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

WILEY

TABLE OF CONTENTS

[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 FOR THE MAIN COMPARISON [*Explanation*]

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 Assessed with: routine and study-specific air quality monitors Follow-up: range 2 months to 2 years	**4 studies**: 1 CBA-EPOC 3 CBA	⊕⊕○○ LOW	2 studies, including 2 CBAs (Titos 2015a**; Titos 2015b**), showed significant decreases in black carbon of 72% and 37% 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.

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.

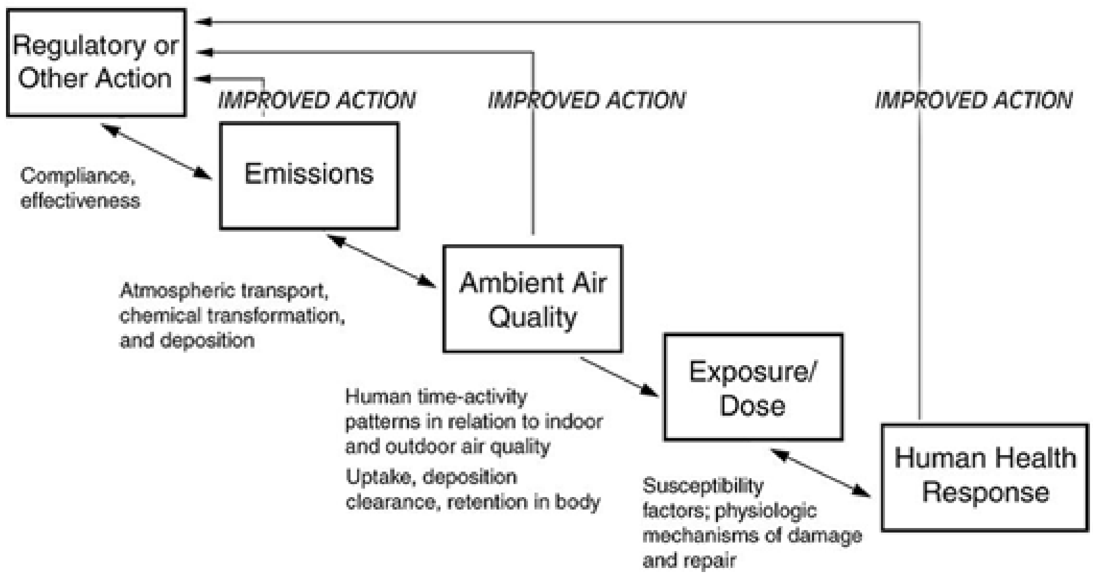
- Industrial: emission standards and regulations for power plants and other industrial sources, fuel changes.
- Residential: stove changeout programmes, banning the sale and use of coal.
- Vehicular: low emission zones, vehicle charging schemes, public transportation expansion; fuel and technology changes; 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).

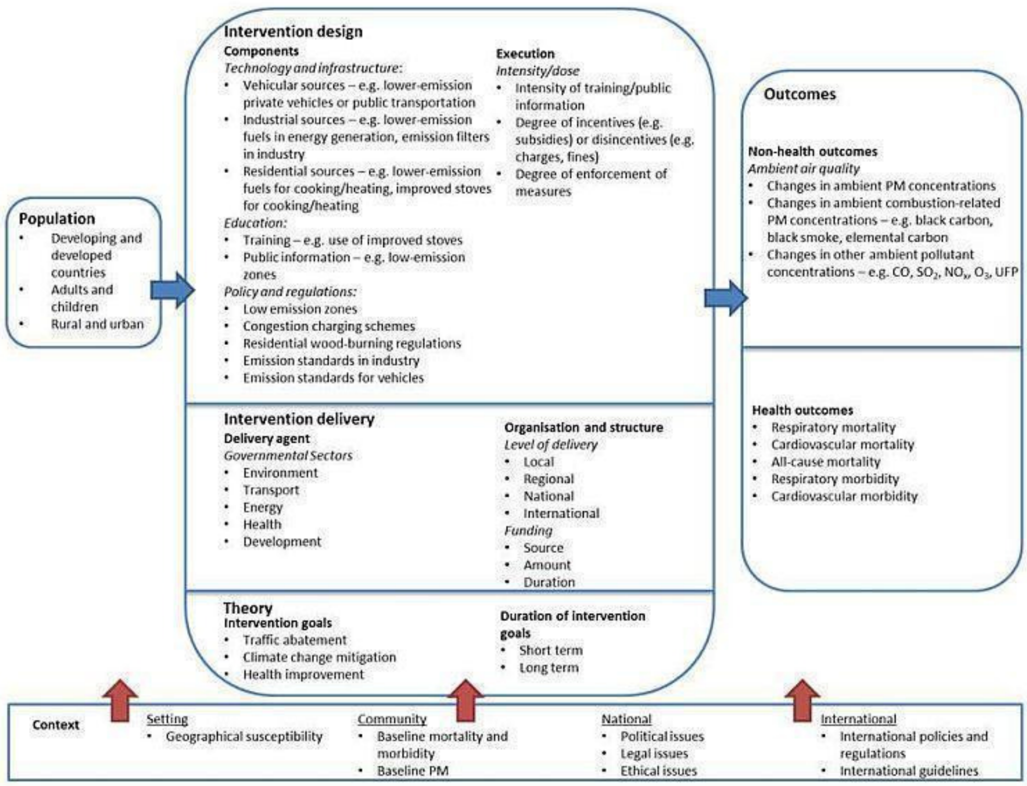
The US National Research Council's Committee on Research Priorities for Airborne Particulate Matter set out a conceptual framework for linking air pollution sources to adverse health effects (NRC 2002). This 'chain of accountability' has been adapted by the Health Effects Institute, as shown in Figure 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.

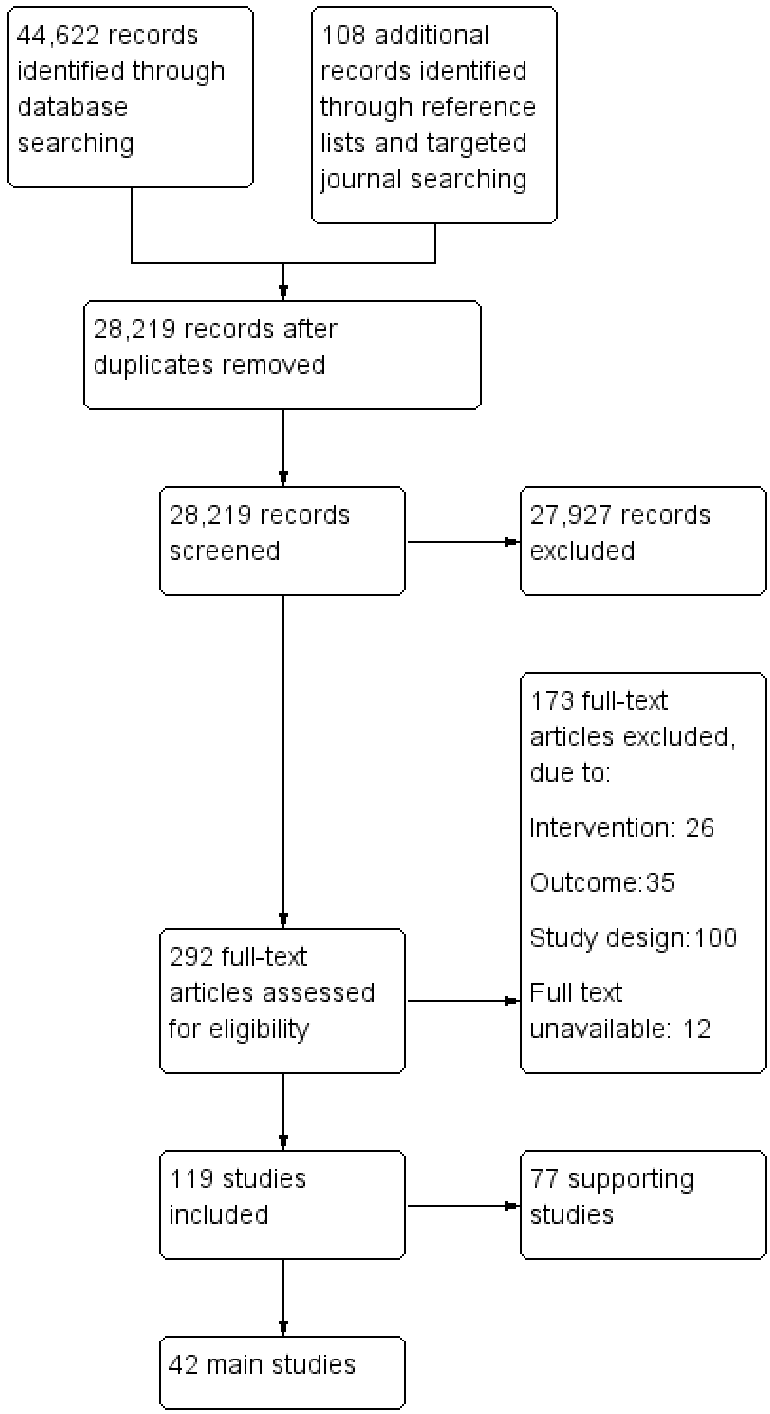
RESULTS

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.

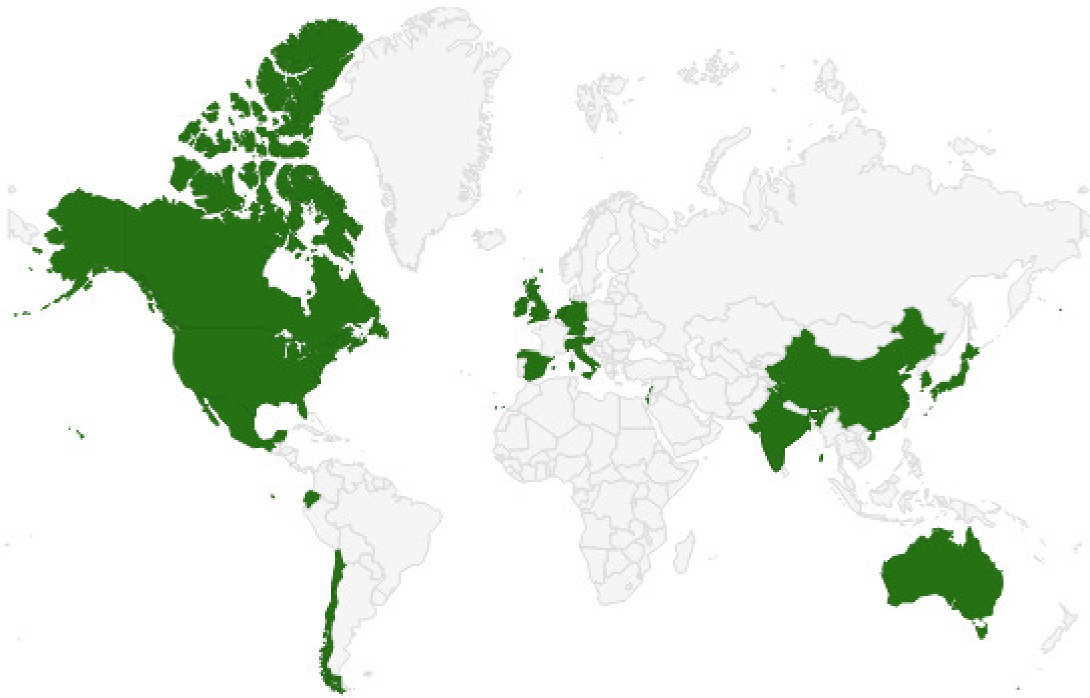
Included studies

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 sub-category**	**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 SO_2 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	Community	2012 - permanent (specific timing unclear)

(*Continued*)

		new stoves		
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)

(*Continued*)

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

(*Continued*)

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, comprehen-	City centre	19 July 1996 to 4 August 1996

(*Continued*)

		sive traveller information and updates, encouraging businesses to provide telecommuting and alternative work hours for employees		
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 restructuring	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
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	Region	October 2003 - permanent;

(*Continued*)

		or be retrofitted to reduce emissions; These standards were then further tightened in some regions.		April 2006 - permanent
Multiple sources				
Giovanis 2015	Repeated coordinated measures	Co-ordinated 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)

(*Continued*)

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	(++)	(++)
Dolislager 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

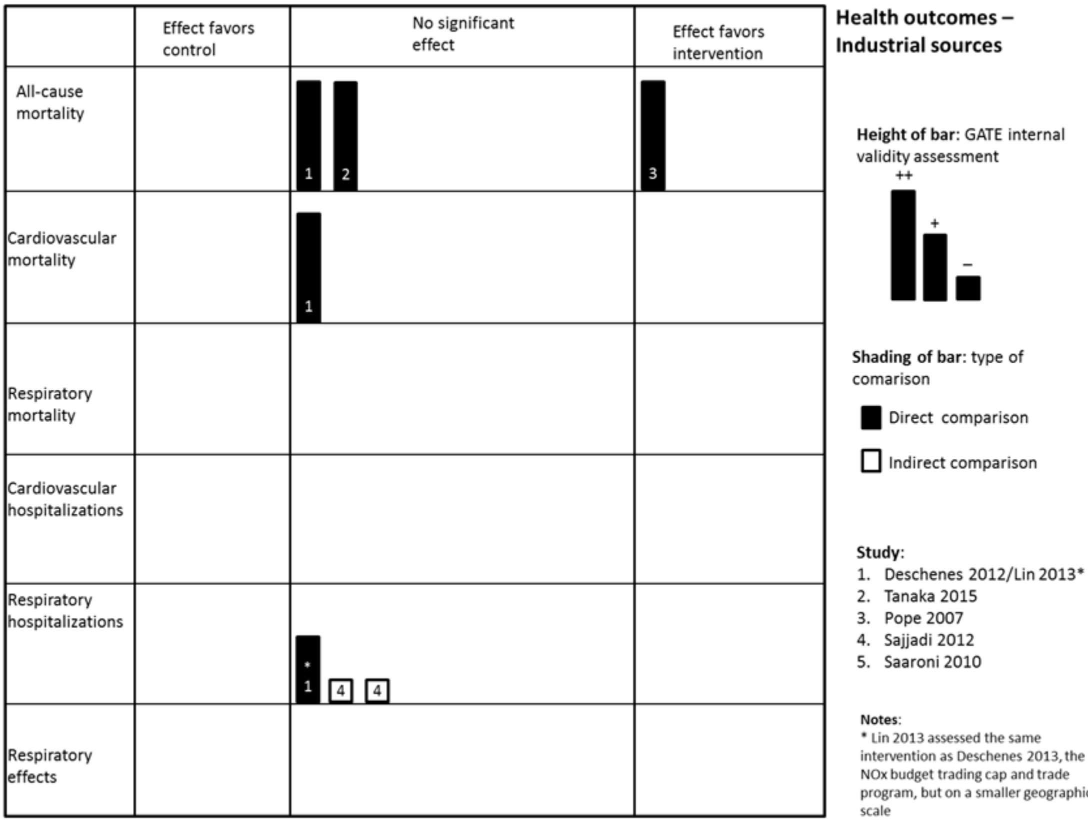
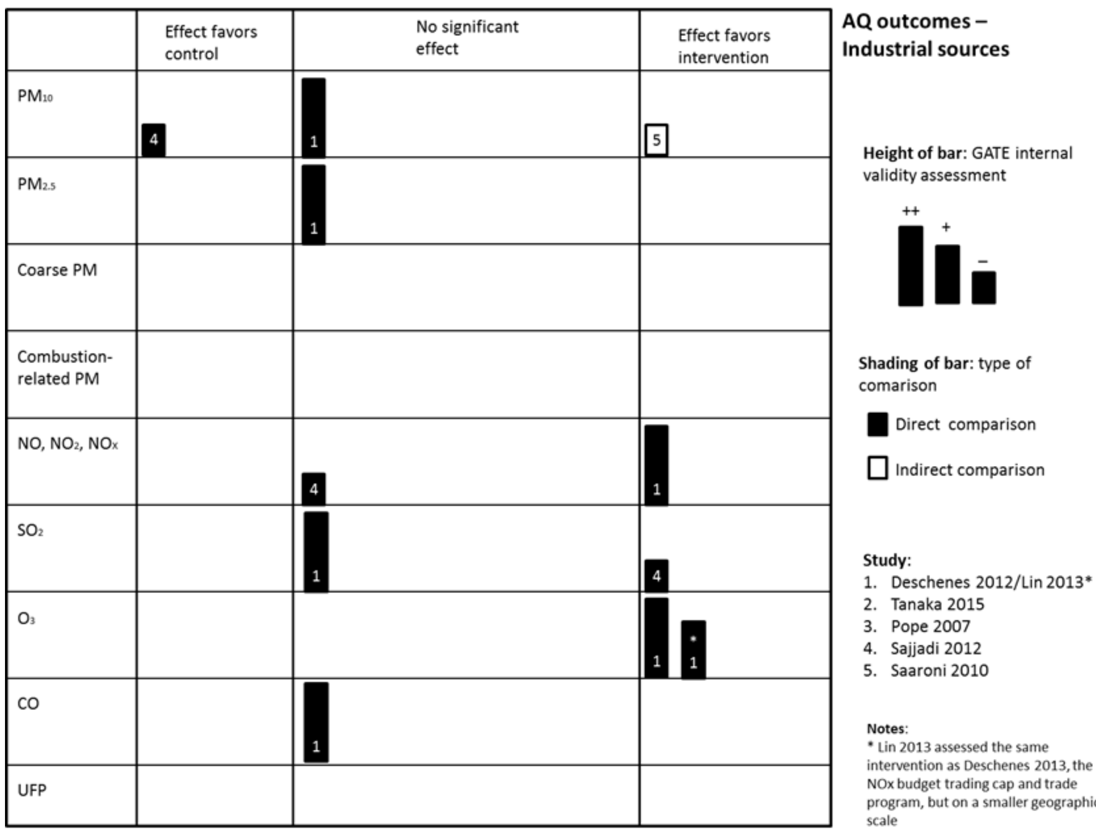


Figure 8. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from industrial sources on AQ outcomes



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

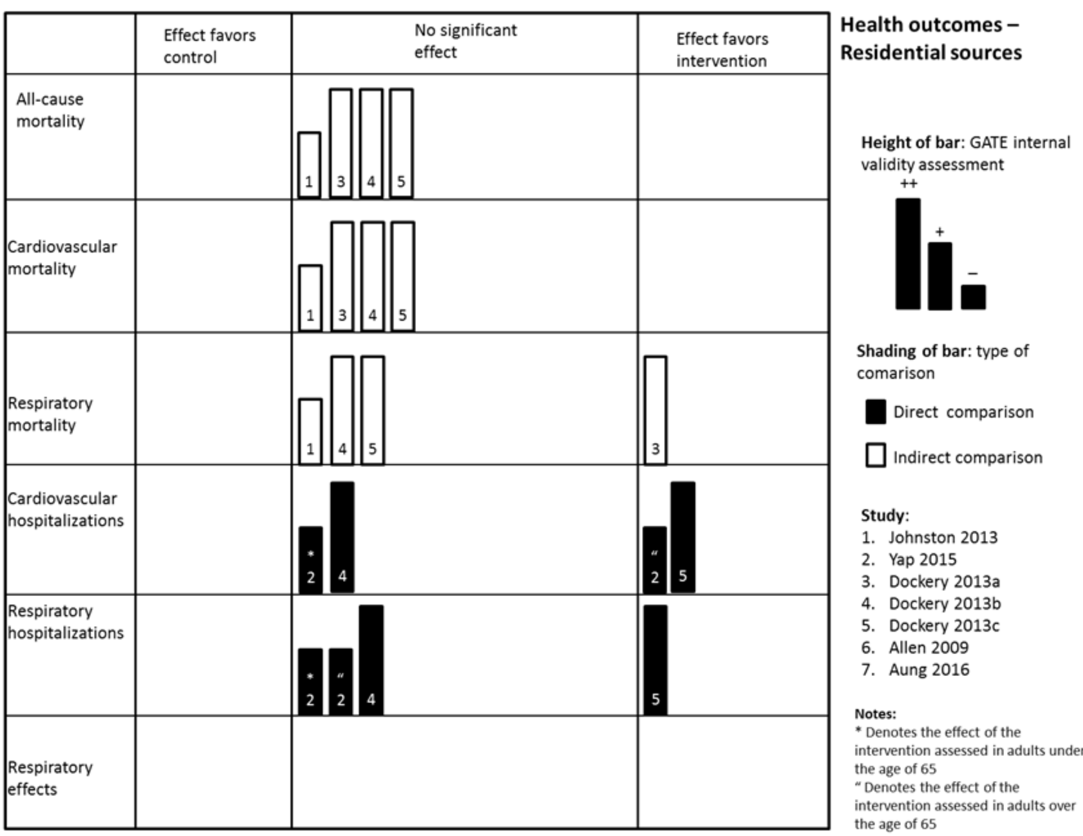
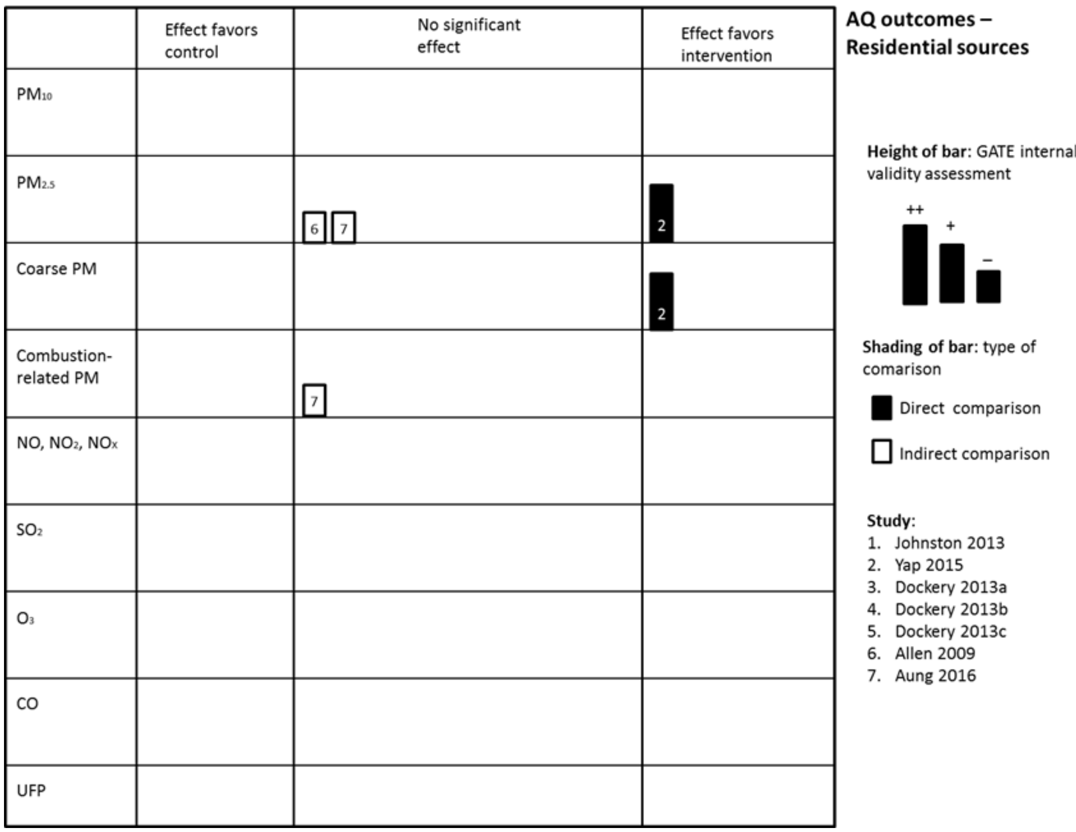


