehicle Supply programme in South Korean cities (Kim 2011) were associated with an increase in PM_{10} concentrations. Associations with regard to secondary outcomes were similarly mixed (Figure 10, Figure 11).

Interventions targeting multiple sources

For interventions targeting multiple sources, the evidence base with respect to primary outcomes was very low certainty for all-cause mortality and PM_{10} (Summary of findings 4). The associations observed in these studies were mixed for both health and air quality outcomes (Figure 12, Figure 13). Coordinated measures to reduce vehicular and industrial pollution enacted in Santiago (Chile) on days for which poor air quality is forecast (Mullins 2014) and the US National Ambient Air Quality Standards non-attainment designation, introduced as part of the US Clean Air Act (Zigler 2016) showed no clear changes in all-cause mortality. The coordinated measures in Santiago (Chile) were associated with a decrease in PM_{10}, while the US National Ambient Air Quality Standards non-attainment designation showed no clear changes in PM_{10} concentrations. Associations with regard to secondary outcomes were mixed (Figure 12, Figure 13).

Overall completeness and applicability of evidence

This systematic review assessed the effectiveness of a broad range of interventions in improving specific air quality and health outcomes, without any geographical or population-related restrictions. The identified evidence base, considering both main and supporting studies, investigates many different interventions in many different contexts and settings, and is largely complete with regard to the systematic review objective. In assessing the overall completeness and applicability of the evidence, we drew from three different sources: 1) the external validity assessment applied using the NICE modified GATE tool; 2) a comparison of the identified evidence with the a priori defined logic model; and 3) relevant gaps as identified using the harvest plots (i.e. where specific intervention types have not been assessed with respect to certain outcomes).

The external validity assessment using the NICE modified GATE tool indicated that identified studies were relevant to a broad range of populations (Figure 4, Figure 5); the routine monitoring data used for both air quality and health outcomes in most studies facilitated the investigation of broad, 'real-world' sample populations.

The system-based logic model illustrates the system in which different types of interventions are implemented, and documents the PICO-related — as well as wider context-related — aspects that may have influenced the effectiveness of interventions (Figure 2). Broadly speaking, included studies covered the majority of aspects populating the logic model. We included studies from across the globe from a variety of contexts and settings (Table 1, Figure 4). Most studies assessed the general population, but we also included studies specifically in infants (Tanaka 2015), children and adolescents (El-Zein 2007; Hasunuma 2014; Sajjadi 2011), and the elderly (Sajjadi 2011). We identified interventions belonging to all four intervention categories; the distribution across intervention categories was imbalanced, however, as a much larger proportion of identified studies were concerned with interventions targeting vehicular sources rather than other sources of ambient air pollution. Within categories several sub-categories were identified; some intervention sub-categories are better represented than others. Within vehicular interventions, for example, a relatively large number of studies reported on LEZs across Europe (Boogaard 2012; Fensterer 2014; Morfeld 2014), and even-odd bans are also well represented by studies in Ecuador, Mexico, China and South Korea (Carrillo 2016; Davis 2008; Gallego 2013a; Viard 2015). Similarly, within the residential interventions category, several studies assessed stove exchanges (Allen 2009; Aung 2016; Johnston 2013). On the other hand some sub-categories, such as the wood burning ban (Yap 2015) and a ban on diesel vehicles (El-Zein 2007), are poorly represented in the evidence base. Although the logic model highlighted the potential influence of various context-related factors, these factors were poorly reported in individual studies, and could not be assessed in a structured manner.

The harvest plots illustrate where evidence is plentiful and where relevant gaps in the evidence base exist. Many studies have, for example, examined the effects of vehicular interventions with respect to most outcomes. There is substantially less evidence regarding the effectiveness of industrial, residential and multiple interventions. The harvest plots indicate that in general across the evidence base for all intervention types, air quality outcomes were assessed much more frequently than health outcomes. Similarly, they illustrate that the evidence base is incomplete with respect to certain outcomes, such as respiratory effects, coarse PM and UPF concentrations.

As described in the Methods section, the final date of searches for this review is August 2016, thus the most current studies are not included in this review. Our Review Advisory Group identified several studies published since then that would potentially be included in the review (Barreca 2017; Font 2016; Gehrsitz 2017; Hales 2016; Han 2018; Li 2017; Lin 2016; Yinon 2017). From their feedback, it is clear that this is a very active field of study, and that an update to this review will be beneficial in the near future. This list of studies is very likely non-comprehensive; however based on an informal survey of these studies, it does not appear that the conclusions of this review would be altered based on this recent evidence.

Quality of the evidence

As described in detail in the 'Summary of findings' tables, applying the GRADE approach to appraise the certainty of evidence yielded low or very low ratings for all primary health and ambient air quality outcomes. These low ratings were primarily driven by the nature of the study designs included in this systematic review, which is exclusively based on non-randomised evidence. Risk of bias of included studies as well as inconsistency in findings — where for certain outcomes we identified studies favouring the intervention, studies favouring the control, as well as studies reporting no or unclear effects — contributed to these ratings and lowered our confidence that the observed effects represent the true effect. In the following we briefly discuss the findings of this systematic review in relation to each of the five criteria for rating down the certainty of evidence — i.e. risk of bias, inconsistency of results, indirectness of evidence, imprecision, and publication bias — and provide examples of each. None of the criteria for rating up the certainty of evidence were applicable.

We assessed whether the main studies included in a given body of evidence were at high risk of bias, and thus would weaken the certainty of that body of evidence. Specific concerns regarding risk of bias differed across the bodies of evidence, but common issues comprised choice of intervention and selection sites and the lack of consideration of potentially important confounders. With regard to industrial interventions, for example, we downgraded the evidence on PM_{10} due to potential selection bias and the lack of consideration of potentially important confounders. One of the three studies contributing to this evidence base, in evaluating the conversion of a Tel Aviv power station from oil to gas, chose only one intervention and one control site based on the prevalent wind patterns with respect to the power station, and did not include any potential confounders in the analysis (Saaroni 2010).

We rated down a body of evidence where effects from included studies varied widely, indicating inconsistency. In some cases, however, given the substantial heterogeneity of the included studies, such inconsistency could be expected. Thus we rated down evidence only when substantial inconsistency was present (i.e. observed effects favouring the intervention and the control), and where this inconsistency could not be readily explained. For vehicular interventions, for example, we rated down the evidence for PM_{10} because effects of similar interventions in similar contexts, for example low emission zones in Dutch cities (Boogaard 2012) and Munich (Germany) (Fensterer 2014), and two speed limit changes in Barcelona (Spain) (Bel 2013a; Bel 2013b), would be expected to be more consistent than observed in these studies.

Considering imprecision in applying GRADE, we rated down a body of evidence where the conduct of the primary studies led to imprecise effect estimates, thus indicating significant uncertainty surrounding the benefits and/or harms of the intervention. For residential interventions, for example, we rated down the evidence for all-cause mortality and respiratory mortality due to imprecision, as one of the four studies reported very wide confidence intervals spanning from a meaningful effect to a potential harmful effect (Johnston 2013). As most studies used routine health and/or air quality data for primary outcomes, we did not rate down any studies for small sample sizes or low numbers of events.

We considered indirectness of evidence in the application of GRADE, but given that the populations, interventions and outcomes of included studies match those of interest for the review, we did not rate any of the evidence down for indirectness.

Given the lack of sufficiently homogeneous studies assessing the same intervention category and outcomes, we were unable to systematically investigate the presence of publication bias. There were generally no stark discrepancies between the described methods and the presented results in the included main studies. However, it is difficult to judge whether all planned analyses were conducted and reported since it is uncommon to publish a study protocol in this research field. Of the 42 main studies, only three cited a study protocol or described study registration (Aung 2016; Morfeld 2013; Morfeld 2014).

It should be emphasized that evaluating the appropriateness and quality of study design and analysis methods for such a heterogeneous body of evidence was challenging. In the absence of randomization, no gold standard exists to guide researchers undertaking such evaluations. Included studies handled key aspects of conduct — such as the definition of intervention and control sites, the incorporation of time in the analysis, and the duration of follow-up — very differently. In assessing changes in air quality associated with low emission zones, for example, some studies drew from intervention and control sites within the same city (Fensterer 2014), while others drew from areas further geographically removed (Boogaard 2012). In fact, two included studies (Friedman 2001; Peel 2010), both of which analyzed the effect of the traffic reduction strategies during the 1996 Atlanta Olympic Games, highlight the importance of some of these methodological aspects on the observed results. Friedman and colleagues assessed changes in acute care visits due to asthma in children in the five central counties of metropolitan Atlanta during the Olympic Games, as compared to four weeks before and four weeks after. They observed a significant decrease in childhood asthma associated with the intervention. However, Peel and colleagues improved upon and expanded the original analysis. They controlled for underlying time trends, assessed 10 years of data, and included control data from immediately outside Atlanta, other areas of Georgia, and other cities located in the US southeast. They observed no change in acute care visits for paediatric cardiorespiratory outcomes, including asthma, associated with the intervention. They found that reductions in ozone levels during the Olympics were due to regional meteorology and that the role of the traffic measures remained unclear. These divergent results illustrate that study design features, like the selection of appropriate control sites and study period, can affect not only the magnitude of the effect estimate, but also the direction of the effect, even when the considered studies are at a low risk of bias. Some studies conducted sensitivity analyses to assess the influence of selected methods on study results, but many studies were limited by available data. Thus some of the reported effect estimates are likely to be very dependent on the specific design and analysis methods applied.

It is important to consider how one might actually achieve higher quality evidence for, and thus a greater confidence in, the effectiveness of interventions to reduce ambient air pollution and their related health outcomes. Choice of study design and analysis methods plays a critical role. When conducting future intervention evaluations, researchers should strive to use the best possible study design and to make the best possible use of any routine or newly collected data. In undertaking evaluations, researchers should also ensure that they analyze their data in the most appropriate way, seeking additional statistical expertise where required. For example, where routine monitoring data are available pre- and post-intervention at both an intervention and control site, researchers should aim to conduct a cITS study. A cITS uses the underlying trend in the outcome to account for temporal changes not associated with the intervention, as well as a geographic control to account for contemporaneous changes occurring on a wider geographical scale not associated with the intervention. ITS, CBA and UBA studies do not inherently apply this level of control. The cITS study can thus ensure a lower risk of bias, as well as a richer understanding of the association between the intervention and various outcomes, compared to other NRS designs and analyses. Regarding the analysis, a range of methods may be applied, and providing general guidance is challenging; however certain aspects could be helpful across most cases. For controlled studies, for example, applying a difference-in-differences analysis approach is appropriate in most cases, as it accounts for any baseline differences in outcomes or other factors and provides a direct statistical comparison between intervention and control sites in calculating the intervention effect, provided an appropriate control population is selected.

When considering the overall summary of findings and the GRADE certainty of evidence ratings, it should be emphasized that difficulties in applying GRADE to complex public health interventions have been documented (Movsisyan 2016; Rehfuess 2013). In this review, for example, where no randomized evidence was identified, all of the primary outcomes assessed with GRADE were automatically rated as either 'low' or 'very low' certainty, which suggests that GRADE does not appropriately differentiate between NRS designs with moderate and low internal validity. These challenges and some criticism have led several ongoing efforts to further develop the GRADE approach, making it more suitable to reviews such as this, where much of the evidence base comprises NRS (Montgomery 2019), accepted for publication). The requirement that all non-randomized study designs begin the GRADE assessment at 'low' certainty, for example, will be relaxed provided the risk of bias of all included studies is rigorously assessed (Schünemann 2018). The newly developed ROBINS-I tool (Sterne 2016), designed specifically for cohort studies of interventions, along with a series of related tools still under development, would allow for a rigorous and appropriate risk of bias assessment. This is likely to better reflect the reality, context and range of study design and analysis methods applied in public health fields such as air pollution intervention research.

Potential biases in the review process

Throughout the conduct of the review, from the initial scoping stages to the interpretation and reporting of the evidence, we applied systematic, robust and transparent methods. We defined our review question and the exact parameters based on a system-based logic model. We conducted multi-disciplinary and multi-database electronic searches, and attempted to locate non-published literature. Our protocol was reviewed by a RAG consisting of air pollution researchers as well as decision makers who represent the potential end-users of this review. In order to better reflect the reality of the air pollution research field, we included a wide range of study designs, including the study designs normally included in EPOC reviews (Cochrane EPOC 2017), but also non-EPOC CBA studies; we included UBA studies as supporting studies. We summarized the heterogeneous evidence base narratively, but also created harvest plots with the aim of more effectively communicating the evidence. All of these methodological aspects were helpful in ensuring that the results reported here are both valid and relevant. There were, however, challenges in the review conduct, and some decisions we made may have led to the introduction of bias into the systematic review.

Although we developed a very broad search strategy, it is still possible that we were unable to identify some studies, especially if those were not published in journals indexed by electronic databases. Additionally, the most recent searches were conducted in August 2016; thus, studies published since then are not included in this review. Newer studies could potentially lead to a more complete and differing evidence base.

As described above, we included a wide range of study designs to ensure that we were capturing those studies considered as relevant and rigorous by air pollution researchers and decision makers. The classification of included studies into one of our included study designs was challenging, and it is possible that potentially eligible studies were misclassified. We aimed, however, to be inclusive at the screening stage with regard to study design and discussed any uncertainties at the full-text screening stage among at least three review authors to avoid such exclusion. Similarly, the distinction between the main studies, which contributed to the data on intervention effectiveness, and supporting studies, which are only reported descriptively, was difficult. However, these decisions were also always made in duplicate, often only after extensive discussion.

Many early accountability studies, as well as several more current studies, have taken an indirect approach to assessing the effects of interventions. Such studies usually apply observational methods, such as the cohort study design, to evaluate changes in outcomes over time, without directly linking these to interventions. One example of such a cohort study is the SAPALDIA study in Switzerland, which has measured changes in air pollution and the associated changes in health for more than two decades (Leuenberger 1994; Schindler 2009). Similar cohort-based studies linking changes in air quality to changes in health have been conducted in California (Gauderman 2015; Gilliland 2017), as well as the entire USA (Correia 2013; Dominici 2007; Pope 2009), and in the Netherlands (Boogaard 2013). Another important type of study, excluded from this review, are those in which participants self-select into lower exposure areas. In Avol 2001, also known as the Movers study, participants who moved from higher to lower pollution areas experienced improvements in respiratory function relative to those who remained in high pollution areas. Although these studies have provided valuable evidence on various interventions, the inclusion criteria of this review required studies to explicitly evaluate a clearly-defined intervention. The decision of whether a study can be explicitly linked to an intervention, however, was occasionally blurry, and it may be questionable whether all of the included studies offer a more direct evaluation of an intervention than several cohort studies that were excluded. Had we included cohort studies, this would have yielded a different evidence base, which may have influenced the results and interpretations of the review.

Assessments of air quality interventions have often relied on concentration-response functions from existing epidemiologic studies to model health outcomes resulting from measured or modelled changes in air quality. There are, however, well-known examples of accountability studies that have used modelled data to assess interventions. Cesaroni 2012, for example, used data on traffic volumes to calculate pollutant concentrations and to assess the effectiveness of the LEZ in Rome after its implementation. Another example evaluated the benefits associated with the US Clean Air Act across the USA by modelling predicted air pollution emissions reductions and the resulting health and cost benefits (US EPA 2011). Such predictive modelling studies were excluded from the current review. If such studies had been included, the resulting evidence base would have been different, and this may have influenced the results and interpretations of the review.

We defined interventions based on four categories, and there are thus certain types of interventions that are not covered by this review. Certain forms of personal protection, including masks and filtration systems, were not included. Additionally, we did not include studies assessing changes to agricultural practices. These types of interventions may also lead to improvements in air quality or reduced exposure to ambient air pollution, thus improvements in health, but this cannot be ascertained by this review.

The harvest plots, though efficient and very accessible for summarizing heterogeneous evidence on effectiveness of interventions, should not be seen as a replacement of the meta-analysis. Readers should be aware that the effects populating the harvest plots are those reported in the individual studies, and could be biased or underpowered, or both. Additionally, graphical summary techniques like the harvest plot have been criticized because they may encourage 'vote-counting' practices, if end-users attempt to quantitatively compare the frequency of effect directions (Thomson 2012; Higgins 2019). This practice is explicitly discouraged in association with harvest plots, and readers are encouraged to carefully read the detailed narrative summary. They also rely on significance testing and P values for arranging the bars into columns, and such practices have also been criticized for relying too heavily on arbitrary significance values (Sterne 2001). We argue, however, that our use of the harvest plots represents a conservative interpretation of effect estimates from individual studies that is biased towards the null, and thus avoids the potential danger of describing misleading changes in outcomes from imprecise and underpowered analyses.

We made several changes after publication of the protocol; these are listed below in the Differences between protocol and review section. Some of these differences, for example the differentiation between main and supporting studies or the use of the NICE-modified GATE tool only, rather than in combination with the Cochrane EPOC 'Risk of bias' tool, may have influenced the results of the review. These decisions, however, were based solely on methodological considerations and problems, and were made without consideration of study results.

Agreements and disagreements with other studies or reviews

Several reviews of air pollution intervention studies have been published recently (Bell 2011; Boogaard 2017; Henneman 2017; Henschel 2012; Rich 2017; van Erp 2012). None of these reviews, however, applied systematic and transparent methods; only one review's authors described their methods for identifying studies (Henschel 2012), and none applied systematic methods for searching and selecting included studies. Rather than aiming to comprehensively describe all interventions that have been evaluated, as we have done, these reviews primarily aimed to describe the current state of knowledge through the use of illustrative examples.

Only one review drew any general conclusions with respect to the effectiveness of interventions, suggesting that based on the evidence, decreases in air pollution due to interventions or other external events were associated with improvements in health outcomes (Henschel 2012). The heterogeneous evidence base we identified did not entirely support this overall conclusion with respect to effectiveness.

Although the scope and methods of these reviews differ, there are several similarities in the results and interpretations that are in line with our systematic review. The reviews, for example, discuss the complexity of the system in which these interventions are implemented, and the resulting challenges researchers face in assessing the effectiveness, including accounting for confounders and underlying trends in the outcomes, as well as decisions around the appropriate length of follow-up and appropriate control populations (Boogaard 2017; Henneman 2017; Rich 2017; van Erp 2012). They also highlight the challenges presented to review authors in comparing across individual studies, due to the heterogeneity of study design and analysis methods (Bell 2011; Henschel 2012). Each review additionally suggested several ways forward, many of which are supported by our findings, including the need for more consistent methodology across studies (Bell 2011; Henschel 2012), prospective evaluations of interventions (Henneman 2017; van Erp 2012), and the further development of methods for intervention evaluation (Boogaard 2017; Henneman 2017).

AUTHORS' CONCLUSIONS

Implications for practice

Air pollutant concentrations are high and still increasing in many parts of the world, in particular in LMICs (van Donkelaar 2015). Even in HICs, where levels have decreased markedly over the past decades, substantial health effects due to air pollution are still being observed (Di 2017; Pinault 2017). The overall burden from outdoor air pollution remains very large (Gakidou 2017), thus it is imperative that policies aiming to improve air quality and associated health outcomes be put in place to protect the health of populations in both HICs and LMICs.

It is especially important for measures to be implemented in areas where few or none exist. We identified few or no studies from several parts of the world, including Africa, the Middle East, Eastern Europe, Central Asia and Southeast Asia. It is likely that some interventions have been implemented and simply not evaluated, but we suspect that this also indicates a general lack of interventions being put into place. Thus decision-makers should prioritize the development and implementation of appropriate interventions in these settings. With the identified evidence base, we were not able to provide a simple answer regarding 'what works'. The choice of specific intervention is context-dependent; in an area where a single pollutant source contributes heavily to concentrations, an intervention aiming to reduce concentrations from this source may be appropriate. In many cases, however, several sources contribute substantially to ambient air pollution, and a more systemic, multi-component approach may be necessary. Indeed in areas where ambient air pollution is still very high and where few or no interventions exist, coordinated and comprehensive measures at the national level are likely to be appropriate. Thus in developing and implementing interventions, decision-makers will need to consult the international evidence, for which the studies included in this review can serve as a valuable resource. In addition, they will need to conduct local analyses to determine what is most appropriate in a given context.

To ensure a better future understanding of 'what works', it is important that decision-makers help ensure high-quality evaluations. Such high-quality evaluations undertaken in different settings and countries should ideally follow an internationally agreed evaluation framework that encourages a more systematic assessment and facilitates comparisons across studies. Air pollution interventions, and especially long-term regulatory programmes, would benefit from having an evaluation component built into them from the start (Boogaard 2017). Such a system of contemporaneous evaluation would also require a system for reliable tracking of both air quality and health outcomes data over the long term, including quality assurance of the data and making them publicly available (Boogaard 2017). Concomitant and potentially more in-depth evaluations could also comprise process evaluations, providing important insights into the fidelity, feasibility, quality of implementation and causal mechanisms related to interventions and their effects for different population groups (Moore 2015).

Implications for research

It is likely that there are many ambient air pollution interventions that have yet to be evaluated, and researchers with experience in accountability research could look for opportunities to evaluate existing and future interventions. Through the conduct of further evaluations the evidence base may become more complete, which may help to further address the ambiguity surrounding what types of interventions work the best, in what populations and in what contexts.

To make future evaluations of ambient air pollution interventions more policy-relevant, it would be helpful if researchers focused on producing more uniform and internally valid evidence that can be readily compared and synthesized with other studies. Researchers should focus on important outcomes widely available through routine data, such as mortality and PM_{10}, $PM_{2.5}$ or other criteria pollutants. Quasi-experimental study designs are increasingly being applied in public health research (Bärnighausen 2017; Craig 2017). Several included studies already employed such designs (Bel 2013a; Carrillo 2016; Deschênes 2012; Giovanis 2015; Mullins 2014; Viard 2015), and more of these evaluations will ensure a more internally valid and methodologically homogeneous evidence base, which can be more readily synthesized (Becker 2017). In addition, new promising methods have been developed for accountability research, including use of causal inference methods (Hubbell 2014; Zigler 2014; Zigler 2016). These and other approaches that would improve the ability to attribute changes in air quality and health directly to an intervention should continue to be advanced and applied.

Similarly, an evaluation of effectiveness may not be sufficient for informing policy; future evaluations should also focus on other important aspects. These include, for example, unintended and adverse events and cost-effectiveness, as well as process-related outcomes, such as intervention fidelity, feasibility and acceptability. This would be helpful for future implementation and adaptation of interventions.

Studies assessing interventions aiming to reduce ambient air pollution are, like other epidemiological studies, susceptible to confounding. In particular, it is challenging to appropriately account for factors other than the intervention that also affect air quality and health. Therefore, the use of appropriate comparison populations or outcomes (i.e. negative controls) unaffected by the intervention and accounting for underlying background trends in outcomes is important for future studies. Specific rigorously conducted included studies accounted for these aspects; Pope 2007, for example, assessed a series of various geographical controls in assessing the intervention effect, Peel 2010 analyzed a 10-year time series to account for underlying trends in hospitalizations, and Yorifuji 2016 assessed changes in non-cardiovascular, non-respiratory deaths, where no change would be expected due to the intervention. Additionally, the conduct and transparent reporting of sensitivity analyses to evaluate, for example, choices of comparison populations and of statistical models adjusting for background trends, should be undertaken, so as to provide readers with an understanding of the uncertainty of the effect (Boogaard 2017).

Future studies should also focus on complete and detailed reporting of all study aspects. In order for studies to effectively inform policy, all aspects should be comprehensively reported, including the populations, intervention, outcomes and study methods. Relevant published reporting guidelines, such as the CONSORT statement for randomized studies (Schulz 2010), the STROBE statement for observational studies (Vandenbroucke 2007) and the TREND statement for non-randomized evaluations (Des Jarlais 2004), are a good starting point, but even these may not be sufficient. Where possible, authors should go beyond describing these aspects in a brief overview; rather than describing the intervention simply as a "low emission zone", for example, authors should describe when the LEZ was implemented, the reach of the LEZ, whether and how the policy was enforced, whether certain vehicle types were excepted, along with any further details that may help readers understand what actually occurred. The TIDier and the TIDier-PHP checklists for better intervention reporting can help facilitate comprehensive intervention description (Hoffmann 2014; Campbell 2018). Similarly, all aspects should be described in detail; where air quality monitors are used, information on the geographic location of monitors, as well as the nature of monitoring sites (e.g. streetside, urban background, suburban background) should be provided. In reporting results authors should provide effect estimates, as well as some measure of variance, such as the 95% confidence interval. Detailed information on context and implementation issues, additionally, can complement traditional evaluations, and may indeed be critical in understanding the effectiveness of interventions (Pfadenhauer 2017); researchers conducting evaluations should strive to include a structured and comprehensive assessment of these aspects. Most journals encourage such detailed reporting, allowing authors to provide additional details in appendices and supplemental material. Additionally, a more concrete conceptualization of the intervention and the system at the onset of research, using, for example, the logic model, may help strengthen the design, conduct and reporting of intervention evaluations (Rehfuess 2017; Rohwer 2017).

From a review perspective, we categorized interventions broadly based on the source targeted, which resulted in us identifying a range of different interventions within each category. Future systematic reviews of interventions aiming to reduce ambient air pollution could consider a more granular categorization of interventions, which may result in a more homogeneous evidence base within categories that could be more readily synthesized.

ACKNOWLEDGEMENTS

We would like to acknowledge the generous comments and suggestions made by our Review Advisory Group at the protocol development stage. Their input has helped to ensure that the review will meet its intended goal of assessing the effectiveness of ambient PM interventions in a systematic and comprehensive way and that the review will appropriately inform policy.

REFERENCES

References to studies included in this review

Allen 2009 *{published data only}*

Allen RW, Leckie S, Millar G, Brauer M. The impact of wood stove technology upgrades on indoor residential air quality. *Atmospheric Environment* 2009;**43**(37):5908-15.

Atkinson 2009 *{published data only}*

Atkinson RW, Barratt B, Armstrong B, Anderson HR, Beevers SD, Mudway IS, et al. The impact of the congestion charging scheme on ambient air pollution concentrations in London. *Atmospheric Environment* 2009;**43**(34):5493-500.

Aung 2016 *{published data only}*

Aung TW, Jain G, Sethuraman K, Baumgartner J, Reynolds C, Grieshop A, et al. Health and climate-relevant pollutant concentrations from a carbon-finance approved cookstove intervention in rural India. *Environmental Science & Technology* 2016;**50**(13):7228-38.

Bel 2013a *{published data only}*

Bel G, Rosell J. Effects of the 80 km/h and variable speed limits on air pollution in the metropolitan area of Barcelona. *Transportation Research Part D: Transport and Environment* 2013;**23**:90-7.

Bel 2013b *{published data only}*

Bel G, Rosell J. Effects of the 80 km/h and variable speed limits on air pollution in the metropolitan area of Barcelona. *Transportation Research Part D: Transport and Environment* 2013;**23**:90-7.

Boogaard 2012 *{published data only}*

Boogaard H, Janssen NAH, Fischer PH, Kos GP, Weijers EP, Cassee FR, et al. Impact of low emission zones and local traffic policies on ambient air pollution concentrations. *Science of the Total Environment* 2012;**435-436**:132-40.

Burr 2004 *{published data only}*

Burr M, Karani G, Davies B, Holmes B, Williams K. Effects on respiratory health of a reduction in air pollution from vehicle exhaust emissions. *Occupational and Environmental Medicine* 2004;**61**(3):212-8.

Butler 2011 *{published data only}*

Butler TJ, Vermeylen FM, Rury M, Likens GE, Lee B, Bowker GE, et al. Response of ozone and nitrate to stationary source NOx emission reductions in the eastern USA. *Atmospheric Environment* 2011;**45**(5):1084-94.

Carrillo 2016 *{published data only}*

Carrillo PE, Malik AS, Yoo Y. Driving restrictions that work? Quito's Pico y Placa Program. *Canadian Journal of Economics [Revue Canadienne d'Economique]* 2016;**49**(4):1536-68.

Clancy 2002 *{published data only}*

Clancy L, Goodman P, Sinclair H, Docker DW. Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *Lancet* 2002;**360**(9341):1210-4.

Cowie 2012 *{published data only}*

Cowie CT, Rose N, Gillett R, Walter S, Marks GB. Redistribution of traffic related air pollution associated with a new road tunnel. *Environmental Science & Technology* 2012;**46**(5):2918-27.

Davis 2008 *{published data only}*

Davis LW. The Effect of Driving Restrictions on Air Quality in Mexico City. *Journal of Political Economy* 2008;**116**(1):38-81.

Deschênes 2012 *{published data only}*

Deschênes O, Greenstone M, Shapiro JS. Defensive investments and the demand for air quality: Evidence from the NOx Budget Program. *American Economic Review* 2012;**107**(10):2958-89.

Dijkema 2008 *{published data only}*

Dijkema MBA, van der Zee, SC, Brunekreef B, van Strien RT. Air quality effects of an urban highway speed limit reduction. *Atmospheric Environment* 2008;**42**(40):9098-105.

Dockery 2013a *{published data only}*

Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. Effect of air pollution control on mortality and hospital admissions in Ireland. *Research Report (Health Effects Institute)* 2013;**176**:3-109.

Dockery 2013b *{published data only}*

Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. Effect of air pollution control on mortality and hospital admissions in Ireland. *Research Report (Health Effects Institute)* 2013;**176**:3-109.