Figure 10. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from residential sources on AQ outcomes



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/m³ median change at intervention sites; −3.4 ug/m³ 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

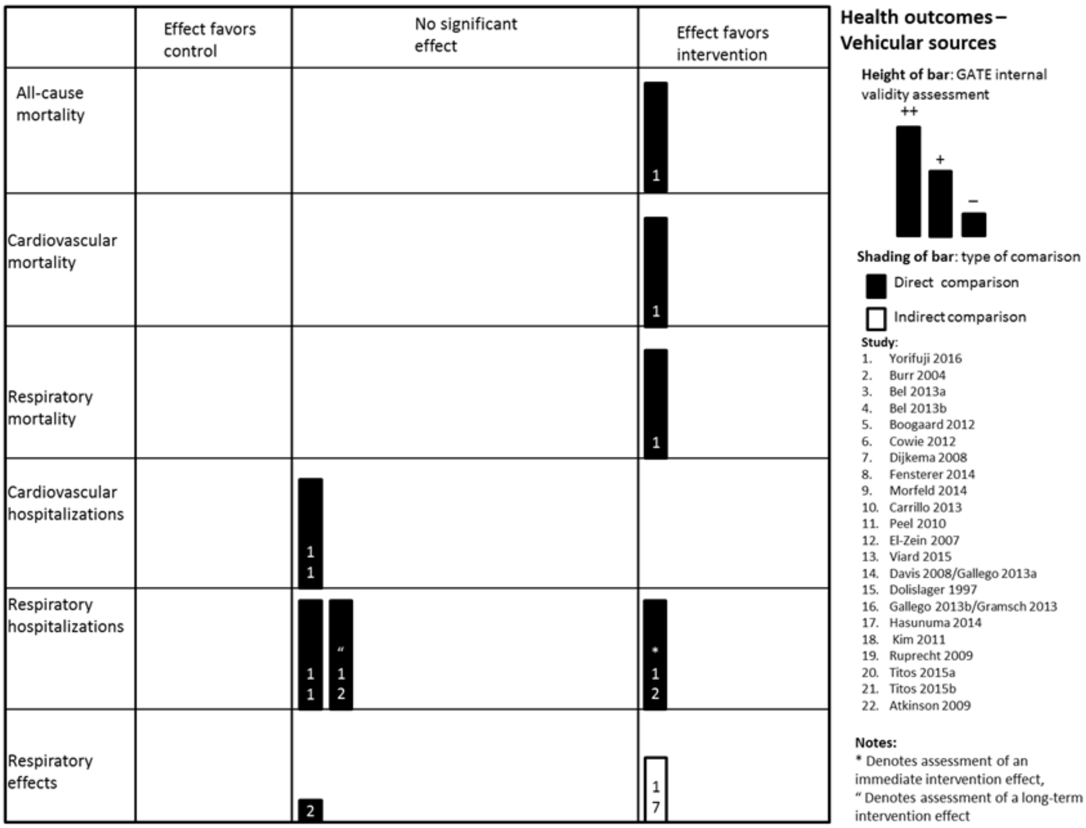
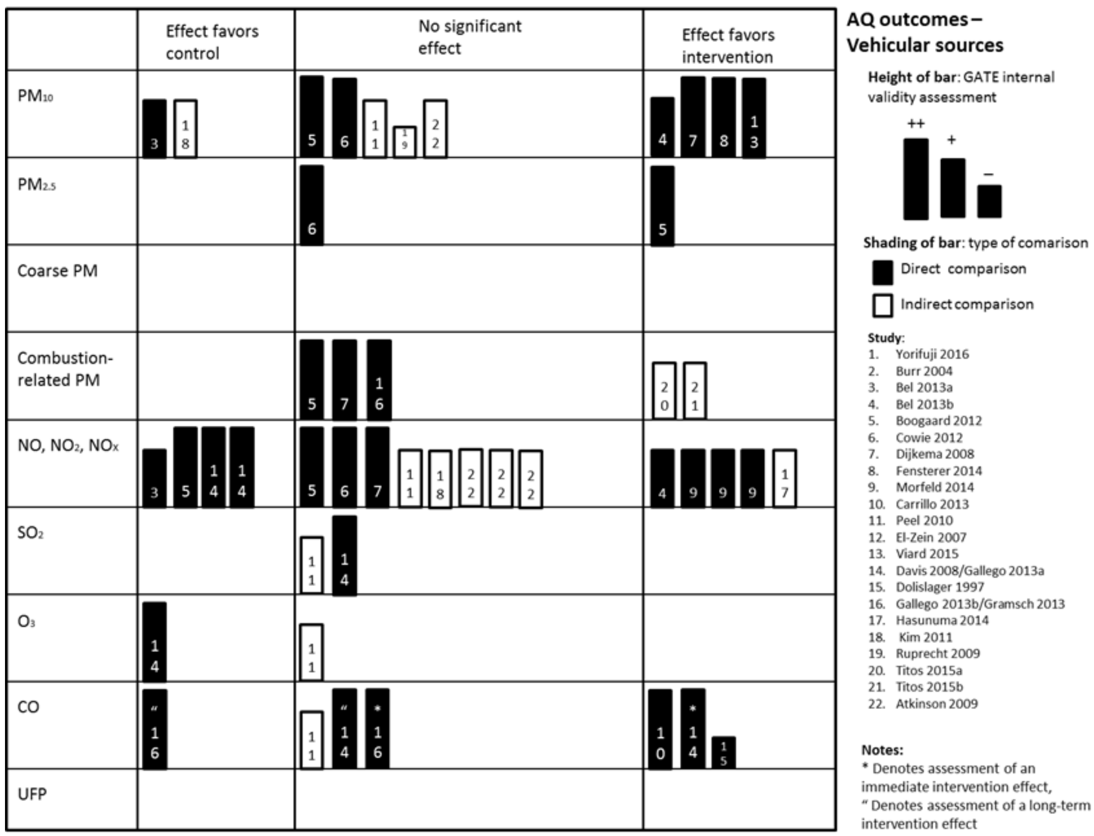


Figure 12. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from vehicular sources on AQ outcomes



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

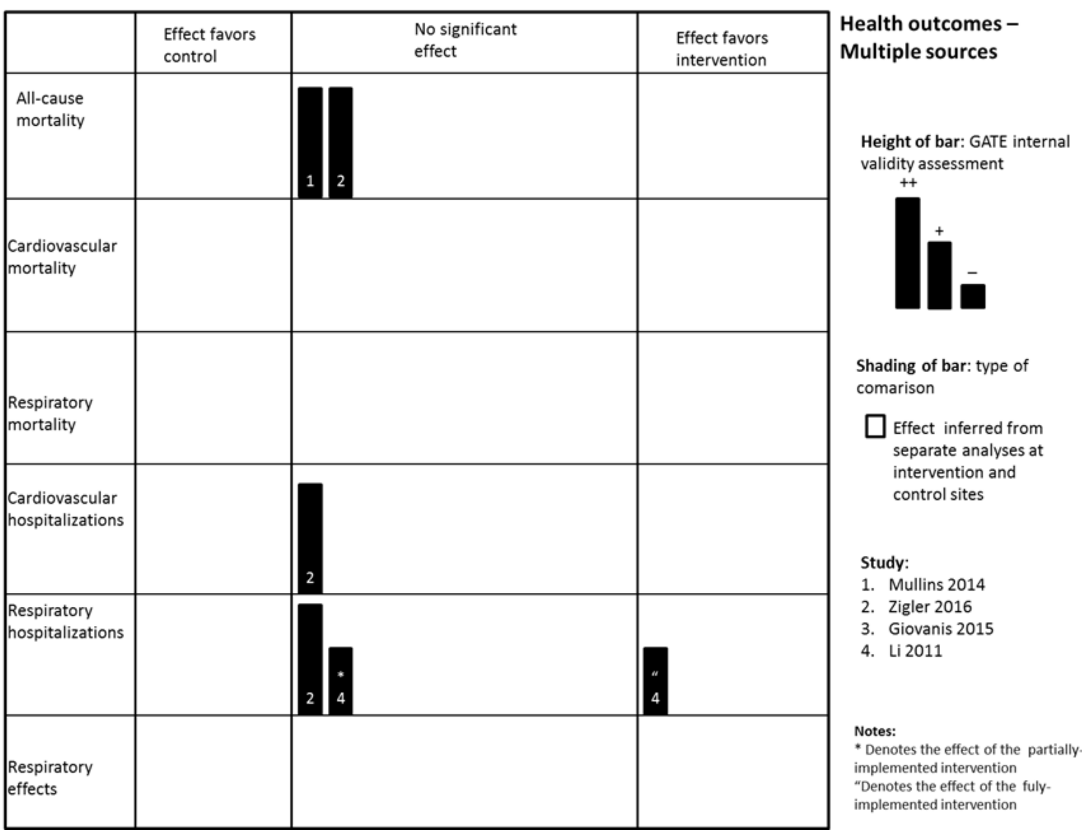
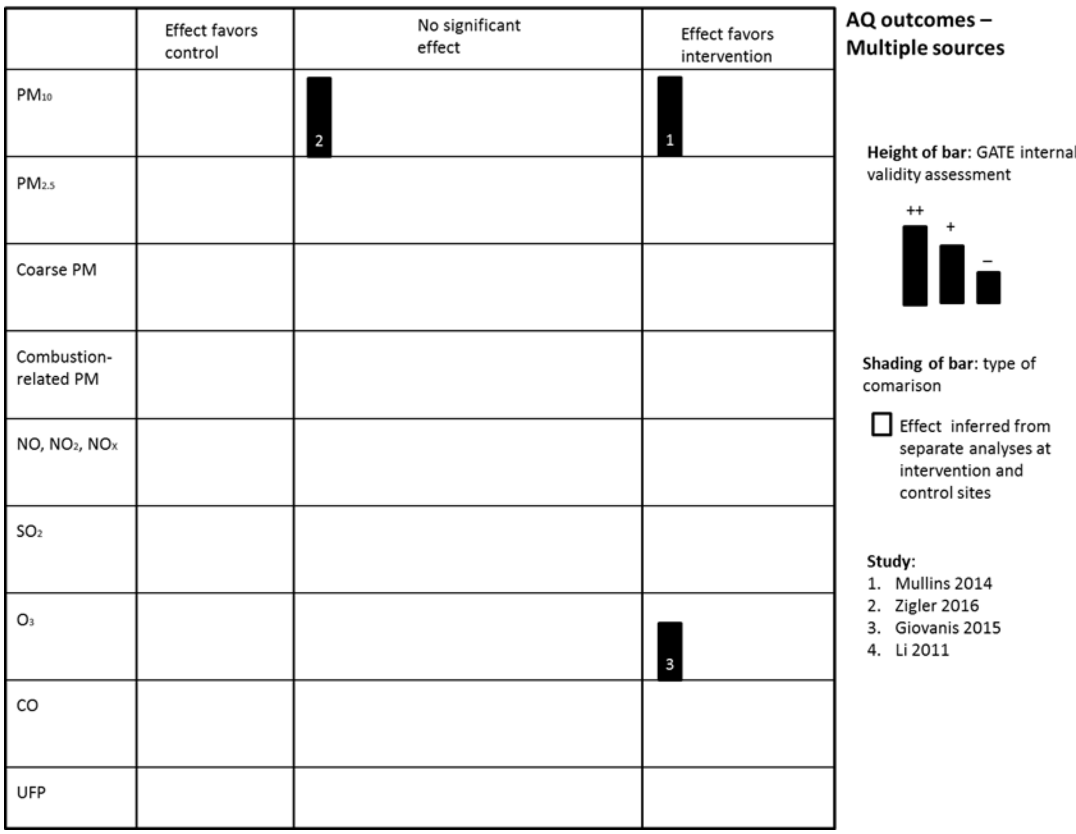


Figure 14. Harvest plot portraying the effects of interventions aiming to reduce ambient air pollution from multiple sources on AQ outcomes



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%).

ADDITIONAL SUMMARY OF FINDINGS [Explanation]

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 [12]	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.

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.

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.

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, multicomponent 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.

Yinon 2017
Yinon L, Thurston G. An evaluation of the health benefits achieved at the time of an air quality intervention in three Israeli cities. *Environment International* 2017;**102**:66–73.

Zamurs 1984
Zamurs J. Assessing the Effect of Transportation Control Strategies on Urban Carbon Monoxide Concentrations. *Journal of the Air Pollution Control Association* 1984;**34**(6): 637–42.

Zhang 2016
Zhang Z, Wang J, Guo M, Xiong M, Zhou Q, Li D, et al. Air quality improvement during 2010 Asian games on blood coagulability in COPD patients. *Environmental Science & Pollution Research* 2016;**23**(7):6631–8.

Zigler 2014
Zigler CM, Dominici F. Point: Clarifying Policy Evidence With Potential-Outcomes Thinking-Beyond Exposure-Response Estimation in Air Pollution Epidemiology. *American Journal of Epidemiology* 2014;**180**(12):1133–40.

* *Indicates the major publication for the study*

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Allen 2009

Methods	Study design: CBA
Participants	Country: Canada Location description: Rural - Smithers and Telkwa, communities in British Columbia Population description: NA Sampling description: NA
Interventions	Category: Residential Sub-category: Stove exchange Level of implementation: Community Description: Stove exchanges, along with financial incentives for purchasing new stoves Timing of introduction and duration: 2012 - permanent (specific timing of introduction unclear)
Outcomes	Health outcomes: NA AQ outcomes: $PM_{2.5}$
Notes	Intervention also assessed by: NA

Atkinson 2009

Methods	Study design: CBA
Participants	Country: UK Location: Urban - London metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Charging scheme Level of implementation: City centre Description: Congestion charging scheme applied to four-wheeled vehicles entering the charging zone on workdays Timing of introduction and duration: First implementation: February 2003 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, NOx, NO_2, NO, CO, O_3
Notes	Intervention also assessed by: Kelly 2011

Aung 2016

Methods	Study design: CBA
Participants	Country: India Location: Rural - Village in Karnataka, southern India Population description: NA Sampling description: NA
Interventions	Category: Residential Sub-category: Stove exchange Level of implementation: Community Description: Removal of traditional stoves from intervention homes, installation of new stoves, assistance with stove operation and maintenance Timing of introduction and duration: 2007 or 2008 - permanent (specific timing unclear)
Outcomes	Health outcomes: NA AQ outcomes: $PM_{2.5}$, BC
Notes	Intervention also assessed by: NA

Bel 2013a

Methods	Study design: cITS-EPOC
Participants	Country: Spain Location: Barcelona Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Speed limit change Level of implementation: City Description: 80 km/h speed limit on motorways Timing of introduction and duration: 1 January 2008 to 31 December 2010
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, NOx
Notes	Intervention also assessed by: NA

Bel 2013b

Methods	Study design: ITS-EPOC
Participants	Country: Spain Location: Barcelona Population description: NA Sampling description: NA

Bel 2013b *(Continued)*

Interventions	Category: Vehicular Sub-category: Speed limit change Level of implementation: City Description: Variable speed limit (minimum 40, maximum 80 km/h) based on traffic density and specific conditions, such as accidents, construction, air pollution, poor weather Timing of introduction and duration: 1 January 2009 to 31 December 2010
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, NOx
Notes	Intervention also assessed by: NA

Boogaard 2012

Methods	Study design: CBA-EPOC
Participants	Country: the Netherlands Location: Urban - City centres of Amsterdam, the Hague, Den Bosch, Tilburg, Utrecht Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Low emission zone Level of implementation: City centre Description: Low emission zones limiting the types of trucks allowed to enter the city centres of the assessed cities. Limits became more stringent over time Timing of introduction and duration: July 2007 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, $PM_{2.5}$, NOx, NO_2, soot
Notes	Intervention also assessed by: NA

Burr 2004

Methods	Study design: CBA
Participants	Country: UK Location: Urban - small town in northern Wales Population description: All residents and workers both in the intervention and a control street Sampling description: Not reported
Interventions	Category: Vehicular Sub-category: Infrastructure changes Level of implementation: Street Description: Opening of bypass around an area subject to heavy traffic congestion Timing of introduction and duration: 1997 or 1998 - permanent (specific timing of introduction unclear)

Burr 2004 *(Continued)*

Outcomes	Health outcomes: Respiratory symptoms, lung function AQ outcomes: PM_{10}, $PM_{2.5}$
Notes	Intervention also assessed by: NA

Butler 2011

Methods	Study design: ITS-EPOC
Participants	Country: USA Location: Mixed urban/rural - areas of the Eastern and Midwestern USA Population description: NA Sampling description: NA
Interventions	Category: Industrial Sub-category: Cap and trade programme Level of implementation: Region Description: 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 utilize control technologies, switch fuels or buy and sell allowances at a free market price Timing of introduction and duration: 2003 to 2008 (ozone season only)
Outcomes	Health outcomes: NA AQ outcomes: O_3
Notes	Intervention also assessed by: Deschênes 2012, Lin 2013

Carrillo 2016

Methods	Study design: CBA-EPOC
Participants	Country: Ecuador Location: Urban - Quito metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Even-odd restriction Level of implementation: City centre Description: Restriction of the city centre during weekday peak traffic hours based on the last digit of a vehicle's licence plate number. Establishment of free parking areas on the periphery of the restriction zone, allowing drivers to utilize public transportation Timing of introduction and duration: 3 May 2010 - permanent (subject to annual reassessment)
Outcomes	Health outcomes: NA AQ outcomes: CO

Carrillo 2016 *(Continued)*

Notes	Intervention also assessed by: NA

Clancy 2002

Methods	Study design: cITS-EPOC
Participants	Country: Ireland Location: Dublin metropolitan area Population description: Residents of Dublin Sampling description: Data on all deaths assessed
Interventions	Category: Residential Sub-category: Coal ban Level of implementation: City Description: Ban on marketing, sale and distribution of coal used for heating Timing of introduction and duration: September 1990 - permanent
Outcomes	Health outcomes: All-cause mortality, cardiovascular mortality, respiratory mortality AQ outcomes: NA
Notes	Intervention also assessed by: Dockery 2013

Cowie 2012

Methods	Study design: cITS-EPOC
Participants	Country: Australia Location: Urban - primary residential area of Sydney Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Infrastructure change Level of implementation: Community Description: 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 Timing of introduction and duration: 25 March 2007 - permanent (tunnel opening); March 2008 - permanent (road changes)
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, $PM_{2.5}$, NOx, NO_2
Notes	Intervention also assessed by: NA

Davis 2008

Methods	Study design: ITS-EPOC
Participants	Country: Mexico Location: Urban - Mexico City metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Even-odd restriction Level of implementation: City Description: Banning of drivers from using their vehicles one day per week based on the last digit of the licence plate Timing of introduction and duration: 20 November 1989 - permanent
Outcomes	Health outcomes: NA AQ outcomes: NOx, NO_2, O_3, SO_2, CO
Notes	Intervention also assessed by: Gallego 2013a

Deschênes 2012

Methods	Study design: cITS-EPOC
Participants	Country: USA Location: Mixed urban/rural - areas of the Eastern and Midwestern USA Population description: NA Sampling description: NA
Interventions	Category: Industrial Sub-category: Cap and trade programme Level of implementation: Region Description: 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 utilize control technologies, switch fuels or buy and sell allowances at a free market price Timing of introduction and duration: 2003 to 2008 (ozone season only)
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, $PM_{2.5}$, NO_2, O_3, SO_2, CO
Notes	Intervention also assessed by: Butler 2011, Lin 2013

Dijkema 2008

Methods	Study design: CBA
Participants	Country: the Netherlands Location: Urban - Amsterdam metropolitan area Population description: NA

Dijkema 2008 *(Continued)*

	Sampling description: NA
Interventions	Category: Vehicular Sub-category: Speed limit change Level of implementation: Street Description: Speed limit reduction on urban traffic ring Timing of introduction and duration: November 2009 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, BS, NOx
Notes	Intervention also assessed by: NA

Dockery 2013a

Methods	Study design: cITS-EPOC - all-cause mortality, cardiovascular mortality, respiratory mortality
Participants	Country: Ireland Location: Urban - Dublin Population description: Residents of Dublin and the Midland and Coastal control counties Sampling description: Data on all deaths and hospital admissions assessed
Interventions	Category: Residential Sub-category: Coal ban Level of implementation: City Description: Ban on marketing, sale and distribution of coal used for heating Timing of introduction and duration: 1990 to 2000 - permanent (specific timing of introduction is city-dependent)
Outcomes	Health outcomes: All-cause mortality, cardiovascular mortality, respiratory mortality, cardiovascular hospitalization, respiratory hospitalization AQ outcomes: NA
Notes	Intervention also assessed by: Clancy 2002