Dockery 2013c *{published data only}*

Dockery DW, Rich DQ, Goodman PG, Clancy L, Ohman-Strickland P, George P, et al. Effect of air pollution control on mortality and hospital admissions in Ireland. *Research Report (Health Effects Institute)* 2013;**176**:3-109.

Dolislager 1997 *{published data only}*

Dolislager LJ. The effect of California's wintertime oxygenated fuels program on ambient carbon monoxide. *Journal of the Air & Waste Management Association* 1997;**47**(7):775-83.

El-Zein 2007 *{published data only}*

El-Zein A, Nuwayhid I, El-Fadel M, Mroueh S. Did a ban on diesel-fuel reduce emergency respiratory admissions for children?. *Science of the Total Environment* 2007;**384**(1-3):134-40.

Fensterer 2014 *{published data only}*

Fensterer V, Kuchenhoff H, Maier V, Wichmann H, Breitner S, Peters A, et al. Evaluation of the impact of low emission zone and heavy traffic ban in Munich (Germany) on the reduction of PM10 in ambient air. *International Journal of Environmental Research and Public Health* 2014;**11**(5):5094-112.

Friedman 2001 *{published data only}*

Friedman MS, Powell KE, Hutwagner L, Graham LM, Teague WG. Impact of changes in transportation and commuting behaviors during the 1996 Summer Olympic Games in Atlanta on air quality and childhood asthma. *Journal of the American Medical Association* 2001;**285**(7):897-905.

Gallego 2013a *{published data only}*

Gallego F, Montero JP, Salas C. The effect of transport policies on car use: Evidence from Latin American cities. *Journal of Public Economics* 2013;**107**:47-62.

Gallego 2013b *{published data only}*

Gallego F, Montero JP, Salas C. The effect of transport policies on car use: Evidence from Latin American cities. *Journal of Public Economics* 2013;**107**:47-62.

Giovanis 2015 *{published data only}*

Giovanis E. Evaluation of ozone smog alerts on actual ozone concentrations: A case study in North Carolina. *International Journal of Environmental Technology and Management* 2014;**18**(5-6):465-77.

Gramsch 2013 *{published data only}*

Gramsch E, Nir GL, Araya M, Rubio MA, Moreno F, Oyola M. Influence of large changes in public transportation (Transantiago) on the black carbon pollution near streets. *Atmospheric Environment* 2013;**65**:153-63.

Hasunuma 2014 *{published data only}*

Hasunuma H, Ishimaru Y, Yoda Y, Shima M. Decline of ambient air pollution levels due to measures to control automobile emissions and effects on the prevalence of respiratory and allergic disorders among children in Japan. *Environmental Research* 2014;**131**:111-8.

Johnston 2013 *{published data only}*

Johnston FH, Hanigan IC, Henderson SB, Morgan GG. Evaluation of interventions to reduce air pollution from biomass smoke on mortality in Launceston, Australia: retrospective analysis of daily mortality, 1994-2007. *BMJ* 2013;**346**:e8446.

Kim 2011 *{published data only}*

Kim KH, Shon ZH. Long-term changes in PM10 levels in urban air in relation with air quality control efforts. *Astmospheric Environment* 2011;**45**(19):3309-17.

Li 2011 *{published data only}*

Li Y, Wang W, Wang J, Zhang X, Lin W, Yang Y. Impact of air pollution control measures and weather conditions on asthma during the 2008 Summer Olympic Games in Beijing. *International Journal of Biometeorology* 2011;**55**(4):547-54.

Lin 2013 *{published data only}*

Lin S, Jones R, Pantea C, Ozkaynak H, Rao ST, Hwang SA, Garcia VC. Impact of NOx emissions reduction policy on hospitalizations for respiratory disease in New York State. *Journal of Exposure Science & Environmental Epidemiology* 2013;**23**(1):73-80.

Morfeld 2013 *{published data only}*

Morfeld P, Stern R, Builtjes P, Groneberg DA, Spallek M. Introduction of a low-emission zone and the effect on air pollutant concentration of particulate matter (PM_{10}) — a pilot study in Munich [Einrichtung einer Umweltzone und ihre Wirksamkeit auf die PM10-Feinstaubkonzentration — eine Pilotanalyse am Beispiel München]. *Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie* 2013;**63**(2):104-15.

Morfeld 2014 *{published data only}*

Morfeld P, Groneberg DA, Spallek MF. Effectiveness of low emission zones: Large scale analysis of changes in environmental NO_2, NO and NOx concentrations in 17 German cities. *PloS One* 2014;**9**(8):e102999.

Mullins 2014 *{published data only}*

Mullins J, Bharadwaj P. Effects of short-term measures to curb air pollution: Evidence form Santiago, Chile. *American Journal of Agricultural Economics* 2014;**97**(4):1107-34.

Peel 2010 *{published data only}*

Peel JL, Klein M, Flanders WD, Mulholland JA, Tolbert PE, HEI Health Review Committee. Impact of improved air quality during the 1996 Summer Olympic Games in Atlanta on multiple cardiovascular and respiratory outcomes. *Research Report (Health Effects Institute)* 2010;**148**:3-23.

Pope 2007 *{published data only}*

Pope CA, Rodermund DL, Gee MM. Mortality effects of a copper smelter strike and reduced ambient sulfate particulate matter air pollution. *Environmental Health Perspectives* 2007;**115**(5):679-83.

Ruprecht 2009 *{published data only}*

Ruprecht AA, Invernizzi G. The effects of the traffic restriction fee (Ecopass) in the center of Milan on urban pollution with particulate matter: The results of a pilot study. *Epidemiologia e Prevenzione* 2009;**33**(1-2):21-6.

Saaroni 2010 *{published data only}*

Saaroni H, Chudnovsky A, Ben-Dor E. Reflectance spectroscopy is an effective tool for monitoring soot pollution in an urban suburb. *Science of the Total Environment* 2010;**408**(5):1102-10.

Sajjadi 2011 *{published data only}*

Sajjadi SA, Bridgman HA. Respiratory Hospital Admissions before and after Closure of a Major Industry in the Lower Hunter Region, Australia. *Iranian Journal of Public Health* 2011;**40**(3):41-54.

Sajjadi 2012 *{published data only}*

Sajjadi SA, Bridgman HA, Sadeghi N. Changes in Air Quality due to Closure of a Major Industry. *International Journal of Collaborate Research on International Medicine & Public Health* 2012;**4**(6):1196-214.

Tanaka 2015 *{published data only}*

Tanaka S. Environmental regulations on air pollution in China and their impact on infant mortality. *Journal of Health Economics* 2015;**42**:90-103.

Titos 2015a *{published data only}*

Titos G, Lyamani H, Drinovec L, Olmo FJ, Mocnik G, Alados-Arboledas A. Evaluation of the impact of transportation changes on air quality. *Atmospheric Environment* 2015;**114**:19-31.

Titos 2015b *{published data only}*

Titos G, Lyamani H, Drinovec L, Olmo FJ, Mocnik G, Alados-Arboledas A. Evaluation of the impact of transportation changes on air quality. *Atmospheric Environment* 2015;**114**:19-31.

Viard 2015 *{published data only}*

Viard VB, Fu S. The effect of Beijing's driving restrictions on pollution and economic activity. *Journal of Public Economics* 2015;**125**(C):98-115.

Yap 2015 *{published data only}*

Yap PS, Garcia C. Effectiveness of residential wood-burning regulation on decreasing particulate matter levels and hospitalizations in the San Joaquin Valley Air Basin. *American Journal of Public Health* 2015;**105**(4):772-8.

Yorifuji 2011 *{published data only}*

Yorifuji T, Kawachi I, Kaneda M, Takao S, Kashima S, Doi H. Diesel vehicle emission and death rates in Tokyo, Japan: a natural experiment. *Science of the Total Environment* 2011;**409**(19):3620-7.

Yorifuji 2016 *{published data only}*

Yorifuji T, Saori K, Hiroyuki D. Fine particulate air pollution from diesel emission control and mortality rates in Tokyo: A quasi-experimental study. *Epidemiology* 2016;**27**(6):769-78.

Zigler 2016 *{published data only}*

Zigler CM, Kim C, Choirat C, Hansen JB, Wang Y, Hund L, et al. Causal inference methods for estimating long-term health effects of air quality regulations. *Research Report (Health Effects Institute)* 2016;**187**:5-49.

References to studies excluded from this review

Adar 2015 *{published data only}*

Adar SD, D'Souza J, Sheppard L, Kaufman JD, Hallstrand TS, Davey ME, et al. Adopting Clean Fuels and Technologies on School Buses Pollution and Health Impacts in Children. *American Journal of Respiratory and Critical Care Medicine* 2015;**191**(12):1413-21.

Ai 2016 *{published data only}*

Ai Z, Mak C, Lee H. Roadside air quality and implications for control measures: A case study of Hong Kong. Atmospheric air quality and implications for control measures: A case study of Hong Kong. *Atmospheric Environment* 2016;**137**:6-16.

Ali 2008 *{published data only}*

Ali M, Athar M. Air pollution due to traffic, air quality monitoring along three sections of National Highway N-5, Pakistan. *Environmental Monitoring and Assessment* 2008;**136**(1-3):219-26.

Altemose 2015 *{published data only}*

Altemose B, Gong J, Zhu T, Hu M, Zhang L, Cheng H, et al. Aldehydes in Relation to Air Pollution Sources: A Case Study around the Beijing Olympics. *Atmospheric Environment* 2015;**109**:61-9.

Alvim-Ferraz 2005 *{published data only}*

Alvim-Ferraz MCM, Pereira MC, Ferraz JM, Almeida e Mello AMC, Martins FG. European directives for air quality: Analysis of the new limits in comparison with asthmatic symptoms in children living in the Oporto Metropolitan Area, Portugal. *Human and Ecological Risk Assessment* 2005;**11**(3):607-16.

Ancelet 2015 *{published data only}*

Ancelet T, Davy P, Trompetter W. Particulate matter sources and long-term trends in a small New Zealand city. *Atmospheric Pollution Research* 2015;**6**(6):1105-12.

Arossa 1987 *{published data only}*

Arossa W, Spinaci S, Bugiani M, Natale P, Bucca C, De Candussio G. Changes in lung function of children after air pollution decrease. *Archives of Environmental Health* 1987;**42**(3):170-4.

Auffhammer 2009 *{published data only}*

Auffhammer M, Bento AM, Lowe SE. Measuring the effects of the Clean Air Act Amendments on ambient PM10 concentrations: The critical importance of a spatially disaggregated analysis. *Journal of Environmental Economics and Management* 2009;**58**(1):15-26.

Auffhammer 2011 *{published data only}*

Auffhammer M, Bento AM, Lowe SE. The city-level effects of the 1990 Clean Air Act Amendments. *Land Economics* 2011;**87**(1):1-18.

Aunan 1998 *{published data only}*

Aunan K, Patzay G, Asbjorn AH, Seip HM. Health and environmental benefits from air pollution reductions in Hungary. *Science of the Total Environment* 1998;**212**(2-3):245-68.

Aunan 2004 *{published data only}*

Aunan K, Fang J, Vennemo H, Oye K, Seip H. Co-benefits of climate policy - lessons learned from a study in Shanxi, China. *Energy Policy* 2004;**32**(4):567-81.

Aydin 2009 *{published data only}*

Aydin CC, Karadavut S. The impact of increasing natural gas use on the environment and population in Ankara, Turkey. *Fresenius Environmental Bulletin* 2009;**18**(9):1559-66.

Baldasano 2010 *{published data only}*

Baldasano JM, Goncalves M, Soret A, Jimenez-Guerrero P. Air pollution impacts of speed limitation measures in large cities: The need for improving traffic data in a metropolitan area. *Atmospheric Environment* 2010;**44**(25):2997-3006.

Barbose 2016 *{published data only}*

Barbose G, Wiser R, Heeter J, Mai T, Bird L, Bolinger M, et al. A retrospective analysis of benefits and impacts of U.S. renewable portfolio standards. *Energy Policy* 2016;**96**:645-60.

Barnes 2015 *{published data only}*

Barnes JH, Hayes T, Longhurst J. Has UK local government action improved local air quality? A Bristol case study. *WIT Transactions on Ecology and the Environment* 2015;**198**:243-54.

Barratt 2014 *{published data only}*

Barratt BM, Fuller GW. Intervention assessments in the control of PM10 emissions from an urban waste transfer station. *Environmental Science: Processes & Impacts* 2014;**16**(6):1328-37.

Bartonova 1999 *{published data only}*

Bartonova A, Clench-Aas J, Gram F, Gronskei KE, Guerreiro C, Larssen S, et al. Air pollution exposure monitoring and estimation. Part V. Traffic exposure in adults. *Journal of Environmental Monitoring* 1999;**1**(4):337-40.

Bauman 1977 *{published data only}*

Bauman RD, Crenshaw JD. Air quality implications of the move toward coal (air pollution impact of increased coal use in the utility sector). Proceedings of the Intersociety Energy Conversion Engineering Conference. Washington D.C., 1977:629-36.

Beevers 2005 *{published data only}*

Beevers SD, Carslaw DC. The impact of congestion charging on vehicle emissions in London. *Atmospheric Environment* 2005;**39**(1):1-5.

Bennett 2010 *{published data only}*

Bennett CM, Dharmage SC, Matheson M, Gras JL, Markos J, Meszaros D, et al. Ambient wood smoke exposure and respiratory symptoms in Tasmania, Australia. *Science of the Total Environment* 2010;**409**(2):294-9.

Berhane 2016 *{published data only}*

Berhane K, Chang CC, McConnell R, Gauderman WJ, Avol E, Rapapport E, et al. Association of changes in air quality with bronchitic symptoms in children in California, 1993-2012. *JAMA* 2016;**315**(14):1491-501.

Bridgman 2002 *{published data only}*

Bridgman HA, Davies TD, Jickells T, Hunova I, Tovey K, Bridges K, et al. Air pollution in the Krusne Hory region, Czech Republic during the 1990s. *Atmospheric Environment* 2002;**36**(21):3375-89.

Buckley 2011 *{published data only}*

Buckley SM, Mitchell MJ. Improvements in urban air quality: Case studies from the New York State, USA. *Water, Air, and Soil Pollution* 2011;**214**(1-4):93-106.

Carvalho 2015 *{published data only}*

Carvalho VSB, Freitas ED, Martins LD, Martins JA, Mazzoli CR, Andrade MF. Air quality status and trends over the Metropolitan Area of São Paulo, Brazil as a result of emission control policies. *Environmental Science & Policy* 2015;**47**:68-79.

Cesaroni 2012 *{published data only}*

Cesaroni G, Boogaard H, Jonkers S, Porta D, Badaloni C, Cattani G, et al. Health benefits of traffic-related air pollution reduction in different socioeconomic groups: The effect of low-emission zoning in Rome. *Occupational and Environmental Medicine* 2012;**69**(2):133-9.

Chalbot 2014 *{published data only}*

Chalbot MCG, Jones TA, Kavouras IG. Trends of Non-Accidental, Cardiovascular, Stroke and Lung Cancer Mortality in Arkansas Are Associated with Ambient PM(2.5) Reductions. *International Journal of Environmental Research and Public Health* 2014;**11**(7):7442-55.

Chang 2007 *{published data only}*

Chang SC, Lee CT. Evaluation of the trend of air quality in Taipei, Taiwan from 1994 to 2003. *Environmental Monitoring and Assessment* 2007;**127**(1-3):87-96.

Chang 2008 *{published data only}*

Chang SC, Lee CT. Evaluation of the temporal variations of air quality in Taipei City, Taiwan, from 1994 to 2003. *Journal of Environmental Management* 2008;**86**(4):627-35.

Chay 2003 *{published data only}*

Chay K, Dobkin C, Greenstone M. The Clean Air Act of 1970 and Adult mortality. *Journal of Risk and Uncertainty* 2003;**27**(3):279-300.

Chen 2014 *{published data only}*

Chen L, Liu CL, Pan T, Chen CC, Li Z, Wang HH, et al. Assessment of the effect of pm2.5 reduction by plain afforestation project in Beijing based on dry deposition model. *Chinese Journal of Ecology* 2014;**33**(11):2897-904.

Chiesa 2014 *{published data only}*

Chiesa M, Perrone MG, Cusumano N, Ferrero L, Sangiorgi G, Bolzacchini E, et al. An environmental, economical and socio-political analysis of a variety of urban air-pollution reduction policies for primary PM10 and NOx: The case study of the Province of Milan (Northern Italy). *Environmental Science & Policy* 2014;**44**:39-50.

Chong 2014 *{published data only}*

Chong U, Yim SSHL, Barrett SRH, Boies AM. Air Quality and Climate Impacts of Alternative Bus Technologies in Greater London. *Environmental Science & Technology* 2014;**48**(8):4613-22.

Chou 2007 *{published data only}*

Chou CM, Chang YM, Lin W, Tseng C, Chen L. Evaluations of street sweeping and washing to reduce ambient PM10. *International Journal of Environment and Pollution* 2007;**31**(3-4):431-48.

Chou 2011 *{published data only}*

Chou CCK, Tsai CY, Chang CC, Lin PH, Liu SC, Zhu T. Photochemical production of ozone in Beijing during the 2008 Olympic Games. *Atmospheric Chemistry and Physics* 2011;**11**(18):9825-37.

Correia 2013 *{published data only}*

Correia AW, Pope CA, Dockery DW, Wang Y, Ezzati M, Dominici F. Effect of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 US Counties for the Period from 2000 to 2007. *Epidemiology* 2013;**24**(1):23-31.

Cox 2015 *{published data only}*

Cox LA, Popken DA. Has reducing fine particulate matter and ozone caused reduced mortality rates in the United States?. *Annals of Epidemiology* 2015;**25**(3):162-73.

Crippa 2016 *{published data only}*

Crippa M, Janssens-Maenhout G, Dentener F, Guizzardi D, Sindelarova K, Muntean M, et al. Forty years of improvements in European air quality: regional policy-industry interactions with global impacts. *Atmospheric Chemistry and Physics* 2016;**16**:3825-41.

Critchley 2015 *{published data only}*

Critchley K, Teather K, Hughes H, Gibson M. Air quality, respiratory health and wood use for women converting from low- to high-efficiency stoves in rural Kenya. *WIT Transactions on Ecology and the Environment* 2015;**198**:205-16.

Cropper 1997 *{published data only}*

Cropper ML, Simon NB, Alberini A, Arora S, Sharma PK. The health benefits of air pollution control in Delhi. *American Journal of Agricultural Economics* 1997;**79**(5):1625-9.

Cruz-Minguillon 2009 *{published data only}*

Cruz-Minguillon M, Monfort E, Querol X, Alastuey A, Celades I, Miro J. Effect of ceramic industrial particulate emission control on key components of ambient PM10. *Journal of Environmental Management* 2009;**90**(8):2558-67.

Cyrys 2014 *{published data only}*

Cyrys J, Peters A, Soentgen J, Wichmann HE. Low emission zones reduce PM10 mass concentrations and diesel soot in German cities. *Journal of the Air & Waste Management Association* 2014;**64**(4):481-7.

Cyrys 2015 *{published data only}*

Cyrys J, Peters A, Soentgen J, Gu J, Wichmann HE. Umweltzonen. *Umweltmedizin - Hygiene - Arbeitsmedizin* 2015;**20**(1):33-57.

Delkash 2016 *{published data only}*

Delkash M, Hussein MM. Examining some potential actions in mitigating gaseous emissions from vehicles, case study: Tehran. *Air Quality, Atmosphere, & Health* 2016;**9**(8):909-21.

DeLuca 2012 *{published data only}*

DeLuca PF, Corr D, Wallace J, Kanaroglou P. Effective mitigation efforts to reduce road dust near industrial sites: Assessment by mobile pollution surveys. *Journal of Environmental Management* 2012;**98**:112-8.

Dickinson 2015 *{published data only}*

Dickinson KL, Kanyomse E, Piedrahita R, Coffey E, Rivera IJ, Adoctor J, et al. Research on Emissions, Air quality, Climate, and Cooking Technologies in Northern Ghana (REACCTING): Study rationale and protocol. *BMC Public Health* 2015;**15**:126.

Dienes 2014 *{published data only}*

Dienes C, Aue A. On-line monitoring of pollution concentrations with autoregressive moving average time series. *Journal of Time Series Analysis* 2014;**35**(3):239-61.

Ding 2016 *{published data only}*

Ding D, Zhu Y, Jang C, Lin C, Wang S, Fu J, et al. Evaluation of health benefit using BenMAP-CE with an integrated scheme of model and monitor data during Guangzhou Asian Games. *Journal of Environmental Sciences* 2016;**42**:9-18.

Dong 2010 *{published data only}*

Dong Y, Liu W, Liu J Liu Y, Han D, et al. Application study of Lidar in urban traffic pollution. *Acta Optica Sinica* 2010;**30**(2):315-20.

Escobedo 2009 *{published data only}*

Escobedo FJ, Nowak DJ. Spatial heterogeneity and air pollution removal by an urban forest. *Landscape and Urban Planning* 2009;**90**(3-4):102-10.

Federal Highway Administration 2014 *{published data only}*

Federal Highway Administration, U.S. Department of Transportation. Congestion Mitigation and Air Quality Improvement (CMAQ) Program. Federal Register 2014:45146-51.

Fernandez-Camacho 2016 *{published data only}*

Fernandez-Camacho R, de la Rosa JD, Sanchez de la Campa AM. Trends and sources vs air mass origins in a major city in South-western Europe: Implications for air quality management. *Science of the Total Environment* 2016;**553**:305-15.

Foster 2011 *{published data only}*

Foster A, Kumar N. Health effects of air quality regulations in Delhi, India. *Atmospheric Environment* 2011;**45**(9):1675-83.

Frye 2003 *{published data only}*

Frye C, Hoelscher B, Cyrys J, Wjst M, Wichmann HE, Heinrich J. Association of lung function with declining ambient air pollution. *Environmental Health Perspectives* 2003;**111**(3):383-7.

Gallagher 2013 *{published data only}*

Gallagher J, Gill LW, McNabola, A. The passive control of air pollution exposure in Dublin, Ireland: A combined measurement and modelling case study. *Science of the Total Environment* 2013;**458-460**:331-43.

Gao 2013 *{published data only}*

Gao X, Nie W, Xue L, Wang T, Wang X, Gao R, et al. Highly time-resolved measurements of secondary ions in pm2.5 during the 2008 Beijing Olympics: the impacts of control measures and regional transport. *Aerosol and Air Quality Research* 2013;**13**(1):367-76.

Gao 2014 *{published data only}*

Gao YB, Mao XQ, Gabriel C. Assessment of co-control effects for air pollutants and greenhouse gases in urban transport: A case study in Urumqi. *China Environmental Science* 2014;**34**(11):2985-92.

Geng 2014 *{published data only}*

Geng Y, Peng C, Zhang J. Contributing role of the 2008 Summer "Green" Olympics to air pollution control in Beijing. WIT Transactions on Engineering Sciences. 2014; Vol. 1:477-84.

Gertler 1999 *{published data only}*

Gertler AW, Sagebiel JC, Dippel WA, O'Connor M. The impact of California Phase 2 reformulated gasoline on real-world vehicle emissions. *Journal of the Air & Waste Management Association* 1999;**49**(11):1339-46.

Gioda 2016 *{published data only}*

Gioda A, Ventura LMB, Ramos MB, Silva MPR. Half century monitoring air pollution in a megacity: A case study of Rio de Janeiro. *Water, Air, and Soil Pollution* 2016;**227**(3):86.

Giuliano 2007 *{published data only}*

Giuliano G, O'Brien T, Hayden S, Dell'aquila P. Reducing port-related truck emissions: The terminal gate appointment system at the Ports of Los Angeles and Long Beach. *Transportation Research Part D: Transport and Environment* 2007;**12**(7):460-73.

Grinshpun 2014 *{published data only}*

Grinshpun SA, Yermakov M, Reponen T, Simmons M, LeMasters GK, Ryan PH. Traffic Particles in Ambient Air of a Major US Urban Area: Has Anything Changed over a Decade?. *Aerosol and Air Quality Research* 2014;**14**:1344-51.

Hao 2006 *{published data only}*

Hao J, Hu J, Fu J. Controlling vehicular emissions in Beijing during the last decade. *Transportation Research Part A: Policy and Practice* 2006;**40**(8):639-51.

Hara 2013 *{published data only}*

Hara K, Homma J, Tamura K, Inoue M, Karita Y, Yano E. Decreasing trends of suspended particulate matter and PM2.5 concentrations in Tokyo, 1990-2010. *Journal of Air & Waste Management Association* 2013;**63**(6):737-48.

Harrison 2015 *{published data only}*

Harrison RM, Pope FD, Shi Z. Trends in Local Air Quality 1970–2014. Still Only One Earth: Progress in the 40 Years Since the First UN Conference on the Environment. The Royal Society of Chemistry, 2015:58-106.

Hedley 2002 *{published data only}*

Hedley AJ, Wong CM, Thach TQ, Mas S, Lam TH, Anderson HR. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: An intervention study. *Lancet* 2002;**360**(9346):1646-52.

Hendryx 2016 *{published data only}*

Hendryx M, Holland B. Unintended consequences of the Clean Air Act: Mortality rates in Appalachian coal mining communities. *Environmental Science & Policy* 2016;**63**:1-6.

Henneman 2015 *{published data only}*

Henneman LRF, Holmes HA, Mulholland JA, Russell AG. Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta. *Atmospheric Environment* 2015;**119**:201-10.

Herrstedt 1992 *{published data only}*

Herrstedt, L. Traffic calming design: A speed management method: Danish experiences on environmentally adapted through roads. *Accident; Analysis and Prevention* 1992;**24**(1):3-16.

Hine 2011 *{published data only}*

Hine DW, Bhullar N, Marks ADC, Kelly P, Scott J. Comparing the effectiveness of education and technology in reducing wood smoke pollution: A field experiment. *Journal of Environmental Psychology* 2011;**31**(4):282-8.

Hirten 1997 *{published data only}*

Hirten J, Beroldo S. Ridesharing programs cost little, do a lot. *Transportation Quarterly* 1997;**51**(2):9-12.

Ho 2015 *{published data only}*

Ho KF, Huang RJ, Kawamura K, Tachibana E, Lee SC, Ho SSH, et al. Dicarboxylic acids, ketocarboxylic acids, α-dicarbonyls, fatty acids and benzoic acid in PM2.5 aerosol collected during CAREBeijing-2007: An effect of traffic restriction on air quality. *Atmospheric Chemistry and Physics* 2015;**15**:3111-23.

Huang 1996 *{published data only}*

Huang BN, Loviscek AL. Assessing the impact of air pollution controls on carbon monoxide reduction: The case of Taiwan. *International Journal of Environment and Pollution* 1996;**6**(1):13-21.

Huang 2015 *{published data only}*

Huang K, Zhang X, Lin Y. The "APEC Blue" phenomenon: Regional emission control effects observed from space. *Atmospheric Research* 2015;**164-165**:65-75.

Hutchinson 2004 *{published data only}*

Hutchinson EJ, Pearson PJ. An evaluation of the environmental and health effects of vehicle exhaust catalysts in the United Kingdom. *Environmental Health Perspectives* 2004;**112**(2):132-41.

Invernizzi 2011 *{published data only}*

Invernizzi G, Ruprecht A, Mazza R, de Marco C, Mocnik G, Sioutas C, et al. Measurement of black carbon concentration as an indicator of air quality benefits of traffic restriction policies within the ecopass zone in Milan, Italy. *Atmospheric Environment* 2011;**45**(21):3522-7.

Jacobi 1999 *{published data only}*

Jacobi P, Segura DB, Kjellen M. Governmental responses to air pollution: summary of a study of the implementation of rodizio in Sao Paulo. *Environment and Urbanization* 1999;**11**(1):79-88.

Jalihal 2006 *{published data only}*

Jalihal SA, Reddy TS. Assessment of the impact of improvement measures on air quality: Case study of Delhi. *Journal of Transportation Engineering* 2006;**132**(6):482-8.

Jenq 1989 *{published data only}*

Jenq FT, Tsai JH. Air pollution control tactics in urban areas. A case study in Taipei. 82nd A&WMA Annual Meeting. Anaheim, CA, US, 1989; Vol. 2.

Jhun 2013 *{published data only}*

Jhun I, Oyola P, Moreno F, Castillo MA, Koutrakis P. PM2.5 mass and species trends in Santiago, Chile, 1998 to 2010: the impact of fuel-related interventions and fuel sales. *Journal of Air & Waste Management Association* 2013;**63**(2):161-9.

Jiang 2015 *{published data only}*

Jiang X, Hong C, Zheng Y, Zheng B, Guan D, Gouldson A, et al. To what extent can China's near-term air pollution control policy protect air quality and human health? A case study of the Pearl River Delta region. *Environmental Research Letters* 2015;**10**(10):1-13.

Jiang 2016 *{published data only}*

Jiang P, Xu B, Geng Y, Dong W, Chen Y, Xue B. Assessing the environmental sustainability with a co-benefits approach: a study of industrial sector in Baoshan District in Shanghai. *Journal of Cleaner Production* 2016;**114**:114-23.

Jin 2013 *{published data only}*

Jin Yanhong, Lin Liguo. China's provincial industrial pollution: the role of technical efficiency, pollution levy and pollution quantity control. *Environment and Development Economics* 2013;**19**(1):111-32.

Karanasiou 2014 *{published data only}*

Karanasiou A, Amato F, Moreno T, Lumbreras J, Borge R, Linares Cristina, et al. Road dust emission sources and assessment of street washing effect. *Aerosol and Air Quality Research* 2014;**14**:734-43.

Kendall 2011 *{published data only}*

Kendall M, Tinker D, Barrat B. UK trends in urban-rural atmospheric nanoparticles, and during a traffic intervention. *AWMA Annual Conference and Exhibition* 2011:2205-9.

Keuken 2012 *{published data only}*

Keuken MP, Jonkers S, Zandveld P, Voogt M, Elshout van den S. Elemental carbon as an indicator for evaluating the impact of traffic measures on air quality and health. *Atmospheric Environment* 2012;**61**:1-8.

Kim 2015 *{published data only}*

Kim H, Kim H, Lee JT. Effects of ambient air particles on mortality in Seoul: Have the effects changed over time?. *Environmental Research* 2015;**140**:684-90.

Kobza 2016 *{published data only}*

Kobza J, Pastuszka JS, Gulis Gabriel. Consideration on the health risk reduction related to attainment of the new particulate matter standards in Poland: A top-down policy risk assessment approach. *International Journal of Occupational Medicine and Environmental Health* 2016;**29**(1):1-14.

Kong 2010 *{published data only}*

Kong Q, Wu Y, Yang L, Fu L. Characteristics and intake dose of gaseous pollutants near a major Beijing road during the 29th Olympic Games. *Huanjing Kexue Xuebao [Acta Scientiae Circumstantiae]* 2010;**30**(2):281-6.

Koutrakis 2005 *{published data only}*

Koutrakis P, Sax SN, Sarnat JA, Coull B, Demokritou P, Oyola P, et al. Analysis of PM10, PM2.5, and PM2.5-10 concentrations in Santiago, Chile, from 1989 to 2001. *Journal of Air & Waste Management Association* 2005;**55**(3):342-51.

Kowalska 2008 *{published data only}*

Kowalska M, Zejda J, Skrzypek M, Osrodka L, Klejnowski K, Krajny E, et al. Air Pollution and Daily Mortality in Urban Katowice, 1994-95 and 2001-02. *Polish Journal of Environmental Studies* 2008;**17**(5):733-8.

Kravchenko 2014 *{published data only}*

Kravchenko J, Akushevich I, Abernethy AP, Holman S, Ross WG, Lyerly HK. Long-term dynamics of death rates of emphysema, asthma, and pneumonia and improving air quality. *International Journal of Chronic Obstructive Pulmonary Disease* 2014;**9**:613-27.

Krawack 1993 *{published data only}*

Krawack S. Traffic management and emissions. *Science of the Total Environment* 1993;**134**(1-3):305-14.

Kuwayama 2012 *{published data only}*

Kuwayama T, Schwartz JR, Harley RA, Kleeman MJ. Particulate matter emissions reductions due to adoption of clean diesel technology at a major shipping port. *Aerosol and Air Quality Research* 2012;**47**(1):29-36.

Lacasana-Navarro 1999 *{published data only}*

Lacasana-Navarro M, Aquilar-Garduno C, Romieu I. Evolution of air pollution and impact on control programs in 3 megacities in Latin America. *Salud Publica de Mexico* 1999;**41**(3):203-15.

Leem 2015 *{published data only}*

Leem JH, Kim ST, Kim HC. Public-health impact of outdoor air pollution for 2nd air pollution management policy in Seoul metropolitan area, Korea. *Annals of Occupational and Environmental Medicine* 2015;**27**(1):7.

Li 2010 *{published data only}*

Li Y, Wang w, Kan H, Xu X, Chen B. Air quality and outpatient visits for asthma in adults during the 2008 Summer Olympic Games in Beijing. *Science of the Total Environment* 2010;**408**(5):1226-7.

Li 2014 *{published data only}*

Li YR, Gibson JM. Health and Air Quality Benefits of Policies to Reduce Coal-Fired Power Plant Emissions: A Case Study in North Carolina. *Environmental Science & Technology* 2014;**48**(17):10019-27.

Li 2015 *{published data only}*

Li WT, Gao QX, Liu JR, Li L, Gao WK, Su BD. [Comparative Analysis on the Improvement of Air Quality in Beijing During APEC]. *Huan jing ke xue= Huanjing kexue* 2015;**36**(12):4340-7.

Li 2016c *{published data only}*

Li S, Williams G, Guo Y. Health benefits from improved outdoor air quality and intervention in China. *Environmental Pollution* 2016;**214**:17-25.

Li 2016d *{published data only}*

Li X, Guo YH, Lu XY, Gulgina H, Wang SL, Zhao KM, et al. Evaluation and analysis on the effects of air pollution control in Urumqi. *Zhongguo Huanjing Kexue/China Environmental Science* 2016;**36**(1):307-13.

Lin 2011b *{published data only}*

Lin Y, Wu X, Hao X, Han C. Influence of green belt structure on the dispersion of particle pollutants in street canyons. *Shengtai Xuebao/ Acta Ecologica Sinica* 2011;**31**(21):6561-6567.

Lin 2016 *{published data only}*

Lin MY, Hagler G, Baldauf R, Isakov V, Lin HY, Khlystov A. The effects of vegetation barriers on near-road ultrafine particle number and carbonmonoxide concentrations. *Science of the Total Environment* 2016;**553**:372-9.

Liu 2015 *{published data only}*

Liu S, Zhang K. Fine particulate matter components and mortality in Greater Houston: Did the risk reduce from 2000 to 2011?. *Science of the Total Environment* 2015;**538**:162-8.

Lomas 2016 *{published data only}*

Lomas, J, Schmitt L, Jones S, McGeorge M, Bates E, Holland M, et al. A pharmacoeconomic approach to assessing the costs and benefits of air quality interventions that improve health: A case study. *BMJ Open* 2016;**6**(6):e010686.

Lopez 2000 *{published data only}*

Lopez AV. Analysis of the air quality-transportation policies in Mexico City. Urban Transportation and Environment. 2000:15-20.

Luechinger 2014 *{published data only}*

Luechinger S. Air pollution and infant mortality: A natural experiment from powerplant desulfurization. *Journal of Health Economics* 2014;**37**:219-31.

Lyons 1993 *{published data only}*

Lyons CE. Quantifying the air pollution emissions reduction effectiveness and costs of oxygenated fuels. *Air and Waste Management Association 86th Annual Meeting and Exhibition* 1993;**25**(8).

Makonese 2015 *{published data only}*

Makonese T, Masekameni DM, Annegarn HJ, Forbes PBC. Influence of fire-ignition methods and stove ventilation rates on gaseous and particle emissions from residential coal braziers. *Journal of Energy in Southern Africa* 2015;**26**(4):16-28.

Mardones 2015 *{published data only}*

Mardones Cristian, Saavedra Andrés, Jiménez Jorge. Cuantificación económica de los beneficios en salud asociados a la reducción de la contaminación por MP10 en Concepción Metropolitano, Chile. *Revista médica de Chile* 2015;**143**:475-83.

Masiol 2014 *{published data only}*

Masiol Mauro, Agostinelli Claudio, Formenton Gianni, Tarabotti Enzo, Pavoni Bruno. Thirteen years of air pollution hourly monitoring in a large city: Potential sources, trends, cycles and effects of car-free days. *Science of the Total Environment* 2014;**494-495**:84-96.

McNabola 2008 *{published data only}*

McNabola A, Broderick BM, Gill LW. Reduced exposure to air pollution on the boardwalk in Dublin, Ireland. Measurement and prediction. *Environmental International* 2008;**34**(1):86-93.

Melkonyan 2012 *{published data only}*

Melkonyan A, Kuttler W. Long-term analysis of NO, NO2, and O3 concentrations in North Rhine-Westphalia, Germany. *Atmospheric Environment* 2012;**60**:316-26.

Minoura 2006 *{published data only}*

Minoura H, Takahashi K, Chow JC, Watson JG. Multi-year trend in fine and coarse particle mass, carbon, and ions in downtown Tokyo, Japan. *Atmospheric Environment* 2006;**40**(14):2478-87.

Minoura 2009 *{published data only}*

Minoura H, Takahashi K, Chow JC, Watson JG. Atmosphere environment improvement in Tokyo by vehicle exhaust purification. *WIT Transactions on Ecology and the Environment* 2009;**123**:129-39.

Mott 2002 *{published data only}*

Mott JA, Wolfe MI, Alverson CJ, Macdonald SC, Bailey CR, Ball LB, et al. National vehicle emissions policies and practices and declining us carbon monoxide – related mortality. *JAMA* 2002;**288**(8):988-95.

Narain 2007 *{published data only}*

Narain U, Krupnick A. The Impact of Delhi's CNG Program on Air Quality. *SSRN Electronic Journal* 2007;**RFF Discussion Paper No. 07-06**:1-51.

Nedellec 2010 *{published data only}*

Nedellec V, Mosqueron L, Desqueyroux H, Jeannee N. Effects of European Euro IV and V standards on the health impact of urban road traffic in France. III. Health impact assessment and comparison of the years 2000 and 2010. *Environnement Risques & Sante* 2010;**9**(6):503-15.

Ngo 2015 *{published data only}*

Ngo NS. Analyzing the relationship between bus pollution policies and morbidity using a quasi-experiment. *Medicine (Baltimore)* 2015;**94**(37):e1499.

Noonan 2011b *{published data only}*

Noonan CW, Ward TJ, Navidi W, Sheppard L, Palmer C, Bergauff M. Changes in respiratory symptoms and infections following a reduction in wood smoke PM. *Epidemiology* 2011;**22**(1):s186.

Norra 2016 *{published data only}*

Norra S, Yu Y, Dietze V, Schleicher N, Fricker M, Kaminski U, et al. Seasonal dynamics of coarse atmospheric particulate matter between 2.5 mm and 80 mm in Beijing and the impact of 2008 Olympic Games. *Atmospheric Environment* 2016;**124**:109-18.

Orozco 2015 *{published data only}*

Orozco D, Delgado R, Wesloh D, Powers RJ, Hoff R. Aerosol particulate matter in the Baltimore metropolitan area: Temporal variation over a six-year period. *Journal of the Air & Waste Management Association* 2015;**65**(9):1050-61.

Pan 2010 *{published data only}*

Pan X. Study on the health effects of air pollution in Beijing. *International Journal of Environmental Studies* 2010;**67**(2):147-54.

Parker 2008 *{published data only}*

Parker JD, Mendola P, Woodruff TJ. Preterm birth after the Utah valley steel mill closure: A natural experiment. *Epidemiology* 2008;**19**(6):820-3.

Pope 1996 *{published data only}*

Pope CA. Adverse health effects of air pollutants in a nonsmoking population. *Toxicology* 1996;**111**(1-3):149-55.

Potoski 2013 *{published data only}*

Potoski M, Prakash A. Do Voluntary Programs Reduce Pollution? Examining ISO 14001's Effectiveness across Countries. *Policy Studies Journal* 2013;**41**(2):273-94.

Qiao 2015 *{published data only}*

Qiao X, Jaffe D, Tang Y, Bresnahan M, Song J. Evaluation of air quality in Chengdu, Sichuan Basin, China: are China's air quality standards sufficient yet?. *Environmental Monitoring and Assessment* 2015;**187**(5):250.

Querol 2014 *{published data only}*

Querol X, Alastuey A, Pandolfi M, Reche C, Pérez N, Minguillón MC, et al. 2001–2012 trends on air quality in Spain. *Science of the Total Environment* 2014;**490**:957-69.

Rafaj 2014 *{published data only}*

Rafaj P, Ammann M, Siri J, Wuester H. Changes in European greenhouse gas and air pollutant emissions 1960–2010: decomposition of determining factors. *Climatic Change* 2014;**124**(3):477-504.

Raman 2008 *{published data only}*

Raman R. Results of implementing aggressive PM reduction on non-road construction equipment at two lower Manhattan project sites. Transportation Land Use, Planning, and Air Quality Congress 2008. 2008; Vol. doi.org/10.1061/40960(320)30.

Rava 2011 *{published data only}*

Rava M, White RH, Dominici F. Does attainment status for the PM10 National Air Ambient Quality Standard change the trend in ambient levels of particulate matter?. *Air Quality, Atmosphere & Health* 2011;**4**(2):133-43.

Recycling 2007 *{published data only}*

Recycling Local Authority Waste &. Hull turns electric to save on fuel. *Local Authority Waste & Recycling* 2007;**15**:14.

Ringquist 1995 *{published data only}*

Ringquist EJ. Is "Effective Regulation" always oxymoronic?: The States and ambient air quality. *Social Science Quarterly* 1995;**76**(1):69-87.

Riveros 2009 *{published data only}*

Riveros HG. Driving restrictions implemented to reduce air pollution in Mexico City. 102nd Air & Waste Management Association Annual Conference and Exhibition 2009. Detroit, MI, 2009.

Roberts 2013 *{published data only}*

Roberts S. Have the short-term mortality effects of particulate matter air pollution changed in Australia over the period 1993-2007?. *Environmental Pollution* 2013;**182**:9-14.

Sabaliauskas 2012 *{published data only}*

Sabaliauskas K, Jeong CH, Yao x, Jun YS, Jadidian P, Evans GJ. Five-year roadside measurements of ultrafine particles in a major Canadian city. *Atmospheric Environment* 2012;**49**:245-56.

Sajjadi 2008 *{published data only}*

Sajjadi SA, Bridgman HA. Changes in air quality in the Lower Hunter Region NSW Australia due to closure of major industry. *Clean Air and Environmental Quality* 2008;**42**(1):27-33.

Shannigrahi 2010 *{published data only}*

Shannigrahi AS, Sharma R, Fukushima T. Air pollution control by optimal green belt development around the Victoria Memorial Monument, Kolkata (India). *International Journal of Environmental Studies* 2010;**60**(3):241-9.

Shu 2014b *{published data only}*

Shu S, Quiros D, Wng R, Zhu Y. Changes of street use and on-road air quality before and after complete street retrofit: An exploratory case study in Santa Monica, California. *Transportation Research Part D: Transport and Environment* 2014;**32**:387-96.

Snowden 2015 *{published data only}*

Snowden JM, Mortimer KM, Dufour MS, Tager IB. Population intervention models to estimate ambient NO2 health effects in children with asthma. *Journal of Exposure Science & Environmental Epidemiology* 2015;**25**(6):567-73.

Song 2015a *{published data only}*

Song GY, Li XZ, Hui D, Jing J, Ting H. The Effects of the Air Pollution Prevention and Control in the Major Cities Around Beijing During the APEC Meeting. 2015 4th International Conference on Energy and Environmental Protection (ICEEP 2015). 2015:2262-70.

Sun 2010 *{published data only}*

Sun ZQ, Ji DS, Song T, Lin H, Wang YS, Jiang CS. Observations and comparison analysis of air pollution in Beijing and nearly surrounding areas during Beijing 2008 Olympic Games. *Huanjing Kexue/Environmental Science* 2010;**31**:2852-9.

Sun 2014 *{published data only}*

Sun C, Zheng S, Wang R. Restricting driving for better traffic and clearer skies: Did it work in Beijing?. *Transport Policy* 2014;**32**:34-41.

Traversi 2008 *{published data only}*

Traversi D, Degan R, De Marco R, Gilli G, Pignata C, Ponzio M, Rava M, et al. Mutagenic properties of PM2.5 air pollution in the Padana Plain (Italy) before and in the course of XX Winter Olympic Games of "Torino 2006". *Environmental International* 2008;**34**(7):966-70.

US EPA 2014 *{published data only}*

US Environmental Protection Agency. Approval and promulgation of implementation plans; State of Missouri; St. Louis inspection and maintenance program. Federal Register. U.S. Environmental Protection Agency, 2014:77996-8.

US EPA 2014a *{published data only}*

US Environmental Protection Agency. Approval and promulgation of implementation plans; Texas; Attainment demonstration for the Houston-Galveston-Brazoria 1997 8- hour Ozone nonattainment area. Federal Register. U.S. Environmental Protection Agency, 2014:57-61.

US EPA 2015 *{published data only}*

US Environmental Protection Agency. Approval and promulgation of air quality implementation plans; Delaware; low emission vehicle program. Federal Register. U.S. Environmental Protection Agency, 2015:61752-7.

van den Elshout 2014 *{published data only}*

van den Elshout S, Molenaar R, Wester B. Adaptive traffic management in cities — Comparing decision-making methods. *Science of the Total Environment* 2014;**488-489**:382-8.

Voorhees 2014 *{published data only}*

Voorhees AS, Wang J, Wang C, Zhao B, Wang S, Kan H. Public health benefits of reducing air pollution in Shanghai: A proof-of-concept methodology with application to BenMAP. *Science of the Total Environment* 2014;**485-486**:396-405.

Wang 2009 *{published data only}*

Wang X, Westerdahl D, Chen LC, Wu Y, Hao J, Pan X, et al. Evaluating the air quality impacts of the 2008 Beijing Olympic Games: On-road emission factors and black carbon profiles. *Atmospheric Environment* 2009;**43**(30):4535-43.

Wang 2010 *{published data only}*

Wang S, Zhao M, Xing J, Wu Y, Zhou Y, Lei Y, et al. Quantifying the air pollutants emission reduction during the 2008 Olympic games in Beijing. *Environmental Science & Technology* 2010;**44**(7):2490-96.

Wang 2014a *{published data only}*

Wang S, Xing J, Zhao B, Jang C, Hao J. Effectiveness of national air pollution control policies on the air quality in metropolitan areas of China. *Journal of Environmental Sciences* 2014;**26**(1):13-22.

Wang 2015 *{published data only}*

Wang Y, Zacharias J. Landscape modification for ambient environmental improvement in central business districts – A case from Beijing. *Urban Forestry & Urban Greening* 2015;**14**(1):8-18.

Westerdahl 2011 *{published data only}*

Westerdahl D, Wang X, Pan X, Zhang M. Traffic-generated pollutants measured on and near roadways and in community air of Beijing: Assessing the effectiveness of controls during the 2008 Olympic games. *Epidemiology* 2011;**22**(1):s146-7.

Wong 1998 *{published data only}*

Wong CM, Lam TH, Peters J, Hedley AJ, Ong SG, Tam AY, et al. Comparison between two districts of the effects of an air pollution intervention on bronchial responsiveness in primary school children in Hong Kong. *Journal of Epidemiology and Community Health* 1998;**52**(9):571-8.

Wood 2015 *{published data only}*

Wood HE, Marlin N, Mudway IS, Bremner SA, Cross L, Dundas I, et al. Effects of Air Pollution and the Introduction of the London Low Emission Zone on the Prevalence of Respiratory and Allergic Symptoms in Schoolchildren in East London: A Sequential Cross-Sectional Study. *PLOS ONE* 2015;**10**(8):e0109121.

Wu 2010 *{published data only}*

Wu S, Deng F, Niu j, Huang Q, Liu Y, Guo X. Association of heart rate variability in taxi drivers with marked changes in particulate air pollution in Beijing in 2008. *Environmental Health Perspectives* 2010;**118**(1):87-91.

Wu 2010a *{published data only}*

Wu D, Xin JY, Sun Y, Wang YS, Wang PC, Wu D, et al. [Change and analysis of background concentration of air pollutants in north China during 2008 Olympic Games]. [Chinese]. *Huanjing Kexue/ Environmental Science* 2010;**31**:1130-8.

Wu 2011 *{published data only}*

Wu Y, Wang R, Zhou Y, Lin B, Fu L, He K, Hao J. On-road vehicle emission control in Beijing: Past, present and future. *Environmental Science & Technology* 2011;**45**(1):147-53.

Xue 2014 *{published data only}*

Xue L, Wang T, Louie PKK, Luk CWY, Blake DR, Xu Z. Increasing External Effects Negate Local Efforts to Control Ozone Air Pollution: A Case Study of Hong Kong and Implications for Other Chinese Cities. *Environmental Science & Technology* 2014;**48**(18):10769-75.

Yang 2011 *{published data only}*

Yang L, Wu J, Davis JM, Hao J. Estimating the effects of meteorology on PM2.5 reduction during the 2008 Summer Olympic Games in Beijing, China. *Frontiers of Environmental Science & Engineering in China* 2011;**5**:331.

Yorifuji 2016b *{published data only}*

Yorifuji T, Kashima S, Doy H. Acute exposure to fine and coarse particulate matter and infant mortality in Tokyo, Japan (2002–2013). *Science of the Total Environment* 2016;**551-552**:66-72.

You 2014 *{published data only}*

You M. Addition of PM2.5 into the National Ambient Air Quality Standards of China and the Contribution to Air Pollution Control: The Case Study of Wuhan, China. *The Scientific World Journal* 2014;**2014**(768405):1-10.

Zhang 2005 *{published data only}*

Zhang J, Hu W, Wei F, Wu G, Cheng WL, Chapman RS. Long-term changes in air pollution and health implications in four Chinese cities. *Energy for Sustainable Development* 2005;**9**(3):67-76.

Zhang 2011 *{published data only}*

Zhang R, Shen Z, Zhang L, Zhang M, Wang X, Zhang K. Elemental composition of atmospheric particles during periods with and without traffic restriction in Beijing: The effectiveness of traffic restriction measure. *Scientific Online Letters on the Atmosphere* 2011;**7**(1):61-4.

Zhang 2014 *{published data only}*

Zhang Q, Yuan B, Shao M, Wang X, Lu S, Lu K, et al. Variations of ground-level O_3 and its precursors in Beijing in summertime between 2005 and 2011. *Atmospheric Chemistry & Physics* 2014;**14**(12):6089-101.

Zhang 2016b *{published data only}*

Zhang J, Zhong C, Yi M. Did Olympic Games improve air quality in Beijing? Based on the synthetic control method. *Environmental Economics and Policy Studies* 2016;**18**(1):21-39.

Zhao 2010 *{published data only}*

Zhao WH, Gong H, Zhao WJ, Li XJ, Ming MT. Evaluating the air quality impacts of the 2008 Beijing Olympic Games: The spatial distribution of inhalable particulate matter and their impact factors. Proceedings of the Symposium Dragon 2 Programme China. 2010; Vol. 684:7.

Zhao 2014 *{published data only}*

Zhao S, Yu Y, Liu N, He J, Chen J. Effect of traffic restriction on atmospheric particle concentrations and their size distributions in urban Lanzhou, Northwestern China. *Journal of Environmental Sciences* 2014;**26**(2):362-70.

Zheng 2015 *{published data only}*

Zheng S, Yi H, Li H. The impacts of provincial energy and environmental policies on air pollution control in China. *Renewable and Sustainable Energy Reviews* 2015;**49**:386-94.

Zhou 2010 *{published data only}*

Zhou Y, Hu G, Wang D, Wang S, Wang Y, Liu Z, et al. Community based integrated intervention for prevention and management of chronic obstructive pulmonary disease (COPD) in Guangdong, China: cluster randomised controlled trial. *BMJ Clinical Research Edition* 2010;**341**:c6387.

Additional references

Amato 2009

Amato F, Querol X, Alastuey A, Pandolfi M, Moreno T, Gracia J, Rodriguez P. Evaluating urban PM10 pollution benefit induced by street cleaning activities. *Atmospheric Environment* 2009;**43**(29):4472-80.

Amato 2010

Amato F, Nava S, Lucarelli F, Querol X, Alastuey A, Baldasano JM, Pandolfi M. A comprehensive assessment of PM emissions from paved roads: Real-world Emission Factors and intense street cleaning trials. *Science of the Total Environment* 2010;**408**(20):4309-18.

Avol 2001

Avol EL, Gauderman WJ, Tan SM, London SJ, Peters JM. Respiratory Effects of Relocating to Areas of Differing Air Pollution Levels. *American Journal of Respiratory and Critical Care Medicine* 2001;**164**(11):2067-72.

Bae 2015

Bae S, Hong Y. Co-benefit on mortality from pro-bicycle policy in Changwon, Korea: A health impact assessment. 6th Regional EST Forum, New Delhi. India. 2015.

Barreca 2017

Barreca AI, Neidell M, Sanders NJ. Long-Run Pollution Exposure and Adult Mortality: Evidence from the Acid Rain Program. National Bureau of Economic Research Working Paper Series 2017; Vol. Working Paper No. 23524.

Barros 2015

Barros N, Fontes T, Silva MP, Manso MC, Carvalho AC. Analysis of the effectiveness of the NEC Directive on the tropospheric ozone levels in Portugal. *Atmospheric Environment* 2015;**106**:80-91.

Becker 2017

Becker BJ, Aloe AM, Duvendack M, Stanley TD, Valentine JC, Fretheim Atle, et al. Quasi-experimental study designs series—paper 10: synthesizing evidence for effects collected from quasi-experimental studies presents surmountable challenges. *Journal of Clinical Epidemiology* 2017;**89**:84-91.

Begum 2008

Begum BA, Biswas SK, Hopke PK. Assessment of trends and present ambient concentrations of PM2.2 and PM10 in Dhaka, Bangladesh. *Air Quality, Atmosphere & Health* 2008;**1**(3):125-33.

Bell 2011

Bell ML, Morgenstern RD, Harrington W. Quantifying the human health benefits of air pollution policies: Review of recent studies and new directions in accountability research. *Environmental Science and Policy* 2011;**14**(4):357-68.

Boogaard 2013

Boogaard H, Fischer PH, Janssen NAH, Kos GPA, Weijers EP, Cassee FR, et al. Respiratory Effects of a Reduction in Outdoor Air Pollution Concentrations. *Epidemiology* 2013;**24**(5):753-61.

Boogaard 2017

Boogaard H, van Erp AM, Walker KD, Shaikh R. Accountability Studies on Air Pollution and Health: the HEI Experience. *Current environmental health reports* 2017;**4**(4):514-22.

Brimblecombe 2015

Brimblecombe P, Ning Z. Effect of road blockages on local air pollution during the Hong Kong protests and its implications for air quality management. *Science of the Total Environment* 2015;**536**:443-8.

Burns 2014

Burns J, Boogaard H, Turley R, Pfadenhauer LM, van Erp AM, Rohwer AC, et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. *Cochrane Database of Systematic Reviews* 2014, (1).

Bärnighausen 2017

Bärnighausen T, Røttingen JA, Rockers P, Shemilt I, Tugwell P. Quasi-experimental study designs series—paper 1: introduction: two historical lineages. *Journal of Clinical Epidemiology* 2017;**89**:4-11.

Campbell 2018

Campbell M, Katikireddi SV, Hoffmann T, Armstrong R, Waters E, Craig P. TIDieR-PHP: a reporting guideline for population health and policy interventions. *BMJ* 2018;**361**:k1079.

Campbell-Lendrum 2019

Campbell-Lendrum D, Prüss-Ustün A. Climate change, air pollution and noncommunicable diseases. *Bulletin of the World Health Organization* 2019;**97**:160-1.

Cesaroni 2012

Cesaroni G, Boogaard H, Jonkers S, Porta D, Badaloni C, Cattani G, et al. Health benefits of traffic-related air pollution reduction in different socioeconomic groups: the effect of low-emission zoning in Rome. *Occupational and Environmental Medicine* 2012;**69**(2):133-9.

Chelani 2011

Chelani AB. Change detection using CUSUM and modified CUSUM method in air pollutant concentrations at traffic site in Delhi. *Stochastic Environmental Research and Risk Assessment* 2011;**25**(6):827-34.

Chin 1996

Chin ATH. Containing air pollution and traffic congestion: Transport policy and the environment in Singapore. *Atmospheric Environment* 1996;**30**(5):787-801.

Chow 1995

Chow JC. Measurement methods to determine compliance with ambient air quality standards for suspended particles. *Journal of the Air and Waste Management Association* 1995;**45**:320-82.

Cirera 2009

Cirera L, Rodriguez M, Gimenez J, Jimenez E, Saez M, Guillen JJ, et al. Effects of public health interventions on industrial emissions and ambient air in Cartagena, Spain. *Environmental science and pollution research international* 2009;**16**(2):152-61.

Cochrane EPOC 2017

EPOC Cochrane. What study designs can be considered for inclusion in an EPOC review and what should they be called?. Cochrane Effective Practice and Organisation of Care (EPOC), 2017:Available at: http://epoc.cochrane.org/resources/epoc-resources-review-authors.

Cohen 2017

Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *The Lancet* 2017;**389**:1907-18.

Correia 2013

Correia AW, Pope CA, Dockery DW, Wang Y, Ezzati M, Dominici F. The Effect of Air Pollution Control on Life Expectancy in the United States: An Analysis of 545 US counties for the period 2000 to 2007. *Epidemiology* 2013;**24**(1):23-31.

Craig 2017

Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. *Annual Review of Public Health* 2017;**38**(1):39-56.

Dadvand 2013

Dadvand P, Parker J, Bell Ml, Bonzini M, Brauer M, Darrow LA, et al. Maternal exposure to particulate air pollution and term birth weight: a multi-country evaluation of effect and heterogeneity. *Environmental Health Perspectives* 2013;**121**(3):267-373.

Dahlgren 1991

Dahlgren G, Whithead M. Policies and strategies to promote social equity in health. Institute for Future Studies 1991; Vol. Arbetsrapport 2007.

Des Jarlais 2004

Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *American Journal of Public Health* 2004;**94**(3):361-6.

Di 2017

Di Q, Dominici F, Schwartz J D. Air Pollution and Mortality in the Medicare Population. *New England Journal of Medicine* 2017;**376**:2513-22.

Ding 2014

Ding L, Chan TW, Ke F, Wang, DKW. Characterization of chemical composition and concentration of fine particulate matter during a transit strike in Ottawa, Canada. *Atmospheric Environment* 2014;**89**:433-42.