Dockery 2013b

Methods	Study design: cITS-EPOC - all-cause mortality, cardiovascular mortality, respiratory mortality ITS-EPOC - cardiovascular hospitalization, respiratory hospitalization
Participants	Country: Ireland Location: Urban - Cork City and County Population description: Residents of Cork City and County and the Midland and Coastal control counties Sampling description: Data on all deaths and hospital admissions assessed

Dockery 2013b (*Continued*)

Interventions	Category: Residential Sub-category: Coal ban Level of implementation: City Description: Ban on marketing, sale and distribution of coal used for heating Timing of introduction and duration: 1990 to 2000 - permanent (specific timing of introduction is city-dependent)
Outcomes	Health outcomes: All-cause mortality, cardiovascular mortality, respiratory mortality, cardiovascular hospitalization, respiratory hospitalization AQ outcomes: NA
Notes	Intervention also assessed by: NA

Dockery 2013c

Methods	Study design: cITS-EPOC - all-cause mortality, cardiovascular mortality, respiratory mortality ITS-EPOC - cardiovascular hospitalization, respiratory hospitalization
Participants	Country: Ireland Location: Urban - Limerick City and County, Louth, Wexford and Wicklow Population description: Residents of Limerick City and County, Louth, Wexford and Wicklow and the Midland and Coastal control counties Sampling description: Data on all deaths and hospital admissions assessed
Interventions	Category: Residential Sub-category: Coal ban Level of implementation: City Description: Ban on marketing, sale and distribution of coal used for heating Timing of introduction and duration: 1990 to 2000 - permanent (specific timing of introduction is city-dependent)
Outcomes	Health outcomes: All-cause mortality, cardiovascular mortality, respiratory mortality, cardiovascular hospitalization, respiratory hospitalization AQ outcomes: NA
Notes	Intervention also assessed by: NA

Dolislager 1997

Methods	Study design: ITS-EPOC
Participants	Country: US Location: Urban - 4 metropolitan areas in California Population description: NA Sampling description: NA

Dolislager 1997 *(Continued)*

Interventions	Category: Vehicular Sub-category: Fuel requirements Level of implementation: Region Description: Requiring gasoline sold during months prone to high CO concentrations to have a low oxygen content Timing of introduction and duration: November 1991 - permanent (winter only)
Outcomes	Health outcomes: NA AQ outcomes: CO
Notes	Intervention also assessed by: NA

El-Zein 2007

Methods	Study design: ITS-EPOC
Participants	Country: Lebanon Location: Urban - Beirut metropolitan area Population description: Children in Beirut under 17 years Sampling description: Data on all hospital admissions from accredited hospitals assessed
Interventions	Category: Vehicular Sub-category: Vehicle restriction Level of implementation: Country Description: Ban on the import of all light- and medium-duty diesel engines Timing of introduction and duration: June 2002 - permanent
Outcomes	Health outcomes: Respiratory hospital admissions AQ outcomes: NA
Notes	Type of effect reported: Indirect Intervention also assessed by: NA

Fensterer 2014

Methods	Study design: CBA
Participants	Country: Germany Location: Urban - Munich metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Low emission zone Level of implementation: City Description: Low emission zone in line with EURO regulations, becoming gradually more stringent Timing of introduction and duration: October 2008 - permanent

Fensterer 2014 *(Continued)*

Outcomes	Health outcomes: NA AQ outcomes: PM_{10}
Notes	Intervention also assessed by: Morfeld 2013

Friedman 2001

Methods	Study design: ITS-EPOC (health outcomes); CBA-EPOC (AQ outcomes)
Participants	Country: USA Location: Atlanta metropolitan area Population description: All residents of Atlanta and control areas Sampling description: Data on all emergency department visits from select hospitals assessed
Interventions	Category: Vehicular Sub-category: Comprehensive traffic reduction strategy Level of implementation: City Description: 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 Timing of introduction and duration:
Outcomes	Health outcomes: Hospital (emergency department) admissions due to asthma AQ outcomes: PM_{10}, NO_2, O_3, SO_2, CO
Notes	Intervention also assessed by: Peel 2010

Gallego 2013a

Methods	Study design: ITS-EPOC
Participants	Country: Mexico Location: Urban - Mexico city metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Even-odd restriction Level of implementation: City Description: Even-odd driving ban: banning of drivers from using their vehicles one day per week based on the last digit of the license plate Timing of introduction and duration: 20 November 1989 - permanent
Outcomes	Health outcomes: NA AQ outcomes: CO

Gallego 2013a *(Continued)*

Notes	Intervention also assessed by: Davis 2008

Gallego 2013b

Methods	Study design: ITS-EPOC
Participants	Country: Chile Location: Santiago metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Public transportation restructuring Level of implementation: City Description: Restructuring of the entire public transport system, including changes to the subway system and bus network Timing of introduction and duration: 10 February 2007 - Permanent
Outcomes	Health outcomes: NA AQ outcomes: CO
Notes	Intervention also assessed by: Gramsch 2013

Giovanis 2015

Methods	Study design: CBA-EPOC
Participants	Country: USA Location: Mixed Urban/Rural - Charlotte, North Carolina and surrounding area Population description: NA Sampling description: NA
Interventions	Category: Multiple Sub-category: Repeated coordinated measures Level of implementation: Region Description: 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 Timing of introduction and duration: March 2006 - permanent (intermittent operation: implemented on days where especially high levels are expected, then relaxed when levels drop)
Outcomes	Health outcomes: NA AQ outcomes: O_3
Notes	Intervention also assessed by: NA

Gramsch 2013

Methods	Study design: ITS-EPOC
Participants	Country: Chile Location: Santiago metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Public transportation restructuring Level of implementation: City Description: Restructuring of the entire public transport system, including changes to the subway system and bus network Timing of introduction and duration: 10 February 2007 - Permanent
Outcomes	Health outcomes: NA AQ outcomes: BC
Notes	Intervention also assessed by: Gramsch 2013

Hasunuma 2014

Methods	Study design: CBA-EPOC
Participants	Country: Japan Location: Mixed Urban/Rural - areas spread across Japan Population description: Children 3 years old living in the 28 survey areas Sampling description: Not reported
Interventions	Category: Vehicular Sub-category: Required vehicle standards Level of implementation: Country Description: Ban on automobiles not conforming to the Automobile NOx/PM Law, in areas designated enforcement areas Timing of introduction and duration: June 2001 - permanent
Outcomes	Health outcomes: Respiratory symptoms AQ outcomes: NA
Notes	Intervention also assessed by: NA

Johnston 2013

Methods	Study design: cITS-EPOC
Participants	Country: Australia Location: Urban - Launceston, Tasmania Population description: Launceston city residents

	Sampling description: Data on all deaths assessed
Interventions	Category: Residential Sub-category: Stove exchange Level of implementation: City Description: Wood Heater Replacement Program, along with an education campaign and adherence monitoring Timing of introduction and duration: July 2001 - June 2004
Outcomes	Health outcomes: Total mortality, cardiovascular mortality, respiratory mortality AQ outcomes: NA
Notes	Intervention also assessed by: NA

Kim 2011

Methods	Study design: CBA-EPOC
Participants	Country: South Korea Location: Urban - Several cities spread across South Korea Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Clean fuel usage Level of implementation: Country Description: Natural Gas Vehicle Supply programme led to the replacement of the entire fleet of diesel-powered city buses with natural gas buses in large cities Timing of introduction and duration: 1 June 2000 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, NO_2
Notes	Intervention also assessed by: Shon 2011 (supporting study)

Li 2011

Methods	Study design: ITS-EPOC
Participants	Country: China Location: Urban - Beijing metropolitan area Population description: All adult residents of Beijing admitted to hospitals for asthma events Sampling description: Data on all admissions assessed
Interventions	Category: Multiple Sub-category: Even-odd restriction; Vehicle restriction; Power plant restriction

Li 2011 *(Continued)*

	Level of implementation: City Description: 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 during the 2008 Beijing Olympic Games Timing of introduction and duration: 1 July 2008 to 7 August 2008
Outcomes	Health outcomes: Asthma hospitalizations AQ outcomes: NA
Notes	Study classification: Main study Type of effect reported: Indirect Intervention also assessed by: Hou 2010, Huang 2012a, Huang 2012b, Lin 2011, Lin 2015, Mu 2014, Rich 2015, Schleicher 2011, Schleicher 2012, Shen 2011, Su 2015, Wang 2014, Xu 2016 (all supporting studies)

Lin 2013

Methods	Study design: ITS-EPOC
Participants	Country: USA Location: Mixed Urban/Rural - State of New York Population description: All residents of New York State hospitalized due to respiratory causes Sampling description: Data on all hospitalzations assessed
Interventions	Category: Industrial Sub-category: Cap and trade programme Level of implementation: Region Description: 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 utilize control technologies, switch fuels or buy and sell allowances at a free market price Timing of introduction and duration: 2003 to 2008 (ozone season only)
Outcomes	Health outcomes: Respiratory hospitalization AQ outcomes: O_3
Notes	Intervention also assessed by: Butler 2011, Deschênes 2012

Morfeld 2013

Methods	Study design: CBA-EPOC
Participants	Country: Germany Location: Urban - Munich city centre Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Low emission zone

Morfeld 2013 *(Continued)*

	Level of implementation: City centre Description: Low emission zone in line with EURO regulations, becoming gradually more stringent Timing of introduction and duration: October 2008 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}
Notes	Intervention also assessed by: Fensterer 2014, Qadir 2013 (supporting study)

Morfeld 2014

Methods	Study design: CBA-EPOC
Participants	Country: Germany Location: Urban - 17 German cities Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Low emission zone Level of implementation: City centre Description: Low emission zone, restricting entrance of diesel cars below Euro II and gasoline cars Euro I standards Timing of introduction and duration: Approximately 2008 - permanent (start date differs for individual cities)
Outcomes	Health outcomes: NA AQ outcomes: NOx, NO_2, NO
Notes	Intervention also assessed by: NA

Mullins 2014

Methods	Study design: ITS-EPOC
Participants	Country: Chile Location: Urban - Santiago metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Multiple Sub-category: Repeated coordinated measures Level of implementation: City Description: 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 Timing of introduction and duration:

Mullins 2014 *(Continued)*

Outcomes	Health outcomes: NA AQ outcomes: PM_{10}
Notes	Intervention also assessed by: NA

Peel 2010

Methods	Study design: ITS-EPOC (health outcomes); CBA-EPOC (AQ outcomes)
Participants	Country: USA Location: Atlanta metropolitan area Population description: All residents of Atlanta and control areas Sampling description: Data on all emergency department visits from select hospitals assessed
Interventions	Category: Vehicular Sub-category: Comprehensive traffic reduction strategy Level of implementation: City Description: 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 Timing of introduction and duration:
Outcomes	Health outcomes: Hospital (emergency department) admissions due to asthma, pneumonia, cardiovascular disease, COPD AQ outcomes: PM_{10}, NO_2, O_3, SO_2, CO
Notes	Intervention also assessed by: Friedman 2001

Pope 2007

Methods	Study design: cITS-EPOC
Participants	Country: USA Location: Mixed Urban/Rural - Southwest US states: Nevada, Utah, New Mexico, Arizona Population description: All residents of the four SW states Sampling description: Data on all hospital admissions assessed
Interventions	Category: Industrial Sub-category: Industry closure Level of implementation: Region Description: National copper smelter strike that was especially relevant in the Southwest US where much copper smelting took place Timing of introduction and duration: 15 July 1967 to early April 1968
Outcomes	Health outcomes: All-cause mortality AQ outcomes: NA

Pope 2007 *(Continued)*

Notes	Intervention also assessed by: NA

Ruprecht 2009

Methods	Study design: CBA
Participants	Country: Italy Location: Urban - Milan city centre Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Charging scheme Level of implementation: City centre Description: Ecopass congestion charging scheme, requiring payment during the week for entering the city centre Timing of introduction and duration: 8 January 2008 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}
Notes	Intervention also assessed by: NA

Saaroni 2010

Methods	Study design: CBA
Participants	Country: Israel Location: Urban - Tel Aviv metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Industrial Sub-category: Power plant conversion Level of implementation: Factory Description: Converting the Tel Aviv power station from oil to gas Timing of introduction and duration: 2005 - permanent (specific timing unclear)
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, NOx, NO_2, NO, SO_2
Notes	Intervention also assessed by: NA

Sajjadi 2011

Methods	Study design: cITS-EPOC
Participants	Country: Australia Location: Mixed Urban/Rural - Lower Hunter region of New South Wales Population description: All residents in the Lower Hunter region hospital catchment area Sampling description: Data on all hospital admissions assessed for relevant outcomes
Interventions	Category: Industrial Sub-category: Factory closure Level of implementation: Factory Description: Closure of the local steel works industry, the major area polluter Timing of introduction and duration: October 1999 - permanent
Outcomes	Health outcomes: Respiratory disease, asthma, asthma (0 to 14 years), COPD (65+ years) AQ outcomes: NA
Notes	Intervention also assessed by: Sajjadi 2012

Sajjadi 2012

Methods	Study design: ITS-EPOC
Participants	Country: Australia Location: Mixed Urban/Rural - Lower Hunter region of New South Wales Population description: NA Sampling description: NA
Interventions	Category: Industrial Sub-category: Factory closure Level of implementation: Factory Description: Closure of the local steel works industry, the major area polluter Timing of introduction and duration: October 1999 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}, $PM_{2.5}$, NO_2, SO_2
Notes	Intervention also assessed by: Sajjadi 2011

Tanaka 2015

Methods	Study design: CBA-EPOC
Participants	Country: China Location: Urban - Several cities spread across China Population description: All infants up to 1 year old from included prefectures Sampling description: Data on all infant deaths assessed

Tanaka 2015 *(Continued)*

Interventions	Category: Industrial Sub-category: Required industry requirements Level of implementation: Country Description: Two Control Zone policy which designated areas exceeding acid rain or SO_2 thresholds as TCZ status. These areas were then subject to more stringent regulations with regard to coal mining and burning Timing of introduction and duration: January 1998 - permanent
Outcomes	Health outcomes: All-cause mortality (age < 1 year old) AQ outcomes: NA
Notes	Intervention also assessed by: NA

Titos 2015a

Methods	Study design: CBA
Participants	Country: Slovenia Location: Urban - Ljubljana metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Infrastructure changes Level of implementation: Street Description: Partial closure and reconstruction of 400 m of a major street. Only public buses and taxis were allowed after implementation Timing of introduction and duration: 22 September 2013 - permanent
Outcomes	Health outcomes: NA AQ outcomes: BC
Notes	Intervention also assessed by: NA

Titos 2015b

Methods	Study design: CBA
Participants	Country: Spain Location: Urban - Granada metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Infrastructure changes Level of implementation: City Description: 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

Titos 2015b *(Continued)*

	Timing of introduction and duration: 29 June 2014 - permanent
Outcomes	Health outcomes: NA AQ outcomes: BC
Notes	Intervention also assessed by: NA

Viard 2015

Methods	Study design: ITS-EPOC
Participants	Country: China Location: Urban - Beijing metropolitan area Population description: NA Sampling description: NA
Interventions	Category: Vehicular Sub-category: Even-odd restriction Level of implementation: City Description: 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 1 day per week Timing of introduction and duration: Two-staged implementation: 20 July 2008 to 20 September 2008; 11 October 2008 - permanent
Outcomes	Health outcomes: NA AQ outcomes: PM_{10}
Notes	Intervention also assessed by: NA

Yap 2015

Methods	Study design: ITS-EPOC
Participants	Country: USA Location: Mixed urban/rural - California's San Joaquin Valley Air Basin Population description: NA Sampling description: NA
Interventions	Category: Residential Sub-category: Wood burning ban Level of implementation: Region Description: 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 Timing of introduction and duration: November 2003 - permanent

Yap 2015 *(Continued)*

Outcomes	Health outcomes: NA AQ outcomes: $PM_{2.5}$, coarse particles
Notes	Intervention also assessed by: NA

Yorifuji 2011

Methods	Study design: ITS-EPOC
Participants	Country: Japan Location: Urban - Tokyo metropolitan area Population description: Residents of Tokyo Sampling description: Data on all deaths assessed
Interventions	Category: Vehicular Sub-category: Required vehicle standards Level of implementation: Region Description: 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 Timing of introduction and duration: two relevant introduction points: • October 2003 - permanent; • April 2006 - permanent
Outcomes	Health outcomes: Total mortality, cardiovascular mortality, respiratory mortality, cerebrovascular mortality, mortality from other causes AQ outcomes: NA
Notes	Intervention also assessed by: Yorifuji 2016

Yorifuji 2016

Methods	Study design: cITS-EPOC
Participants	Country: Japan Location: Urban - Tokyo metropolitan area Population description: Residents of Tokyo Sampling description: Data on all deaths assessed
Interventions	Category: Vehicular Sub-category: Required vehicle standards Level of implementation: Region Description: 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

Yorifuji 2016 *(Continued)*

	Timing of introduction and duration: Two relevant introduction points: October 2003 - permanent; April 2006 - permanent
Outcomes	Health outcomes: Total mortality, cardiovascular mortality, respiratory mortality, cerebrovascular mortality, mortality from other causes AQ outcomes: NA
Notes	Intervention also assessed by: Yorifuji 2011

Zigler 2016

Methods	Study design: CBA-EPOC
Participants	Country: USA Location: Mixed Urban/Rural - Western United States Population description: Residents of the Western United States assessed in the study Sampling description: Data on all deaths and hospitalizations from individuals on Medicare assessed
Interventions	Category: Multiple Sub-category: Tailored selection of measures Level of implementation: Region Description: 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 Timing of introduction and duration: 1990 - permanent
Outcomes	Health outcomes: All-cause mortality, cardiovascular hospital admissions, respiratory hospital admissions AQ outcomes: PM_{10}
Notes	Intervention also assessed by: NA

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Adar 2015	No eligible outcome assessed
Ai 2016	Ineligible study design applied
Ali 2008	Ineligible study design applied
Altemose 2015	No eligible outcome assessed

(*Continued*)

Alvim-Ferraz 2005	No relevant intervention assessed
Ancelet 2015	No relevant intervention assessed
Arossa 1987	No relevant intervention assessed
Auffhammer 2009	Ineligible study design applied
Auffhammer 2011	No relevant intervention assessed
Aunan 1998	No relevant intervention assessed
Aunan 2004	Ineligible study design applied
Aydin 2009	No eligible outcome assessed
Baldasano 2010	Ineligible study design applied
Barbose 2016	No eligible outcome assessed
Barnes 2015	Ineligible study design applied
Barratt 2014	Ineligible study design applied
Bartonova 1999	Ineligible study design applied
Bauman 1977	Full text not available; conference proceedings with no associated full publication
Beevers 2005	No eligible outcome assessed
Bennett 2010	Ineligible study design applied
Berhane 2016	No relevant intervention assessed
Bridgman 2002	Ineligible study design applied
Buckley 2011	No relevant intervention assessed
Carvalho 2015	Ineligible study design applied
Cesaroni 2012	Ineligible study design applied
Chalbot 2014	Ineligible study design applied
Chang 2007	No relevant intervention assessed
Chang 2008	Ineligible study design applied

(*Continued*)

Chay 2003	Ineligible study design applied
Chen 2014	No eligible outcome assessed
Chiesa 2014	Ineligible study design applied
Chong 2014	Ineligible study design applied
Chou 2007	Ineligible study design applied
Chou 2011	Ineligible study design applied
Correia 2013	Ineligible study design applied
Cox 2015	Ineligible study design applied
Crippa 2016	Ineligible study design applied
Critchley 2015	No eligible outcome assessed
Cropper 1997	No relevant intervention assessed
Cruz-Minguillon 2009	Ineligible study design applied
Cyrys 2014	Ineligible study design applied
Cyrys 2015	Ineligible study design applied
Delkash 2016	Ineligible study design applied
DeLuca 2012	No eligible outcome assessed
Dickinson 2015	Ineligible study design applied
Dienes 2014	Ineligible study design applied
Ding 2016	Ineligible study design applied
Dong 2010	Full text not available; English abstract for Chinese publication
Escobedo 2009	Ineligible study design applied
Federal Highway Administration 2014	Ineligible study design applied
Fernandez-Camacho 2016	No relevant intervention assessed

(*Continued*)

Foster 2011	Ineligible study design applied
Frye 2003	Ineligible study design applied
Gallagher 2013	No eligible outcome assessed
Gao 2013	No eligible outcome assessed
Gao 2014	No eligible outcome assessed
Geng 2014	Full text not available; conference proceedings on the intervention during the 2008 Beijing Olympics (also assessed in Hou 2010 and Li 2011, among others)
Gertler 1999	No eligible outcome assessed
Gioda 2016	No relevant intervention assessed
Giuliano 2007	No eligible outcome assessed
Grinshpun 2014	Ineligible study design applied
Hao 2006	Ineligible study design applied
Hara 2013	Ineligible study design applied
Harrison 2015	Ineligible study design applied
Hedley 2002	Ineligible study design applied
Hendryx 2016	No eligible outcome assessed
Henneman 2015	No relevant intervention assessed
Herrstedt 1992	No eligible outcome assessed
Hine 2011	No eligible outcome assessed
Hirten 1997	Ineligible study design applied
Ho 2015	Ineligible study design applied
Huang 1996	Full text not available; evaluation of a range of measures undertaken in Taiwan (also assessed in Kuo 2009)
Huang 2015	No eligible outcome assessed
Hutchinson 2004	Ineligible study design applied

(*Continued*)