Dominici 2007

Dominici Francesca, Peng RD, Zeger SL, White RH, Samet JM. Particulate Air Pollution and Mortality in the United States: Did the Risks Change from 1987 to 2000?. *American Journal of Epidemiology* 2007;**166**(8):880-8.

Ebelt 2001

Ebelt S, Brauer M, Cyrys J, Tuch T, Kreyling WG, Wichmann HE, Heinrich J. Air quality in postunification Erfurt, East Germany: associating changes in pollutant concentrations with changes in emissions. *Environmental Health Perspective* 2001;**109**(4):325-33.

Engelbrecht 1999

Engelbrecht JP, Swanepoel L, Chow JC, Watson JG, Egami RT. PM2.5 and PM10 concentrations from the Qalabotjha low-smoke fuels macro-scale experiment in South Africa. *Environmental monitoring and assessment* 2001;**69**(1):1-15.

Ferreira 2015

Ferreira F, Gomes P, Tente H, Carvalho AC, Pereira P, Monjardino J. Air quality improvements following implementation of Lisbon's Low Emission Zone. *Atmospheric Environment* 2015;**122**:373-81.

Font 2016

Font A, Fuller GW. Did policies to abate atmospheric emissions from traffic have a positive effect in London?. *Environmental Pollution* 2016;**218**:463-74.

Fransen 2013

Fransen M, Perodin J, Hada J, He X, Sapkota A. Impact of vehicular strike on particulate matter air quality: results from a natural intervention study in Kathmandu valley. *Environmental Research* 2013;**122**:52-7.

Gakidou 2017

Gakidou E, Afshin A, Abajobir AA, Abate KH, Abbafati C, Abbas KM, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet 2017; Vol. 390, issue 10100:1345-422.

Gauderman 2015

Gauderman WJ, Urman R, Avol E, Berhane K, McConnell R, Rappaport E, et al. Association of Improved Air Quality with Lung Development in Children. *New England Journal of Medicine* 2015;**372**(10):905-13.

Gehrsitz 2017

Gehrsitz M. The effect of low emission zones on air pollution and infant health. *Journal of Environmental Economics and Management* 2017;**83**:121-44.

Gilliland 2017

Gilliland F, Avol E, McConnell R, Berhane K, Gauderman WJ, Lurmann FW, et al. Effects of Policy-Driven Air Quality Improvements on Children's Respiratory Health. *Research Report (Health Effects Institute)* 2017;**190**:5-50.

Goodman 2009

Goodman PG, Rich DQ, Zeka A, Clancy L, Dockery DW. Effect of Air Pollution Controls on Black Smoke and Sulfur Dioxide Concentrations across Ireland. *Journal of the Air & Waste Management Association* 2009;**59**(2):207-13.

Graham 2016

Graham H, White P C L. Social determinants and lifestyles: integrating environmental and public health perspectives. *Public Health* 2016;**141**:270-8.

Guo 2016

Guo J, He J, Liu H, Miao Y, Liu H, Zhai P. Impact of various emission control schemes on air quality using WRF-Chem during APEC China 2014. *Atmospheric Environment* 2016;**140**:311-19.

Guyatt 2008

Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal* 2008;**336**:924-6.

Hales 2016

Hales NM, Barton CC, Ransom MR, Allen RT, Pope CA. A Quasi-Experimental Analysis of Elementary School Absences and Fine Particulate Air Pollution.. *Medicine (Baltimore)* 2016;**95**(9):e2916.

Han 2018

Han C, Lim YH, Yorifuji T, Hong YC. Air quality management policy and reduced mortality rates in Seoul Metropolitan Area: A quasi-experimental study. *Environment International* 2018;**121**:600-9.

HEI 2003

HEI Accountability Working Group. Assessing health impact of air quality regulations: concepts and methods for accountability research. HEI Communication 11 September 2003.

Henneman 2017

Henneman LR, Liu C, Mulholland JA, Russell AG. Evaluating the effectiveness of air quality regulations: A review of accountability studies and frameworks. *Journal of the Air & Waste Management Association* 2017;**67**(2):144-72.

Henschel 2012

Henschel S, Atkinson R, Zeka A, Le Tertre A, Analitis A, Kasouyanni K, et al. Air pollution interventions and their impact on public health. *International Journal of Public Health* 2012;**57**(5):757-68.

Henschel 2015

Henschel S, Le Tertre A, Atkinson RW, Querol X, Pandolfi M, Zeka A, et al. Trends of nitrogen oxides in ambient air in nine European cities between 1999 and 2010. *Atmospheric Environment* 2015;**117**:234-41.

Higgins 2012

Higgins JPT, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. *Research Synthesis Methods* 2013;**4**:12-25.

Higgins 2019

Higgins JPT, Lopez-Lopez JA, Becker BJ, Davies SR, Dawson S, et al. Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ Global Health* 2019;**4**:e000858.

Hoek 2013

Hoek G, Krishnan R, Beelen R, Peters A, Ostro B, Brunekreef B, et al. Long-term air pollution exposure and cardio- respiratory mortality: a review. *Environmental Health* 2013;**12**(1):43.

Hoffmann 2014

Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ : British Medical Journal* 2014;**348**.

Hong 2015

Hong A, Schweitzer L, Yang W, Marr LC. Impact of temporary freeway closure on regional air quality: A lesson from Carmageddon in Los Angeles, United States. *Environmental Science & Technology* 2015;**49**(5):3211-8.

Hou 2010

How Q, An XQ, Wang Y, Guo JP. An evaluation of resident exposure to respirable particulate matter and health economic loss in Beijing during Beijing 2008 Olympic Games. *Science of the Total Environment* 2010;**408**(19):4026-32.

Huang 2012a

Huang W, Wang G, Lu SE, Kipen H, Wang Y, Hu M, et al. Inflammatory and oxidative stress responses of healthy young adults to changes in air quality during the Beijing Olympics. *American Journal of Respiratory and Critical Care Medicine* 2012;**186**(11):1150-9.

Huang 2012b

Huang W, Zhu T, Pan X, Hu M, Lu SE, Lin Y, et al. Air pollution and autonomic and vascular dysfunction in patients with cardiovascular disease: interactions of systemic inflammation, overweight, and gender. *American Journal of Epidemiology* 2012;**176**(2):117-26.

Hubbell 2014

Hubbell B, Greenbaum D. Counterpoint: Moving From Potential-Outcomes Thinking to Doing—Changing Research Planning to Enable Successful Health Outcomes Research. *American Journal of Epidemiology* 2014;**180**(12):1141-4.

Ibarra-Berastegi 2002

Ibarra-Berastegi G, Elias A, Madariaga I, Agirre E, Uria J. Environmental impact assessment of the underground railway in Bilbao (Spain). Development and application of computer techniques to environmental studies. WIT Press, 2002; Vol. 51:193-200.

Jackson 2006

Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, et al. The GATE frame: critical appraisal with pictures. *Evidence Based Medicine* 2006;**11**(2):35-8.

James 2012

James K, Farrell RE, Siciliano SD. Comparison of human exposure pathways in an urban brownfield: reduced risk from paving roads. *Environmental Toxicology and Chemistry* 2012;**31**(10):2423-30.

Janssen 2011

Janssen NAH, Hoek G, Simic-Lawson M, Fischer P, van Bree L, Brink H, et al. Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM_{10} and $PM_{2.5}$. *Environmental Health Perspectives* 2011;**119**(12):1691-9.

Johansson 2009

Johansson C, Burman L, Forsberg B. The effects of congestions tax on air quality and health. *Atmospheric Environment* 2009;**43**(31):4843-54.

Jones 2012

Jones AM, Harrison RM, Barratt B, Fuller G. A large reduction in airborne particle number concentrations at the time of the introduction of “sulphur free” diesel and the London Low Emission Zone. *Atmospheric Environment* 2012;**50**:129-38.

Karagulian 2015

Karagulian F, Belis CA, Dora CFC, Prüss-Ustün AM, Bonjour S, et al. Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmospheric Environment* 2015;**120**:475-83.

Karanasiou 2011

Karanasiou A, Moreno T, Amato F, Lumbreras J, Narros A, Borge R, et al. Road dust contribution to PM levels – Evaluation of the effectiveness of street washing activities by means of Positive Matrix Factorization. *Atmospheric Environment* 2011;**45**(13):2193-201.

Karanasiou 2012

Karanasiou A, Moreno T, Amato F, Tobias A, Boldo E, Linares C, et al. Variation of PM2.5 concentrations in relation to street washing activities. *Atmospheric Environment* 2012;**54**:465-9.

Kelly 2011

Kelly F, Anderson HR, Armstrong B, Atkinson R, Barratt B, Beevers S, et al. The Impact of the congestion charging scheme on air quality in London. *Research Report (Health Effects Institute)* 2011;**155**:5-71.

Keuken 2010a

Keuken M, Denier van der Gon H, van der Valk K. Non-exhaust emissions of PM and the efficiency of emission reduction by road sweeping and washing in the Netherlands. *Science of the Total Environment* 2010;**408**(20):4591-9.

Keuken 2010b

Keuken MP, Jonkers S, Wilmink IR, Wesseling J. Reduced NOx and PM10 emissions on urban motorways in The Netherlands by 80 km/h speed management. *Science of the Total Environment* 2010;**408**(12):2517-26.

Khillare 2008

Khillare PS, Agarwal T, Shridhar V. Impact of CNG implementation on PAHs concentration in the ambient air of Delhi: a comparative assessment of pre- and post-CNG scenario. *Environmental Monitoring and Assessment* 2008;**147**(1-3):223-33.

Kotchenruther 2015

Kotchenruther, RA. The effects of marine vessel fuel sulfur regulations on ambient PM2.5 along the west coast of the U.S. *Atmospheric Environment* 2015;**103**:121-8.

Kuo 2009

Kuo PH, Ni PC, Keats A, Tsuang BJ, Lan YY, Lin MD, et al. Retrospective assessment of air quality management practices in Taiwan. *Atmospheric Environment* 2009;**43**(25):3925-34.

Latha 2004

Latha KM, Badarinath KVS, Moorthy KK. Impact of diesel vehicular emission on ambient black carbon concentration at an urban location in India. *Current Science* 2004;**86**:451-3.

Le Tertre 2014

Le Tertre A, Henschel S, Atkinson RW, Analitis A, Zeka A, Katsouyanni K, et al. Impact of legislative changes to reduce the sulphur content in fuels in Europe on daily mortality in 20 European cities: An analysis of data from the Aphekom project. *Air Quality, Atmosphere & Health* 2014;**7**(1):83-91.

Lee 2005

Lee BK, Jun NY, Lee HK. Analysis of impacts on urban air quality by restricting the operation of passenger vehicles during Asian Game events in Busan, Korea. *Atmospheric Environment* 2005;**39**(12):2323-38.

Lee 2007

Lee JT, Son JY, Cho YS. Benefits of mitigated ambient air quality due to transportation control on childhood asthma hospitalization during the 2002 summer Asian games in Busan, Korea. *Journal of the Air & Waste Management Association* 2007;**57**(8):968-73.

Leuenberger 1994

Leuenberger P, Schwartz J, Ackermann-Liebrich U, Blaser K, Bolognini G, Bongard JP, et al. Passive smoking exposure in adults and chronic respiratory symptoms (SAPALDIA Study). Swiss Study on Air Pollution and Lung Diseases in Adults, SAPALDIA Team. *American Journal of Respiratory and Critical Care Medicine* 1994;**150**(5):1222-8.

Levy 2006

Levy JI, Baxter LK, Clougherty JE. The air quality impacts of road closures associated with the 2004 Democratic National Convention in Boston. *Environmental Health* 2006;**26**(5):16.

Lewin 2017

Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). *BMC Medical Research Methodology* 2017;**17**(1):76.

Li 2016a

Li R, Mao H, Wu L, He J, Ren P, Li X. The evaluation of emission control to PM concentration during Beijing APEC in 2014. *Atmospheric Pollution Research* 2016;**7**(2):363-69.

Li 2016b

Li SW, Li HB, Luo J, Li HM, Qian X, Bi J, et al. Influence of pollution control on lead inhalation bioaccessibility in PM2.5: A case study of 2014 Youth Olympic Games in Nanjing. *Environmental International* 2016;**Sep**(94):69-75.

Li 2017

Li X, Qiao Y, Zhu J, Shi L, Wang Y. The “APEC blue” endeavor: Causal effects of air pollution regulation on air quality in China. *Journal of Cleaner Production* 2017;**168**:1381-8.

Lin 2011

Lin W, Huang W, Zhu T, Hu M, Brunekreef B, Zhang Y, et al. Acute respiratory inflammation in children and black carbon in ambient air before and during the 2008 Beijing Olympics. *Environmental Health Perspectives* 2011;**119**(10):1507-12.

Lin 2014

Lin H, Zhang Y, Liu T, Xiao J, Xu Y, Xu X, et al. Mortality reduction following the air pollution control measures during the 2010 Asian Games. *Atmospheric Environment* 2014;**91**:24-31.

Lin 2015

Lin W, Zhu T, Xue T, Peng W, Brunekreef B, Gehring U, et al. Association between changes in exposure to air pollution and biomarkers of oxidative stress in children before and during the Beijing Olympics. *American Journal of Epidemiology* 2015;**181**(8):575-83.

Lippmann 2013

Lippmann M, Chen L-C, Gordon T, Ito K, Thurston GD. National Particle Component Toxicity (NPACT) initiative: integrated epidemiologic and toxicologic studies of the health effects of particulate matter component. *Research Report (Health Effects Institute)* 2013;**177**:5-13.

MacNeill 2009

MacNeill SJ, Goddard F, Pitman R, Tharme S, Cullinan P. Childhood peak flow and the Oxford Transport Strategy. *Thorax* 2009;**64**(8):651-6.

Montgomery 2019

Montgomery P, Movsisyan A, Grant S, Macdonald G, Rehfuess E. Considerations of complexity in rating certainty of evidence in systematic reviews: a primer on using the GRADE approach in global health. *BMJ Global Health* 2019;**4**:e000848.

Moore 2015

Moore GF, Audrey S, Barker M, Bond L, Bonell C, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ* 2015;**350**:h1258.

Movsisyan 2016

Movsisyan A, Melendez-Torres GJ, Montgomery P. Users identified challenges in applying GRADE to complex interventions and suggested an extension to GRADE. *Journal of Clinical Epidemiology* 2016;**70**:191-9.

Mu 2014

Mu L, Deng F, Tian L, Li Y, Swanson M, Ying J, et al. Peak expiratory flow, breath rate and blood pressure in adults with changes in particulate matter air pollution during the Beijing Olympics: A panel study. *Environmental Research* 2014;**Aug**(133):4-11.

Nguyen 2010

Nguyen HT, Kim KH, Ma CH, Cho SJ, Sohn JR. A dramatic shift in CO and CH4 levels at urban locations in Korea after the implementation of the Natural Gas Vehicle Supply (NGVS) program. *Environmental Research* 2010;**110**(4):396-409.

NICE 2012

National Institute for Health and Clinical Excellence. Methods for the development of NICE public health guidance. Third Edition. NICE, September 2012.

Nidhi 2007

Nidhi GJ. Air quality and respiratory health in Delhi. *Environmental Monitoring and Assessment* 2007;**135**(1-3):313-25.

Noonan 2011

Noonan CW, Ward TJ, Navidi W, Sheppard L, Bergauff M, Palmer C, et al. Assessing the impact of a wood stove replacement program on air quality and children's health. *Research Report (Health Effects Institute)* 2011;**162**:3-37.

NRC 2002

National Research Council. Estimating the public health benefits of proposed air pollution regulations. National Academy Press 2002.

Ogilvie 2008

Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, et al. The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Medical Research Methodology* 2008;**8**:8.

Panteliadis 2014

Panteliadis P, Strak M, Hoek G, Weijers E, Zee S, Dijkema M. Implementation of a low emission zone and evaluation of effects on air quality by long-term monitoring. *Atmospheric Environment* 2014;**86**:113-9.

Pereira 2007

Pereira MC, Santos RC, Alvim-Ferraz MC. Air quality improvements using European environment policies: a case study of SO2 in a coastal region in Portugal. *Journal of Toxicology and Environmental Health* 2007;**70**(3-4):347-51.

Peters 1996

Peters J, Hedley A, J Wong CM, Lam T, Ong SG, Liu J, Spiegelhalter DJ. Effects of an ambient air pollution intervention and environmental tobacco smoke on children's respiratory health in Hong Kong. *International Journal of Epidemiology* 1996;**25**(4):821-8.

Pfadenhauer 2017

Pfadenhauer LM, Gerhardus A, Mozygemba K, Bakke Lysdahl K, Booth A, Hofmann B, et al. Making sense of complexity in context and implementation: the Context and Implementation of Complex Interventions (CICI) framework. *Implementation Science* 2017;**12**(1):21.

Pinault 2017

Pinault LL, Weichenthal S, Crouse DL, Brauer M, Erickson A, van Donkelaar A, et al. Associations between fine particulate matter and mortality in the 2001 Canadian Census Health and Environment Cohort. *Environmental Research* 2017;**159**:406-15.

Pope 1989

Pope CA. Respiratory disease associated with community air pollution and a steel mill, Utah Valley. *American Journal of Public Health* 1989;**79**(5):623-8.

Pope 2006

Pope CA, Dockery DW. Health effects of fine particulate airpollution: lines that connect. *Journal of the Air & Waste Management Association* 2006;**56**:709-42.

Pope 2009

Pope CA, Ezzati M, Dockery DW. Fine-Particulate Air Pollution and Life Expectancy in the United States. *New England Journal of Medicine* 2009;**360**(4):376-86.

Qadir 2013

Qadir RM, Abbaszade G, Schnelle-Kreis J, Chow JC, Zimmerman R. Concentrations and source contributions of particulate organic matter before and after implementation of a low emission zone in Munich, Germany. *Environmental Pollution* 2013;**175**:158-67.

Quiros 2013

Quiros DC, Zhang Q, Choi W, He M, Paulson SE, Winer A, et al. Air quality impacts of a scheduled 36-h closure of a major highway. *Atmospheric Environment* 2013;**67**:404-14.

Ravindra 2006

Ravindra K, Wauters E, Tyagi SK, Mor S, van Grieken R. Assessment of air quality after the implementation of compressed natural gas (CNG) as fuel in public transport in Delhi, India. *Environmental Monitoring and Assessment* 2006;**115**(1-3):405-17.

Rehfuess 2013

Rehfuess Eva A, Akl Elie A. Current experience with applying the GRADE approach to public health interventions: an empirical study. *BMC Public Health* 2013;**13**(1):9.

Rehfuess 2017

Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. *Research Synthesis Methods* 2017;**9**(1):13-24.

Ribeiro 2003

Ribeiro H, Cardoso MR. Air pollution and children's health in Sao Paulo (1986-1998). *Social Science & Medicine* 2003;**57**(11):2013-22.

Rich 2015

Rich DQ, Liu K, Zhang J, Thurston SW, Stevens TP, Pan Y, et al. Differences in birth weight associated with the 2008 Beijing Olympics air pollution reduction: Results from a natural experiment. *Environmental Health Perspectives* 2015;**123**(9):880-7.

Rich 2017

Rich DQ. Accountability studies of air pollution and health effects: lessons learned and recommendations for future natural experiment opportunities. *Environment International* 2017;**100**:62-78.

Rohwer 2017

Rohwer AC, Pfadenhauer LM, Burns J, Brereton L, Gerhardus A, Booth A, et al. Series: Clinical Epidemiology in South Africa. Paper 3: Logic models help make sense of complexity in systematic reviews and health technology assessments. *Journal of Clinical Epidemiology* 2017;**83**:37-47.

Rutter 2017

Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. *Lancet* December 2017;**390**(10112):2602-4.

Rückerl 2011

Rückerl R, Schneider A, Breitner S, Cyrys J, Peters A. Health effects of particulate air pollution: A review of epidemiological evidence. *Inhalation Toxicology* 2011;**23**(10):555-92.

Schindler 2009

Schindler C, Keidel D, Gerbase MW, Zemp E, Bettschart R, Brändli O, et al. Improvements in PM10 Exposure and Reduced Rates of Respiratory Symptoms in a Cohort of Swiss Adults (SAPALDIA). *American Journal of Respiratory and Critical Care Medicine* 2009;**179**(7):579-87.

Schleicher 2011

Schleicher N, Norra S, Dietze V, Yu Y, Fricker M, Kaminski U, et al. The effect of mitigation measures on size distributed mass concentrations of atmospheric particles and black carbon concentrations during the Olympic Summer Games 2008 in Beijing. *Science of the Total Environment* 2011;**412-413**:185-93.

Schleicher 2012

Schleicher N, Norra S, Chen Y, Chai F, Wang S. Efficiency of mitigation measures to reduce particulate air pollution--a case study during the Olympic Summer Games 2008 in Beijing, China. *Science of the Total Environment* 2012;**427-8**:146-58.

Schmitt 2016

Schmitt LHM. QALY gain and health care resource impacts of air pollution control: A Markov modelling approach. *Environmental Science & Policy* 2016;**63**:35-43.

Schulz 2010

Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;**340**:c332.

Schünemann 2018

Schünemann HJ, Cuello C, Akl EA, Mustafa RA, Meerpohl JJ, Thayer K, et al. GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. Journal of Clinical Epidemiology (in press).

Shen 2011

Shen J, Tang A, Liu X, Kopsch J, Fangmeier A, Goulding K, Zhang F. Impacts of pollution controls on air quality in Beijing during the 2008 Olympic Games. *Journal of Environmental Quality* 2011;**40**(1):37-45.

Shon 2011

Shon ZH, Kim KH. Impact of emission control strategy on NO2 in urban areas of Korea. *Atmospheric Environment* 2011;**45**(3):808-12.

Shu 2014

Shu S, Quiros DC, Wang R, Zhu Y. Changes of street use and on-road air quality before and after complete street retrofit: An exploratory case study in Santa Monica, California. *Transportation Research Part D: Transport and Environment* 2014;**32**:387-96.

Shu 2016

Shu S, Batteate C, Cole B, Froines J, Zhu Y. Air quality impacts of a CicLAvia event in downtown Los Angeles, CA. *Environmental Pollution* 2016;**208**:170-6.

Song 2015

Song W, Chang Y, Liu X, Li X, Li K, Gong Y, et al. A multiyear assessment of air quality benefits from China's emerging shale gas revolution: Urumqi as a case study. *Environmental Science & Technology* 2015;**49**(4):2066-72.

Sterne 2001

Sterne JAC, Cox DR, Smith GD. Sifting the evidence—what's wrong with significance tests?Another comment on the role of statistical methods. *BMJ* 2001;**322**(7280):226.

Sterne 2016

Sterne JAC, Hernán MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919.

Su 2015

Su C, Hampel R, Franck U, Wiedensohler A, Cyrys J, Pan X, et al. Assessing responses of cardiovascular mortality to particulate matter air pollution for pre-, during- and post-2008 Olympics periods. *Environmental Research* 2015;**142**:112-22.

Tao 2015

Tao J, Zhang L, Zhang Z, Huang R, Wu Y, Zhang R, et al. Control of PM2.5 in Guangzhou during the 16th Asian Games period: Implication for hazy weather prevention. *Science of the Total Environment* 2015;**508**:57-66.

Thomson 2012

Thomson HJ, Thomas S. The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Research Synthesis Methods* 2012;**4**(1):95-101.

Thornbush 2015

Thornbush MJ. Building health assessed through environmental parameters after the OTS in the city centre of Oxford, UK. *Area* 2015;**47**(4):354-9.

Tonne 2008

Tonne C, Beevers S, Armstrong B, Kelly F, Wilkinson P. Air pollution and mortality benefits of the London Congestion Charge: spatial and socioeconomic inequalities. *Occupational and Environmental Medicine* 2008;**65**(9):620-7.

Tonne 2017

Tonne C. A call for epidemiology where the air pollution is. *Lancet Planetary Health* 2016;**1**(9):E355-6.

Turley 2013

Turley R, Saith R, Bhan N, Rehfuess E, Carter B. Slum upgrading strategies involving physical environment and infrastructure interventions and their effects on health and socio-economic outcomes. *Cochrane Database of Systematic Reviews* 2013, Issue 1. [DOI: 10.1002/14651858.CD010067.pub2]

Turner 2017

Turner MC, Nieuwenhuijsen M, Anderson K, Balshaw D, Cui Y, Dunton G, Hoppin JA, Koutrakis P, Jerrett M. Assessing the exposome with external measures: commentary on the state of the science and research recommendations. *Annual Review of Public Health* 2017;**38**:215-39.

US EPA 2011

United States Environmental Protection Agency. The benefits and costs of the Clean Air Act 1990 to 2020. EPA 2011.

Valencia 2002

Valencia A, Huertas J. Reducing Air Pollution Of A Large City By Organizing Public Transportation. Air Pollution X. UK: WIT Press, 2002.

van Donkelaar 2015

van Donkelaar A, Martin RV, Brauer M, Boys BL. Use of Satellite Observations for Long-Term Exposure Assessment of Global Concentrations of Fine Particulate Matter. *Environmental Health Perspectives* 2015;**123**(2):135-43.

van Erp 2012

van Erp AM, Kelly FJ, Demerjian KL, Pope A III, Cohen AJ. Progress in research to assess the effectiveness of air quality interventions towards improving public health. *Air Quality, Atmosphere, & Health* 2012;**5**:217-30.

Vandenbroucke 2007

Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLOS Medicine* 2007;**4**(10):e297.

Voss 2013

Voss PH, Rehfuess EA. Quality appraisal in systematic reviews of public health interventions: an empirical study on the impact of choice of tool on meta-analysis. *Journal of Epidemiology and Community Health* 2013;**67**(1):98-104.

Wang 2014

Wang S, Gao J, Zhang Y, Zhang J, Cha F, Wang T, et al. Impact of emission control on regional air quality: An observational study of air pollutants before, during and after the Beijing Olympic Games. *Journal of Environmental Sciences (China)* 2014;**26**(1):175-80.

Wang 2016

Wang H, Zhao L, Xie Y, Hu Q. APEC blue-The effects and implications of joint pollution prevention and control program. *Science of the Total Environment* 2016;**553**:429-38.

Ward 2009

Ward TJ, Palmer CP, Houck JE, Navidi WC, Geinitz S, Noonan CW. Community woodstove changeout and impact on ambient concentrations of polycyclic aromatic hydrocarbons and phenolics. *Environmental Science & Technology* 2009;**43**(14):5345-50.

Ward 2010

Ward TJ, Palmer CP, Noonan CW. Fine particulate matter source apportionment following a large woodstove changeout program in Libby, Montana. *Journal of the Air & Waste Management Association* 2010;**60**(6):688-93.

Wells 2012

Wells M, Williams B, Treweek S, Coyle J, Taylor J. Intervention description is not enough: evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. *Trials* 2012;**13**:95.

WHO 2016

World Health Organization. Ambient air pollution: A global assessment of exposure and burden of disease. 2016.

WHO Europe 2013

Europe WHO. Review of evidence on health aspects of air pollution – REVIHAAP Project: Technical Report. Copenhagen: WHO Regional Office for Europe, 2013.

Wong 2012

Wong CM, Rabl A, Thach TQ, Chau YK, Chan KP, Cowling BJ, et al. Impact of the 1990 Hong Kong legislation for restriction on sulfur content in fuel. *Research Report (Health Effects Institute)* 2012;**170**:5-91.

Xu 2013

Xu HM, Tao J, Ho SSH, Ho KF, Cao JJ, Li N, et al. Characteristics of fine particulate non-polar organic compounds in Guangzhou during the 16th Asian Games: Effectiveness of air pollution controls. *Atmospheric Environment* 2013;**76**:94-101.

Xu 2016

Xu R, Tang G, Wang Y, Tie X. Analysis of a long-term measurement of air pollutants (2007–2011) in North China Plain (NCP); Impact of emission reduction during the Beijing Olympic Games. *Chemosphere* 2016;**159**:647-58.

Supplemental material

The full supplemental material for this Cochrane Review is available at:

https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD010919.pub2/full

6. Paper II

Received: 27 April 2017 | Revised: 4 October 2017 | Accepted: 23 October 2017

DOI: 10.1002/jrsm.1284

BRIEF METHOD NOTE

WILEY Research Synthesis Methods

Looking beyond the forest: Using harvest plots, gap analysis, and expert consultations to assess effectiveness, engage stakeholders, and inform policy

J. Burns[1] | S. Polus[1] | L. Brereton[2,3] | J. Chilcott[2] | S.E. Ward[2] | L.M. Pfadenhauer[1] | E.A. Rehfuess[1]

[1] Institute for Medical Informatics, Biometry and Epidemiology, LMU Munich, Marchioninistrasse 15, 81377 Munich, Germany

[2] The School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, UK

[3] College of Health and Social Sciences, University of Lincoln, Lincoln, UK

Correspondence
J Burns, Institute for Medical Informatics, Biometry and Epidemiology, LMU Munich, Marchioninistrasse 15, 81377 Munich, Germany.
Email: burns@ibe.med.uni-muenchen.de

Funding information
FP7 Health, Grant/Award Number: 306141; European Union Seventh Framework Programme, Grant/Award Number: 306141

We describe a combination of methods for assessing the effectiveness of complex interventions, especially where substantial heterogeneity with regard to the population, intervention, comparison, outcomes, and study design of interest is expected. We applied these methods in a recent systematic review of the effectiveness of reinforced home-based palliative care (rHBPC) interventions, which included home-based care with an additional and explicit component of lay caregiver support. We first summarized the identified evidence, deemed inappropriate for statistical pooling, graphically by creating harvest plots. Although very useful as a tool for summary and presentation of overall effectiveness, such graphical summary approaches may obscure relevant differences between studies. Thus, we then used a gap analysis and conducted expert consultations to look beyond the aggregate level at how the identified evidence of effectiveness may be explained. The goal of these supplemental methods was to step outside of the conventional systematic review and explore this heterogeneity from a broader perspective, based on the experience of palliative care researchers and practitioners. The gap analysis and expert consultations provided valuable input into possible underlying explanations in the evidence, which could be helpful in the further adaptation and testing of existing rHBPC interventions or the development and evaluation of new ones. We feel that such a combination of methods could prove accessible, understandable, and useful in informing decisions and could thus help increase the relevance of systematic reviews to the decision-making process.

KEYWORDS
complex interventions, evidence synthesis, expert consultation, harvest plot, systematic review

Disclaimer: The sole responsibility for the content of this presentation lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein.

 |

1 | EVIDENCE SYNTHESIS IN THE ABSENCE OF META-ANALYSIS: NEED FOR METHODS DEVELOPMENT

The challenges associated with conducting systematic reviews of complex interventions have been well documented; a potentially broad research question that requires intricate, multidisciplinary searches may lead to the collection of very heterogeneous evidence, with a potentially wide range of methodological characteristics, included populations, interventions, comparisons, outcomes, and results.[1] A range of novel meta-analytical and other statistical methods exist to address and assess such heterogeneity,[2,3] but a critical decision for the reviewer is nevertheless whether the identified evidence is sufficiently homogenous to be statistically combined in a meta-analysis.[4] In systematic reviews of complex interventions, the a priori expectation of substantial heterogeneity among studies often leads reviewers to forgo a meta-analysis, deciding instead for a narrative synthesis. A narrative synthesis of effectiveness evidence on its own, however, may prove lengthy and inaccessible to the end user[5] and in fact may leave the decision maker to make further sense of the evidence on his or her own. Thus, evidence in this form may not be ideal for informing decision makers.[6,7]

In contrast, a clear, accessible summary is particularly important to decision makers, and non–meta-analytical graphical summary methods have been shown to be an informative and comprehensible mode of presenting results of systematic reviews. The forest plot without a pooled effect estimate, for example, provides an overview of the effects for all studies assessing a given outcome and is likely already familiar to various stakeholders.[4,8] Other graphical methods, like the harvest plot,[9-11] the effect direction plot,[5] and the bubble plot,[12,13] can summarize large bodies of information, usually facilitating the arrangement of various intervention types, outcomes, and other aspects in a single structure. The albatross plot is more statistical in nature and attempts to illustrate the relationship between the *P* value, the effect size, and the size of the assessed population.[14,15] Each method is unique in its presentation of results from primary studies, but all aim to summarize and present intervention effects across studies in an accessible and user-friendly manner. The lack of a meta-analysis, nevertheless, means that most systematic reviews using a graphical summary method will fail to provide the precise quantitative answer that decision-makers may desire and may lead to further questions about included studies and how aspects of these studies may influence intervention effectiveness. Thus, a way to extract more detailed information from the underlying systematic review, which facilitates a better understanding of included participants, interventions, outcomes, context, or other aspects, could be a valuable complement to graphical summary methods and could thus help increase the relevance of systematic reviews to the decision-making process.

As part of the recently completed European Union-funded INTEGRATE-HTA project,[16] we developed concepts and methods for the comprehensive, integrated assessment of complex interventions. These concepts and methods were then applied in a demonstration health technology assessment (HTA) on reinforced home-based palliative care (rHBPC), which refers to home palliative care with an additional component of lay caregiver support.[17] We chose to assess rHBPC within INTEGRATE-HTA because, based on the current literature on complexity of health interventions and health systems, it can be considered highly complex: There are multiple interacting components, a unit of care composed of the patients and their lay caregivers, as well as multiple service providers and other stakeholders at various levels, a range of physical, psychological, and spiritual outcomes, and the need for a degree of tailoring.[18] Additionally, the interactions between the intervention, context, setting, and implementation likely influence effectiveness,[19] and these various aspects may change in adaptive ways over time.[2,20] The methods applied in the effectiveness assessment of rHBPC are presented in this paper.

In the following section, we will briefly describe the scope of the systematic review of effectiveness, including the population, intervention, comparison, outcomes, and study designs of interest. In the subsequent section, we will describe the combination of methods applied for summarizing, presenting, and further exploring the evidence included in this review, which included creating harvest plots and a subsequent gap analysis and expert consultations. In the final sections, we will present the results from this combined approach to evidence synthesis and interpretation and briefly discuss the strengths and limitations of the applied methods.

2 | SCOPE AND OVERVIEW OF METHODS OF SYSTEMATIC REVIEW OF RHBPC

Following a Cochrane review that showed mixed results regarding the effectiveness of identified interventions,[21] we aimed to update the evidence base and assess the effectiveness of rHBPC interventions across a range of health outcomes in adult patients and their lay caregivers. The review scope is summarized in Box 1.

Box 1: Clinical and methodological scope of systematic review

Population: We included all adults (≥18 years) with any life-limiting condition receiving rHBPC. We included all lay caregivers, as the lay caregiving role may be taken on by any number of individuals and is by no means limited to family.[22]
Intervention: rHBPC encompasses a wide range of services. For the purpose of this review, we included any intervention, which allowed patients to receive care primarily at home, and which additionally used an explicit component focusing on supporting the lay caregiver. This additional support included any psycho-educational intervention aimed at providing assistance to lay caregivers (eg, individual or group counseling, education, advice, or respite services, which alleviate burden).
Comparison: We included any comparator, as during protocol development, it became clear that services offered to patients and caregivers as part of usual care were very heterogeneous.
Outcomes: Patient outcomes included pain, symptom control, quality of life (QoL), psychological health, death at home, hospitalization, response (eg, coping, preparedness, and mastery), and satisfaction with care. Lay caregiver outcomes included QoL, psychological health, response, and satisfaction with care.
Study designs: We included studies applying any of the following designs.

- Patient or cluster randomized controlled trials
- Patient or cluster nonrandomized controlled trials
- Controlled before-after studies with at least 2 interventions and 2 control sites.[23]
- Interrupted time series studies with at least 3 data points both before and after a clearly defined intervention.[23]

We searched for and selected studies and appraised the quality of included studies in line with Gomes et al[21] and guidance published by Cochrane.[4] A more detailed description of the scope and methods can be found in the review protocol, available online.[24]

3 | EVIDENCE SYNTHESIS AND BEYOND: HARVEST PLOTS, GAP-ANALYSIS, AND EXPERT CONSULTATIONS

At the evidence synthesis stage, we diverged from the methods applied in the original review by Gomes et al,[21] where a narrative synthesis and a limited number of meta-analyses were performed. Based on the expected clinical and methodological heterogeneity of studies, we decided a priori to forgo meta-analysis and to present findings graphically through harvest plots. We arranged studies on a matrix in columns according to direction of effect-favors control, no difference or favors intervention, and in rows according to the outcome category. Additionally, information regarding study design—represented by the height of the bar, and where no statistical analysis was performed—indicated with a dotted border, was portrayed. The color of the bar designates whether that study was originally included in Gomes et al[21] (white) or newly identified through our review update (black).

We recognized that while harvest plots are a good means of providing an overview of the evidence of effectiveness, decision makers tend to be interested in more detailed and concrete information regarding the various populations, interventions, and outcomes. Systematic review authors increasingly engage content experts, both at the planning and execution stage, in the hopes to increase the relevance and utility of review results.[4,25-27] Thus, in an attempt to engage with experts in palliative care practice, we subsequently conducted a gap analysis and expert consultations to further explore the review results. “Gap analysis” is a catchall term used to describe a range of methods applied in many scenarios.[28-30] In this study, gap analysis refers to the process by which the entire review team, with expertise in palliative care, effectiveness and cost-effectiveness research, and evidence synthesis, sought to examine the main findings related to rHBPC effectiveness in an open and iterative discussion. Gaps could, for example, be open questions or inconsistencies around study methods, included populations, interventions, comparisons, or outcomes, as well as about the effects observed in the included studies. These identified gaps, which we refer to as “emerging aspects,” were used as a flexible structure for the one-on-one consultations with palliative care practitioners and researchers, as explained below and for summarizing the insights obtained.

Following the gap analysis, 4 palliative care professionals, including researchers and practitioners with knowledge and experience in palliative care from England, Germany, and the Netherlands, were consulted individually via telephone or Skype. These individuals were purposively selected from a group of experts that had previously expressed interest in contributing to INTEGRATE-HTA. Each expert was provided the opportunity to study the review protocol and the harvest plots and was asked to discuss methodological or palliative care-related issues relevant to the emerging aspects arising from the gap analysis. For example, if a certain type of intervention seemed to be comparatively effective, the experts would discuss, based on their knowledge and experiences, why this particular intervention may be observed as effective. As well as discussing the emerging aspects, experts were invited to contribute other relevant questions, comments, or topics. Each consultation was audiotaped to ensure fidelity. We reviewed consultation findings descriptively using the emerging aspects to structure the findings. As an author team, we then aimed to further distill the insights into potential implications for research and practice.

4 | RESULTS

The results of the study selection process can be seen in Figure S1. We included 9 studies assessing rHBPC, 5 included in the original review[31-35] and 4 newly identified through our updated searches.[36-39] The studies differed widely with regard to the study setting, population, intervention, comparison, and outcomes, and detailed information on these aspects is provided in Table S1.

The harvest plots provide an overall summary of the effect estimates of included primary studies across all outcomes. For caregiver outcomes, most of the 9 studies showed no greater benefit for rHBPC than for standard nonreinforced home-based interventions; a small number of studies showed some positive effects (Figure 1). Although rHBPC interventions focused mostly on lay caregivers, 5 studies also assessed patient outcomes (Figure 2). For pain, QoL, hospitalization, patient response, and satisfaction of care, there appeared to be no difference between rHBPC and non-rHBPC interventions. Symptom control and psychological health displayed a mix of positive intervention effects and no effect.

Through the gap analysis, the review team identified 4 emerging aspects, which potentially influenced the effectiveness of the included rHBPC interventions or the assessment of effectiveness (Table 1, "Emerging Aspect" column). These included (1) the heterogeneity and ambiguity of the primary study comparator, nonreinforced care, against which rHBPC interventions were compared; (2) the potential lack of individually tailored care based on patient and caregiver needs; (3) the appropriateness of outcomes used in the review, as well as in primary studies; and (4) the primary study designs with which these interventions are usually evaluated.

In the subsequent consultations, experts highlighted both clinical and methodological aspects, such as the need to embrace more tailored, evolving care, the use of more responsive outcomes and more appropriate study designs, and overall better reporting in primary research. A summary of the findings of these consultations is provided in Table 1 ("Expert Consultations" column).

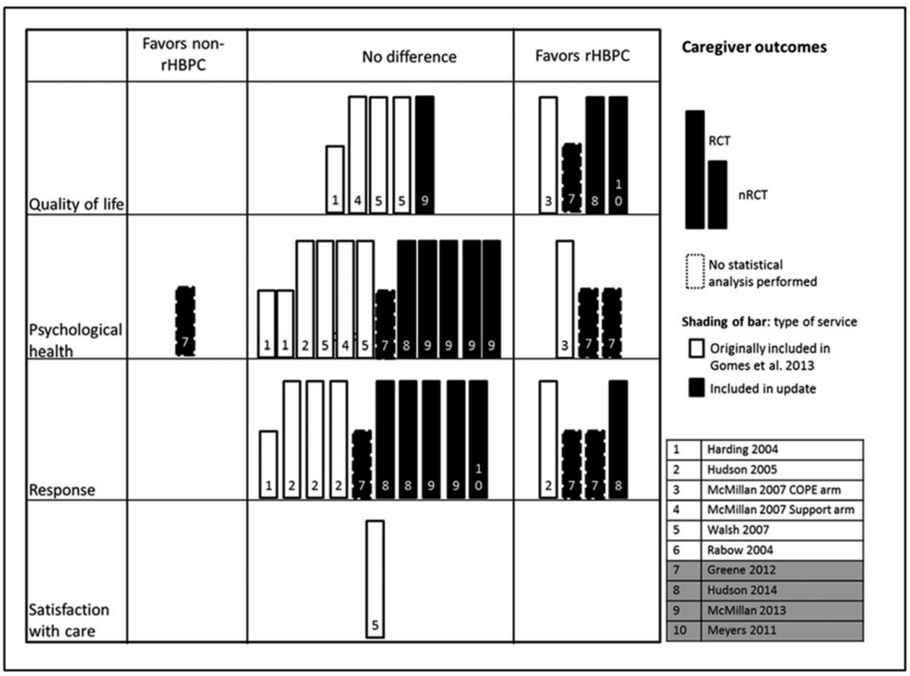


FIGURE 1 Effect estimates of included rHBPC interventions for lay caregiver outcomes

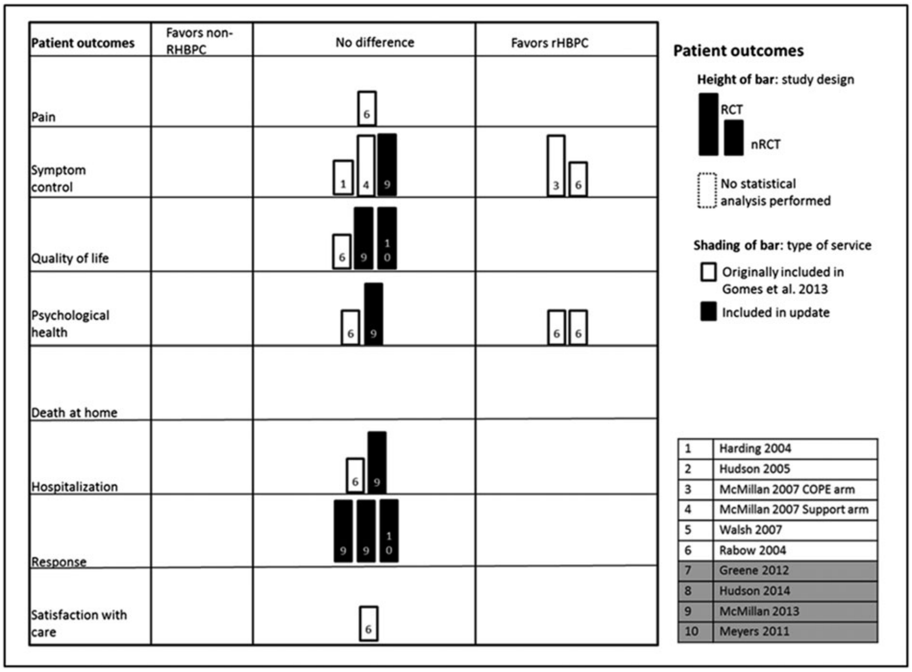


FIGURE 2 Effect estimates of included rHBPC interventions for patient outcomes

5 | DISCUSSION

As we felt that statistical aggregation was unlikely to be appropriate given expected heterogeneity in populations, interventions, outcomes, and study methods, the graphical summary provided by the harvest plots allowed us to produce evidence on effectiveness that is accessible to and informative for decision makers.[6] The harvest plots show, for example, that rHBPC interventions were, for the most part, not effective in improving patient or lay caregiver outcomes. Harvest plots, however, do not tell the entire story; presenting the evidence in such an overview manner can obscure critical differences in individual studies, and recognizing such differences may require looking beyond the overall summary of evidence. At this stage, rather than concluding that rHBPC does not work, it may be valuable to examine both the factors that may have led to some interventions being more effective than others[40,41] and to step outside of the conventional systematic review and explore this heterogeneity from a broader perspective.

Thus, we consulted experts with the aim of supplementing the summary of evidence generated through the systematic review with more detailed information regarding the assessed populations, interventions and outcomes, and potentially, the effects (or lack thereof) of included studies. The inclusion of expert input in systematic reviews of effectiveness has been shown to add value, both at the planning stage to define the scope of the review[4,25,26] and at the evidence synthesis stage, for example, through Bayesian meta-analysis.[27] Our application of gap analysis and expert consultations as a supplement to the more traditional evidence synthesis likewise aimed to go beyond the conduct and reporting of any individual study, to enrich the results of the review with the knowledge and experience of experts. These experts did, in fact, highlight both clinical and methodological aspects, which could potentially be helpful in the further adaptation and testing of existing rHBPC interventions or the development and evaluation of new ones. For example, rHBPC interventions could be designed to be more tailor-fit to patients and their lay caregivers; or in evaluating interventions, researchers could look at outcomes and study designs that are more responsive in this population.

There were, of course, limitations in the application of this combination of methods. Harvest plots allow the presentation of a bulk of evidence, but readers may need some time to "orient" themselves. Another criticism of graphical summary techniques is that they could potentially encourage "vote-counting" practices, if readers or decision makers attempt to quantitatively compare the frequency of effect directions,[5] but this should be explicitly discouraged in association with harvest plots. The information gained from the expert consultations is useful but is based on personal experiences and is exploratory in nature, and thus should not be taken as hard evidence. Additionally, because of time and resource constraints, we were only able to conduct 4 consultations with experts from 3 countries, and we did not involve other stakeholders (ie, patients, lay caregivers, or other interested parties). Hence, further applications of these or similar

TABLE 1 Findings of gap analysis and expert consultations

Emerging Aspect	Gap Analysis	Expert Consultations	Potential Implications for Research and Practice
Primary study comparator (nonreinforced care)	The type of care, against which reinforced care was assessed, was poorly described in most included studies. Caregivers may be receiving substantial support through standard home-based services. Usual care, and especially the extent to which caregivers are supported as part of usual care, likely varies widely among included studies.	Usual care varies from place to place—not only from country to country—although there are very substantial differences to be seen at that level, but also within countries from one location to another. The support that caregivers receive as part of usual care is extremely heterogeneous. Some caregivers receive structured support throughout the illness trajectory, while others receive help only when they are overwhelmed by problems and seek care themselves. The extent to which caregivers are involved in decisions regarding patient care differs within "usual care," from virtually none, to playing a part in care-planning discussions.	In determining what care may be appropriate in a given setting, a clear understanding of what type of support patients and caregivers receive as part of usual care is likely to be critical to identifying whether rHBPC could be effective, and which additional, alternative, or complementary services could be warranted.
Lack of tailored care	Although some of the included interventions did offer some flexibility, it could be that for reinforced palliative care to be effective, targeted and tailored care should be more strongly emphasized and delivered to those patients and caregivers assessed as needing it most.	Care tailored to the individual patient and caregiver, at least to a certain extent, is seen as the best practice—this could be based on diagnosis, age, illness trajectory, social surrounding, etc—and the recognition of such indicators is important. Being able to assess the needs of patients and/or lay caregivers and to inform them of (evidence-based) options is essential. The health and social care professional, however, should not make assumptions about what patients and/or caregivers need or want, and they should be involved in these discussions. The needs of caregivers and patients are not static and will likely change over time and trajectory of the illness. This makes repeated assessments through ongoing communication important.	As changes in patient and caregiver needs occur frequently in relation to the illness trajectory, assessing these needs frequently and reacting to them through tailored care may be an important means to design more effective interventions.

(Continues)

TABLE 1 (Continued)

Emerging Aspect	Gap Analysis	Expert Consultations	Potential Implications for Research and Practice
Appropriateness of assessed outcomes	All of the outcomes assessed in this review have been used in the primary literature and are thought to be important for patients and caregivers. It should be considered, nevertheless, whether these are most appropriate, and whether certain additional or alternative outcomes should have been assessed, both in the primary literature and in this review.	Hard outcomes used in palliative care may only tell part of the story, and meaningful effects can potentially be hidden among the noise, eg, in a population so severely burdened, it may be unrealistic to expect clinically significant differences in quality of life. It is important, therefore, to ask patients and caregivers if their care has improved, and specifically what the benefits of care were. Outcome importance may differ between subgroups, and it is important to recognize this when evaluating services.	The outcomes used to assess rHPBC interventions should also be revisited, and standardized health outcomes such as QoL and psychological health should be supplemented with more qualitative accounts of patients' and caregivers' perceptions and experiences.
Primary study design	Included studies encountered a range of problems when implementing and assessing palliative care services—eg, attrition. Study designs, other than those included, may be more appropriate for assessing the effectiveness of reinforced home-based palliative care services.	Mixed methods and qualitative research should play a large role in assessing the effectiveness of services in a meaningful way—it is important to see what exactly is happening, to hear what patients and caregivers feel they are receiving, as opposed to assuming, based on the intervention design. If care is truly based on caregiver/patient assessment and therefore truly tailored to the individual, and because the goals of individual participants will be different, evaluation of care becomes very difficult, especially in a randomized trial. Other study designs—eg, process evaluations, qualitative studies, participative approaches, N-of-1 studies—should be considered.	Researchers should also revisit which research approaches are most appropriate for answering a given question in primary studies and systematic reviews. For effectiveness, they could consider designs other than the RCT, such as N-of-1 studies; for questions beyond effectiveness, qualitative studies, mixed-method studies, or process evaluations are likely to be valuable.

methods would benefit from consulting a larger, more diverse base of stakeholders.

For the purposes of assessing the effectiveness of rHBPC, this combination of harvest plots, followed by a gap analysis and expert consultations proved to be useful both in summarizing the evidence and identifying evidence gaps, as well as in looking beyond the aggregate level at how these findings may be explained. We would welcome applications of this approach or similar approaches to a range of interventions in health and other disciplines, potentially consulting a larger, more diverse base of stakeholders, to learn from the insights gained. In addition, it would be worth examining whether decision makers find such a combination of methods accessible, understandable, and useful in informing decisions.

FUNDING INFORMATION

The research leading to this publication is part of the project INTEGRATE-HTA and has received funding from the European Union Seventh Framework Programme under grant agreement 306141.

DECLARATION OF CONFLICTING INTERESTS

The authors declare that they have no conflict of interest.

ORCID

J. Burns http://orcid.org/0000-0003-4015-6862
S.E. Ward http://orcid.org/0000-0002-3344-0913
L.M. Pfadenhauer http://orcid.org/0000-0001-5038-8072
E.A. Rehfuess http://orcid.org/0000-0002-4318-8846

REFERENCES

1. Burford B, Lewin S, Welch V, Rehfuess E, Waters E. Assessing the applicability of findings in systematic reviews of complex interventions can enhance the utility of reviews for decision making. *J Clin Epidemiol*. 2013;66:1251-1261.
2. Petticrew M, Rehfuess E, Noyes J, et al. Synthesizing evidence on complex interventions: how meta-analytical, qualitative, and mixed-method approaches can contribute. *J Clin Epidemiol*. 2013;66(11):1230-1243. https://doi.org/10.1016/j.jclinepi.2013.06.005
3. Pigott T, Shepperd S. Identifying, documenting, and examining heterogeneity in systematic reviews of complex interventions. *J Clin Epidemiol*. 2013;66(11):1244-1250. https://doi.org/10.1016/j.jclinepi.2013.06.013
4. Higgins JPT, Green S, (Eds.). *Cochrane Handbook for Systematic Reviews of Interventions, Version [5.1.0]*. The Cochrane Collaboration; 2011. Available from www.cochrane-handbook.org.
5. Thomson HJ, Thomas S. The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Res Syn Meth*. 2013;4(1):95-101. https://doi.org/10.1002/jrsm.1060
6. Pettman TL, Hall BJ, Waters E, de Silva-Sanigorski A, Armstrong R, Doyle J. Communicating with decision-makers through evidence reviews. *J Public Health (Oxf)*. 2011;33(4):630-633. https://doi.org/10.1093/pubmed/fdr092
7. Sweet M, Moynihan R. *Improving Population Health: The Uses of Systematic Reviews*. New York, USA: Milbank Memorial Fund; Centers for Disease Control and Prevention; 2007.
8. Valentine JC, Thompson SG. Issues relating to confounding and meta-analysis when including non-randomized studies in systematic reviews on the effects of interventions. 2012.
9. Nehring I, Kostka T, von Kries R, Rehfuess EA. Impacts of in utero and early infant taste experiences on later taste acceptance: a systematic review. *J Nutr*. 2015;145(6):1271-1279. https://doi.org/10.3945/jn.114.203976
10. Ogilvie D, Fayter D, Petticrew M, et al. The harvest plot: a method for synthesising evidence about the differential effects of interventions. *BMC Med Res Methodol*. 2008;8(1):8. https://doi.org/10.1186/1471-2288-8-8
11. Turley R, Saith R, Bhan N, Rehfuess E, Carter B. Slum upgrading strategies involving physical environment and infrastructure interventions and their effects on health and socio-economic outcomes. *Cochrane Database Syst Rev*. 2013;(1):Cd010067. https://doi.org/10.1002/14651858.CD010067.pub2
12. Erasmus Y, Tannous N, Langer L. A systematic map of evidence on the links between ecosystem services and poverty alleviation in Africa: a user guide. Africa Centre for Evidence; 2017.
13. Totten AM, Womack DM, Eden KB, et al. *Telehealth: Mapping the Evidence for Patient Outcomes From Systematic Reviews*. Rockville (MD): Agency for Healthcare Research and Quality; 2016.
14. Harrison S, Jones HE, Martin RM, Lewis SJ, Higgins JPT. The albatross plot: a novel graphical tool for presenting results of diversely reported studies in a systematic review. *Res Syn Meth*. 2017;8(3):281-289. https://doi.org/10.1002/jrsm.1239
15. Hunter BM, Harrison S, Portela A, Bick D. The effects of cash transfers and vouchers on the use and quality of maternity care services: a systematic review. *Plos One*. 2017;12(3):e0173068. https://doi.org/10.1371/journal.pone.0173068
16. Gerhardus A. Integrated health technology assessment for evaluating complex technologies (INTEGRATE-HTA): an introduction to the guidances. 2016.
17. Brereton L, Wahlster P, Lysdahl KB, et al. Integrated assessment of home based palliative care with and without reinforced caregiver support: 'a demonstration of INTEGRATE-HTA methodological guidances'—Executive Summary. 2016. http://www.integrate-hta.eu/downloads/
18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337. https://doi.org/10.1136/bmj.a1655
19. Pfadenhauer LM, Gerhardus A, Mozygemba K, et al. Making sense of complexity in context and implementation: the context and implementation of complex interventions (CICI) framework. *Implement Sci*. 2017;12(1):21. https://doi.org/10.1186/s13012-017-0552-5
20. Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ*. 2001;323(7313):625-628. https://doi.org/10.1136/bmj.323.7313.625
21. Gomes B, Calanzani N, Curiale V, McCrone P, Higginson IJ. Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. *Cochrane Database Syst Rev*. 2013;(6):Cd007760. https://doi.org/10.1002/14651858.CD007760.pub2
22. NICE. *Guidance on Cancer Services: Improving Supportive and Palliative Care for Adults with Cancer. The Manual*. London: National Institute for Clinical Excellence; 2004.
23. EPOC, N. S. o. t. C. E. P. a. O. o. C. G. What study designs should be included in an EPOC review and what should they be called? [online]. 2014. Norwegian Satellite of the Cochrane Effective Practice and Organisation of Care Group EPOC. Available: http://epocoslo.cochrane.org/sites/epocoslo.cochrane.org/files/uploads/What%20study%20designs%20should%20be%20included%20in%20an%20EPOC%20reviews%20and%20what%20should%20they%20be%20called.pdf [Accessed 20 Feb, 2016.
24. Burns J, Polus S, Brereton L, et al. Home-based palliative care: assessing effectiveness through harvest plots and post-review consultations—effectiveness review protocol. 2014b. http://