Invernizzi 2011	Ineligible study design applied
Jacobi 1999	Ineligible study design applied
Jalihal 2006	Ineligible study design applied
Jenq 1989	Full text not available; conference proceedings with no associated full publication
Jhun 2013	No relevant intervention assessed
Jiang 2015	No eligible outcome assessed
Jiang 2016	No eligible outcome assessed
Jin 2013	Ineligible study design applied
Karanasiou 2014	No eligible outcome assessed
Kendall 2011	Ineligible study design applied
Keuken 2012	Ineligible study design applied
Kim 2015	No relevant intervention assessed
Kobza 2016	Ineligible study design applied
Kong 2010	Ineligible study design applied
Koutrakis 2005	Ineligible study design applied
Kowalska 2008	Ineligible study design applied
Kravchenko 2014	Ineligible study design applied
Krawack 1993	No eligible outcome assessed
Kuwayama 2012	No eligible outcome assessed
Lacasana-Navarro 1999	Ineligible study design applied
Leem 2015	Ineligible study design applied
Li 2010	Ineligible study design applied
Li 2014	Ineligible study design applied

(*Continued*)

Li 2015	Ineligible study design applied
Li 2016c	Ineligible study design applied
Li 2016d	Full text not available; English abstract for Chinese publication on the natural gas for heating intervention taking place in Urumqi, China (also assessed by Song 2015)
Lin 2011b	Full text not available; English abstract for Chinese publication
Lin 2016	Ineligible study design applied
Liu 2015	Ineligible study design applied No relevant intervention assessed
Lomas 2016	No eligible outcome assessed
Lopez 2000	Full text not available; conference proceedings with no associated full publication
Luechinger 2014	Ineligible study design applied
Lyons 1993	Ineligible study design applied
Makonese 2015	No eligible outcome assessed
Mardones 2015	Ineligible study design applied
Masiol 2014	Ineligible study design applied
McNabola 2008	Ineligible study design applied
Melkonyan 2012	No relevant intervention assessed
Minoura 2006	Ineligible study design applied
Minoura 2009	Ineligible study design applied
Mott 2002	No eligible outcome assessed
Narain 2007	Ineligible study design applied
Nedellec 2010	Ineligible study design applied
Ngo 2015	No relevant intervention assessed
Noonan 2011b	Ineligible study design applied
Norra 2016	No eligible outcome assessed

(*Continued*)

Orozco 2015	Ineligible study design applied
Pan 2010	Ineligible study design applied
Parker 2008	No eligible outcome assessed
Pope 1996	Ineligible study design applied
Potoski 2013	Ineligible study design applied
Qiao 2015	Ineligible study design applied
Querol 2014	Ineligible study design applied
Rafaj 2014	No relevant intervention assessed
Raman 2008	No eligible outcome assessed
Rava 2011	Ineligible study design applied
Recycling 2007	Full text not available; non-quantitative report
Ringquist 1995	No relevant intervention assessed
Riveros 2009	Ineligible study design applied
Roberts 2013	Ineligible study design applied
Sabaliauskas 2012	Ineligible study design applied
Sajjadi 2008	Ineligible study design applied
Shannigrahi 2010	Ineligible study design applied
Shu 2014b	No eligible outcome assessed
Snowden 2015	No relevant intervention assessed
Song 2015a	Full text not available; conference publication on the intervention during the 2014 APEC convention in Beijing (also assessed in Guo 2016, among others)
Sun 2010	Full text not available; English abstract for Chinese publication on the intervention during the 2008 Beijing Olympics (also assessed in Hou 2010 and Li 2011, among others)
Sun 2014	Ineligible study design applied
Traversi 2008	No relevant intervention assessed

(*Continued*)

US EPA 2014	Ineligible study design applied
US EPA 2014a	Ineligible study design applied
US EPA 2015	Ineligible study design applied
van den Elshout 2014	Ineligible study design applied
Voorhees 2014	Ineligible study design applied
Wang 2009	Ineligible study design applied
Wang 2010	No eligible outcome assessed
Wang 2014a	Ineligible study design applied
Wang 2015	No eligible outcome assessed
Westerdahl 2011	Ineligible study design applied
Wong 1998	Ineligible study design applied
Wood 2015	Ineligible study design applied
Wu 2010	No eligible outcome assessed
Wu 2010a	Full text not available; English abstract for Chinese publication on the intervention during the Beijing Olympics
Wu 2011	No relevant intervention assessed
Xue 2014	Ineligible study design applied
Yang 2011	Ineligible study design applied
Yorifuji 2016b	No relevant intervention assessed
You 2014	Ineligible study design applied
Zhang 2005	No relevant intervention assessed
Zhang 2011	No eligible outcome assessed
Zhang 2014	No relevant intervention assessed
Zhang 2016b	No eligible outcome assessed

(*Continued)*

Zhao 2010	Ineligible study design applied
Zhao 2014	Ineligible study design applied
Zheng 2015	No eligible outcome assessed
Zhou 2010	No relevant intervention assessed

ADDITIONAL TABLES

Table 1. Summary of the PICO aspects of included studies

Study ID	Setting: country and location	Population description and sampling	Intervention sub-category	AQ outcomes	Health outcomes	Study design
Industrial sources						
Butler 2011/ Deschênes 2012/ Lin 2013	USA Mixed Urban/ Rural Areas of the Eastern and midwestern US	Population: Residents of the states of interest Sampling: Data on all deaths assessed	Cap and trade programme	O_3	All-cause mortality; Cardiovascular mortality; Respiratory mortality/ Respiratory hospital admissions	ITS-EPOC/ cITS-EPOC / ITS-EPOC
Pope 2007	USA Mixed Urban/ Rural Southwest US states: Nevada, Utah, New Mexico, Arizona	Population: Residents of the four SW states Sampling: Data on all hospital admissions assessed	Factory closure	NA	All-cause mortality	cITS-EPOC
Saaroni 2010	Israel Urban Tel Aviv metropolitan area	NA	Power plant conversion	PM_{10}; NOx; NO_2; NO SO_2	NA	CBA
Sajjadi 2011/ Sajjadi 2012	Australia Mixed Urban/ Rural Lower Hunter	Population: Residents in the Lower Hunter region hospital	Factory closure	PM_{10}; $PM_{2.5}$; NO_2; SO_2	Respiratory disease hospital admissions; Asthma hospital	cITS-EPOC [AQ]/ ITS-EPOC [health]

Table 1. Summary of the PICO aspects of included studies *(Continued)*

	region of New South Wales	catchment area Sampling: Data on all hospital admissions assessed All ages: respiratory disease; 0 to 14 yr: asthma 65+ yr: COPD			admissions; COPD hospital admissions	
Tanaka 2015	China Urban Several cities spread across China	Population: Infants up to 1 year old from included prefectures Sampling: Data on all infant deaths assessed	Required industry requirements	NA	All-cause mortality	CBA-EPOC
Residential sources						
Allen 2009	Canada Rural Smithers and Telkwa, communities in British Columbia	NA	Stove exchange	$PM_{2.5}$	NA	CBA
Aung 2016	India Rural Village in Karnataka, southern India	NA	Stove exchange	$PM_{2.5}$; BC	NA	CBA
Dockery 2013a/ Clancy 2002	Ireland Urban Dublin	Population: Residents Dublin and the Midland and Coastal control counties Sampling: Data on all deaths assessed	Coal ban	NA	All-cause mortality; Cardiovascular mortality; Respiratory mortality	cITS-EPOC/ cITS-EPOC

Table 1. Summary of the PICO aspects of included studies *(Continued)*

Dockery 2013b	Ireland Mixed Urban/Rural Cork City and County	Population: Residents Cork City and County and the Midland and Coastal control counties Sampling: Data on all deaths and hospital admissions assessed	Coal ban	NA	All-cause mortality; Cardiovascular mortality; Respiratory mortality; Cardiovascular hospital admissions; Respiratory hospital admission	cITS-EPOC; ITS-EPOC [hospital admissions]
Dockery 2013c	Ireland Mixed Urban/Rural Limerick City and County, Louth, Wexford and Wicklow	Population: Residents Limerick City and County, Louth, Wexford and Wicklow and the Midland and Coastal control counties Sampling: Data on all deaths assessed	Coal ban	NA	All-cause mortality; Cardiovascular mortality; Respiratory mortality; Cardiovascular hospital admissions; Respiratory hospital admission	cITS-EPOC; ITS-EPOC [hospital admissions]
Johnston 2013	Australia Urban Launceston, Tasmania city-wide	Population: Residents of Launceston city Sampling: Data on all deaths assessed	Stove exchange	NA	All-cause mortality; Cardiovascular mortality; Respiratory mortality	cITS-EPOC
Yap 2015	USA Mixed urban/rural California's San Joaquin Valley Air Basin	Population: Adult residents of the San Joaquin Valley Air Basin Sampling: Data on all hospitalizations assessed	Wood burning ban	$PM_{2.5}$; Coarse particles	Cardiovascular hospital admissions; Respiratory hospital admissions	ITS-EPOC
Vehicular sources						
Atkinson 2009	UK Urban Lon-	NA	Charging scheme	PM_{10}; NOx; NO_2;	NA	CBA

Table 1. **Summary of the PICO aspects of included studies** *(Continued)*

	don metropolitan area			NO CO; O_3		
Bel 2013a	Spain Urban Barcelona metropolitan area	NA	Speed limit change	PM_{10}; NOx	NA	cITS-EPOC
Bel 2013b	Spain Urban Barcelona metropolitan area	NA	Speed limit change	PM_{10}; NOx	NA	ITS-EPOC
Boogaard 2012	The Netherlands Urban City centres of Amsterdam, the Hague, Den Bosch, Tilburg, Utrecht	NA	Low emission zone	PM_{10}; $PM_{2.5}$; NOx; NO_2; Soot	NA	CBA-EPOC
Burr 2004	UK Urban Small town in northern Wales	Population: Residents and workers both in the intervention and a control street Sampling: Not specified	Infrastructure changes	PM_{10}; $PM_{2.5}$	Respiratory symptoms; Lung function	CBA
Carrillo 2016	Ecuador Urban Quito metropolitan area	NA	Even-odd restriction	CO	NA	CBA-EPOC
Cowie 2012	Australia Urban Local, primarily residential area of Sydney	NA	Tunnel construction; Road restructuring	PM_{10}; $PM_{2.5}$; NOx; NO_2	NA	cITS-EPOC
Davis 2008/ Gallego 2013a	Mexico Urban Mexico City metropolitan area	NA	Even-odd restriction	NOx; NO_2; O_3; SO_2; CO	NA	ITS-EPOC/ ITS-EPOC

Table 1. Summary of the PICO aspects of included studies *(Continued)*

Dijkema 2008	The Netherlands Urban Amsterdam metropolitan area	NA	Speed limit change	PM_{10}; BS; NOx	NA	CBA
Dolislager 1997	USA Urban Four metropolitan areas in California	NA	Fuel requirements	CO	NA	ITS-EPOC
El-Zein 2007	Lebanon Urban Beirut city-wide	Population: Children in Beirut under 17 years Sampling: All hospital admissions from accredited hospitals assessed	Vehicle ban	NA	Respiratory hospital admissions	ITS-EPOC
Gallego 2013b/ Gramsch 2013	Chile Urban Santiago metropolitan area	NA	Public transport restructuring	CO; BC	NA	ITS-EPOC/ CBA
Hasunuma 2014	Japan Mixed Urban/ Rural Areas spread across Japan	Population: Children 3 years old living in the 28 survey areas Sampling: Not specified	Required vehicle standards	NO_2	Respiratory symptoms	CBA-EPOC
Kim 2011	South Korea Urban Several cities spread across South Korea	NA	Clean fuel use	PM_{10}; NO_2	NA	CBA-EPOC
Morfeld 2013/ Fensterer 2014	Germany Urban Munich city centre	NA	Low emission zone	PM_{10}	NA	CBA-EPOC/ CBA

Table 1. Summary of the PICO aspects of included studies *(Continued)*

Morfeld 2014	Germany Urban 17 German cities	NA	Low emission zone	NOx; NO_2; NO	NA	CBA-EPOC
Peel 2010/ Friedman 2001	USA Urban Atlanta metropolitan area	Population: Residents of Atlanta and control areas Sampling: Data on all emergency department visits assessed	Comprehensive traffic reduction strategy	NOx; NO_2; O_3; SO_2; CO	Asthma emergency department (ED) visits; Pneumonia ED visits; COPD ED visits; CVD ED visits	cITS-EPOC [health] CBA-EPOC [AQ]/ cITS-EPOC [health] CBA-EPOC [AQ]
Ruprecht 2009	Italy Urban Milan city centre	NA	Charging scheme	PM_{10}	NA	CBA
Titos 2015a	Slovenia Urban Ljubljana metropolitan area	NA	Road restructuring	BC	NA	CBA
Titos 2015b	Spain Urban Granada metropolitan area	NA	Public transport restructuring	BC	NA	CBA
Viard 2015	China Urban Beijing metropolitan area	NA	Even-odd restriction	PM_{10}	NA	ITS-EPOC
Yorifuji 2016/ Yorifuji 2011	Japan Urban Tokyo metropolitan area	Population: Residents of Toyko Sampling: Data on all deaths assessed	Required vehicle standards	$PM_{2.5}$; NO_2	All-cause mortality; Cardiovascular mortality; Respiratory mortality; Cerebrovascular mortality; Mortality from other causes	cITS-EPOC/ ITS-EPOC
Multiple sources						

Table 1. Summary of the PICO aspects of included studies *(Continued)*

Giovanis 2015	USA Mixed Urban/ Rural Charlotte, North Carolina and surrounding area	NA	Repeated coordinated measures	O_3	All-cause mortality	CBA-EPOC
Li 2011	China Urban Beijing metropolitan area	Population: Adult residents of Beijing admitted to hospitals for asthma events Sampling: Data on all admissions assessed	Even-odd restriction; Vehicle restriction; Power plant restriction	NA	Asthma	ITS-EPOC
Mullins 2014	Chile Urban Santiago metropolitan area	NA	Repeated coordinated measures	PM_{10}	NA	ITS-EPOC
Zigler 2016	USA Mixed Urban/ Rural Western USA	NA	Tailored selection of measures	PM_{10}	All-cause mortality; Cardiovascular hospital admissions; Respiratory hospital admissions	CBA-EPOC

Table 2. Description of study design and analysis methods for included main studies assessing health outcomes

Study ID	Description of intervention and control sites	Outcomes	Temporal aspects	Analysis
cITS-EPOC studies				
Deschênes 2012	**Intervention** 20 states located in the US Midwest and Northeast **Control** 22 states located in the US Southeast, Midwest, and West	All-cause, cardiovascular and respiratory mortality	Quarterly data analyzed; 1997 to 2007	Triple difference-indifferences estimated using a non-specified regression technique (comparing pre-vs. post-intervention, intervention vs. control site, summer-operating seasons vs. winter);

				Underlying time trend and seasonality accounted for the inclusion of county-by-year, season-by-year, county-by-season fixed effects; Underlying time trend and seasonality accounted for by the inclusion of the county-by-year; season-by-year; county-by-season fixed effects; Various temporal and geographical autocorrelation schemes assessed through sensitivity analyses
Dockery 2013a	**Intervention** Dublin **Control** 12 Midlands counties not affected by the bans	All-cause, cardiovascular and respiratory mortality	Yearly data; 1981 to 2004	Time-series Poisson regression; Underlying time trend accounted for through inclusion of Loess smooth term for mortality in the reference Coastal counties; Autocorrelation considered by authors to account for autocorrelation; Controlled through similar analyses performed for Midland counties not affected by the ban; Adjusted for influenza epidemics, weekly mean temperature.
Dockery 2013b	**Intervention** Cork City and County **Control** 12 Midlands counties not affected by the bans	All-cause, cardiovascular and respiratory mortality; Cardiovascular and respiratory hospitalization	Yearly data; 1981 to 2004	Time-series Poisson regression; Underlying time trend accounted for through inclusion of Loess smooth term for mortality in the reference Coastal counties; Autocorrelation considered by authors to account for autocorrelation; Controlled through similar analyses performed for Midland counties not affected by the ban;

Table 2. Description of study design and analysis methods for included main studies assessing health outcomes *(Continued)*

				Adjusted for influenza epidemics, weekly mean temperature.
Dockery 2013c	**Intervention** Limerick City and County, Louth, Wexford, Wicklow **Control** 12 Midlands counties not affected by the bans	All-cause, cardiovascular and respiratory mortality; Cardiovascular and respiratory hospitalization	Yearly data; 1981 to 2004	Time-series Poisson regression; Underlying time trend accounted for through inclusion of Loess smooth term for mortality in the reference Coastal counties; Autocorrelation considered by authors to account for autocorrelation; Controlled through similar analyses performed for Midland counties not affected by the ban; Adjusted for influenza epidemics, weekly mean temperature.
Johnston 2013	**Intervention** City of Launceston **Control** City of Hobart	All-cause, cardiovascular and respiratory mortality	Yearly data; January 1994 to November 2007	Time-series Poisson regression; Seasonality accounted for by the inclusion of mortality in the rest of Tasmania; Controlled through identical analysis conducted for control city; Adjusted for meteorology, respiratory epidemics, and secular trends in daily mortality in the rest of Tasmania
Pope 2007	**Intervention** 4 southwest US states (Arizona, New Mexico, Nevada, Utah) where large effect due to copper smelter strike was expected **Control** States where little or no effect due to the copper smelter strike was expected - 7 bordering states; 6 neighboring states; 46 non-southwest states	All-cause mortality	Yearly data; 1969 to 1974	Poisson regression; Underlying time trend accounted for using spline smoother; Seasonality accounted for by the inclusion of nationwide influenza and pneumonia counts; Controlled through inclusion of mortality counts of bordering states, neighbouring states and non-southwest states

Table 2. Description of study design and analysis methods for included main studies assessing health outcomes *(Continued)*

Sajjadi 2011	**Intervention** The city of Newcastle located in the Lower Hunter area of New South Wales **Control** The city of Port Stephens lo- cated in the Lower Hunter area of New South Wales; furthest region in the area from the intervention	Respiratory, asthma and COPD hospitalizations	Monthly data; January 1996 to June 2004	Mixed model regression; Underlying time trend accounted for using month-of-year dummies; Autoregressive effects accounted for through compound symmetry covariance structure; Controlled through parallel analysis at intervention and control sites;
Tanaka 2015	**Intervention** 61 Chinese prefectures designated as part of the Two Control Zone **Control** 84 Chinese prefectures not designated as part of the Two Control Zone	All-cause mortality (infant)	Yearly data; 1991 to 2000	Difference-in-differences regression; Underlying time trend accounted for by year fixed effects; Adjusted for city fixed effects; Adjusts for birth, parental and city characteristics
Yorifuji 2016	**Intervention** City of Tokyo (23 wards) **Control** City of Osaka	All-cause, cardiovascular, ischemic heart disease, cerebrovascular, pulmonary disease and lung cancer mortality	Daily data; 2000 to 2012	Time-series Poisson regression; Con- trolled through weighting of mortality rates in Tokyo with mortality rates in the reference city Osaka; Adjusted for relevant meteorological variables, influenza deaths, day of the week, public holiday
ITS-EPOC studies				
Clancy 2002	**Intervention** City of Dublin **Control** NA	All-cause, cardiovascular and respiratory mortality	Yearly data; Sep. 1984 to August 1996	Time-series Poisson regression; Unclear whether underlying time trend was accounted for; Autoregressive effects not

				accounted for; Adjusted for temperature, humidity, respiratory epidemics, death rates in the rest of Ireland
El-Zein 2007	**Intervention** City of Beirut **Control** NA	Respiratory hospitalizations	Monthly data October 2000 to February 2004	Time-series Poisson regression; Underlying time trend not considered; Autoregressive effects not considered; Adjusted for temperature, humidity and rainfall
Friedman 2001	**Intervention** City of Atlanta (5 counties making up metropolitan area) **Control** NA	Asthma hospitalization (child emergency department visits)	Daily data; June to September 1996	Time-series Poisson regression; Underlying time trend not considered (authors state this is due to short study period); Autoregressive effect of 1 for daily correlation Adjusted for minimum temperature and day of the week
Li 2011	**Intervention** City of Beijing **Control** NA	Asthma hospitalizations (outpatient visits)	Daily data 1 June to 17 September 2008	Time-series Poisson regression; Underlying time trend not considered (authors state this is due to short study period); Autoregressive effect of 1 for daily correlation Adjusted for relevant meteorological variables and day of the week

Table 2. Description of study design and analysis methods for included main studies assessing health outcomes *(Continued)*

Mullins 2014	**Intervention** City of Santiago, Chile **Control** NA	All-cause and respiratory mortality (age > 64)	Daily data; 1989 to 2008	Difference-in difference regression technique, comparing changes before to after an Episode (after the intervention was introduced) to changes before to after another (similar) day (before the intervention was introduced); Propensity score matching for choosing appropriate pre-intervention comparison days; No underlying time trends assessed; Seasonality accounted for through month-level fixed effects; Adjusted for relevant meteorological variables
Peel 2010	**Intervention** City of Atlanta (5 counties making up metropolitan area) **Control** NA	Asthma, COPD, CVD, pneumonia hospitalizations	Daily data; 21 June to 1 September 1995 to 2004	Time-series Poisson regression; Underlying time trend accounted for through inclusion of day of the summer variable; Autoregressive effects explored in sensitivity analyses through GEE analysis; Adjusted for relevant meteorological variables, day of the week
Yap 2015	**Intervention** California's San Joaquin Valley Air Basin **Control** NA	Cardiovascular and respiratory hospitalizations	Daily data; November to February, 2000 to 2006	Multivariate Poisson regression; Unclear to what extent underlying the time trend was considered; Unclear to what extent autoregressive effects were considered;

Table 2. Description of study design and analysis methods for included main studies assessing health outcomes *(Continued)*