www.ibe.med.uni-muenchen.de/mitarbeiter/mitarbeiter/jakob_burns/index.htm

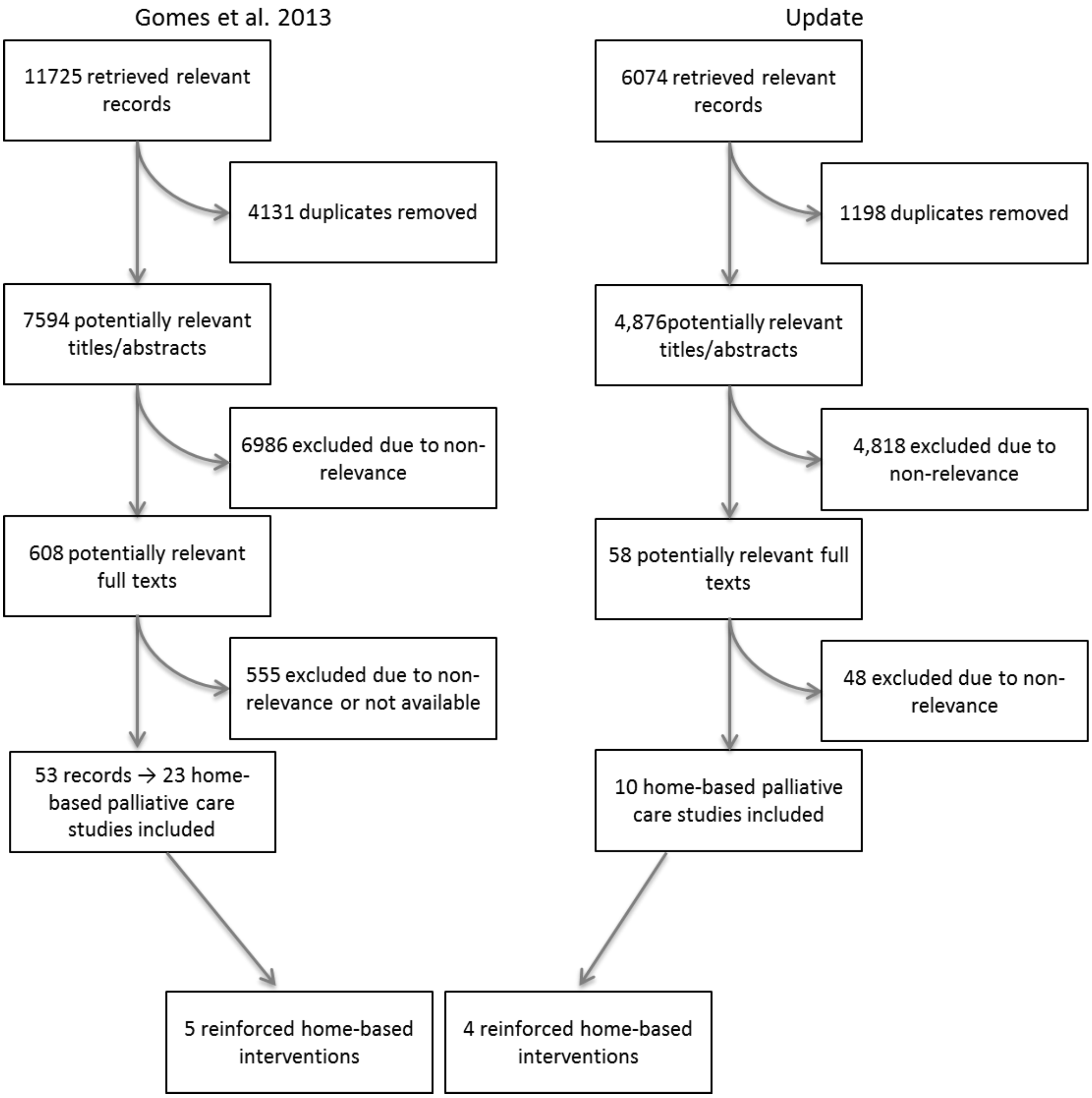
25. Burns J, Boogaard H, Turley R, et al. Interventions to reduce ambient particulate matter air pollution and their effect on health (protocol). *Cochrane Database Syst Rev*. 2014a;1. https://doi.org/10.1002/14651858.CD010919
26. von Philipsborn P, Stratil JM, Burns J, et al. Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health. *Cochrane Database Syst Rev*. 2016;(7).
27. Woertman W, Vermeulen B, Groenewoud H, van der Wilt GJ. Evidence based policy decisions through a Bayesian approach: the case of a statin appraisal in The Netherlands. *Health Policy*. 2013;112(3):234-240. https://doi.org/10.1016/j.healthpol.2013.06.015
28. Al-Momani MM. Gap analysis between perceptions and expectations of medical-surgical patients in a public hospital in Saudi Arabia. *J Clin Epidemiol*. 2016;25:79-84.
29. Bidulescu A, Ferguson TS, Hambleton I, et al. Educational health disparities in hypertension and diabetes mellitus among African descent populations in the Caribbean and the USA: a comparative analysis from the Spanish town cohort (Jamaica) and the Jackson heart study (USA). *Int J Equity Health*. 2017;16(1):33. https://doi.org/10.1186/s12939-017-0527-9
30. Zeng W, Shepard DS, Avila-Figueroa C, Ahn H. Resource needs and gap analysis in achieving universal access to HIV/AIDS services: a data envelopment analysis of 45 countries. *Health Policy Plan*. 2016;31(5):624-633. https://doi.org/10.1093/heapol/czv109
31. Harding R, Higginson IJ, Leam C, et al. Evaluation of a short-term group intervention for informal carers of patients attending a home palliative care service. *J Pain Symptom Manage*. 2004;27(5):396-408. https://doi.org/10.1016/j.jpainsymman.2003.09.012
32. Hudson PL, Aranda S, Hayman-White K. A psycho-educational intervention for family caregivers of patients receiving palliative care: a randomized controlled trial. *J Pain Symptom Manage*. 2005;30(4):329-341. https://doi.org/10.1016/j.jpainsymman.2005.04.006
33. McMillan SC, Small BJ, Weitzner M, et al. Impact of coping skills intervention with family caregivers of hospice patients with cancer: a randomized clinical trial. *Cancer*. 2006;106(1):214-222. https://doi.org/10.1002/cncr.21567
34. Rabow MW, Dibble SL, Pantilat SZ, McPhee SJ. The comprehensive care team: a controlled trial of outpatient palliative medicine consultation. *Arch Intern Med*. 2004;164(1):83-91. https://doi.org/10.1001/archinte.164.1.83
35. Walsh K, Jones L, Tookman A, et al. Reducing emotional distress in people caring for patients receiving specialist palliative care. Randomised trial. *Br J Psychiatry*. 2007;190(2):142-147. https://doi.org/10.1192/bjp.bp.106.023960
36. Greene A, Aranda S, Tieman JJ, Fazekas B, Currow DC. Can assessing caregiver needs and activating community networks improve caregiver-defined outcomes? A single-blind, quasi-experimental pilot study: community facilitator pilot. *Palliat Med*. 2012;26(7):917-923. https://doi.org/10.1177/0269216311421834
37. Hudson P, Trauer T, Kelly B, et al. Reducing the psychological distress of family caregivers of home-based palliative care patients: short-term effects from a randomised controlled trial. *Psychooncology*. 2013;22(9):1987-1993. https://doi.org/10.1002/pon.3242
38. McMillan SC, Small BJ, Haley WE, Zambroski C, Buck HG. The COPE intervention for caregivers of patients with heart failure: an adapted intervention. *J Hosp Palliat Nurs*. 2013;15:207-208.
39. Meyers FJ, Carducci M, Loscalzo MJ, Linder J, Greasby T, Beckett LA. Effects of a problem-solving intervention (COPE) on quality of life for patients with advanced cancer on clinical trials and their caregivers: simultaneous care educational intervention (SCEI): linking palliation and clinical trials. *J Palliat Med*. 2011;14(4):465-473. https://doi.org/10.1089/jpm.2010.0416
40. Anderson LM, Fau OS, Fau MS, Fau RE, Fau NJ, Shemilt I. Investigating complexity in systematic reviews of interventions by using a spectrum of methods. *J Clin Epidemiol*. 2013;66:1223-1229.
41. Squires JE, Valentine JC, Grimshaw JM. Systematic reviews of complex interventions: framing the review question. *J Clin Epidemiol*. 2013;66:1215-1222.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Burns J, Polus S, Brereton L, et al. Looking beyond the forest: Using harvest plots, gap analysis, and expert consultations to assess effectiveness, engage stakeholders, and inform policy. *Res Syn Meth*. 2018;9:132–140. https://doi.org/10.1002/jrsm.1284

Supplementary Figure 1 – Study selection flowchart



Supplementary Table 1 – Included Study characteristics

Study ID	Study design	Country	Population	Life expectancy	Intervention	Patient outcomes (x)* indicates the multiple outcomes in the same category were assessed	Caregiver outcomes (x)* indicates the multiple outcomes in the same category were assessed	Time points assessed (after baseline)
rHBPC interventions								
Originally included studies								
Harding 2004 (45)	CCT	UK	73 lay caregivers of patients with mixed diagnoses	–	Thematic group discussions; Pragmatic support	Symptom control	Quality of life Psychological health (2)* Caregiver response	2, 5 months
Hudson 2005 (40)	RCT	Australia	106 lay caregivers of patients with of advanced cancer	–	Comprehensive care plans; support in executing plan and preparing for patient death		Psychological health Caregiver response (4)*	5 weeks, 2 months after patient death
McMillan 2006 (36)	RCT	USA	329 lay caregivers of patients with advanced cancer.	–	COPE problem-solving approach	Symptom control	Quality of life Psychological health	2, 4 weeks
Rabow 2004 (26)	CCT	USA	90 patients and lay caregivers of patients with cancer, advanced COPD, or advanced CHF	1-5 years	Integrated case management; formal classes and individual consultations for caregivers	Pain Symptom control Quality of life Psychological health (3)* Hospitalization Satisfaction with care		6, 12 months
Walsh 2007 (46)	RCT	UK	228 lay caregivers of patients with advanced cancer	4 months	Supportive visits focusing on advice, discussion and emotional support		Quality of life (2)* Psychological health (2)* Satisfaction with care	4, 9 , 12 weeks
Newly included studies								
Greene 2012 (41)	CCT	Australia	66 lay caregivers of patients in a designated geographical area	–	Training and support for caregiver role; identification and utilization of social		Quality of life Psychological health (4)*	1, 2 months

					support network		Caregiver response	
Hudson 2014 (42)	RCT	Australia	298 lay caregivers of advanced cancer patients	–	Care plan; preparation for patient death; bereavement support; referral to other services		Quality of life Psychological health Caregiver response (3)*	5 weeks, 8 weeks after patient death
McMillan 2013 (37)	RCT	USA	40 lay caregivers of patients with heart disease	–	COPE problem-solving approach	Symptom control Quality of life Psychological health (2)* Hospitalization Patient response (2)*	Quality of life Psychological health (4)* Caregiver response (2)*	4, 5 weeks
Meyers 2011 (38)	RCT	USA	476 advanced cancer patients participating in clinical trials and their caregivers (study dyads)	–	COPE problem-solving approach	Quality of life Patient response	Quality of life Caregiver response	1, 2, 3, 4, 6 months

7. Paper III

Atmospheric Environment 246 (2021) 118089

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: http://www.elsevier.com/locate/atmosenv

COVID-19 mitigation measures and nitrogen dioxide – A quasi-experimental study of air quality in Munich, Germany

Jacob Burns[a,c,*], Sabine Hoffmann[a,c], Christoph Kurz[b,c,d], Michael Laxy[b,c], Stephanie Polus[a,c], Eva Rehfuess[a,c]

[a] *Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Germany*
[b] *Institute of Health Economics and Health Care Management, Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Germany*
[c] *Pettenkofer School of Public Health, Munich, Germany*
[d] *Munich School of Management and Munich Center of Health Sciences, LMU Munich, Germany*

HIGHLIGHTS

- The effect of COVID-19 mitigation measures on NO_2 in Munich was unclear.
- We applied two robust quasi-experimental approaches.
- All hypotheses, as well as main and additional analyses were defined a priori.
- As hypothesized, we observed largest reductiotns in NO_2 at traffic sites.

ARTICLE INFO

Keywords:
Accountability
Quasi-experimental study
Nitrogen dioxide
Air quality
COVID-19

ABSTRACT

Background: In response to the COVID-19 pandemic, the Bavarian State government announced several COVID-19 mitigation measures beginning on March 16, 2020, which likely led to a reduction in traffic and a subsequent improvement in air quality. In this study, we evaluated the short-term effect of COVID-19 mitigation measures on NO_2 concentrations in Munich, Germany.

Methods: We applied two quasi-experimental approaches, a controlled interrupted time-series (c-ITS) approach and a synthetic control (SC) approach. Each approach compared changes occurring in 2020 to changes occurring in 2014–2019, and accounted for weather-related and other potential confounders. We hypothesized that the largest reductions in NO_2 concentrations would be observed at traffic sites, with smaller reductions at urban background sites, and even small reductions, if any, at background sites. All hypotheses, as well as the main and additional analyses were defined a priori. We also conducted post-hoc analyses to ensure that observed effects were not due to factors other than the intervention.

Results: Main analyses largely supported our hypotheses. Specifically, at the two traffic sites, using the c-ITS approach we observed reductions of 9.34 μg/m^3 (95% confidence interval: −23.58; 4.90) and 10.02 μg/m^3 (−19.25; −0.79). Using the SC approach we observed reductions of 15.65 μg/m^3 (−27.58; −4.09) and 15.1 μg/m^3 (−24.82; −9.83) at these same sites. We observed effects ranging from smaller in magnitude to no effect at urban background and background sites. Additional analyses showed that the effect was largest in the first two weeks following introduction of measures, and that a 3-day lagged intervention time also showed a larger effect. Post-hoc analyses suggested that at least some of the observed effects may have been attributable to changes in air quality occurring before the intervention, as well as unusually high concentrations in January 2020.

Conclusion: We applied two quasi-experimental approaches in assessing the impact of the COVID-19 mitigation measures on NO_2 concentrations in Munich. Taking the 2020 pre-intervention average concentrations as a reference, we observed reductions in NO_2 concentrations of approximately 15–25% and 24–36% at traffic sites, suggesting that reducing traffic may be an effective measure to reduce NO_2 concentrations in heavily trafficked areas by margins which could translate to public health benefits.

* Corresponding author. Institute for Medical Information Processing, Biometry, and Epidemiology – IBE, LMU Munich, Marchioninistr. 17, 81377, Munich, Germany.

E-mail address: burns@ibe.med.uni-muenchen.de (J. Burns).

https://doi.org/10.1016/j.atmosenv.2020.118089
Received 27 June 2020; Received in revised form 17 November 2020; Accepted 19 November 2020
Available online 22 November 2020
1352-2310/

1. Background

In December 2019 the first cases of the novel coronavirus, SARS-CoV-2, were observed in Wuhan, China. Over the next days and weeks the virus, and the associated respiratory disease referred to as COVID-19, spread further into China and by mid-January cases were documented in Thailand, Japan and South Korea (WHO, 2020a). By March 11, 2020, when the World Health Organization declared COVID-19 a global pandemic, cases had been observed in over 100 countries and territories across the globe (WHO, 2020b).

To slow the spread of this viral respiratory infection, the effects of which range from limited or no symptoms to death, national and sub-national governments have implemented numerous mitigation measures (Health System Response MONITOR, 2020). These mitigation measures differ between countries, but include, for example, social distancing recommendations and requirements, school closures, border closures, non-essential business closures and required wearing of masks.

Such external shocks can be conceptualized as natural experiments to explore the short-term effect that decreased automobile traffic or industrial activity has on ambient pollutant concentrations. A recent systematic review (BURNS et al., 2019, BURNS et al., 2020) identified a range of such studies, for example, evaluating the effect on air quality or health of the closure of a main highway for construction in California, US (HONG et al., 2015), the US Democratic National Convention in Boston (LEVY et al., 2006), the suspension of the public transportation system due to a strike in Ottawa, Canada (DING et al., 2014), the suspension of trucking operations due to a nationwide strike in India (Latha et al., 2004), political demonstrations in Nepal (Fransen et al., 2013) and Hong Kong (Brimblecombe and Ning, 2015), and the closure of a copper smelter due to a strike in the Southwest US (POPE et al., 2007).

Limited evidence already suggests that COVID-19 mitigation measures may have led to reductions in air pollution. For example satellite imagery has shown that concentrations of nitrogen dioxide (NO_2), a pollutant largely stemming from automobile traffic, have decreased in China (ESA, 2020c), India (ESA, 2020a) and across several European cities (ESA, 2020b). Monitor-based measurements have also implied decreased NO_2 concentrations in some European cities (EEA, 2020). Researchers on each of these projects, however, have been quick to emphasize the influential role that weather and other factors, such as celebration of the Chinese New Year, have on NO_2 concentrations, and that fully adjusting for the effects of such measures using standard epidemiological approaches is challenging.

Embedded in the national COVID-19 response, the Bavarian State government announced several COVID-19 mitigation measures beginning on March 16, 2020 (Bayerische STAATSREGIERUNG, 2020). As several of these measures could plausibly lead to reduced automobile traffic, it provided a unique opportunity to assess the effects of these measure on air quality, and to do so using rigorous quasi-experimental.

2. Objective

In this study, we applied two quasi-experimental approaches, a controlled interrupted time-series (c-ITS) approach and a synthetic control (SC) approach, to evaluate the short-term effect of COVID-19 mitigation measures on NO_2 concentrations in Munich, Germany.

3. Methods

3.1. Intervention and context

On March 13, 2020 the Bavarian state government announced that, aiming to mitigate the spread of COVID-19, schools across Bavaria would be closed from 16 March until at least April 19, 2020. This initial announcement was followed by several further mitigation measures over the next week. The measures implemented in Bavaria included:

March 2020 – closure of schools and daycare facilities

March 17, 2020 – closure of public facilities, ban on gatherings and events, closure of retail stores, restrictions on the restaurant industry;

March 18, 2020 – closure of institutes of higher education, ban on visits to hospitals and care facilities;

March 21, 2020 – ban on dine-in services for restaurant industry, partial lockdown (Bayerische Staatsregierung, 2020).

As a part of these various measures, individuals were encouraged, and then required, where possible, to remain home. Evidence suggests that Bavarian residents largely adhered to these measures. An analysis of the effective reproduction number of the virus, i.e. the expected number of cases generated by an infected individual, showed that the number fell from approximately 3.5 to 1.0 between 16 March and 3 April in Bavaria. In Munich, the effective reproduction number fell from approximately 3.0 to 0.5 over the same time period (Khailaie et al., 2020). Given that these measures initially encouraged and later on required people to stay at home, it is likely that these measures led to a decrease in people's movements. Indeed, mobility data released by Apple and Google show a clear reduction in traffic during this time (APPLE (2020); GOOGLE (2020)). Traffic critical to essential supply chains, as well as some local traffic related to grocery shopping or outdoor recreational activity, likely did not decrease or decreased to a lesser extent, so any change was likely driven by a reduction in driving by those commuting to work and/or driving their children to school. We assume that this reduction in traffic likely also subsequently led to reduced concentrations of automobile-related pollutants like NO_2.

3.2. Study design overview

The study uses an approach that compares the trend in NO_2 concentrations in 2020, i.e. the intervention year, with the trend in several years in which no mitigation measures for COVID-19 control were implemented, i.e. the control years. The use of historical controls is advantageous in this study, because, given that virtually all European cities implemented mitigation measures in March of 2020, no appropriate geographical control was available. The study period for the intervention year includes Monday, January 6, 2020 (2nd calendar week) – Sunday, April 12, 2020 (15th calendar week). March 16, 2020, the date on which the first major COVID-19 measure was implemented divides this period into pre- and post-intervention periods. The study period for the control years includes this same time period (Monday of the 2nd calendar week – Sunday of the 15th calendar week) in 2014–2019, with the Monday of the 12th calendar week splitting each year into pre- and post-intervention periods.

Both the c-ITS and SC approaches allow for the comparison of serial changes to an intervention unit receiving the intervention with changes to one or multiple control units not receiving the intervention (CRAIG et al., 2017). Thus each approach utilizes serial data from intervention and control units to create a 'counterfactual', i.e. what would have happened had the intervention not been implemented. This allowed us to ensure that any effect observed in 2020 is neither due to the current trend in NO_2 concentrations nor due to yearly seasonal fluctuations.

The main difference between the two approaches, however, relates to how data from control units are utilized. The c-ITS study utilizes data from all control units in full. Specifically, we compared the change in NO_2 concentrations between the pre- and post-intervention periods in 2020, the intervention year, to changes in concentrations between the pre- and post-intervention periods in 2014–2019, the control years (LOPEZ BERNAL et al., 2018). The SC study can be utilized when there are multiple controls to draw from, but no clear rationale for choosing which is the most appropriate. Specifically, we compared the change in NO_2 concentrations between the pre- and post-intervention periods in 2020 to changes between the pre- and post-intervention periods in a weighted average of 2014–2019. This data-driven weighted average is calculated to provide the most similar comparison, with respect to the pre-intervention outcome trend and a pre-defined set of covariates (Bouttell et al., 2018).

3.3. Data

3.3.1. Outcome

The Bavarian Environmental Administration (Bayerisches Landesamt für Umwelt) is charged with the monitoring of air quality in Bavaria, and data for the 50 monitoring stations are freely available (LfU BAYERN, 2020). We obtained NO_2 data for the five stations located in Munich, which included two classified as urban traffic monitors – Landshuter Allee (LAN) and Stachus (STA), one as urban background – Lothstrasse (LOT), and two as background – Allach (ALL) and Johanneskirchen (JOH). Hourly data were provided, which we converted to daily averages.

3.3.2. Covariates

We obtained data for other factors that are associated with NO_2 concentrations, including several weather-related variables – daily averages of temperature, rain fall, air pressure, humidity, and wind speed (Peel, 2010). These data were freely available from the German Weather Service (Deutscher Wetterdienst) (DWD, 2020). We also used publicly available information indicating when school holidays were in place – these included the Christmas, winter and Easter holidays. Within these time periods, relevant days were defined as either holiday high travel days (i.e. specific holidays or holiday weekends – Friday and Saturday, on which people tend to travel more) or holiday low travel days (i.e. during the week when people tend to travel less – Sunday through Thursday).

3.4. Statistical analyses

We registered a study protocol on May 3, 2020 through OSF (https://osf.io/7vkfc); all hypotheses and methods for main and additional analyses were defined a priori in the protocol. We designed and piloted these analyses using data from 2014 to 2019. The data for the intervention year, 2020, were downloaded and analyzed only after registration of the protocol.

3.4.1. Main analysis

As part of the main analyses we applied a c-ITS and SC approach. For both of these approaches, it is important to define the impact model, i.e. how the intervention would impact the outcome if it were effective – this subsequently shapes decisions made in defining the analysis parameters (Lopez BERNAL et al., 2016). With regard to the timing of the effect, we assumed that the COVID-19 mitigation measures began influencing NO_2 concentrations immediately after implementation of the first of the measures on March 16, 2020, thus we defined this day as the first day of the post-intervention time period. Given that the mitigation measures could have led to an immediate drop in NO_2 concentrations and that we are interested in the effect of the measures over the entire post-intervention period, we assumed and tested for a level change. This level change represents an immediate change, which is sustained across the post-intervention period.

As described above in section 3.3, we obtained data from five air quality monitoring stations. Our a priori hypothesis was that the observed effect would be greatest at the two traffic monitors LAN and STA, a smaller effect at the urban background monitor LOT, and the smallest effect, if any, at the two background monitors ALL and JOH.

For the c-ITS approach, we fitted a linear model using the general least squares method. The model took the following form:

$$NO_2 = \beta_0 + \beta_1 Day + \beta_2 Year + \beta_3 Post + \beta_4 Int + \beta_5 Post{*}Int + \beta_{6-13} Covs$$

where, NO_2 represents the outcome, NO_2 concentrations in μg/m^3 at a given monitor; *Day* is a continuous variable from 1 to 98, from the first to the last day of the study period in each year, thus capturing the underlying trend in the outcome over the study period; *Year* is a categorical variable taking the value of the year between 2014 and 2020; *Post* is a dummy variable taking the value of 0 in the pre-intervention period and 1 in the post-intervention period in each year (March 16, 2020, or Monday of the 12th calendar week in all years was treated as the first day of the post-intervention period), thus capturing the change in the outcome in the post-intervention period relative to the pre-intervention period; *Int* is a dummy variable taking the value for the control years 2014–2019 and 1 in the intervention year 2020; *Post*Int* is an interaction term which captures the change in *Post* in 2020 compared to in 2014–2019; *Covs* includes the potentially important covariates, including temperature, rain fall, air pressure, humidity, wind speed, holiday high and low travel days and day of the week. β_5, the change in NO_2 concentrations between the pre- and post-intervention periods in 2020 relative to the change in 2014–2019, represents the level change described above in section 3.3, and is thus the effect estimate of interest. Given the serially correlated nature of the data, we used auto-correlation and partial auto-correlation plots to determine an appropriate correlation structure for each model. For the site ALL, substantial data were missing for the year 2014 (23%); because of this, 2014 was excluded from the c-ITS analysis for ALL only.

The SC approach was structured similarly. However, instead of comparing changes in 2020 to changes in 2014–2019, the method allows for the construction of a synthetic control, ensuring that the intervention year and synthetic control year were similar with regard to the pre-intervention outcome trend and potentially important covariates. Specifically, this synthetic control was constructed using input data from the pre-intervention NO_2 concentrations, as well as the covariates listed above, from 2014 to 2019. Based on a linear interactive fixed effects model, we calculated the effect of interest. This effect is the average difference between the observed time series, i.e. the post-intervention outcome trend observed in 2020, and the synthetic control time series, i.e. the post-intervention counterfactual series. This approach allows for a treatment effect to be calculated for each post-intervention time point, allowing an investigation of how the intervention effect changes over time, as well as for the entire post-intervention time period, allowing an investigation of the average effect of the intervention, or the average treatment effect (ATT).

3.4.2. Additional analyses specified a priori

We conducted a series of sensitivity analyses to evaluate the extent to which our results were robust to changes to our assumptions, and to further explore how the intervention effect developed and changed over time.

The mitigation measures were dependent on individuals changing their behavior, and this behavior may have been adapted over time. We suspected that the effect in the two weeks immediately following the intervention may have been larger than the effect in the subsequent two weeks. We investigated this using both the c-ITS and SC approaches. For the c-ITS approach, we modelled two intervention effects separately, one specifically for the first two-week period, and the other for the second two-week period. For the SC approach, we shortened the post-intervention time period to two weeks.

It is also plausible that individuals did not immediately change their behavior on March 16, 2020, but instead slowly adapted as further mitigation measures were announced. To investigate this possibility, we mimicked the main analyses, treating March 19, 2020 as the first day of the post-intervention period, under the assumption that behaviors changed measurably after a lag of three days.

3.4.3. Post hoc analyses

After conducting the a priori specified main and additional analyses, we further conducted three sets of analyses to ensure that observed changes were not due to factors other than the mitigation measures. To ensure that concentration changes occurring prior to the intervention were not driving observed changes, we conducted all analyses with a series of backdated intervention start points 2, 4 and 6 weeks prior to March 16, 2020. Each of these 'placebo analyses' assessed whether changes occurred within two weeks of the respective intervention point, although no intervention actually occurred. Next, to assess whether high concentrations in January 2020 may have biased the pre-intervention trend and thus the calculated effects, we conducted all analyses with a shortened pre-intervention period lasting 6 weeks. Finally, to ensure that the noisy nature of daily air quality data, characterized by serial correlation as well as random noise, was not driving observed concentration changes, we repeated all analyses with smoothed NO_2 data. To do so, we analyzed only the trend component of the decomposed data.

All data processing and analyses were conducted using R version 3.6.3. The c-ITS approach was conducted using the Fit Linear Model Using Generalized Least Squares (nlme) (WEISBERG and FOX, 2015) and the SC approach was conducted using the Generalized Synthetic Control Method (gsynth) package (Xu, 2017).

4. Results

4.1. NO_2 concentrations in munich

Concentrations of NO_2 improved in Munich over the period 2014–2020, as illustrated by Fig. 1. For all five monitoring sites, concentrations in 2020 were lower than during any of the previous years, with the greatest differences observed for the year 2014. Taking March as an example, concentrations in 2020 were lower than in 2014 by 47% at LAN, 55% at STA, 43% at Loth, 45% at ALL and 51% at JOH. It is also evident that concentrations at traffic sites (LAN and STA) were, as expected, higher than at urban background (LOT) and background sites (ALL and JOH).

4.2. Effect of COVID-19 mitigation measures on NO_2 concentrations

Regarding the effect of the COVID-19 mitigation measures on NO_2 concentrations across the post-intervention period, our main analyses are summarized in Fig. 2 (panel A) and Table 1.

At traffic sites, where we hypothesized the largest reduction in NO_2 concentrations, reductions of 9.34 μg/m^3 (95% confidence interval: −23.58; 4.90) and 10.02 μg/m^3 (−19.25; −0.79) were observed at LAN and STA, respectively, using the c-ITS approach, and 15.65 μg/m^3 (−27.58; −4.09) and 15.1 μg/m^3 (−24.82; −9.83) using the SC approach. At LOT, the urban background site, where we hypothesized a smaller reduction, small decreases of 1.94 μg/m^3 (−11.90; 8.03) and 8.84 μg/m^3 (−20.04; −2.51) were observed using the c-ITS and SC approach, respectively. At background sites, where we hypothesized a small effect if any, a small reduction of 1.37 μg/m^3 (−12.77; 10.02) was observed at ALL, while a slight increase of 0.75 μg/m^3 (−8.79; 10.29) was observed at JOH using the c-ITS approach; using the SC approach slight decreases were observed at both sites, −3.08 μg/m^3 (−12.59; 5.39) at ALL and −4.69 μg/m^3 (−11.65; 1.86) at JOH. Confidence intervals for all estimates should be noted; for the c-ITS approach, a significant effect was observed only at STA, while for the SC approach significant effects were observed at LAN, STA and LOT. For all other estimates, confidence intervals included 0, indicating some uncertainty regarding the direction of these effects.

Fig. 3 illustrates, based on the SC approach, how the effect of the mitigation measures changed over time. Across all sites, a reduction in NO_2 concentrations shortly after the implementation of the measures can be seen. It is also evident across sites that concentrations began

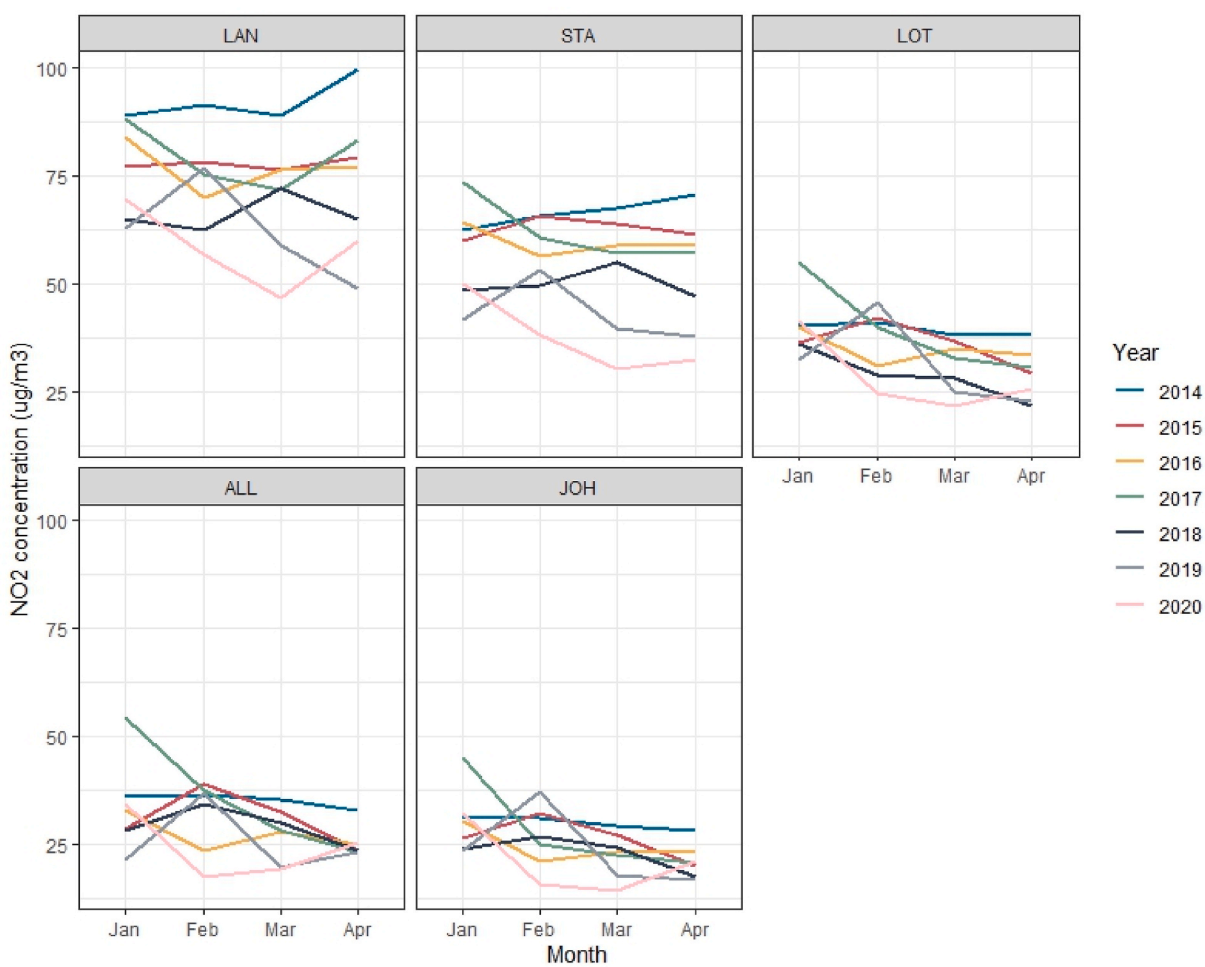


Fig. 1. NO_2 concentrations from January–April in 2014–2020 at LAN, STA, LOT, ALL and JOH.

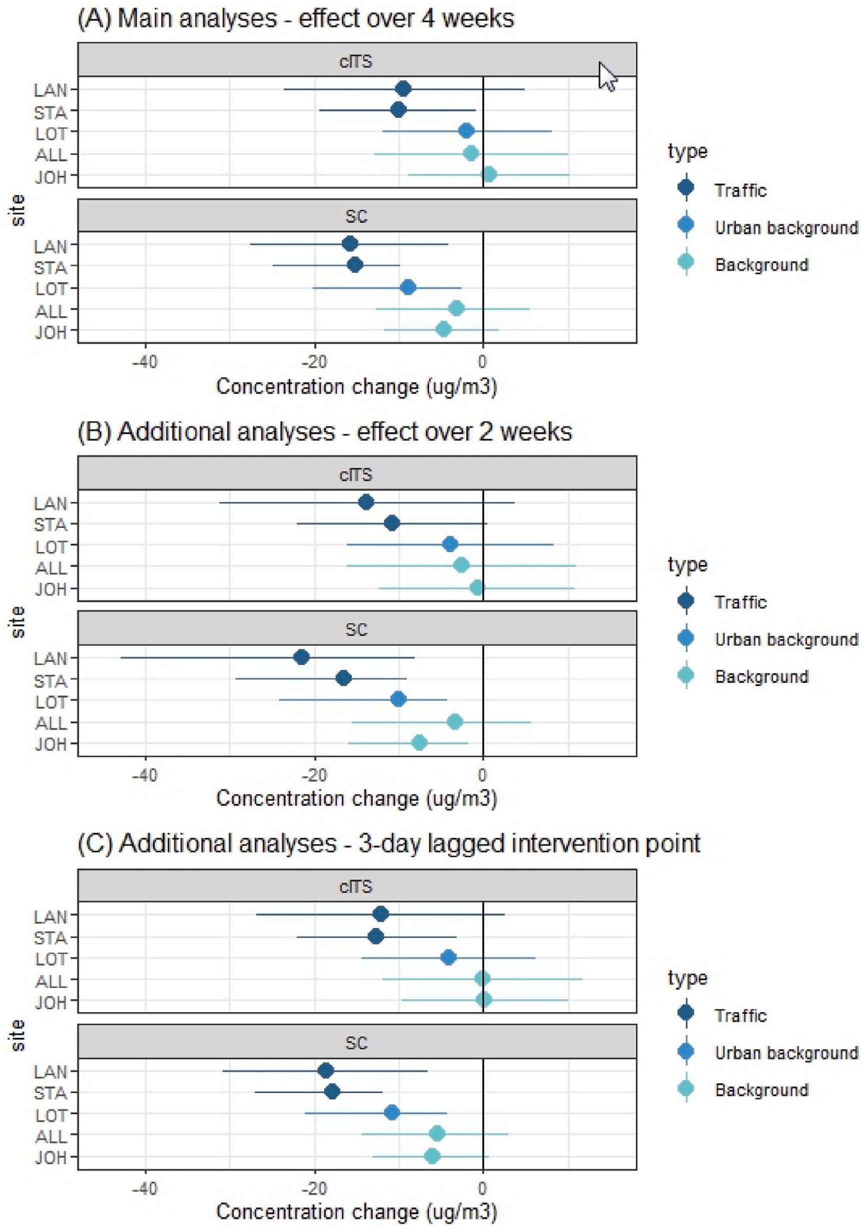


Fig. 2. Effect of the COVID-19 mitigation measures on NO_2 concentrations at the five sites from (A) main analyses, and additional analyses of (B) a two-week post-intervention period and (C) a 3-day lagged intervention point.

creeping upward again after approximately 1.5–2 weeks. Additional analyses below explore how the intervention effect developed and changed over time.

Fig. 3 shows that the SC approach was not able to calculate an optimal counterfactual – for an optimal counterfactual, the pre-intervention ATT would lie very close to 0 at all points along the time series. Additionally, one can see that the average NO_2 concentration approximately 4 weeks prior to the intervention appears to lie below the 0 ATT line, meaning that the observed effects may in part be attributable to changes occurring before the intervention. Post hoc analyses, described below, explore whether these aspects may have biased observed effects.

Additional analyses specified a priori.

Regarding the timing of the effect, we further investigated whether the effect in the first two-week post-intervention period was larger in magnitude than the effect over the entire four weeks. These results are summarized in Fig. 2 (panel B) and Table 1. As hypothesized, across sites effects were slightly larger when considering a two-week post-intervention period rather than a four-week period. Regarding the second two-week post-intervention period, which we assessed using the c-ITS approach, observed effects were smaller at all sites than in the first two-week period. Confidence intervals for all estimates should be noted; for the c-ITS approach, a significant effect was observed only at STA, while for the SC approach significant effects were observed at LAN, STA and LOT. For all other estimates, confidence intervals contained 0, indicating some uncertainty regarding the direction of these effects.

Additionally, we investigated whether the effect differed if the intervention start was delayed for three days from 16 March to March 19, 2020. These results are summarized in Fig. 2 (panel C) and Table 1. As hypothesized, a lagged intervention start resulted in a slightly larger effect at traffic sites. At urban background and background sites similar to slightly larger effects were observed. Confidence intervals for all estimates should be noted; for the c-ITS approach, a significant effect was observed only at STA, while for the SC approach significant effects were

Table 1
Summary of results from main and additional analyses.

	4-week post-intervention (main analyses)		2-week post-intervention period, period 1		2-week post-intervention period, period 2		3-day lagged intervention start	
Site	**Effect**[a] (μg/m^3)	**95% CI**	**Effect** (μg/m^3)	**95% CI**	**Effect** (μg/m^3)	**95% CI**	**Effect** (μg/m^3)	**95% CI**
cITS approach								
LAN (T)	−9.34	−23.58; 4.90	−13.73	−31.24; 3.78	−4.72	−22.95; 13.51	−12.08	−26.73; 2.57
STA (T)	**−10.02**	**−19.25; -0.79**	−10.78	−22.04; 0.48	−9.17	−21.07; 2.72	**−12.61**	**−22.00; -3.21**
LOT (UB)	−1.94	−11.90; 8.03	−3.82	−16.01; 8.37	0.27	−12.59; 13.13	−4.08	−14.32; 6.17
ALL (B)	−1.37	−12.77; 10.02	−2.57	−16.09; 10.95	0.17	−14.44; 14.78	−0.06	−11.84; 11.73
JOH (B)	0.75	−8.79; 10.29	−0.73	−12.21; 10.74	2.52	−9.74; 14.78	0.15	−9.69; 9.99
SC approach								
LAN (T)	**−15.65**	**−27.58; -4.09**	**−21.46**	**−42.82; -8.19**	-[b]	–	**−18.49**	**−30.73; - 6.55**
STA (T)	**−15.1**	**−24.82; -9.83**	**−16.52**	**−29.30; -9.01**	–	–	**−17.82**	**−26.91; -11.92**
LOT (UB)	**−8.84**	**−20.04; -2.51**	**−10.04**	**−24.09; -4.32**	–	–	**−10.82**	**−21.06; -4.37**
ALL (B)	−3.08	−12.59; 5.39	−3.35	−15.53; 5.71	–	–	−5.47	−14.40; 2.93
JOH (B)	−4.69	−11.65; 1.86	−7.46	−15.92; −1.71		–	−6.02	−12.98; 0.67

Bold: denotes statistical significance at an alpha level of 5%.
AbbreviationscITS: controlled ITS; SC: synthetic control; (T): traffic site; (UB): urban background site; (B): background site.
[a] Effects are expressed as the effect over the post-intervention time period, e.g. −9.34 corresponds to a reduction in NO_2 concentration of 9.34 μg/m^3 between the pre- and post-intervention periods in 2020 relative to the control year(s).
[b] The SC approach did not allow for testing the second 2-week post-intervention period, thus no results are reported.

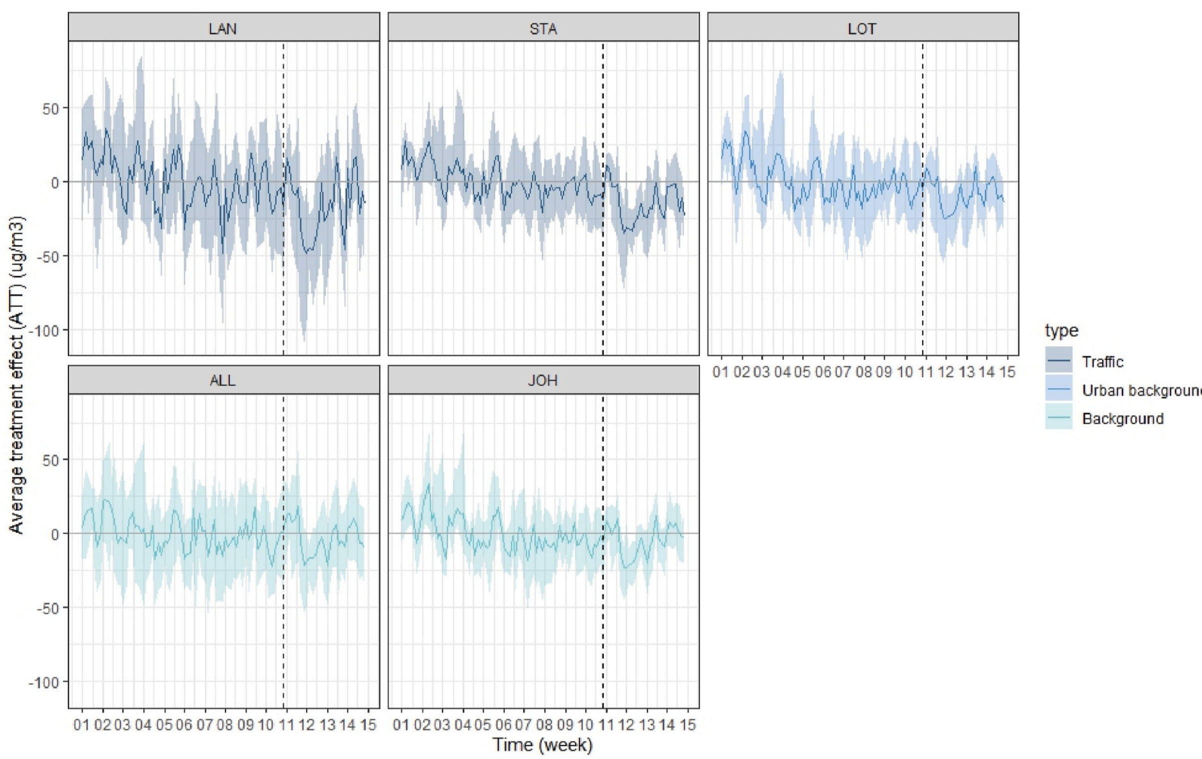


Fig. 3. Difference between the observed NO_2 concentrations in 2020 and those from the SC counterfactual (based on the years 2014–2019) at all investigated sites. The vertical dotted line represents the point at which the intervention was implemented.

observed at LAN, STA and LOT. For all other estimates, confidence intervals contained 0.

Analyses of a backdated intervention points at 3 February, 17 February and March 2, 2020 are summarized in Fig. 4 panels A–C, respectively, and Appendix Table 1 in the supplementary material. Compared to the main analyses, effects at traffic sites are smaller when either 3 February or 2 March is taken as the intervention point. However, for 17 February, observed effects are actually larger than those observed in the main analyses. At urban background and background sites, where we expected a small or no effect due to the COVID-19 measures, larger effects were observed for almost all backdated analyses compared to main analyses. Taken together, this suggests that the effect observed in main analyses may be at least partially attributable to changes in air quality across Munich (i.e. not only in heavily trafficked areas) already occurring prior to March 16, 2020.

Analyses of a shortened pre-intervention period allowed us to assess whether the high concentrations observed in January 2020 influenced the observed effect; these results are summarized in Fig. 5 (panel A) and Appendix Table 2 in the Supplementary material. Smaller effects at traffic sites were observed for the shortened pre-intervention period than for main analyses, potentially suggesting that observed effects are at least partially attributable to high concentrations observed in January 2020. Analyses of smoothed NO_2 data are summarized in Fig. 5 (panel B) and Appendix Table 2 in the Supplementary material. The smoothed data allowed for the calculation of a better counterfactual than the raw data (Appendix Fig. 1). Compared to results from the main analyses, a

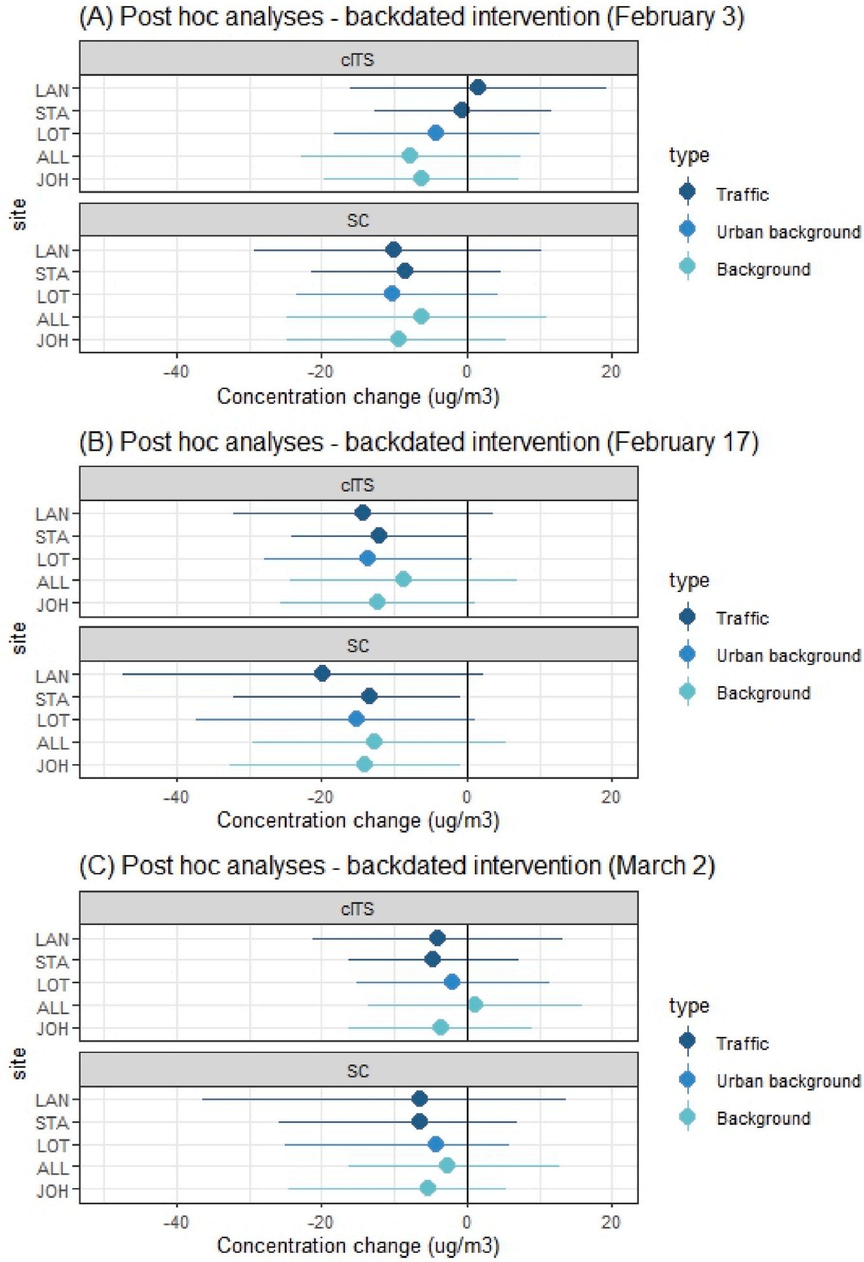


Fig. 4. Effect of the COVID-19 mitigation measures on NO_2 concentrations at the five sites from post hoc analyses assessing backdated intervention points, including (A) February 3, 2020 (B) February 17, 2020 and (C) March 2, 2020.

slightly smaller effect across sites was observed. This suggests that some of the effect observed in main analysis may be attributable to random noise or serial correlation, although at the same time, it is possible that the smoothing of the data smoothed away part of an actual effect.

5. Discussion

In this study, we applied a c-ITS and SC approach to evaluate the short-term effect of COVID-19 mitigation measures on NO_2 in Munich. Main and additional analyses suggest a consistent pattern – after introduction of the mitigation measures decreases in NO_2 concentrations were observed at traffic sites, while little to no change was observed at urban background and background sites. As expected, reductions were largest in magnitude in the two weeks immediately following the introduction; a lagged intervention start suggests that the effect became more pronounced as additional measures were implemented. Post-hoc analyses, however, point to other aspects to which effects may have been partially attributable; these include reductions in NO2 concentrations occurring prior to 16 March, as well as high concentrations observed in January.

Events such as the COVID-19 pandemic with the resulting mitigation measures are natural experiments that provide a unique opportunity to assess how specific policies may influence air quality. This study, for example, provides information on whether policies reducing traffic at heavily-trafficked sites could lead to improved air quality. Reductions in NO_2 of 9.34 μg/m^3 and 15.65 μg/m^3 at LAN and 10.02 μg/m^3 and 15.10 μg/m^3 at STA, corresponding to the c-ITS and SC approach from the main analyses, represent meaningful changes given the current air quality in Munich. Taking, for example, the 2020 pre-intervention average concentrations at LAN and STA of 60.94 μg/m^3 and 41.61 μg/m^3, respectively, these equate to reductions of approximately 15–25% and 24–36%. In Munich and other German cities, where debates around

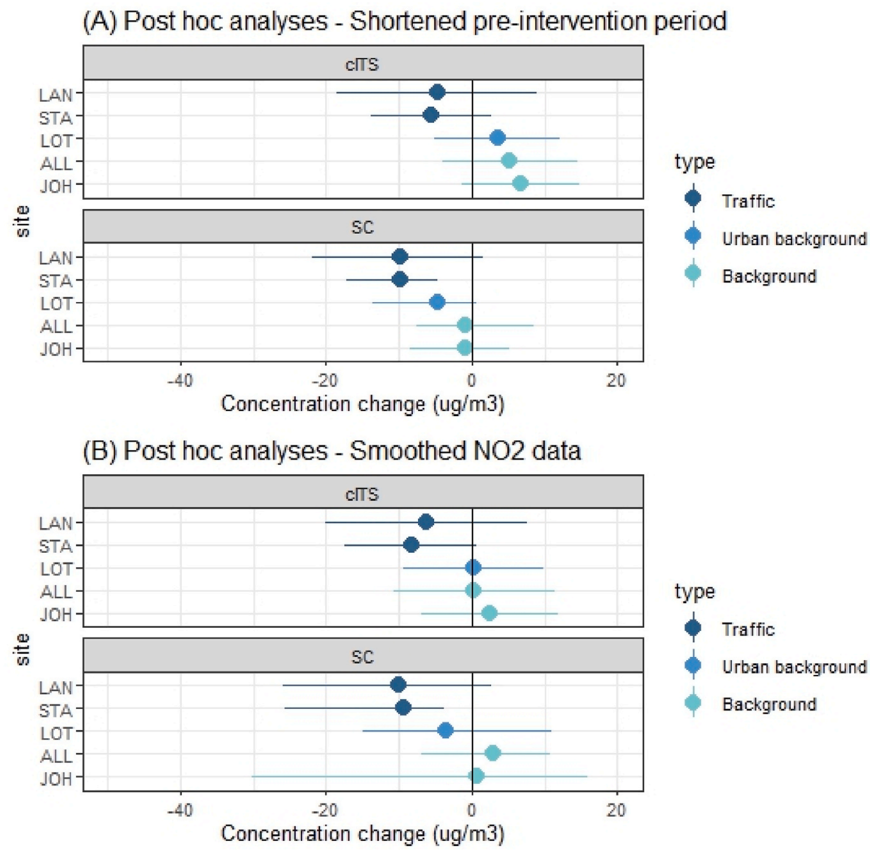


Fig. 5. Effect of the COVID-19 mitigation measures on NO2 concentrations at the five sites from post hoc analyses assessing (A) a shortened pre-intervention period (6 weeks), and (B) analyses of smoothed NO_2 data.

air quality in cities and, in particular, how to further reduce NO2 concentrations are common both in the scientific and political communities, this is an important finding (Bayerische STAATSREGIERUNG, 2019; Leopoldina, 2019).

Other studies have shown decreases in air pollution linked to COVID-19 measures. Data from satellites have suggested reductions in NO_2 concentrations, ranging from a similar to slightly larger magnitude, in China (30–40%) (ESA, 2020c, MUHAMMAD et al., 2020), India (40–53%) (ESA, 2020a, MAHATO et al., 2020) and in cities across Europe (20–55%) (ESA, 2020b, MUHAMMAD et al., 2020). Data from regulatory monitors across Europe have been somewhat less consistent, although they have also shown decreases ranging from 15% to 50% in cities in Western Europe (EEA 2020). Studies applying similar methods to ours, i.e. quasi-experimental approaches using regulatory monitors and historical controls, also identified reductions in Beijing and Wuhan, China as well as Milan, Italy (Malpede et al., 2020), Rio de Janeiro, Brazil (Dantas et al., 2020), and in Munich (FULLERTON, 2020) ranging between approximately 5 μg/m^3 and 50 μg/m^3. One study, comparing areas of China where lockdowns were implemented to areas where no lockdown was implemented, observed decrease in fine particulate matter of approximately 15% (He et al., 2020). Another study estimated what changes in air quality across Europe could mean for public health, calculating that 11,000 deaths, including approximately 2000 deaths in Germany, may have been avoided due to decreases in air pollution during this time (Myllyvirta and Thieriot, 2020).

Our recent systematic review of ambient air pollution interventions, as well as multiple other reviews have emphasized important limitations of existing studies, including lack of control for underlying outcome trends and lack of control for confounding through appropriate selection of control conditions and assessment of confounding factors (Boogaard et al., 2017, BURNS et al., 2020; HENNEMAN et al., 2017; RICH, 2017). The use of two approaches, each of which represents an internally valid quasi-experimental approach, strengthens the rigour of our study. Both the c-ITS and the SC approaches are appropriate study designs for evaluating changes over time; they utilize the temporal nature of the data to establish a counterfactual (LOPEZ BERNAL et al., 2018; BOUTTELL et al., 2018). Each approach also utilizes data from a control condition, in this study historical controls, to ensure that any observed change in the outcome trend is not due to seasonal patterns. The use of historical controls can add a level of control to studies where no appropriate geographical controls exist; in this study, for example, all urban (as well as rural) areas in Germany and Europe implemented COVID-19 mitigation measures roughly at the same time. Specifically, the c-ITS approach allows comparison of trends in 2020 to the average of trends over the time period of 2014–2019 so that the comparison will not be heavily skewed by any one year that does not fit the true long-term trend. The SC study complements this approach by creating a control condition from 2014 to 2019 that most closely matches the intervention time trend. We further accounted for potentially important confounders in both approaches: the c-ITS model was adjusted for temperature, rainfall, air pressure, humidity, wind speed, day of the week and holidays; the SC approach used these factors in creating an appropriately weighted synthetic control. We defined most hypotheses and analyses a priori and registered a study protocol, before downloading the data for 2020. Only the analyses of a backdated intervention point, a shortened pre-intervention period and smoothed NO_2 data were defined post hoc; these were added to ensure that observed effects were not attributable to other factors.

Nevertheless, there are limitations to this study. We assume that the COVID-19 mitigation measures led to reductions in traffic, which subsequently led to reductions in NO_2 concentrations. Lacking reliable data on traffic, however, we cannot assess to what extent this assumption of effects along the causal chain are appropriate. Mobility data from smartphones made available by GOOGLE, 2020 and APPLE, 2020 suggest that mobility was starkly reduced during these weeks; however the current study would have benefited from the incorporation of long-term, representative routine traffic data. Post hoc analyses suggest that effects observed in main analyses may at least partly stem from factors other than the mitigation measures, including reductions in NO2 concentrations occurring prior to 16 March, high concentrations observed in January and the noisy nature of the data. However, the large decrease immediately after March 16, 2020 is observable across all main and additional analyses, meaning it is unlikely that observed effects are due only to factors other than the mitigation measures. This large decrease is consistent with the reduction in traffic reported in the mobility data described above. While the monitoring sites assessed represent all regulatory sites available for Munich during the study period, it is possible that these are not fully representative of air quality across Munich. Additionally, for the ALL site, the year 2014 was excluded from the c-ITS approach because much of the data from that year were missing. However, we consider it unlikely that this substantially influenced our results. We assessed changes only in NO_2 concentrations, as this allowed us to most closely assess whether changes to air quality were likely due to changes in traffic reductions. Nevertheless, a more comprehensive assessment of the impact of a specific intervention, measure or event would entail the assessment of multiple pollutants. To the best of our knowledge, this is the first use of historical controls within a SC study, and we feel that this is an appropriate use of the available data. Nevertheless, our study highlights challenges associated with calculating an optimal counterfactual using a SC study in the context of air quality data. However, given that the c-ITS approach, which can better account for time-varying confounders, and the analyses of smoothed data yielded similar results, if slightly smaller in magnitude, we think it unlikely that our results are biased by this limitation.

Given that traffic is only one source of NO_2 and other air pollutants, continuing to improve air quality will likely require multiple control measures targeting multiple sources. However, this study suggests that reducing traffic may be an effective measure to reduce NO_2 concentrations in heavily trafficked areas by margins which could translate to public health benefits.