				Adjusted for day of the week, no-burn days, and percentage of poverty
Yorifuji 2011	**Intervention** City of Tokyo (23 wards) **Control** NA	All-cause, cardiovascular and respiratory mortality	Daily data April 2003 to December 2008	Time-series Poisson regression; Underlying time trend accounted for through inclusion of a natural spline smoothing function; Adjusted for relevant meteorological variables, influenza deaths, day of the week, public holiday
CBA-EPOC studies				
Hasunuma 2014	**Intervention** 16 regions of Japan designated as PM-law enforcement areas **Control** 12 regions of Japan designated as non-PM-law enforcement areas	Respiratory symptoms	Yearly data; 1997 to 2009	t-tests comparing pre- and post-intervention averages conducted for intervention and control sites
CBA studies				
Burr 2004	**Intervention** Urban area in Northern Wales with heavy traffic congestion **Control** NA	Respiratory symptoms and lung function	Single pre-, post-intervention observations; July 1996 to November 1997; July 1998 to November 1999	Comparison of pre- and post-intervention concentrations calculated for intervention and control areas separately

Table 3. Description of study design and analysis methods for included main studies assessing air quality outcomes

Study ID	Description of intervention sites	Out-comes	Time points analyzed	Analysis methods
cITS-EPOC studies				
Bel 2013a	**Intervention** 15 regulatory monitors in the Barcelona city centre, within the 80 km/h speed limit area;	PM_{10}; NOx	Daily data analyzed; 2006 to 2010	Difference-in-differences regression; Time-specific fixed effects control for municipal

	Control 15 regulatory monitors in the Barcelona city centre, outside of the 80 km/h speed limit area Note: unclear how many of the 15 were intervention and control sites			trends; Municipal-specific fixed effects control for time-invariant, non-observed variables; Adjusted for relevant meteorological variables
Cowie 2012	**Intervention** 4 study urban background monitors located in the area surrounding the Lane Cove Tunnel **Control** 3 regulatory suburban background monitors located in the suburban area surrounding Sydney, Australia	PM_{10}; $PM_{2.5}$; NOx; NO_2	Daily data analyzed; Mar 2006 to Mar 2009	Step-wise regression approach comparing changes in concentrations 1 and 2 years after the intervention; Autocorrelation accounted for through an autoregressive error model using the Yule - Walker method; Adjusted for relevant meteorological variables Controlled through adjustment for regional background air quality
Deschênes 2012	**Intervention** Regulatory monitors located in 20 states in the US midwest and northeast; **Control** Regulatory monitors located in 22 states in the US southeast, midwest and west; Note: Total number of counties for which data is available ranges from 39-298 depending on pollutant	PM_{10}; $PM_{2.5}$; NO_2; O_3; SO_2; CO	Quarterly data analyzed: 1997 to 2007	Triple difference-in-differences regression (comparing pre- vs. post-intervention; intervention vs. control site; summer - operating season vs. winter); Underlying time trend and seasonality accounted for by the inclusion of county-by-year; season-by-year; county-by-season fixed effects; Various temporal and geographical autocorrelation schemes assessed through sensitivity analyses
ITS-EPOC studies				
Bel 2013b	**Intervention** 15 regulatory monitors in the Barcelona city centre, within the 80 km/h speed limit area;	PM_{10}; NOx	Daily data analyzed; 2006 to 2010	Difference-in-differences regression; Time-specific fixed effects control for municipal trends;

				Municipal-specific fixed effects control for time-invariant, non-observed variables; Adjusted for relevant meteorological variables
Butler 2011	**Intervention** 42 regulatory regional background monitors located in rural areas of the northeastern, mid-Atlantic, southeastern, and midwestern US **Control** NA	O_3	Daily 8-hour max data analyzed; 2000 to 2002; 2006 to 2008	Autoregressive Integrated Moving Average (ARIMA) models comparing changes in trends before and after the intervention
Davis 2008	**Intervention** Regulatory monitors - between 5-15 monitors, depending on pollutant - located in the greater Mexico City area **Control** NA	NOx; NO_2; O_3; SO_2; CO	Hourly data analyzed; 1986 to 1993	Ordinary least squares (OLS) regression comparing concentrations pre- and post-intervention; No underlying time trends assessed; Seasonality accounted for using month-dummies; Adjusted for relevant meteorological variables
Dolislager 1997	**Intervention** 16 regulatory monitors located in the state of California; Exact site characteristics not described **Control** NA	CO;	Peak traffic data analyzed 7:00 a.m. to 9:00a.m., 7:00 p.m. to 10:00 p.m.; 1985 to 1994	Regression-based prediction of post-intervention concentrations, based on pre-intervention measurements, compared to actual measured post-intervention concentrations; NOx included as a negative pollutant control, as it was not expected that the Oxyfuels Program being evaluated would have affected its concentrations
Gallego 2013a	**Intervention** Regulatory monitors located in the greater Mexico City area; Exact site characteristics	CO	Peak 2-hour data analyzed; 1987 to 1991	Regression-based comparison of pre- and post-intervention concentrations; Accounted for pre-intervention trend using linear

	not described **Control** NA			trend; Seasonality addressed through inclusion of hour of the day, day of the week and month of the year fixed effects; Adjusted for relevant meteorological variables; Adjusted for background CO and SO2 pollution
Gallego 2013b	**Intervention** Regulatory monitors located in the greater Santiago area; Exact site characteristics not described **Control** NA	CO	Peak 2-hour analyzed; 2005 to 2009	Regression-based comparison of pre- and post-intervention concentrations; Accounted for pre-intervention trend using linear trend; Seasonality addressed through inclusion of hour of the day, day of the week and month of the year fixed effects; Adjusted for relevant meteorological variables; Adjusted for background CO and SO_2 pollution
Mullins 2014	**Intervention** 3 regulatory urban background monitors located in the greater Santiago area **Control** NA	PM_{10}	Daily data analyzed; 1989 to 2008	Difference-in difference regression technique, comparing changes before to after an Episode (after the intervention was introduced) to changes before to after another (similar) day (before the intervention was introduced); Propensity score matching for choosing appropriate pre-intervention comparison days; No underlying time trends assessed; Seasonality accounted for through month-level fixed effects; Adjusted for relevant meteorological variables

Table 3. Description of study design and analysis methods for included main studies assessing air quality outcomes *(Continued)*

Sajjadi 2012	**Intervention** 1 regulatory regional background monitors located in the Lower Hunter area of New South Wales **Control** NA	PM_{10}; $PM_{2.5}$; NO_2; SO_2	Monthly data analyzed; Jan 1996 to June 1999; Jan 2001 to June 2004	Mixed model regression; Underlying time trend accounted for using month-of-year dummies Autoregressive effects accounted for through compound symmetry covariance structure
Viard 2015	**Intervention** 27 regulatory monitors located in the greater Beijing area; Exact site characteristics not described **Control** NA	PM_{10}	Daily data analyzed; 2007 to 2009	Regression discontinuity technique (analogous to interrupted time series in this case); Underlying time trend accounted for through week-of-year dummies; Adjusted for relevant meteorological variables, weekends and holidays
Yap 2015	**Intervention** Regulatory monitors located in California's San Joaquin Valley Air Basin; Exact site characteristics not described **Control** NA	$PM_{2.5}$; Coarse particles	Daily data analyzed; Nov to Feb, 2000 to 2006	Generalized linear mixed model regression; Underlying time trend assessed through the inclusion of year-dummies; Seasonality not considered, as only wintertime was analyzed; Adjusted for relevant meteorological variables and for regulatory "no burn" days
CBA-EPOC studies				
Boogaard 2012	**Intervention** 13 study monitors - 8 streetside and 5 urban background - located in five Dutch cities **Control** 4 study suburban background monitors located in Dutch suburban areas (one near each intervention city)	PM_{10}; $PM_{2.5}$; NO_x; NO_2; Soot	Weekly data analyzed; July to Dec 2008; July to Dec 2010	t-tests comparing pre- and post-intervention averages conducted for each site; t-tests comparing changes at urban street and urban background sites with changes at the matching suburban locations conducted

Table 3. Description of study design and analysis methods for included main studies assessing air quality outcomes *(Continued)*

Carrillo 2016	**Intervention** 3 regulatory streetside monitors located in the Quito city centre **Control** 2 regulatory streetside monitors located in the Quito city centre	CO	Peak traffic data analyzed: 7:00 a.m. to 9:30 a.m.; 4:00 p.m. to 7:30 p.m.; 2008 to 2012	Triple difference-in-differences ordinary least squares regression (comparing pre- vs. post-intervention; intervention vs. control site; peak vs. non-peak hours); Serial correlation as well as contemporaneous correlation in pollution across stations are accounted for by clustering (robust) standard errors at the quarter level
Hasunuma 2014	**Intervention** Regulatory monitors located in 16 regions of Japan designated as PM-law enforcement areas; **Control** Regulatory monitors located in 12 regions of Japan designated as non-PM-law enforcement areas Note: Total number of monitors was 106 (unclear how many of these were in intervention regions)	NO_2	Yearly data analyzed; 1996 to 2000; 2006 to 2009	t-tests comparing pre- and post-intervention averages conducted for intervention and control sites
Giovanis 2015	**Intervention** 4 regulatory regional background monitors located in counties participating in the intervention of interest **Control** 7 regulatory regional background monitors located in counties not participating in the intervention of interest	O_3	Monthly data analyzed; 2000 to 2010	Difference-in-differences regression; Underlying time trend and seasonality accounted for through monthly dummies; Models adjusted for a range of relevant covariates
Kim 2011	**Intervention** 16 regulatory streetside monitors located in 7 major and minor cities in South Korea **Control**	PM_{10}	Monthly data analyzed; 1998 to 2008	t-tests comparing pre- and post-intervention averages conducted for intervention and control sites

	4 regulatory regional background monitors located in non-urban regions of South Korea			
Morfeld 2013	**Intervention** 5 regulatory monitoring sites located in the Munich city centre **Control** 1 regulatory monitoring site located in the greater Munich area	PM_{10}	30-minute data analyzed; Oct 2007 to Jan 2008; Oct 2008 to Jan 2009	Linear regression of pre-post intervention differences at intervention sites on pre-post intervention differences at the control site; Adjusted for relevant meteorological variables
Morfeld 2014	**Intervention** 53 regulatory monitoring sites located in areas of 17 German cities within the LEZs **Control** 55 regulatory monitoring sites located in areas of 17 German cities outside of the LEZs	NOx; NO_2; NO	30-minute data analyzed; 2005 to 2009	Linear regression of pre-post intervention differences at intervention sites on pre-post intervention differences at the control site; Adjusted for baseline concentrations at intervention sites, baseline concentrations at control sites, changes at reference stations (proxy for meteorological changes)
Zigler 2016	**Intervention** 219 regulatory monitors located in areas of the western US designated "Non-attainment" **Control** 276 regulatory monitors located in areas of the western US designated "Attainment"	PM_{10}	Yearly data analyzed; 1990; 1999 to 2001	Propensity score matching to create more appropriately comparable subsets of intervention and control monitors; Pruning of monitors based on outlying propensity scores; Regression-based comparison of pre- and post-intervention concentrations; Adjusted using propensity score matching
CBA studies				
Allen 2009	**Intervention** Study monitors at 17 study homes in Smithers and Telkwa, British Columbia	$PM_{2.5}$	Frequency of data analyzed not specified; November 2007 to April 2008	t-tests assessing changes in concentrations pre- and post-intervention separately at intervention and control sites

	Control 2 study regional background monitors in Smithers and Telkwa, British Columbia			
Atkinson 2009	**Intervention** 2 regulatory monitors - 1 streetside, 1 urban background - located within the charging zone **Control** 19 regulatory monitors - 14 streetside, 5 urban background - located in greater London, but at least 8km from the charging zone	PM_{10}; NOx; NO_2; NO; CO; O_3	Daily data analyzed; 2001 to 2005	Calculation of geometric means for pre- and post-intervention at each site
Aung 2016	**Intervention** 1 study monitor located in the centre of the Southern Indian study village **Control** 1 study monitor located 1km in the predominant upwind direction of the village	$PM_{2.5}$; BC	Daily data analyzed; Sep 2011; July to Aug 2012	Wilcoxon rank-sum test for unpaired samples comparing concentrations between upwind and village centre sites for pre- and post-intervention time periods;
Burr 2004	**Intervention** 1 study streetside monitor located in the North Wales city affected by heavy traffic congestion **Control** 1 study streetside monitor located in the North Wales city not affected by heavy traffic congestion	PM_{10}; $PM_{2.5}$	Frequency of data analyzed not specified; July 1996 to Nov 1997; July 1998 to Nov 1999	Calculation of means for pre- and post-intervention periods at the intervention and control sites, as well as percent change at each site;
Dijkema 2008	**Intervention** 1 regulatory streetside monitor located on a section of ring highway in Amsterdam where the intervention was implemented **Control** 1 regulatory	PM_{10}; BS; NOx	Daily data analyzed; Nov 2004 to Nov 2006	Multivariate linear regression comparing pre- and post-intervention concentrations; Adjusted for concentrations at urban background sites to obtain "traffic contribution"; Adjusted for

	streetside monitor located on a section of ring highway in Amsterdam where no effect due to the intervention was expected			traffic flow, traffic congestion and wind direction
Fensterer 2014	**Intervention** 2 regulatory monitors - 1 streetside, 1 urban background - located in the Munich city centre **Control** 1 regulatory regional background monitor located in the Greater Munich area	PM_{10}	Hourly data analyzed; Feb 2006 to Jan 2008; Oct 2008 to Sep 2010	A semiparametric regression model comparing pre- and post-intervention concentrations at intervention sites; Controlled through adjustment for concentrations at the control site; Autocorrelation accounted for through the inclusion of first-order autoregressive errors; Adjusted for wind direction, season, time throughout a week, and public holidays
Gramsch 2013	**Intervention** 3 streetside monitors located in the Santiago city centre where changes due to the intervention were made **Control** 1 study streetside monitor located in the Santiago city centre where no changes due to the intervention were made	BC	Hourly data analyzed; June to July 2005; June to July 2007	Comparison of concentrations pre- and post-intervention at each site using the Wilcoxon rank-sum test; Multiple linear regression; Adjusted for several relevant meteorological variables
Peel 2010	**Intervention** 5 regulatory monitors located in 5 counties of Metropolitan Atlanta **Control** Regulatory monitors located in counties of Metropolitan Atlanta outside of the 5 central counties; Other areas of Georgia; Metropolitan areas in other parts of the	NO_x; NO_2; O_3; SO_2; CO	Daily data analyzed; 21 June to 1 September, 1995 to 2004	Regression-based comparison of pre- and post-intervention concentrations separately for intervention and control sites

	US southeast; Note: Number of monitors varies per pollutant between 2-20			
Ruprecht 2009	**Intervention** 1 regulatory monitor in the Milan city centre within the Ecopass zone **Control** 1 regulatory monitor in the Milan city centre outside of the Ecopass zone	PM_{10}	Daily data analyzed; November 2007 to February 2008	t-tests comparing changes in concentrations between the intervention and control sites both pre- and post-intervention
Saaroni 2010	**Intervention** 1 study urban background monitor located in a residential suburban area of Tel Aviv downwind of power plant **Control** 2 study urban background monitors located in the greater Tel Aviv area upwind of power plant	PM_{10}	Monthly data analyzed; July to October 2004; July to October 2006	t-tests comparing changes in concentrations pre- and post-intervention at the intervention site only; Concentrations before and after intervention at intervention and control sites compared graphically
Titos 2015a	**Intervention** 1 study streetside monitor located in the Ljubljana city centre **Control** 2 study monitors - 1 streetside, 1 urban background - located in Ljubljana outside of the driving restriction zone	BC	Frequency of data analyzed not specified; August to October 2013	t-tests comparing changes in concentrations pre- and post-intervention separately for intervention and control sites
Titos 2015b	**Intervention** 2 study monitors - 1 streetside, 1 urban background - located in the Granada city centre **Control** 1 study urban background monitor located in Granada outside the immediate city centre	BC	30-minute data analyzed; June to July 2014	t-tests comparing changes in concentrations pre- and post-intervention separately for intervention and control sites

A P P E N D I C E S

Appendix 1. Search strategy, as adapted for each database

CENTRAL

1. ((air NEAR/2 (pollut* OR quality OR ambient)) OR (atmospher* NEAR/2 pollut*) OR ("particulate matter" OR "ambient particulate" OR "ultrafine particulate*" OR "ultrafine particle*" or UFP) OR ("coarse particle*" OR "black smoke" or "black carbon" or "elemental carbon" OR "wood smoke")):ti,ab,kw
2. ((mortalit* OR death*) OR ((cardiovascular OR respiratory OR pulmonary OR lung) NEAR/3 (mortality OR death* OR fatal* OR "hospital admission*" OR event* OR disease OR outcome*)) OR (asthma OR pneumonia OR "lung cancer" OR "lung function") OR ((improv* OR reduc* OR lower* OR increas* OR adverse OR measure* OR outcome* OR effect* OR impact* OR concentration OR level* OR absor* OR exposure* OR exposed) NEAR/3 ("air pollution" OR "particulate matter" OR "ambient particulate" OR "coarse particule*" OR "black smoke" OR "black carbon" OR "elemental carbon"))) :ti,ab,kw
3. (((emission* OR air OR "particulate matter" OR "ambient particulate" OR "ultrafine particulate*" OR "ultrafine particle*" OR UFP) NEAR/4 (control* OR regulation* OR policy OR policies OR guideline OR intervention OR act OR directive* OR vehicle OR transport* OR traffic OR automobile* OR car* OR industr* OR fuel OR "emission filter*" OR cooking OR heating OR cookstove* OR stove* OR "power generat*" OR zone* OR Olympic OR residential OR "wood burning" OR mobile OR Low* OR reduc* OR improv* OR clean* OR congestion* OR "coal burning" OR ban OR bans)) OR ((improved or clean* or "low emission" or efficient*) NEAR/1 (cookstove* or stove or stoves or heater))) :ti,ab,kw
4. #1 AND #2 AND #3

MEDLINE & MEDLINE In-Process

1. exp Air Pollution/
2. exp Particulate Matter/
3. (Air adj2 (pollut* or quality or ambient)).ti,ab.
4. (atmospher* adj2 pollut*).ti,ab.
5. (Particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or ultrafine particulate* or ultrafine particle* or UFP).ti,ab.
6. (Coarse particle* or Soot or Black smoke or Black carbon or Elemental carbon or wood smoke).ti,ab.
7. ((Emission* or air or atmospher*) adj2 (anthropogenic or motor or vehicle or road or power generation or indust* or combustion or smelting or construction or demolition or burning or residential)).ti,ab.
8. or/1-7
9. exp Mortality/ or Cardiovascular Diseases/mo or Respiratory Tract Diseases/mo
10. (Mortalit* or Death*1).ti,ab.
11. (Cardiovascular adj3 (mortality or death* or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
12. (Respiratory adj3 (mortality or death or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
13. (Heart attack* or stroke or strokes).ti,ab.
14. (asthma or Pneumonia or lung cancer or Lung function* or lung disease* or pulmonary function* or pulmonary disease*).ti,ab.
15. (exp air pollution/sn, td or exp particulate matter/sn, td) and (Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level*).ti,ab.
16. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (air pollution or particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or coarse particule* or soot or black smoke or black carbon or elemental carbon or combustion)).ti,ab.
17. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (carbon monoxide or SO2 or sulphur dioxide or sulfur dioxide or NO2 or nitrogen dioxide or O3 or ozone or UFP or ultrafine particle*)).ti,ab.
18. or/9-17
19. exp air pollution/pc or exp particulate matter/pc
20. ((emission* or air or PM or PM1* or PM2* or PM10* or particulate matter or ambient particulate or ultrafine particulate* or ultrafine particle* or UFP or climate or green or smoke) adj8 (control* or regulation* or policy or policies or guideline or intervention

or act or directive* or vehicle or transport* or traffic or automobile* or car*1 or industr* or fuel or emission filter* or cooking or heating or cookstove* or stove* or power generat* or energy or zone* or Olympic or residential or wood burning or mobile or Low* or reduc* or improv* or clean* or congestion* or coal burning or ban or bans)).ti,ab.
21. air pollution/pc or smoke/pc
22. ((Improved or clean* or low emission or efficient*) adj1 (cookstove* or stove or stoves or heater)).ti,ab.
23. Wood burning regulation*.ti,ab.
24. or/19-23
25. 8 and 18 and 24
26. randomized controlled trial.pt.
27. controlled clinical trial.pt.
28. comparative study.pt.
29. intervention studies/
30. evaluation studies/
31. program evaluation/
32. random allocation/ or clinical trial/ or single-blind method/ or double-blind method/ or control groups/
33. (randomized or randomised or placebo or randomly or groups).ab.
34. trial.ti,ab.
35. (time adj series).ab,ti. or (interrupted* adj2 series).ti,ab.
36. quasi-experiment$.ab,ti.
37. (pre test or pretest or pre-intervention or post-intervention or posttest or post test).ab,ti.
38. (controlled before or "before and after stud$" or follow-up-assessment).ab,ti.
39. ((evaluat$ or intervention or interventional or treatment) and (control or controlled or study or program$ or comparison or "before and after" or comparative)).ab,ti.
40. ((intervention or interventional or process or program) adj8 (evaluat$ or effect$ or outcome$)).ab,ti.
41. (program or programme or secondary analys$).ti,ab.
42. ecological study.ti,ab.
43. (Case study or observational study or cohort or uncontrolled study or observational research).ti,ab. or exp Epidemiologic Studies/
44. or/26-43
45. exp animals/ not humans.sh.
46. 44 not 45
47. 25 and 46

Embase

1. exp Air Pollution/
2. exp Particulate Matter/
3. (Air adj2 (pollut* or quality or ambient)).ti,ab.
4. (atmospher* adj2 pollut*).ti,ab.
5. (Particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or ultrafine particulate* or ultrafine particle* or UFP).ti,ab.
6. (Coarse particle* or Soot or Black smoke or Black carbon or Elemental carbon or wood smoke).ti,ab.
7. ((Emission* or air or atmospher*) adj2 (anthropogenic or motor or vehicle or road or power generation or indust* or combustion or smelting or construction or demolition or burning or residential)).ti,ab.
8. or/1-7
9. exp Mortality/ or Cardiovascular Disease/et, pc, di, ep or Respiratory Tract Disease/et, pc, di, ep
10. (Mortalit* or Death*1).ti,ab.
11. (Cardiovascular adj3 (mortality or death* or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
12. (Respiratory adj3 (mortality or death or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
13. (Heart attack* or stroke or strokes).ti,ab.
14. (asthma or Pneumonia or lung cancer or Lung function* or lung disease*).ti,ab.
15. (pulmonary function* or pulmonary disease*).ti,ab.

16. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (air pollution or particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or coarse particule* or soot or black smoke or black carbon or elemental carbon or combustion)).ti,ab.
17. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (carbon monoxide or SO2 or sulphur dioxide or sulfur dioxide or NO2 or nitrogen dioxide or O3 or ozone or UFP or ultrafine particle*)).ti,ab.
18. or/9-17
19. exp air pollution/pc or exp particulate matter/pc)
20. ((emission* or air or PM or PM1* or PM2* or PM10* or particulate matter or ambient particulate or ultrafine particulate* or ultrafine particle* or UFP or climate or green or smoke) adj3 (control* or regulation* or policy or policies or guideline or intervention or act or directive* or vehicle or transport* or traffic or automobile* or car*1 or industr* or fuel or emission filter* or cooking or heating or cookstove* or stove* or power generat* or energy or zone* or Olympic or residential or wood burning or mobile or Low* or reduc* or improv* or clean* or congestion* or coal burning or ban or bans)).ti,ab.
21. air pollution/pc or smoke/pc
22. ((Improved or clean* or low emission or efficient*) adj1 (cookstove* or stove or stoves or heater)).ti,ab.
23. Wood burning regulation*.ti,ab.
24. or/19-23
25. 8 and 18 and 24
26. “randomized controlled trial (topic)”/
27. exp clinical trial/)
28. epidemiology/
29. intervention study/
30. evaluation/
31. randomization/
32. control group/
33. (randomized or randomised or placebo or randomly or groups).ab. (2299564)
34. trial.ti,ab.
35. (time adj series).ab,ti. or (interrupted* adj2 series).ti,ab.
36. quasi-experiment$.ab,ti.
37. (pre test or pretest or pre-intervention or post-intervention or posttest or post test).ab,ti.
38. (controlled before or “before and after stud$” or follow-up-assessment).ab,ti.
39. ((evaluat$ or intervention or interventional or treatment) and (control or controlled or study or program$ or comparison or “before and after” or comparative)).ab,ti.
40. ((intervention or interventional or process or program) adj8 (evaluat$ or effect$ or outcome$)).ab,ti.
41. (program or programme or secondary analys$).ti,ab.
42. ecological study.ti,ab.
43. (Case study or observational study or cohort or uncontrolled study or observational research).ti,ab.
44. or/26-43
45. exp animal/ not human/
46. 44 not 45
47. 25 and 46

PsycINFO

1. exp Pollution/
2. atmospheric conditions/
3. (Air adj2 (pollut* or quality or ambient)).ti,ab.
4. (atmospher* adj2 pollut*).ti,ab.
5. (Particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or ultrafine particulate* or ultrafine particle* or UFP).ti,ab.
6. (Coarse particle* or Soot or Black smoke or Black carbon or Elemental carbon or wood smoke).ti,ab.
7. ((Emission* or air or atmospher*) adj2 (anthropogenic or motor or vehicle or road or power generation or indust* or combustion or smelting or construction or demolition or burning or residential)).ti,ab.

8. or/1-7
9. “death and dying”/ or Cardiovascular Disorders/ or Respiratory Tract Disorders/
10. (Mortalit* or Death*1).ti,ab.
11. (Cardiovascular adj3 (mortality or death* or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
12. (Respiratory adj3 (mortality or death or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
13. (Heart attack* or stroke or strokes).ti,ab.
14. (asthma or Pneumonia or lung cancer or Lung function* or lung disease* or pulmonary function* or pulmonary disease*).ti,ab.
15. ((air pollution or atmospheric conditions) and (Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level*)).ti,ab.
16. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (air pollution or particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or coarse particule* or soot or black smoke or black carbon or elemental carbon or combustion)).ti,ab.
17. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (carbon monoxide or SO2 or sulphur dioxide or sulfur dioxide or NO2 or nitrogen dioxide or O3 or ozone or UFP or ultrafine particle*)).ti,ab.
18. or/9-17
19. (air pollution or particulate matter).ti,ab.
20. ((emission* or air or PM or PM1* or PM2* or PM10* or particulate matter or ambient particulate or ultrafine particulate* or ultrafine particle* or UFP or climate or green or smoke) adj8 (control* or regulation* or policy or policies or guideline or intervention or act or directive* or vehicle or transport* or traffic or automobile* or car*1 or industr* or fuel or emission filter* or cooking or heating or cookstove* or stove* or power generat* or energy or zone* or Olympic or residential or wood burning or mobile or Low* or reduc* or improv* or clean* or congestion* or coal burning or ban or bans)).ti,ab.
21. air pollution.ti,ab.
22. ((Improved or clean* or low emission or efficient*) adj1 (cookstove* or stove or stoves or heater)).ti,ab.
23. Wood burning regulation*.ti,ab.
24. or/19-23
25. 8 and 18 and 24
26. randomised controlled trial.ti,ab.
27. (comparative study or program evaluation or intervention study or evaluation study or random allocation or clinical trial or single-blind or double-blind or epidemiol$ stud$).ti,ab.
28. (randomized or randomised or placebo or randomly or groups).ab.
29. trial.ti,ab.
30. (time adj series).ab,ti. or (interrupted* adj2 series).ti,ab.
31. quasi-experiment$.ab,ti.
32. (pre test or pretest or pre-intervention or post-intervention or posttest or post test).ab,ti.
33. (controlled before or “before and after stud$” or follow-up-assessment).ab,ti.
34. ((evaluat$ or intervention or interventional or treatment) and (control or controlled or study or program$ or comparison or “before and after” or comparative)).ab,ti.
35. ((intervention or interventional or process or program) adj8 (evaluat$ or effect$ or outcome$)).ab,ti.
36. (program or programme or secondary analys$).ti,ab.
37. ecological study.ti,ab. (
38. (Case study or observational study or cohort or uncontrolled study or observational research).ti,ab.
39. or/26-38
40. exp animals/ not humans/
41. (25 and 39) not 40

Scopus

1. (TITLE-ABS-KEY(air w/2 ambient OR “air pollut*” OR “air quality”) OR(“particulate matter” OR “ambient particulate” OR “ultrafine particule*” OR “ultrafine particle*” OR UFP OR “coarse particle”) OR (“black smoke” OR “black carbon” OR “elemental carbon”))
2. (TITLE-ABS-KEY((mortalit* OR death*) OR ((cardiovascular OR respiratory) w/1 (mortality OR death OR fatal* OR “hospital admission*” OR event* OR disease OR outcome*)) OR (“heart attack” OR stroke OR strokes) OR (asthma OR pneumonia OR “lung

cancer" OR "lung function*" OR "lung disease*" OR "pulmonary function*" OR "pulmonary disease*") OR ((improv* OR reduc* OR lower* OR increas* OR adverse OR measure* OR outcome* OR effect* OR impact* OR concentration OR level* OR absor* OR exposure* OR exposed) w/2 ("air pollution" OR "particulate matter" OR "ambient particulate" OR "coarse particle*" OR "black smoke" OR "black carbon" OR "elemental carbon" OR UFP OR "ultrafine particle*"))))
3. (TITLE-ABS-KEY(((air OR "particulate matter" OR "ambient particulate" OR "ultrafine particulate*" OR "ultrafine particle*" OR UFP OR "coarse particle*" OR "black smoke" OR "black carbon" OR "elemental carbon") w/4 (control* OR regulation* OR policy OR policies OR guideline* OR intervention* OR act or directive* OR vehicle OR transport* OR traffic OR automobile* OR car* OR industr* OR "emission filter" OR cooking OR heating OR cookstove* OR stove* OR zone* OR olympic OR residential OR "wood burning" OR mobile OR low* OR reduc* OR improv* OR clean* OR congestion* OR "coal burning" OR ban OR bans)) OR ((improved OR clean* OR "low emission" OR efficient*) w/1 (cookstove* OR stove* OR stove OR stoves OR heater))))
4. (TITLE-ABS-KEY((randomized OR randomised OR placebo OR ramdomly OR groups) OR trial OR ("time series" OR interrupted w/2 series) OR "quasi-experiment" OR ("pre test" OR pretest OR "pre-intervention" OR "post-intervention" OR posttest OR "post test") OR ("controlled before" OR "before and after stud*" OR "follow-up-assessment") OR ((evaluat* OR intervention OR interventional OR treatment) AND (control OR controlled OR study OR program* OR comparison OR "before and after" OR comparative)) OR ((intervention OR interventional OR process OR program) w/8 (evaluat* OR effect* OR outcome*)) OR (program OR programme OR "secondary analys*") OR "ecological study" OR ("case study" OR "observational study" OR cohort OR "uncontrolled study" OR "observational research")))
5. 1 AND 2 AND 3 AND 4

Science Citation Index and Social Science Citation Index

1. TS = ((air NEAR/2 (pollut* OR quality OR ambient)) OR (atmospher* NEAR/2 pollut*) OR ("particulate matter" OR "ambient particulate" OR PM OR PM1* OR PM2* OR PM10* OR "ultrafine particulate*" OR "ultrafine particle*" or UFP) OR ("coarse particle*" OR soot OR "black smoke" or "black carbon" or "elemental carbon" OR "wood smoke"))
2. TS = ((mortalit* OR death*) OR ((cardiovascular OR respiratory OR pulmonary OR lung) NEAR/3 (mortality OR death* OR fatal* OR "hospital admission*" OR event* OR disease OR outcome*)) OR (asthma OR pneumonia OR "lung cancer" OR "lung function") OR ((improv* OR reduc* OR lower* OR increas* OR adverse OR measure* OR outcome* OR effect* OR impact* OR concentration OR level* OR absor* OR exposure* OR exposed) NEAR/3 ("air pollution" OR "particulate matter" OR "ambient particulate" OR PM OR PM1* OR PM2* OR PM10* OR "coarse particule*" OR soot OR "black smoke" OR "black carbon" OR "elemental carbon" OR combustion)) OR ((improv* OR reduc* OR lower* OR increas* OR adverse OR measure* OR outcome* OR effect* OR impact* OR concentration OR level* OR absor* OR exposure* OR exposed) NEAR/3 ("carbon monoxide" OR SO2 OR "sulphur dioxide" OR "sulfur dioxide" OR NO2 OR "nitrogen dioxide" OR O3 OR ozone OR UFP OR "ultrafine particle*")))
3. TS =(((emission* OR air OR PM OR PM1* OR PM2* OR PM10* OR "particulate matter" OR "ambient particulate" OR "ultrafine particulate*" OR "ultrafine particle*" OR UFP OR climate OR green OR smoke) NEAR/8 (control* OR regulation* OR policy OR policies OR guideline OR intervention OR act OR directive* OR vehicle OR transport* OR traffic OR automobile* OR car* OR industr* OR fuel OR "emission filter*" OR cooking OR heating OR cookstove* OR stove* OR "power generat*" OR energy OR zone* OR Olympic OR residential OR "wood burning" OR mobile OR Low* OR reduc* OR improv* OR clean* OR congestion* OR "coal burning" OR ban OR bans)) OR ((improved or clean* or "low emission" or efficient*) NEAR/1 (cookstove* or stove or stoves or heater)))
4. TS =(("comparative study" OR "intervention study" OR "evaluation study" OR "program evaluation") OR ("random allocation" OR "clinical trial" OR "single-blind" OR "double-blind" or "control group*") OR (randomized OR randomized OR placebo OR randomly OR groups) OR (trial) OR ("time series" OR interrupted NEAR/2 series) OR ("quasi-experiment*") OR ("pre test" OR pretest or "pre-intervention" OR "post-intervention" OR posttest OR "post test") OR ("controlled before" OR "before and after stud*" OR "follow-up-assessment") OR ((evaluat* OR intervention OR interventional OR treatment) AND (control OR controlled OR study OR program$ OR comparison OR "before and after" OR comparative)) OR ((intervention OR interventional OR process OR program) NEAR/8 (evaluat* OR effect* OR outcome*)) OR (program OR programme OR secondary analys*) OR ("case study" OR "observational study" OR cohort OR "uncontrolled study" OR "observational research") OR ("epidemiologic* study" OR "ecological study"))
5. 1 AND 2 AND 3 AND 4

GREENFILE

S1: TX ("Air pollution" or "airborne particles" or "particulate matter" or "ambient particulate" or "black smoke" or PM) (problem)

S2: TX (Mortality or cardiovascular or cardiac or death or “hospital admission*” or asthma or Pneumonia or “lung cancer” or “Lung function*” or “lung disease*” or “pulmonary function*” or “pulmonary disease*”)
S3: (Reduc* or improve* or decreas*)
S4: S2 AND S3
S5: TX (“Air pollution” or “airborne particles” or “particulate matter” or “ambient particul*”)
S6: S3 AND S5
S7: S4 OR S6
S8: S1 AND S7
S9: TX (clean air or emission* or PM or PM1* or PM2* or PM10* or “particulate matter” or “ambient particulate” or “ultrafine particulate*” or “ultrafine particle*” or UFP or climate policy or climate control or climate act or green policy or black smoke)
S10: TX (control* or regulation* or policy or policies or guideline or intervention or act or directive* or vehicle or transport* or traffic or automobile* or car*1 or industr* or fuel or emission filter* or cooking or heating or cookstove* or stove* or power generat* or energy or zone* or Olympic or residential or wood burning or mobile or Low* or reduc* or improv* or clean* or congestion* or coal burning or ban or bans)
S11: S9 AND S10
S12: S8 AND S11
S13: TX (Trial or randomization or randomisation or random allocation or “evaluation study” or “program evaluation” or control group* or epidemiol* study or “comparative study” or “intervention study” or intervention evaluation or “before and after” or “time series”)
S14: S12 AND S13

WHO GHL regional Indexes, GHL WHOLIS

1. (“air pollution” OR “particulate matter” OR “air quality” OR PM1* OR PM2* OR “ultrafine particulate” OR “ultrafine particle*” OR UFP OR “coarse particle” OR combustion OR soot OR “black smoke” OR “black carbon” OR “elemental carbon” OR “wood smoke”)
2. (moralit* OR death* OR “hospital admission” OR ((cardiovascular OR respiratory OR lung OR pulmonary) AND (fatal* OR event* OR disease* OR outcome*)) OR “heart attack” OR stroke OR asthma OR pneumonia OR “lung cancer”)
3. (control* OR regulation* policy OR policies OR guideline* OR intervention* OR act OR directive* OR vehicle OR transport* OR traffic OR automobile* OR car* OR industr* OR fuel OR “emission filter*” OR cooking OR heating OR cookstove* OR stove* OR “power generat*” OR energy OR zone* OR olympic OR residential OR “wood burning” OR mobile OR low* OR reduc* OR improv* OR clean* OR congestion* OR “coal burning” OR ban OR bans)
4. 1 AND 2 AND 3

HMIC

1. exp Air Pollution/
2. exp airborne particles/
3. (Air adj2 (pollut* or quality or ambient)).ti,ab.
4. (atmospher* adj2 pollut*).ti,ab.
5. (Particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or ultrafine particulate* or ultrafine particle* or UFP).ti,ab.
6. (Coarse particle* or Soot or Black smoke or Black carbon or Elemental carbon or wood smoke).ti,ab.
7. ((Emission* or air or atmospher*) adj2 (anthropogenic or motor or vehicle or road or power generation or indust* or combustion or smelting or construction or demolition or burning or residential)).ti,ab.
8. or/1-7
9. Mortality/ or Cardiovascular Diseases/ or Respiratory Tract Diseases/
10. (Mortalit* or Death*1).ti,ab.
11. (Cardiovascular adj3 (mortality or death* or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
12. (Respiratory adj3 (mortality or death or fatal* or hospital admission* or event*1 or disease or outcome*)).ti,ab.
13. (Heart attack* or stroke or strokes).ti,ab.
14. (asthma or Pneumonia or lung cancer or Lung function* or lung disease* or pulmonary function* or pulmonary disease*).ti,ab.

15. (air pollution/ or airborne particles/) and (Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level*).ti,ab.
16. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (air pollution or particulate matter or ambient particulate or PM or PM1* or PM2* or PM10* or coarse particule* or soot or black smoke or black carbon or elemental carbon or combustion)).ti,ab.
17. ((Improv* or reduc* or lower* or increas* or adverse or measure* or outcome* or effect* or impact* or concentration or level* or absor* or exposure* or exposed) adj3 (carbon monoxide or SO2 or sulphur dioxide or sulfur dioxide or NO2 or nitrogen dioxide or O3 or ozone or UFP or ultrafine particle*)).ti,ab.
18. or/9-17
19. air pollution/ or airborne particules/
20. ((emission* or air or PM or PM1* or PM2* or PM10* or particulate matter or ambient particulate or ultrafine particulate* or ultrafine particle* or UFP or climate or green or smoke) adj8 (control* or regulation* or policy or policies or guideline or intervention or act or directive* or vehicle or transport* or traffic or automobile* or car*1 or industr* or fuel or emission filter* or cooking or heating or cookstove* or stove* or power generat* or energy or zone* or Olympic or residential or wood burning or mobile or Low* or reduc* or improv* or clean* or congestion* or coal burning or ban or bans)).ti,ab.
21. smoke/
22. ((Improved or clean* or low emission or efficient*) adj1 (cookstove* or stove or stoves or heater)).ti,ab.
23. Wood burning regulation*.ti,ab.
24. or/19-23
25. 8 and 18 and 24
26. randomised controlled trials/
27. clinical trials/
28. comparative methods/
29. intervention study.ti,ab.
30. evaluation/
31. longitudinal studies/
32. (random allocation or clinical trial or single-blind method or double-blind method or control groups).ti,ab.
33. (randomized or randomised or placebo or randomly or groups).ab.
34. trial.ti,ab.
35. (time adj series).ab,ti. or (interrupted* adj2 series).ti,ab.
36. quasi-experiment$.ab,ti.
37. (pre test or pretest or pre-intervention or post-intervention or posttest or post test).ab,ti. (538)
38. (controlled before or "before and after stud$" or follow-up-assessment).ab,ti.
39. ((evaluat$ or intervention or interventional or treatment) and (control or controlled or study or program$ or comparison or "before and after" or comparative)).ab,ti.
40. ((intervention or interventional or process or program) adj8 (evaluat$ or effect$ or outcome$)).ab,ti.
41. (program or programme or secondary analys$).ti,ab. (17869)
42. ecological study.ti,ab.
43. (Case study or observational study or cohort or uncontrolled study or observational research or epidemiol* stud*).ti,ab.
44. or/26-43
45. exp animals/ not people/
46. 44 not 45
47. 25 and 46

WHO ICTRP

air pollution OR particulate matter OR air quality OR PM1* OR PM2*

Clinical Trials.gov

"air pollution" OR "clean air" OR "particulate matter" | Child, Adult, Senior

IDEAS

("air pollution" | "particulate matter" | "air quality" | PM10 | PM2.5 | "ultrafine particulate" | "ultrafine particle" | UFP | "coarse particle" | combustion | soot | "black smoke" | "black carbon" | "elemental carbon")

JOLIS

Keyword "air pollution OR particulate matter OR clean air" AND Keyword "improve OR improved OR improving OR reduce OR reducing OR reduction OR reduced" AND all "study OR intervention OR evaluation OR policy OR trial"

Appendix 2. Data extraction form

Interventions to reduce ambient particulate matter air pollution and their effect on health - Data extraction form

1. Study details

Study ID:
Study title:
Date of extraction:
Extractor:
Publication type

□ Journal

□ Book

□ Other (specify):
Funding source of study:
Potential conflict of interest from funding?

□ Yes

□ No

□ Unclear
Country of study:
List any other studies included in the review documenting the same intervention:

Study design

In cases where multiple study designs (e.g. ITS, CBA) or statistical analyses (e.g. for all versus a subset of monitors) are contained within the same study, the following criteria should be used in hierarchical order in order to help in assigning a study design:
1. If study authors describe the theory behind the intervention, i.e. how they expect the intervention will influence ambient air quality and/or health temporally and/or spatially, **the study design most closely matching this intervention theory should be assigned**.
2. For studies with multiple monitoring stations, yet no clear rationale as to where changes are expected or not, **the study design utilizing city-wide averages should be assigned**.
3. If two or more study designs are possible, and neither of the above criteria applies, **the study design representing the highest quality evidence should be assigned**.

□ Individual or cluster randomized controlled trial (RCT)

□ Individual or cluster controlled clinical trial (CCT)

□ Controlled before-and-after study adhering to EPOC criteria (CBA-EPOC):
· Contemporaneous data collection;

· comparable control site;
· at least 2 intervention and 2 control sites

□ Interrupted time series study adhering to the following EPOC criteria (ITS-EPOC):
· clearly defined intervention point;
· at least 3 time points before and 3 after the intervention

□ Controlled before-and-after study not adhering to EPOC criteria (CBA)

□ Uncontrolled before-and-after study (UBA)

□ Interrupted time series study, with clear intervention point, not adhering to EPOC criteria (ITS)

□ Repeated CSS with at clearly defined intervention point, and data collected at least once before and after intervention (CSS)
Notes regarding study design:
Total duration of study (in weeks, months, days - please specify exact dates where possible):
Where did the study take place?
Be as detailed as possible, and include eg. geographic location, specific setting, etc.
For controlled studies, do authors provide a rationale for intervention and control site selection?

2. Intervention

What is the pollutant target source of the intervention?

□ Vehicular

□ Industrial

□ Residential

□ Multiple
Description of the intervention:

Intervention theory

What is the specific goal(s) of the intervention?
In what timeframe was the intervention expected to influence air quality (e.g. short-term, long-term - be as specific as possible)?
Is the effect of the intervention itself expected to remain constant over time or might it evolve over time?
In what geographical or spatial area is the intervention expected to influence air quality (e.g. street-side, local, regional, national)?

Intervention components

List all intervention components. **If specific temporal or spatial information is relevant to the specific component, include this as well.**

Policy measure(s)	*Technology/infrastructure change(s)*	*Training/education*

List any incentives and/or penalties, which were introduced along with the intervention.
List any individuals or groups that were responsible for the implementation or delivery of the intervention?
List any funding sources important in the delivery of the intervention. What was the amount and/or duration of this funding?

3. Outcomes and Results

Note: for all included outcomes, this section, *3. Outcomes and Results, should be copied and pasted, and filled out.*

Outcome 1

List the assessed outcome
Is the assessed outcome a primary or secondary outcome according to the systematic review?

□ Primary

□ Secondary
How is the outcome defined and/or measured in the study?
For what geographical area(s) are the data representative?
At what time points was the outcome assessed?
Describe the time points at which the outcome was analyzed
For AQ outcomes:
At how many monitoring sites was the outcome measured?
Is it clear, either from the description of the specific monitoring sites or the intervention itself, at which monitors changes are expected and at which no (or lesser) changes are expected? *Elaborate on this point if possible*
Were before-intervention and after-intervention measurements taken from the same monitors, with the same timing?

□ Yes

□ No - Please describe below
For controlled studies, were baseline pollutant levels similar between intervention and control sites?
For health outcomes:
Were outcome data collected as part of the study or taken from (an) existing database(s)

□ Collected

□ Existing data
If data were taken from single or multiple databases, describe the source(s) in detail
Were before-intervention and after-intervention measurements taken from the same database(s)?

□ Yes

□ No
Were any data excluded based on specific factors (e.g. age, previous condition, etc.)?
For how many individuals were data available at baseline?

	Intervention Group	Control Group
T0		

For controlled studies, were individuals at intervention and control sites similar with regard to the outcome?
For controlled studies, were individuals at intervention and control sites similar with regard to other factors, which could potentially influence the outcome (e.g. participant age, comorbidities)?
Statistical analysis
Describe the statistical method applied
Describe any methods used for adjustment
Describe the method by which time was adjusted for in the analysis
Results
Pre- and Post-intervention means
· *Include variance measure and indicate where statistical testing showed significant differences*

· *If necessary, copy and paste table to include data for both unadjusted and adjusted values, or for multiple monitors (e.g. if area-wide average not provided OR not consistent with intervention theory) or multiple databases*

	Intervention Group	Comparison Group
T0		
T1		

Specify any resulting effect estimate(s), with variance measure, as reported in study
eg. odds ratio, risk ratio, mean difference, percent change, regression coefficients:
Graphical portrayal of the data included in the paper (e.g. time-series, bar graphs):
Describe any sensitivity analyses related to the outcome that were performed:
Did authors describe any specific weather events (e.g. extended rainy periods, uncharacteristically windy periods, etc.) either before or after intervention, which may have disproportionately influenced air quality?
* Narrative summary for this outcome by extractor:
Note: as described above, for all other included outcomes, the above section should be copied and pasted, and filled out.
Other important outcomes
List any potentially relevant indicators that might shed additional light on intervention effectiveness (e.g. traffic flow; specific source apportionment; etc.)

4. Subgroups

Participant subgroup
Which participant subgroups from paper can be analyzed?
Intervention subgroups
Which intervention subgroups from paper can be analyzed?
Context subgroups
Which subgroups dealing with contextual factors from paper can be analyzed?
Inequality subgroups
Which subgroups dealing with inequality from paper can be analyzed?

5. Context

Setting
Locational: which locational characteristics influence the intervention, its implementation, its population reach and its effectiveness?
Geographical: which geographical characteristics influence the intervention, its implementation, its population reach and its effectiveness?
Community
Epidemiological: which epidemiological characteristics of the community influence the intervention, its implementation, its population reach and its effectiveness?
Socio-economic: which socio-economic characteristics of the community influence the intervention, its implementation, its population reach and its effectiveness?
Socio-cultural: which socio-cultural characteristics of the community influence the intervention, its implementation, its population reach and its effectiveness?
Political: what aspects of the political environment influence the intervention, its implementation, its population reach and its effectiveness?
Legal: what aspects of the legal environment influence the intervention, its implementation, its population reach and its effectiveness?
Ethical: what aspects of the political environment influence the intervention, its implementation, its population reach and its effectiveness?
International

International: what aspects of the international environment influence the intervention, its implementation, its population reached and its effectiveness?

6. Contact authors

Should authors be contacted for further details?

□ Yes à contact details of author:

□ No

What type of further information is needed?

□ PICO description

□ Graph or figure details

□ Table details

Describe in detail what information should be obtained from study authors.

Appendix 3. GATE tool for correlation studies, as modified and employed by NICE

The Centre for Public Health Excellence at NICE provides guidance for using this modified GATE tool (NICE 2012). Individual criteria within sections 1-4, listed below, were rated 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

Section 1: Population (external validity)

1.1 Is the source population or source area well described?
1.2 Is the eligible population or area representative of the source population or area?
1.3 Do the selected participants or areas represent the eligible population or area?

Section 2: Method of selection of exposure (or comparison) group

2.1 Selection of exposure (and comparison) group. How was selection bias minimised?
2.2 Was the selection of explanatory variables based on sound theoretical basis?
2.3 Was the contamination acceptably low?
2.4 How well were likely confounding factors identified and controlled?

Section 3: Outcomes

3.1 Were the outcome measures and procedures reliable?
3.2 Were the outcome measurement complete?
3.3 Were all important outcomes assessed?
3.4 Was there a similar follow-up time in exposure & comparison groups?
3.5 Was follow-up time meaningful?

Section 4: Analyses

4.1 Was the study sufficiently powered to detect an effect if one exists?
4.2 Were multiple explanatory variables considered in the anlayses?
4.3 Were the analytical methods appropriate?
4.4 Was the precision of association given or calculable? Is association meaningful?

Section 5: Summary

Criteria for the summary section 5, listed below, were rated 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
5.1 Are the study results internally valid (i.e unbiased)?
5.2 Are the results generalisable to the source population (i.e externally valid)?

Appendix 4. Narrative description of supporting studies

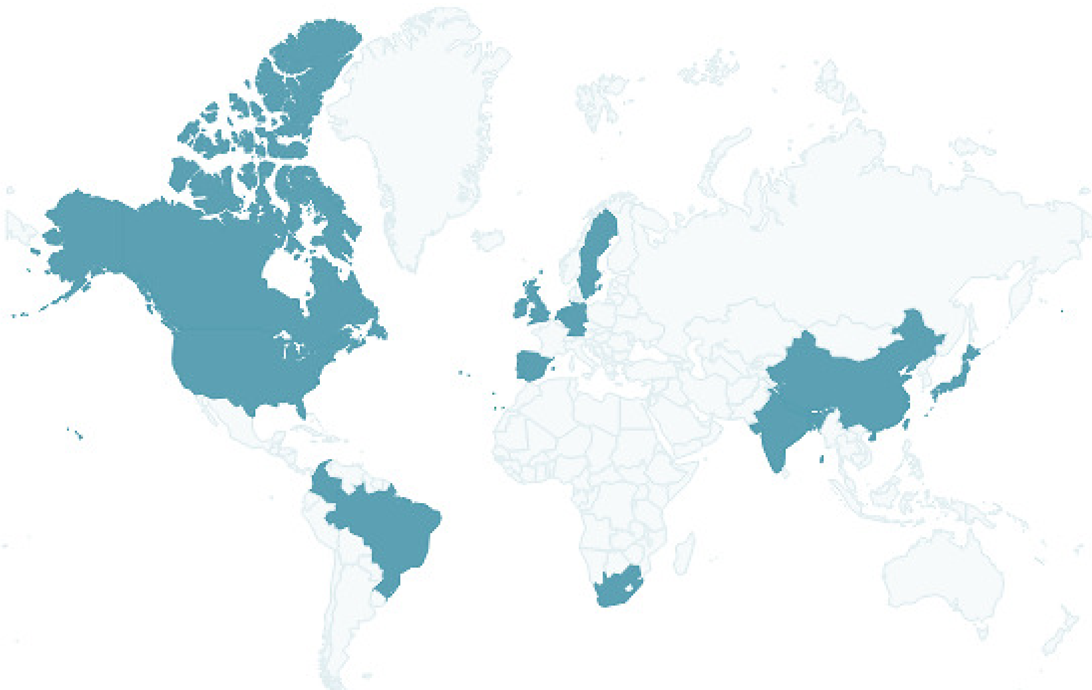
Description of supporting studies

The characteristics of each of the 77 supporting studies are summarized below and described in detail in the Characteristics of included studies.

Setting

Overall, the settings of supporting studies were similar to those of the main studies. Included supporting studies examined interventions in 19 different countries (Figure 15). Of the 50 interventions, the majority (n = 34) were implemented in HICs (Amato 2009; Bae 2015; Barros 2015; Chin 1996; Cirera 2009; Ding 2014; Ebelt 2001; Ferreira 2015; Goodman 2009; Henschel 2015; Hong 2015; Ibarra-Berastegi 2002; James 2012; Johansson 2009; Jones 2012; Karanasiou 2012; Kelly 2011; Keuken 2010a; Keuken 2010b; Kotchenruther 2015; Lee 2007; Levy 2006; MacNeill 2009; Noonan 2011; Panteliadis 2014; Pereira 2007; Pope 1989; Qadir 2013; Quiros 2013; Shon 2011; Shu 2014; Shu 2016; Le Tertre 2014; Zamurs 1984). Two other regions were fairly well represented, with eight interventions assessed in the Southeast Asia, East Asia and Oceania region (Brimblecombe 2015; Guo 2016; Hou 2010; Kuo 2009; Li 2016b; Peters 1996; Song 2015; Xu 2013) and five interventions in the South Asia region (Begum 2008; Chelani 2011; Fransen 2013; Khillare 2008; Latha 2004). The other world regions were poorly represented, with only two interventions in the Latin America and the Caribbean region (Ribeiro 2003; Valencia 2002), one in the Sub-Saharan Africa region (Engelbrecht 1999) and no interventions in the North Africa and the Middle East region or the Central Europe, Eastern Europe and Central Asia region. Comparing main and supporting studies the latter evaluated substantially more interventions in Southeast Asia, East Asia, Oceania and South Asia.

Figure 15. Geographic location of the 50 interventions evaluated in the supporting studies.



As with the main studies, most interventions evaluated in included supporting studies were implemented and assessed in an urban or community setting (n = 45). A further seven interventions were implemented in mixed urban/rural settings (Barros 2015; Guo 2016; Hou 2010; James 2012; Kotchenruther 2015; Latha 2004; Pope 1989). No interventions were implemented in rural settings.

Population

Some supporting studies assessed subsets of the population, including primary school children (Lin 2011; Lin 2015; MacNeill 2009; Peters 1996), children 11- to 13-years-old (Ribeiro 2003), children less than 15 years old (Lee 2007), and school children of any age (Noonan 2011).

Interventions and comparisons

Among the 50 unique interventions included in the supporting studies, 2 aimed to reduce ambient air pollution from industrial sources, 4 from residential sources, 34 from vehicular sources, and 10 from multiple sources.
In all studies, the comparison against which the intervention was compared can be considered practice as usual.
A description of each of the interventions from supporting studies is included in Appendix 5.

Interventions targeting industrial sources

Among the supporting studies, we included compulsory standards applied to the main industrial polluters in the city of Cartagena, Spain (Cirera 2009) and the temporary closure of a steel mill in the Utah Valley area of the US (Pope 1989).

Interventions targeting residential sources

Supporting studies covered the use of clean fuels for cooking in rural South Africa (Engelbrecht 1999), further evidence on the ban on the marketing, sale and distribution of coal for heating purposes across Ireland (Goodman 2009), a wood stove exchange programme in Libby, Montana, USA (Noonan 2011), and the replacement of coal-based with natural gas-based heating in the Urumqi region of northern China (Song 2015).

Interventions targeting vehicular sources

Supporting studies comprised vehicle charging schemes in Stockholm (Johansson 2009), further evidence for London (Kelly 2011) and Singapore, where an individual vehicle quota scheme was also introduced (Chin 1996). Three interventions focused on the use of cleaner fuels in vehicles, including measures in Dhaka, Bangladesh that banned two-stroke vehicles and converted public buses to natural gas engines (Begum 2008), the conversion of three- and four-wheeled vehicles to natural gas in Delhi, India (Chelani 2011) and further evidence on the Natural Gas Vehicle Supply (NGVS) programme in urban areas of South Korea (Shon 2011). Similarly, one intervention was concerned with the introduction of the EURO vehicle emission standards in Europe (Henschel 2015). Six interventions comprised temporary road closures due to one-time events, including political demonstrations in Kathmandu, Nepal (Fransen 2013) and in Hong Kong (Brimblecombe 2015), road construction on streets in California, USA (Hong 2015; Quiros 2013), the promotion of active transport and exercise in Los Angeles, USA (Shu 2016), and the 2004 Democratic National Convention in Boston, USA (Levy 2006). Two interventions focused on the public transport system, one the temporary closure of the system due to a strike in Ottawa, Canada (Ding 2014), and the other the construction of an underground railway system in Bilbao, Spain (Ibarra-Berastegi 2002). Two interventions targeted the speed and flow of traffic, through a speed limit change in Rotterdam and Amsterdam (Keuken 2010b) and through an increase in the duration of 'green time' for traffic signals in Syracuse, New York, US (Zamurs 1984). Two interventions involved various requirements for fuel, including a restriction on sulphur in vehicle fuel in Europe (Le Tertre 2014), as well as the California Ocean-Going Vessel Clean Fuel regulation and the North American Emissions Control Area, which reduced the use of sulphur in marine fuels (Kotchenruther 2015). Several interventions consisted of some form of vehicle restriction, including an even-odd ban during the 2002 Summer Asian Games in Busan, South Korea (Lee 2007), a one day per week restriction on all vehicles in Bogota, Colombia (Valencia 2002), and the Oxford Transport Strategy restricting traffic in the city centre of Oxford, UK (MacNeill 2009). We included further evidence on low emission zones in Munich, Germany (Qadir 2013), Amsterdam (Panteliadis 2014), Lisbon (Ferreira 2015), and London (Jones 2012). Mechanical street sweeping and cleaning measures were implemented in Rotterdam and Amsterdam in the Netherlands (Keuken 2010a), in Madrid, Spain (Karanasiou 2012), and in Barcelona, Spain (Amato 2010). Two interventions focused on long-term infrastructure changes to roads, including the paving of all roads in a rural area of northern Canada (James 2012); and the complete redesign of a street in Santa Monica, California, USA (Shu 2014). One intervention consisted of the installation of a public bicycle rental system in Changwon City, South Korea (Bae 2015). One intervention comprised a natural experiment surrounding the suspension of all trucking operations in response to a nationwide strike in India (Latha 2004).

Interventions targeting multiple sources

As supporting studies, we included several further interventions comprising coordinated measures to reduce pollution from industrial and vehicular sources surrounding short-term events. These include the Youth Olympic Games in Nanjing, China (Li 2016b), the 2010 Asian Games in Guangzhou, China (Xu 2013), the 2014 Asia-Pacific Economic Cooperation (APEC) convention in Beijing, China (Guo 2016), and further evidence on the 2008 Olympic Games in Beijing (Hou 2010). Further interventions included city-wide coordinated measures targeting industrial and vehicular polluters. Such coordinated measures were carried out in Erfurt, Germany (Ebelt 2001), and in Delhi, India (Khillare 2008). Others included overarching national policies aiming to reduce pollution from multiple sources in Brazil (Ribeiro 2003), Taiwan (Kuo 2009) and in Portugal (Barros 2015; Pereira 2007). One intervention specifically targeted the sulphur content of vehicle and industrial fuels in Hong Kong (Peters 1996).

Level of implementation of interventions

The pattern for supporting studies was similar to that of the main studies, with most interventions being implemented at the city level. Supporting studies covered all levels, however, including international level (Henschel 2015; Le Tertre 2014), national level (Barros 2015; Begum 2008; Chin 1996; Goodman 2009; Kuo 2009; Pereira 2007; Ribeiro 2003; Shon 2011), regional level (Xu 2013; Kotchenruther 2015), and city/community level (Bae 2015; Chelani 2011; Cirera 2009; Ding 2014; Ebelt 2001; Engelbrecht 1999; Ferreira 2015; Fransen 2013; Guo 2016; Hong 2015; Ibarra-Berastegi 2002; James 2012; Johansson 2009; Jones 2012; Kelly 2011; Keuken 2010a; Keuken 2010b; Khillare 2008; Latha 2004; Lee 2007; Li 2016b; Lin 2014; MacNeill 2009; Noonan 2011; Panteliadis 2014; Peters 1996; Pope 1989; Qadir 2013; Song 2015; Valencia 2002). In contrast to the main studies, several supporting studies were implemented at the street-level (Amato 2009; Brimblecombe 2015; Karanasiou 2012; Keuken 2010a; Keuken 2010b; Levy 2006; Quiros 2013; Shu 2014; Shu 2016; Zamurs 1984).

Timing and duration of interventions

As for the main studies, the timing and duration of interventions varied. We included supporting studies assessing interventions aiming to permanently improve air quality, such as the conversion of all public buses to natural gas (Begum 2008), the paving of roads (James 2012), and the redesign of a street (Shu 2014). We also included interventions with a temporary impact on air quality, such as measures during the 2002 and 2008 Asian Games and street sweeping and cleaning measures (Amato 2009; Karanasiou 2012; Keuken 2010a). Additionally we included interventions that were implemented or only expected to influence air quality during the higher pollution winter season, such as those targeting heating practices (Noonan 2011; Song 2015).

Outcomes

With regard to the outcomes assessed, the pattern for supporting studies was similar to that of the main studies.

Health outcomes

Of the 50 unique interventions, only 12 were evaluated with respect to their effect on health outcomes; five with regard to the primary health outcomes of the review, and 10 with regard to secondary health outcomes.

Air quality outcomes

Air quality outcomes were assessed for all of the included 50 unique interventions, 42 with regard to the primary air quality outcomes and 41 with regard to secondary outcomes.

Appendix 5. Characteristics of supporting studies

Study ID	Setting: country and location	Description of intervention	Level of allocation	Study period	Duration of intervention	Outcomes
Industrial sources						
Cirera 2009	Spain Urban Areas surrounding three factories within the city	Required abatement of industrial pollution at three major factories - with possibility of complete shut down	City	January 1992 to January 2002	Intermittent	Health NA AQ NO_2; SO_2
Pope 1989	USA Mixed Urban/ Rural Area of Utah County	Geneva steel mill shut down due to labour dispute	City	April 1985 to Feb 1988	Temporary	Health Respiratory hospital admissions AQ PM_{10};
Residential sources						
Engelbrecht 1999	South Africa Urban	3 low-smoke fuels (Flame Africa, Chartech,	City	21 June 1997 to 20 July 1997	Temporary	Health NA AQ

(*Continued*)

	Town of Qalabotjha and surrounding suburban area	and AFC), combusted in domestic stoves and braziers by the residents of Qalabotjha				PM_{10}; $PM_{2.5}$
Goodman 2009	Ireland Urban Urban and suburban areas across Ireland	Ban of the marketing, sale, and distribution of coal in Dublin (1990); in Cork (1995); extended to other cities Arklow, Drogheda, Dundalk, Limerick, and Wexford (1998), and Celbridge, Galway, Leixlip, Naas, and Waterford (2000)	City	Specific period varies for individual cities; 1980 to 2005	Permanent	Health NA AQ BS; SO_2
Noonan 2011	USA Urban Town of Libby, Montana	Woodstove changeout programme exchanging older woodstoves to EPA certified woodstoves	City	August 2003 to February 2009	Permanent	Health Respiratory symptoms AQ $PM_{2.5}$; EC
Ward 2009	USA Urban Town of Libby, Montana	Woodstove changeout programme exchanging older woodstoves to EPA certified woodstoves	City	November to February 2004 to 2008	Permanent	Health NA AQ $PM_{2.5}$
Ward 2010	USA Urban Town of Libby, Montana	Woodstove changeout programme exchanging older woodstoves to EPA certified woodstoves	City	2003 to 2008	Permanent	Health NA AQ $PM_{2.5}$

(*Continued*)

Song 2015	China Urban Urumqi city area in northern China	Replacement of coal-based heating systems with natural gas heating systems	City	January 2011 to 2014	Permanent	Health NA AQ $PM_{2.5}$; NO_x; SO_2
Vehicular sources						
Johansson 2009	Sweden Urban Stockholm metropolitan area	Congestion charging system in Stockholm: vehicles travelling into and out of the charge zone were charged during weekdays	City centre	January 2003 to July 2007	Permanent	Health NA AQ PM_{10}; NO_x; NO_2; CO
Kelly 2011	UK Urban London metropolitan area	Congestion charging scheme applied to four-wheeled vehicles entering the charging zone on workdays	City centre	17 February 2001 to 16 February 2005	Permanent	Health NA AQ PM10; NOx; NO2; NO CO; O3
Chin 1996	China - Singapore Urban Singapore metropolitan area	Reducing traffic air pollution through controlling of congestion and automobile ownership by using road pricing and vehicle quota schemes (VQS)	Country	1974 to 1993	Permanent	Health NA AQ NOx; NO2; CO
Chelani 2011	India Urban Delhi metropolitan area	Change of 3- and 4-wheeled vehicles to compressed natural gas engines	City	January 2000 to December 2004	Permanent	Health NA AQ PM_{10}; NO_2; CO

(*Continued*)