CRediT authorship contribution statement

Jacob Burns: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing - original draft. **Sabine Hoffmann:** Methodology, Validation, Writing - review & editing. **Christoph Kurz:** Methodology, Validation, Visualization, Writing - review & editing. **Michael Laxy:** Methodology, Validation, Writing - review & editing. **Stephanie Polus:** Methodology, Validation, Writing - review & editing. **Eva Rehfuess:** Methodology, Writing - review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing for financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors also submitted the same study in the Indian patent office for consideration.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2020.118089.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

DWD, 2020. Open data. Deutscher Wetterdienst.
Leopoldina, 2019. Saubere Luft. Stickstoffoxide und Feinstaub in der Atemluft: Grundlagen und Empfehlungen. Halle (Saale): Nationale Akademie der Wissenschaften Leopoldina.
APPLE, 2020. Mobility Trends Reports. Apple.
Bayern, L., 2020. Lufthygienisches Landesüberwachungssystem Bayern. Bayerisches Landesamt für Umwelt.
Bernal, J.L., Cummins, S., Gasparrini, A., 2016. Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int. J. Epidemiol. 46, 348–355.
Boogaard, H., Van Erp, A.M., Walker, K.D., Shaikh, R., 2017. Accountability studies on air pollution and health: the HEI experience. Current Environmental Health Reports 4, 514–522.
Bouttell, J., Craig, P., Lewsey, J., Robinson, M., Popham, F., 2018. Synthetic control methodology as a tool for evaluating population-level health interventions. J. Epidemiol. Community 72, 673.
Brimblecombe, P., Ning, Z., 2015. Effect of road blockages on local air pollution during the Hong Kong protests and its implications for air quality management. Sci. Total Environ. 536, 443–448.
Burns, J., Boogaard, H., Polus, S., Pfadenhauer, L., Rohwer, A., Van Erp, A., Turley, R., Rehfuess, E., 2019. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database Syst. Rev. 5, 1469, 493X (Electronic).
Burns, J., Boogaard, H., Polus, S., Pfadenhauer, L.M., Rohwer, A.C., Van Erp, A.M., Turley, R., Rehfuess, E.A., 2020. Interventions to reduce ambient air pollution and their effects on health: an abridged Cochrane systematic review. Environ. Int. 135, 105400.
Craig, P., Katikireddi, S.V., Leyland, A., Popham, F., 2017. Natural experiments: an overview of methods, approaches, and contributions to public health intervention research. Annu. Rev. Publ. Health 38, 39–56.
Dantas, G., Siciliano, B., França, B.B., Da Silva, C.M., Arbilla, G., 2020. The impact of COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil. Sci. Total Environ. 729, 139085.
Ding, L., Chan, T., Ke, F., Wang, D., 2014. Characterization of chemical composition and concentration of fine particulate matter during a transit strike in Ottawa, Canada. Atmos. Environ. 89, 433–442.
EEA, 2020. Air Quality and COVID-19. European Environment Agency.
ESA, 2020a. Air Pollution Drops in India Following Lockdown. European Space Agency.
ESA, 2020b. Air Pollution Remains Low as Europeans Stay at Home. European Space Agency.
ESA, 2020c. COVID-19: Nitrogen Dioxide over China. European Space Agency.
Fransen, M., Pérodin, J., Hada, J., He, X., Sapkota, A., 2013. Impact of vehicular strike on particulate matter air quality: results from a natural intervention study in Kathmandu valley. Environ. Res. 122, 52–57.
Fullerton, B., 2020. NO2 Konzentrationen Während Covid-19 Lock-Downs in München. HawaDawa.
GOOGLE, 2020. COVID-19 Community Mobility Report. Google, Munich.
He, G., Pan, Y., Tanaka, T., 2020. The short-term impacts of COVID-19 lockdown on urban air pollution in China. Nat. Sustain. https://doi.org/10.1038/s41893-020-0581-y.
Henneman, L.R.F., Liu, C., Mulholland, J.A., Russell, A.G., 2017. Evaluating the effectiveness of air quality regulations: a review of accountability studies and frameworks. J. Air Waste Manag. Assoc. 67, 144–172.
Hong, A., Schweitzer, L., Yang, W., Marr, L., 2015. Impact of temporary freeway closure on regional air quality: a lesson from carmageddon in los angeles, United States. Environ. Sci. Technol. 49, 3211–3218.
Khailaie, S., Mitra, T., Bandyopadhyay, A., Schips, M., Mascheroni, P., Vanella, P., Lange, B., Binder, S., Meyer-Hermann, M., 2020. Estimate of the development of the epidemic reproduction number Rt from Coronavirus SARS-CoV-2 case data and implications for political measures based on prognostics. medRxiv 36–37, 04.04.2005.
Latha, K., Badarinath, K., Moorthy, K., 2004. Impact of diesel vehicular emission on ambient black carbon concentration at an urban location in India. Curr. Sci. 86, 451–453.
Levy, J., Baxter, L., Clougherty, J., 2006. The air quality impacts of road closures associated with the 2004 Democratic National Convention in Boston. Environ. Health 5, 16.
Lopez Bernal, J., Cummins, S., Gasparrini, A., 2018. The use of controls in interrupted time series studies of public health interventions. Int. J. Epidemiol. 47, 2082–2093.
Mahato, S., Pal, S., Ghosh, K.G., 2020. Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. Sci. Total Environ. 730, 139086.
Malpede, M., Percoco, M., Bosetti, B., 2020. Social Distancing Measures Following COVID-19 Epidemics Had Positive Environmental Consequences. GREEN Working Paper Series. Centre for Research on Geography, Resources, Environment, Energy and Networks.
Monitor, H.S.R., 2020. COVID-19 health system response monitor [online]. Available: https://www.covid19healthsystem.org/mainpage.aspx. (Accessed 20 March 2020).
Muhammad, S., Long, X., Salman, M., 2020. COVID-19 pandemic and environmental pollution: a blessing in disguise? Sci. Total Environ. 728, 138820.
Myllyvirta, L., Thieriot, H., 2020. 11,000 Air Pollution-Related Deaths Avoided in Europe as Coal, Oil Consumption Plummet. Centre for Research on Energy and Clean Air.
Pope, C.R., Rodermund, D., Gee, M., 2007. Mortality effects of a copper smelter strike and reduced ambient sulfate particulate matter air pollution. Environ. Health Perspect. 115, 679–683.

Rich, D.Q., 2017. Accountability studies of air pollution and health effects: lessons learned and recommendations for future natural experiment opportunities. Environ. Int. 100, 62–78.
Staatsregierung, B., 2019. Bericht aus der Kabinettssitzung.
Staatsregierung, B., 2020. Bericht aus der Kabinettssitzung.
Weisberg, S., Fox, J., 2015. An R Companion to Applied Regression, Appendix: Mixed Effects Models in R. SAGE.
Who, 2020a. Novel Coronavirus (2019-nCoV) Situation Report - 1. World Health Organization.
Who, 2020b. Novel Coronavirus (2019-nCoV) Situation Report - 51. World Health Organization.
Xu, Y., 2017. Generalized synthetic control method: causal inference with interactive fixed effects models. Polit. Anal. 25, 57–76.

Appendix 1

	Intervention point: March 16 (main analyses)		Intervention point: March 2 (post-hoc analyses)		Intervention point: February 17 (post-hoc analyses)		Intervention point: February 3 (post-hoc analyses)	
Site	**Effect**** (µg/m^3)	**95% CI**	**Effect** (µg/m^3)	**95% CI**	**Effect** (µg/m^3)	**95% CI**	**Effect** (µg/m^3)	**95% CI**
cITS approach								
LAN	-9.34	-23.58; 4.90	-3.95	-21.22; 13.33	-14.30	-32.26; 3.66	1.60	-16.11; 19.30
STA	**-10.02**	**-19.25; -0.79**	-4.58	-16.29; 7.12	-12.07	-24.20; 0.06	-0.56	-12.75; 11.63
LOT	-1.94	-11.90; 8.03	-1.94	-15.24; 11.37	-13.63	-27.88; 0.63	-4.11	-18.28; 10.06
ALL	-1.37	-12.77; 10.02	1.23	-13.58; 16.03	-8.66	-24.27; 6.96	-7.73	-22.77; 7.32
JOH	0.75	-8.79; 10.29	-3.60	-16.24; 9.04	-12.19	-25.60; 1.22	-6.17	-19.57; 7.24
SC approach								
LAN	**-15.65**	**-27.58; -4.09**	-6.55	-36.32; 13.79	-19.95	-47.37; 2.32	-9.92	-29.27; 10.25
STA	**-15.1**	**-24.82; -9.83**	-6.43	-25.86; 7.07	**-13.49**	**-32.20; -0.96**	-8.39	-21.49; 4.69
LOT	**-8.84**	**-20.04; -2.51**	-4.29	-24.92; 5.76	-15.14	-37.33; 1.21	-10.21	-23.56; 4.37
ALL	-3.08	-12.59; 5.39	-2.74	-16.35; 12.90	-12.79	-29.60; 5.34	-6.13	-24.76; 11.07
JOH	-4.69	-11.65; 1.86	-5.43	-24.67; 5.47	**-13.99**	**-32.59; -0.88**	-9.33	-24.84; 5.41

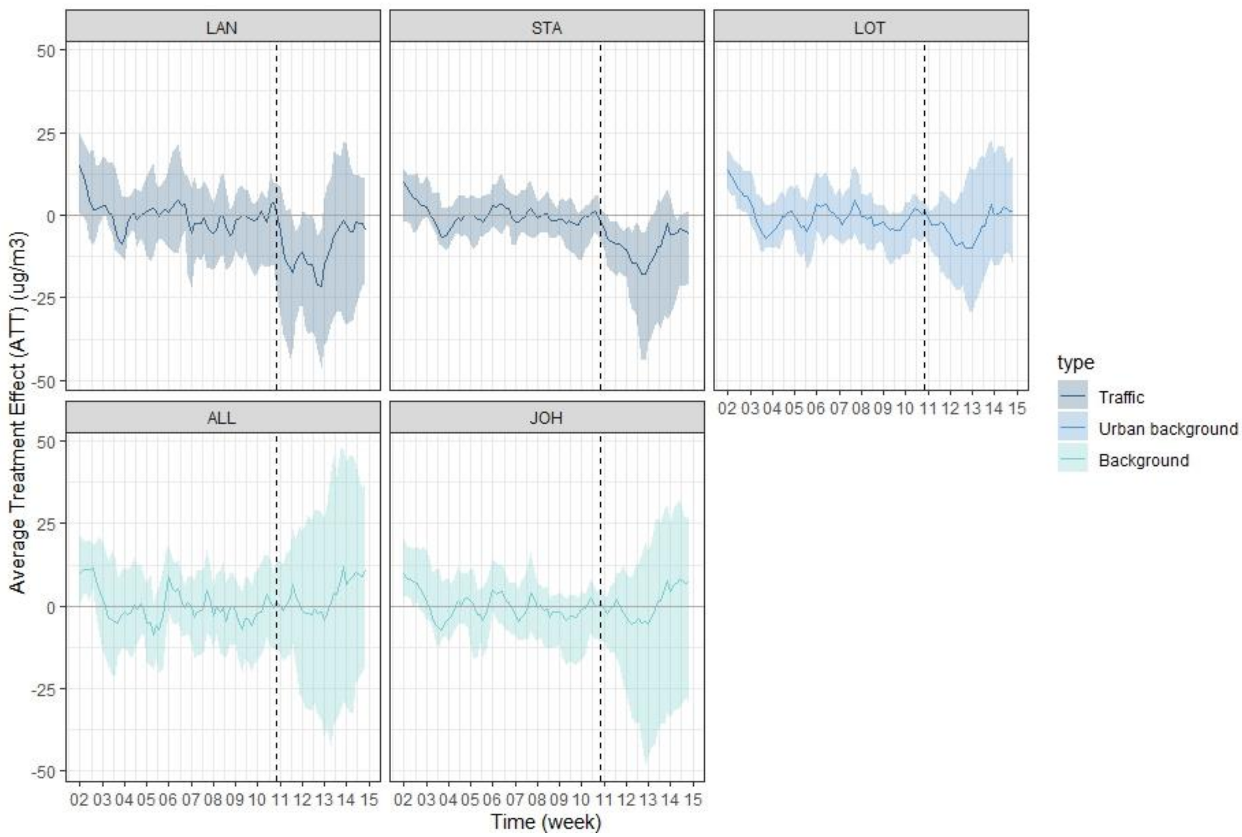
Appendix Table 1: Summary of results from main (as reported in the main study – here for reference) and post hoc analyses assessing backdated intervention points.
Bold: denotes statistical significance at an alpha level of 5%.
*Effects are expressed as the effect over the post-intervention time period, e.g. -9.34 corresponds to a reduction in NO_2 concentration of 9.34 µg/m^3 between the pre- and post-intervention periods in 2020 relative to the control year(s).

	4-week post-intervention (main analyses)		Smoothed data (post hoc analyses)		Shortened pre-intervention period (post hoc analyses)	
Site	**Effect*** (μg/m^3)	**95% CI**	**Effect** (μg/m^3)	**95% CI**	**Effect** (μg/m^3)	**95% CI**
cITS approach						
LAN	-9.34	-23.58; 4.90	-6.26	-20.19; 7.66	-4.75	-18.51; 9.00
STA	**-10.02**	**-19.25; -0.79**	-8.33	-17.33; 0.67	-5.53	-13.89; 2.82
LOT	-1.94	-11.90; 8.03	0.33	-9.31; 9.98	3.52	-5.17; 12.20
ALL	-1.37	-12.77; 10.02	0.33	-10.77; 11.43	5.29	-3.97; 14.54
JOH	0.75	-8.79; 10.29	2.51	-6.90; 11.91	6.74	-1.32; 14.79
SC approach						
LAN	**-15.65**	**-27.58; -4.09**	-10.02	-25.94; 2.82	-9.89	-21.78; 1.71
STA	**-15.1**	**-24.82; -9.83**	**-9.37**	**-25.59; -3.79**	**-9.83**	**-17.14; -4.69**
LOT	**-8.84**	**-20.04; -2.51**	-3.46	-15.02; 10.98	-4.76	-13.55; 0.64
ALL	-3.08	-12.59; 5.39	2.96	-6.79; 2.96	-0.08	-7.48; 8.47
JOH	-4.69	-11.65; 1.86	0.69	-30.07; 15.90	-0.93	-8.53; 5.13

Appendix Table 2: Summary of results from main (as reported in the main study – here for reference) and post hoc analyses assessing smoothed NO2 data and a shortened pre-intervention period.

Bold: denotes statistical significance at an alpha level of 5%.

*Effects are expressed as the effect over the post-intervention time period, e.g. -9.34 corresponds to a reduction in NO_2 concentration of 9.34 μg/m^3 between the pre- and post-intervention periods in 2020 relative to the control year(s).



Appendix Figure 1: Difference between the observed NO_2 concentrations in 2020 and those from the SC counterfactual (based on the years 2014-2019) at all investigated sites using smoothed NO_2 data.

8. References

1. Burns J, Boogaard H, Polus S, Pfadenhauer LM, Rohwer AC, van Erp AM, et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database of Systematic Reviews. 2019(5).

2. Burns J, Polus S, Brereton L, Chilcott J, Ward SE, Pfadenhauer LM, et al. Looking beyond the forest: Using harvest plots, gap analysis, and expert consultations to assess effectiveness, engage stakeholders, and inform policy. Research Synthesis Methods. 2018;9(1):132-40.

3. Burns J, Hoffmann S, Kurz C, Laxy M, Polus S, Rehfuess E. COVID-19 mitigation measures and nitrogen dioxide – A quasi-experimental study of air quality in Munich, Germany. Atmospheric Environment. 2021;246:118089.

4. Boylston A. Thomas Nettleton and the dawn of quantitative assessments of the effects of medical interventions 2010 [Available from: https://www.jameslindlibrary.org/articles/thomas-nettleton-and-the-dawn-of-quantitative-assessments-of-the-effects-of-medical-interventions/.

5. Cochrane. About Cochrane Reviews: Cochrane Library; 2021 [Available from: https://www.cochranelibrary.com/about/about-cochrane-reviews.

6. McKenzie J, Brennan S, Ryan R, Thomson H, Johnston R, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions2021.

7. Armitage P. Fisher, Bradford Hill, and randomization. International Journal of Epidemiology. 2003;32(6):925-8.

8. Fisher RA. The Arrangement of Field Experiments. In: Kotz S, Johnson NL, editors. Breakthroughs in Statistics: Methodology and Distribution. New York, NY: Springer New York; 1992. p. 82-91.

9. STREPTOMYCIN treatment of pulmonary tuberculosis. Br Med J. 1948;2(4582):769-82.

10. Shadish W, Cook T, Campbell D. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Belmont, CA, USA: Wadsworth Cengage Learning; 2002.

11. Bhargava A. Randomized controlled experiments in health and social sciences: Some conceptual issues. Economics & Human Biology. 2008;6(2):293-8.

12. Styles B, Torgerson C. Randomised controlled trials (RCTs) in education research – methodological debates, questions, challenges. Educational Research. 2018;60(3):255-64.

13. de Souza Leão L, Eyal G. The rise of randomized controlled trials (RCTs) in international development in historical perspective. Theory and Society. 2019;48(3):383-418.

14. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, et al. 2020 International Society of Hypertension Global Hypertension Practice Guidelines. Hypertension. 2020;75(6):1334-57.

15. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Hypertension. 2018;71(6):1269-324.

16. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension: The Task Force for the management of arterial hypertension of the European Society of Cardiology and the European Society of Hypertension. Journal of Hypertension. 2018;36(10).

17. Mackenzie M, O'Donnell C, Halliday E, Sridharan S, Platt S. Evaluating complex interventions: one size does not fit all. BMJ. 2010;340.

18. Petticrew M, Chalabi Z, Jones DR. To RCT or not to RCT: deciding when 'more evidence is needed' for public health policy and practice. Journal of Epidemiology and Community Health. 2012;66(5):391.

19. Victora CG, Habicht J-P, Bryce J. Evidence-based public health: moving beyond randomized trials. Am J Public Health. 2004;94(3):400-5.

20. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ (Clinical research ed). 2000;321(7262):694-6.

21. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655.

22. Hawe P, Shiell A, Riley T. Theorising Interventions as Events in Systems. American Journal of Community Psychology. 2009;43(3-4):267-76.

23. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, et al. The need for a complex systems model of evidence for public health. The Lancet. 2017;390(10112):2602-4.

24. Petticrew M, Knai C, Thomas J, Rehfuess EA, Noyes J, Gerhardus A, et al. Implications of a complexity perspective for systematic reviews and guideline development in health decision making. BMJ Global Health. 2019;4(Suppl 1):e000899.

25. Morfeld P, Groneberg DA, Spallek MF. Effectiveness of Low Emission Zones: Large Scale Analysis of Changes in Environmental NO2, NO and NOx Concentrations in 17 German Cities. PLOS ONE. 2014;9(8):e102999.

26. HEI. Assessing Health Impact of Air Quality Regulations: Concepts and Methods for Acountability Research. Health Effects Institute; 2003.

27. Bärnighausen T, Tugwell P, Røttingen J-A, Shemilt I, Rockers P, Geldsetzer P, et al. Quasi-experimental study designs series - paper 4: uses and value. Journal of Clinical Epidemiology. 2017;89:21-9.

28. Anderson LM, Petticrew M, Rehfuess E, Armstrong R, Ueffing E, Baker P, et al. Using logic models to capture complexity in systematic reviews. Research Synthesis Methods. 2011;2(1):33-42.

29. Higgins JPT, Ramsay C, Reeves BC, Deeks JJ, Shea B, Valentine JC, et al. Issues relating to study design and risk of bias when including non-randomized studies in systematic reviews on the effects of interventions. Research Synthesis Methods. 2013;4(1):12-25.

30. Noyes J, Booth A, Moore G, Flemming K, Tunçalp Ö, Shakibazadeh E. Synthesising quantitative and qualitative evidence to inform guidelines on complex interventions: clarifying the purposes, designs and outlining some methods. BMJ Global Health. 2019;4(Suppl 1):e000893.

31. Petticrew M, Anderson L, Elder R, Grimshaw J, Hopkins D, Hahn R, et al. Complex interventions and their implications for systematic reviews: a pragmatic approach. Journal of Clinical Epidemiology. 2013;66(11):1209-14.

32. Schünemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Research Synthesis Methods. 2013;4(1):49-62.

33. Shepperd S, Lewin S, Straus S, Clarke M, Eccles MP, Fitzpatrick R, et al. Can We Systematically Review Studies That Evaluate Complex Interventions? PLOS Medicine. 2009;6(8):e1000086.

34. Squires JE, Valentine JC, Grimshaw JM. Systematic reviews of complex interventions: framing the review question. Journal of Clinical Epidemiology. 2013;66(11):1215-22.

35. Becker BJ, Aloe AM, Duvendack M, Stanley TD, Valentine JC, Fretheim A, et al. Quasi-experimental study designs series—paper 10: synthesizing evidence for effects collected from quasi-experimental studies presents surmountable challenges. Journal of Clinical Epidemiology. 2017;89:84-91.

36. Higgins JPT, López-López JA, Becker BJ, Davies SR, Dawson S, Grimshaw JM, et al. Synthesising quantitative evidence in systematic reviews of complex health interventions. BMJ Global Health. 2019;4(Suppl 1):e000858.

37. Reeves B, Deeks J, Higgins J, Shea B, Tugwell P, Wells G. Chapter 24: Including non-randomized studies on intervention effects. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions: Cochrane; 2021.

38. Thomas J, Petticrew M, Noyes J, Chandler J, Rehfuess E, Tugwell P, et al. Chapter 17: Intervention complexity. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Intervention2021.

39. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annual Review of Public Health. 2017;38(1):39-56.

40. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. Journal of Epidemiology and Community Health. 2020;74(2):203.

41. Handley MA, Lyles CR, McCulloch C, Cattamanchi A. Selecting and Improving Quasi-Experimental Designs in Effectiveness and Implementation Research. Annual Review of Public Health. 2018;39(1):5-25.

42. Minary L, Trompette J, Kivits J, Cambon L, Tarquinio C, Alla F. Which design to evaluate complex interventions? Toward a methodological framework through a systematic review. BMC Medical Research Methodology. 2019;19(1):92.

43. Bärnighausen T, Oldenburg C, Tugwell P, Bommer C, Ebert C, Barreto M, et al. Quasi-experimental study designs series—paper 7: assessing the assumptions. Journal of Clinical Epidemiology. 2017;89:53-66.

44. Chen J, Hoek G. Long-term exposure to PM and all-cause and cause-specific mortality: A systematic review and meta-analysis. Environment International. 2020;143:105974.

45. Huangfu P, Atkinson R. Long-term exposure to NO2 and O3 and all-cause and respiratory mortality: A systematic review and meta-analysis. Environment International. 2020;144:105998.

46. Lee KK, Spath N, Miller MR, Mills NL, Shah ASV. Short-term exposure to carbon monoxide and myocardial infarction: A systematic review and meta-analysis. Environment International. 2020;143:105901.

47. Orellano P, Reynoso J, Quaranta N. Short-term exposure to sulphur dioxide (SO2) and all-cause and respiratory mortality: A systematic review and meta-analysis. Environment International. 2021;150:106434.

48. Orellano P, Reynoso J, Quaranta N, Bardach A, Ciapponi A. Short-term exposure to particulate matter (PM10 and PM2.5), nitrogen dioxide (NO2), and ozone (O3) and all-cause and cause-specific mortality: Systematic review and meta-analysis. Environment International. 2020;142:105876.

49. Zheng X-y, Orellano P, Lin H-l, Jiang M, Guan W-j. 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. Environment International. 2021;150:106435.

50. Gakidou E, Afshin A, Abajobir AA, Abate KH, Abbafati C, Abbas KM, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390(10100):1345-422.

51. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane; 2011.

52. Burns J, Boogaard H, Turley R, Pfadenhauer LM, van Erp AM, Rohwer AC, et al. Interventions to reduce ambient particulate matter air pollution and their effect on health. Cochrane Database of Systematic Reviews. 2014(1).

53. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336(7650):924.

54. Gomes B, Calanzani N, Curiale V, McCrone P, Higginson IJ. Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. Cochrane Database of Systematic Reviews. 2013(6).

55. Hong QN, Pluye P, Bujold M, Wassef M. Convergent and sequential synthesis designs: implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. Systematic Reviews. 2017;6(1):61.

56. Woertman W, Vermeulen B, Groenewoud H, van der Wilt GJ. Evidence based policy decisions through a Bayesian approach: The case of a statin appraisal in the Netherlands. Health Policy. 2013;112(3):234-40.

57. Glenton C, Colvin CJ, Carlsen B, Swartz A, Lewin S, Noyes J, et al. Barriers and facilitators to the implementation of lay health worker programmes to improve access to maternal and child health: a qualitative evidence synthesis. Cochrane Database of Systematic Reviews. 2013(10).

58. ESA. COVID-19: Nitrogen Dixode over China. European Space Agency; 2020.

59. He G, Pan Y, Tanaka T. The short-term impacts of COVID-19 lockdown on urban air pollution in China. Nature Sustainability. 2020.

60. Muhammad S, Long X, Salman M. COVID-19 pandemic and environmental pollution: A blessing in disguise? Science of The Total Environment. 2020;728:138820.

61. Dantas G, Siciliano B, França BB, da Silva CM, Arbilla G. The impact of COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil. Science of The Total Environment. 2020;729:139085.

62. ESA. Air Pollution Drops in India Following Lockdown. European Space Agency; 2020.

63. Mahato S, Pal S, Ghosh KG. Effect of lockdown amid COVID-19 pandemic on air quality of the megacity Delhi, India. Science of The Total Environment. 2020;730:139086.

64. ESA. Air Pollution Remains Low as Europeans Stay at Home. European Space Agency; 2020.

65. EEA. Air Quality and COVID-19. European Environment Agency; 2020.

66. WHO. WHO Expert Consultation: Available evidence for the future update of the WHO Global Air Quality Guidelines (AQGs). World Health Organization Regional Office for Europe; 2016 29 September - 1 October 2015.

67. Burns J, Movsisyan A, Stratil JM, Coenen M, Emmert-Fees KMF, Geffert K, et al. Travel-related control measures to contain the COVID-19 pandemic: a rapid review. Cochrane Database of Systematic Reviews. 2020(9).

68. Burns J, Movsisyan A, Stratil JM, Biallas RL, Coenen M, Emmert-Fees KMF, et al. International travel-related control measures to contain the COVID-19 pandemic: a rapid review. Cochrane Database of Systematic Reviews. 2021(3).

69. von Philipsborn P, Stratil JM, Burns J, Busert LK, Pfadenhauer LM, Polus S, et al. Environmental interventions to reduce the consumption of sugar-sweetened beverages and their effects on health. Cochrane Database of Systematic Reviews. 2019(6).

70. Pfadenhauer LM, Burns J, Rohwer A, Rehfuess EA. Effectiveness of interventions to reduce exposure to lead through consumer products and drinking water: A systematic review. Environmental Research. 2016;147:525-36.

71. Koch A, Burns J, Catchpole K, Weigl M. Associations of workflow disruptions in the operating room with surgical outcomes: a systematic review and narrative synthesis. BMJ Quality & Safety. 2020;29(12):1033.

72. Durao S, Schmidt BM, Burns J, Tumusiime D, Pfadenhauer LM, Rehfuess E, et al. Population-level interventions to increase physical activity to prevent cardiovascular diseases and diabetes in low- and middle-income countries: a systematic review protocol. PROSPERO. 2018;CRD42018093429.

73. Kobusingye O, Muni K, Bayiga E, Biribawa C, Kyamanywa P, Mutto M, et al. Interventions to prevent road traffic injuries and deaths among pedestrians: a protocol for a global systematic review. PROSPERO. 2018;CRD42018104287.

74. Krishnaratne S, Pfadenhauer LM, Coenen M, Geffert K, Jung-Sievers C, Klinger C, et al. Measures implemented in the school setting to contain the COVID-19 pandemic: a rapid scoping review. Cochrane Database of Systematic Reviews. 2020(12).

75. Stratil JM, Biallas RL, Burns J, Arnold L, Geffert K, Kunzler AM, et al. Non-pharmacological measures implemented in the setting of long-term care facilities to prevent SARS-CoV-2 infections and their consequences: a rapid review. Cochrane Database of Systematic Reviews. 2021(5).

76. Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. Research Synthesis Methods. 2018;9(1):13-24.

77. Rohwer A, Pfadenhauer L, Burns J, Brereton L, Gerhardus A, Booth A, et al. Series: Clinical Epidemiology in South Africa. Paper 3: Logic models help make sense of complexity in systematic reviews and health technology assessments. Journal of Clinical Epidemiology. 2017;83:37-47.

78. Polus S, Burns J, Hoffmann S, Mathes T, Mansmann U, Been JV, et al. Interrupted time series study found mixed effects of the impact of the Bavarian smoke-free legislation on pregnancy outcomes. Scientific Reports. 2021;11(1):4209.

79. Pedron S, Hanselmann M, Rich A, Burns J, Laxy M. The effect of population-based blood pressure screening and feedback on hypertension status on long-term cardio-metabolic morbidity and mortality: A regression discontinuity analysis. Forthcoming.

80. Burns J, Kurz C, Laxy M. Effectiveness of the German Disease Management Programs: quasi-experimental analyses assessing the population-level health impact. Forthcoming.

81. Polus S, Pieper D, Burns J, Fretheim A, Ramsay C, Higgins JPT, et al. Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews. Journal of Clinical Epidemiology. 2017;91:56-69.

Acknowledgements

After such a long process, there are naturally many people who deserve a big thanks.

I had the luxury of being taken up into a very welcoming (and at the time very small) team, and for that I would like to sincerely thank Eva Rehfuess – from day one of my Master's internship until today, she has been and continues to be such a kind and dependable supervisor.

I had the most incredible work environment one can imagine – in fact, I believe most probably cannot imagine what it is like to have your closest colleagues also be your closest friends – so for that I would like to say thanks to Lisa Pfadenhauer and Stephanie Polus. Our time in Freimann was interesting to say the least, and I always look forward to continuing our Freefolk traditions.

On a personal level, it is difficult to know where to begin, as so many people wonderful people have been with me through this entire process. To all of the Burns and Lawson crew, at home in Arkansas, in Oregon and in Kentucky – thanks for everything; and there is so much that falls under 'everything'. Finally, and importantly, thanks so much to Kathrin Mößmer, I could not have asked for a better companion over the last many years.