Shon 2011	South Korea Urban Several cities spread across South Korea	Natural Gas Vehicle Supply programme led to the replacement of the entire fleet of diesel-powered city buses with natural gas buses in large cities	Country	January 1998 to December 2008	Permanent	Health NA AQ PM10; NO2
Nguyen 2010	South Korea Urban Seoul metropolitan area	Natural Gas Vehicle Supply programme led to the replacement of the entire fleet of diesel-powered city buses with natural gas buses in large cities	Country	1996 to 2006	Permanent	Health NA AQ CO
Henschel 2015	Multiple - Europe Urban 9 European cities: Athens, Barcelona, Lisbon, Glasgow, London, Brussels, Vienna, Frankfurt and Leipzig	EURO vehicle emission standard regulations	International	1999 to 2010	Permanent	Health NA AQ NO_x; NO NO_2;
Lee 2005/ Lee 2007	South Korea Urban Busan metropolitan area	Even-odd day vehicle ban, restricting all cars from entering the city every other day based on the licence plate number	City	8 September 2002 to 4 November 2002	Temporary	Health Childhood asthma hospital admissions AQ PM_{10}; NO_2; O_3; SO_2; CO
Le Tertre 2014	Multiple - Europe	European directive on reducing sulphur content	International	1990 to 2008	Permanent	Health All-cause mortality;

(*Continued*)

	Urban 20 European cities	in fuels				Respiratory mortality; Cardiovascular mortality AQ SO_2
Lin 2011b	China Urban City of Taiyuan	Establishment of 'green belt' consisting of trees and hedges parallel to a non-motorized vehicle road	Street	Chinese	Permanent	Health NA AQ PM_{10}
Ibarra-Berastegi 2002	Spain Urban Port city of Bilbao	Construction of an underground railway system	City	1993 to 1998	Permanent	Health NA AQ O_3; SO_2; CO;
Ding 2014	Canada Urban City of Ottawa	Public transit services strike in Ottawa, Ontario: strike by transit workers that paralyzed public transport	City	10 December 2008 to 9 February 2009	Permanent	Health NA AQ $PM_{2.5}$; EC;
Ferreira 2015	Portugal Urban Lisbon metropolitan area	Low emission zone, which gradually increased its size and vehicle coverage; additionally road changes at two main traffic areas	City	2001 to 2013	Permanent	Health NA AQ PM_{10}; NO_2
Jones 2012	UK Urban London city centre	Low emission zone enforced initially for heavy goods vehicles, and thereafter for other goods vehicles, buses, and coaches of cer-	City centre	October 2009 to January 2009	Permanent	Health NA AQ UFP; NO_x

		tain weights				
Qadir 2013	Germany Urban Munich metropolitan area	Low emission zone in line with EURO regulations, becoming gradually more stringent	City	2006 to 2010	Permanent	Health NA AQ EC
Panteliadis 2014	Netherlands Urban Amsterdam metropolitan area	Restriction of heavy duty vehicles from entering the Amsterdam Low Emission Zone	City	2007 to 2010	Permanent	Health NA AQ PM_{10}; EC; NO_2; NO_x
Bae 2015	South Korea Urban Changwon urban area	A pro-bicycle campaign, including a public rental system, encouraging city dwellers in Changwon City to travel by bicycle	City	1991 to 2009	Permanent	Health All-cause mortality AQ PM_{10}; NO_2 SO_2; O_3; CO
James 2012	Canada Rural Community in Canada	Paving of the roads in a small rural town	City	Exact timing is unclear - 2008 to 2009	Permanent	Health NA AQ PM_{10}; $PM_{2.5}$
Keuken 2010b	The Netherlands Urban Metropolitan areas of Amsterdam and Rotterdam	Speed reduction: 80 km/h zones with strict enforcement of trajectory speed control enforced through camera surveillance and automatic fining systems	City	April 2005 to November 2006	Permanent	Health NA AQ PM_{10} NO_x
Brimblecombe 2015	China (Hong Kong) Urban Hong Kong metropoli-	Political strike led to road blockages caused by protesters and also imple-	Street	June to December, 2014	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$;

(*Continued*)

	tan area	mented by police				O_3
Fransen 2013	Nepal Urban Kathmandu metropolitan area	Establishment of band- has (roadblocks), restricting transportation (motor vehicles and busses)	City	1 January 2003 to 18 February 2008	Intermittent	Health NA AQ PM_{10};
Hong 2015	USA Urban Los Angeles metropolitan area	Closure of a 15 km segment of Highway I-405 for construction	City	June to August 2011	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$ NO_2; 0_3; CO
Levy 2006	USA Urban Boston metropolitan area	Road closures affecting approximately 40 miles of roads during the Democratic National Convention	Partial-city	19 July 2006 to 2 August 2006	Temporary	Health NA AQ $PM_{2.5}$; EC; NO_2
Quiros 2013	USA Urban Los Angeles localized street environment	Temporary closure of I-405	Street	8 July 2011 to 24 July 2011	Temporary	Health NA AQ PM10; UFP
Shu 2016	USA Urban Downtown and Eastern Los Angeles	Closure of 10 km of streets in Los Ange- les to road vehicles, where people were invited to use bicycles, scooters, or walk and run along these routes	Street	28 September 2014 to 12 October 2014	Temporary	Health NA AQ $PM_{2.5}$; UFP
Shu 2014	USA Urban Santa Monica, California metropolitan	Restructuring of entire street area: widened sidewalks, street furniture, marking of crosswalks and	Street	March, April 2011, 2013	Permanent	Health NA AQ $PM_{2.5}$; UFP

(*Continued*)

	area	bicycle lanes, improved landscaping, light poles, and improved storm-water management				
Amato 2009/ Amato 2010	Spain Urban Barcelona city centre	Road washing followed by mechanical sweeping	Street	February to March 2009	Intermittent	Health NA AQ PM_{10}; EC
Karanasiou 2011/ Karanasiou 2012	Spain Urban Metropolitan area of Madrid	Localized street washing followed by mechanical sweeping	Street	June 17 - July 20, 2009	Intermittent	Health NA AQ PM_{10}
Keuken 2010a	The Netherlands Urban Metropolitan areas of Amsterdam and Rotterdam	Road sweeping and washing vacuuming, high pressure washing, road cleaning and washing	City	July to November 2008	Intermittent	Health NA AQ $PM_{2.5-10}$; (coarse fraction)
Zamurs 1984	USA Urban Metropolitan area of Syracuse, New York	Traffic control strategies (TCS) - increasing green time on the traffic signal and strict parking restrictions	Street	November 1980 to April 1981	Temporary	Health NA AQ CO
Latha 2004	India Urban Hyderabad metropolitan area	Truck operations over the entire country temporarily suspended in response to a nationwide strike call by the operators	City	1 April to 25 April 2003	Temporary	Health NA AQ BC
Valencia 2002	Colombia Urban	One day of the week restriction to the circulation	City	July 2001 to December 2001	Permanent	Health NA AQ

(*Continued*)

	Bogotá metropolitan area	of public and private transportation vehicles in Bogotá				PM_{10}; NO_x; SO_2; CO
Thornbush 2015	UK Urban Metropolitan area of Oxford	The Oxford Transport Strategy (OTS) involved a wide range of changes focused primarily on the city centre from which all traffic was barred from some streets and private vehicles from others	City	1997 to 2012	Permanent	Health NA AQ PM_{10}; NO_x; NO_2; O_3 SO_2; CO
MacNeill 2009	UK Urban Metropolitan area of Oxford	The Oxford Transport Strategy (OTS) involved a wide range of changes focused primarily on the city centre from which all traffic was barred from some streets and private vehicles from others	City	1998, 2000	Temporary	Health Lung function; Respiratory symptoms AQ NA
Begum 2008	Bangladesh Urban Metropolitan area of Dhaka	Banning of commercial two-stroke vehicles and replacement with compressed natural gas or 4-stroke engines. Conversion of buses to compressed natural gas engines	Country	May 2000 to November 2005	Permanent	Health NA AQ PM_{10}; BC
Kotchenruther 2015	USA Mixed urban/rural	Implementation of California's Ocean-Going Vessel Clean	Regional	1 June 2006 to 31 August 2013	Permanent	Health NA AQ $PM_{2.5}$;

(*Continued*)

	Large area within the US west coast states of California, Oregon and Washington	Fuel regulation (CA-CFR) and North American Emissions Control Area (NA-ECA): Intervention targeted use of clean fuels through the reduction of sulfur in marine fuels				EC
Multiple sources						
Li 2016b	China Urban Nanjing metropolitan area	Approximately 2630 construction sites were closed; heavy-industry factories e.g. iron and steel, petrochemical industries required to reduce production by 20%; vehicles with high emissions e.g. trucks, as well as vehicles transporting hazardous materials banned from city	City	1 June 2014 to 20 October 2014	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$
Kuo 2009	Taiwan Urban Three major cities in central Taiwan: Taichung, Chaiyi, and Tainan	Tightened exhaust emission standards; reduced sulfur in fuels; reinforced control of fugitive particulate emissions; tax fees for pollutant emissions; increase license tax and fuel tax for older vehicles; better man-	Country	1996 to 2002	Permanent	Health NA AQ PM_{10}; NO_x; SO_2;

(*Continued*)

		agement of construction sites, road-dust sweeping				
Barros 2015	Portugal Mixed urban/rural Mainland Portugal as well as Azores and Madeira regions	NEC Directive aiming to limit emissions of acidifying and eutrophication pollutants as well as ground-level ozone precursors	Country	1990 to 2011	Permanent	Health NA AQ NO_2; NO_x; O_3
Hou 2010	China Urban Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	1 May 2008 to 31 October 2008	Temporary	Health All-cause mortality; Respiratory hospital admissions; Cardiovascular hospital admissions; Childhood asthma hospital admissions AQ PM_{10};
Huang 2012a	China Urban Beijing metropolitan area	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 during the 2008 Beijing	City	1 July 2008 to 29 August 2008	Temporary	Health NA AQ $PM_{2.5}$; BC; NO_2; SO_2; O_3; CO

(*Continued*)

		Olympic Games				
Huang 2012b	China Urban Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	2 June 2008 to 31 October 2008	Temporary	Health NA AQ $PM_{2.5}$; EC; NO_2; O_3 SO_2; CO
Lin 2011	China Urban Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	June 2007 to September 2008	Temporary	Health Acute respiratory inflammation (childhood) AQ $PM_{2.5}$; BC; NO_x SO_2; CO;
Schleicher 2011	China Urban Beijing metropolitan area	Alternative transportation strategy banning trucks not meeting emission standards, even-odd ban on private vehicles every other day, and strict restrictions on pol-	City	21 July 2008 to 26 September 2008	Temporary	Health NA AQ $PM_{2.5}$

(*Continued*)

		luting industries in Beijing and the surrounding provinces during the 2008 Beijing Olympic Games				
Schleicher 2012	China Urban Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	Oct 2007 - Feb 2009	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$; BC
Shen 2011	China Urban Suburban site in Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	8 August to 23 October, 2005 to 2009	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$; NO_2; SO_2
Lin 2015	China Urban Beijing metropolitan area	Alternative transportation strategy banning trucks not meeting emission standards, even-	City	June 2007 to September 2008	Temporary	Health NA AQ $PM_{2.5}$; BC; NO;

(*Continued*)

		odd ban on private vehicles every other day, and strict restrictions on polluting industries in Beijing and the surrounding provinces during the 2008 Beijing Olympic Games				SO_2; CO;
Mu 2014	China Urban Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	8 August 2008 to 17 September 2008	Temporary	Health Peak Expiratory flow AQ PM_{10}
Rich 2015	China Urban Beijing metropolitan area	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 during the 2008 Beijing Olympic Games	City	8 August to 24 September, 2007 to 2009	Temporary	Health NA AQ $PM_{2.5}$; SO_2; NO2; CO;

(*Continued*)

Wang 2014	China Mixed urban/ Rural Beijing Metropolitan area and a rural site in Hebei province	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 during the 2008 Beijing Olympic Games	City	Extended Olympic period (exact dates not specified), 2007 to 2009	Temporary	Health NA AQ O_3; SO_2; CO; NO_x; BC; $PM_{2.5}$
Xu 2016	China Mixed urban/ Rural Beijing Metropolitan area with 6 sites from urban and rural settings	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 during the 2008 Beijing Olympic Games	City	20 July to 20 September 2007-2011	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$; NO; NO_x;
Su 2015	China Urban Beijing metropolitan area	Alternative transportation strategy banning trucks not meeting emission standards, even-odd ban on private vehicles every other day, and strict restrictions on polluting industries	City	20 May to 1 December 2008	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$ NO_2;

(*Continued*)

		in Beijing and the surrounding provinces during the 2008 Beijing Olympic Games				
Zhang 2016	China Urban Guangzhou metropolitan area including surrounding suburban area	Even-odd ban on private vehicles every other day, ban of heavy vehicles and emission control from heavy industrial polluters	Regional	1 November 2009 to 21 December 2011	Temporary	Health NA AQ PM_{10}; NO_2; SO_2;
Pereira 2007	Portugal Urban Porto metropolitan area	Auto-oil directives reducing sulfur content in fuels for industrial and vehicular sources	Country	1999 to 2003	Permanent	Health NA AQ PM_{10}; SO_2;
Ribeiro 2003	Brazil Urban Sao Paulo metropolitan area	Standards regarding the maximum sulfur levels in fuel oil, and of the substitution of fuel oil by natural gas	Country	1984 to 1998	Permanent	Health Respiratory symptoms AQ SO_2;
Ebelt 2001	Germany Urban Erfurt, Germany metropolitan area	Shut down of old plants, transition from coal to liquid and gaseous fuels, reduction of sulfur content in coal, renewal of vehicle fleet	City	October 1991 to March 1999	Permanent	Health NA AQ $PM_{2.5}$; UFP; NO_x; SO_2; CO
Khillare 2008	India Urban Delhi metropolitan area	Renewal of public transport to 10, 000 busses; replacing pre-1990 autos with new vehicles, cleaner fuels, financial incentives to purchase	City	1998 to 2004	Permanent	Health NA AQ PM10;

(*Continued*)

		new autos; imposing CNG for buses older than 8 years; converting city bus fleet to single fuel mode; increase CNG supply outlets from 9 to 80				
Nidhi 2007	India Urban Delhi metropolitan area	Renewal of public transport to 10,000 busses; replacing pre-1990 autos with new vehicles, cleaner fuels, financial incentives to purchase new autos; imposing CNG for buses older than 8 years; converting city bus fleet to single fuel mode; increase CNG supply outlets from 9 to 80	City	January 1998 to December 2004	Permanent	Health NA AQ SO_2
Ravindra 2006	India Urban Delhi metropolitan area	Renewal of public transport to 10,000 busses; replacing pre-1990 autos with new vehicles, cleaner fuels, financial incentives to purchase new autos; imposing CNG for buses older than 8 years; converting city bus fleet to single fuel mode; in-	City	1998 to 2003	Permanent	Health NA AQ PM_{10}; NO_x; CO; SO_2

(*Continued*)

		crease CNG supply outlets from 9 to 80				
Guo 2016	China Mixed urban/rural Urban Beijing, Huairou, Tianjin, Hebei and other districts	Emission control measures during APEC China, 2014: Temporary closure of factories and restriction of motor vehicles in Beijing	City	29 October 2014 to 19 November 2014	Temporary	Health NA AQ $PM_{2.5}$; PM_{10}; NO_2; O_3; SO_2; CO
Li 2016a	China Urban Beijing metropolitan area	Emission control measures during APEC China, 2014: Temporary closure of factories and restriction of motor vehicles in Beijing	City	1 November 2014 to 12 November 2014	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$
Wang 2016	China Mixed urban/ Rural Beijing Metropolitan area and surrounding rural area	Emission control measures during APEC China, 2014: Temporary closure of factories and restriction of motor vehicles in Beijing	City	20 October 2014 to 24 November 2014	Temporary	Health NA AQ PM_{10}; $PM_{2.5}$; NO_2; O_3; SO_2; CO
Peters 1996	China - Hong Kong Urban Two districts of Hong Kong: Kwai Tsing and Southern	1990 restriction on sulfur fuel , limited to 0. 5% of sulfur by weight	City	1985 to 1995	Permanent	Health Respiratory symptoms AQ NA
Wong 2012	China - Hong Kong Urban Hong Kong metropoli-	1990 restriction on sulfur fuel , limited to 0. 5% of sulfur by weight	City	1985 to 1995	Permanent	Health All-cause mortality; Respiratory mortality; Cardiovascular

(*Continued*)

	tan area					mortality AQ PM_{10}; NO_2; O_3; SO_2;
Lin 2014	China Urban Guangzhou metropolitan area	Even-odd ban on private vehicles every other day, ban of heavy vehicles and emission control from heavy industrial polluters	City	1 November to 21 December, 2006 to 2011	Temporary	Health All-cause mortality: Respiratory mortality; Cardiovascular mortality AQ PM_{10}; NO_2; SO_2;
Xu 2013	China Urban Guangzhou metropolitan area	Even-odd ban on private vehicles every other day, ban of heavy vehicles and emission control from heavy industrial polluters	Regional	9 November 2010 to 30 November 2010	Temporary	Health NA AQ $PM_{2.5}$; O_3; SO_2 CO;
Tao 2015	China Mixed urban/rural Guangzhou metropolitan area and surrounding rural areas	Even-odd ban on private vehicles every other day, ban of heavy vehicles and emission control from heavy industrial polluters	Regional	1 November 2010 to 21 December 2010;	Temporary	Health NA AQ $PM_{2.5}$; EC; NO_2; O_3; SO_2 CO

Appendix 6. Modified GATE tool: Judgements for individual criteria for each included main study assessing health outcomes

Figure 16

Figure 16.

	1.1	1.2	1.3	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	5.1	5.2
Industrial interventions																		
Deschenes 2012	(+)	(++)	(++)	(++)	(-)	(++)	(++)	(++)	(++)	(++)	(-)	(++)	(+)	(++)	(++)	(++)	(++)	(++)
Lin 2013	(++)	(++)	(++)	(++)	(++)	(++)	(+)	(-)	(++)	(+)	(+)	(++)	(++)	(+)	(+)	(++)	(+)	(++)
Pope 2007	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	NR	(+)	(++)	(++)	(++)	(++)	(++)	(+)	(++)	(++)
Sajjadi 2011	(+)	(+)	(++)	(+)	(-)	(+)	(-)	(-)	NR	(++)	(++)	(+)	(+)	(-)	(-)	(-)	(-)	(+)
Tanaka 2015	(++)	(++)	(++)	(+)	(++)	(+)	(++)	(+)	(+)	(+)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)
Residential interventions																		
Dockery 2013	(++)	(++)	(++)	(++)	(-)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Johnston 2013	(++)	(+)	(++)	(++)	(+)	NA	(++)	(++)	(++)	(++)	(++)	(++)	(-)	(++)	(++)	(++)	(+)	(++)
Yap 2015	(++)	(++)	(+)	(++)	(-)	(++)	(++)	(++)	(+)	(++)	(++)	(+)	(+)	(++)	(-)	(++)	(-)	(+)
Vehicular interventions																		
Burr 2004	(+)	(+)	(+)	(+)	NA	(-)	NA	(-)	(+)	(++)	(++)	(++)	(-)	(-)	(+)	(-)	(-)	(+)
El-Zein 2007	(+)	(+)	(+)	(-)	(-)	NA	(+)	(+)	(+)	(+)	(++)	(++)	(-)	(++)	(-)	(-)	(-)	(+)
Hasunuma 2014	(++)	(+)	(++)	(+)	(++)	(++)	(+)	(+)	(+)	(++)	(++)	(++)	(++)	(-)	(-)	(-)	(+)	(++)
Yorifuji 2016	(+)	(++)	(++)	(++)	(-)	(++)	(++)	(+)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Peel 2010	(+)	(+)	(++)	(+)	(-)	NA	(++)	(++)	(++)	(++)	(-)	(++)	(+)	(++)	(++)	(++)	(++)	(++)
Multiple interventions																		
Giovanis 2015	(+)	(+)	(-)	(+)	(++)	(+)	(++)	(-)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(+)	(+)
Mullins 2014	(++)	(++)	(++)	(++)	(++)	(+)	(+)	(+)	(+)	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)
Li 2011	(++)	(++)	(++)	(++)	(++)	(+)	(+)	(+)	(++)	(++)	(++)	(++)	(+)	(++)	(+)	(++)	(+)	(++)
Zigler 2016	(+)	(+)	(++)	(++)	(+)	(-)	(++)	(++)	(+)	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)

See Appendix 3 for a detailed description of the individual criteria, and Appendix 8 for the support for the individual judgements.

Appendix 7. Modified GATE tool: Judgements for individual criteria for each included main study assessing AQ outcomes

Figure 17

Figure 17.

	1.1	1.2	1.3	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	5.1	5.2
Industrial interventions																		
Butler 2011	(++)	(++)	(+)	(++)	(+)	NA	(+)	(++)	(++)	(+)	(++)	(++)	(++)	(-)	(+)	(+)	(+)	(++)
Deschenes 2012	(+)	(++)	(+)	(+)	(-)	(++)	(++)	(++)	(++)	(++)	(-)	(++)	(+)	(++)	(++)	(++)	(++)	(++)
Lin 2013	(++)	(++)	(++)	(+)	(++)	(+)	(-)	(+)	(++)	(++)	(+)	(++)	(++)	(+)	(+)	(++)	(+)	(++)
Sajjadi 2012	(+)	(+)	(+)	(+)	(-)	NA	(+)	(+)	(-)	(++)	(++)	(++)	(++)	(++)	(+)	(++)	(-)	(+)
Saaroni 2010	(++)	(++)	(-)	(-)	(-)	(++)	(-)	(++)	(++)	(+)	(++)	(+)	(++)	(-)	(+)	(+)	(-)	(+)
Residential interventions																		
Allen 2009	(++)	(++)	(++)	(+)	NA	(-)	(-)	NA	NA	(-)	(++)	(-)	(++)	(-)	(+)	(+)	(-)	(++)
Aung 2016	(++)	(++)	(-)	(-)	(-)	(+)	(++)	(++)	(-)	(+)	(+)	(+)	(-)	(-)	(-)	(-)	(-)	(+)
Yap 2015	(++)	(++)	(+)	(+)	(-)	(++)	(++)	(+)	(+)	(++)	(+)	(+)	(++)	(++)	(+)	(++)	(+)	(+)
Vehicular interventions																		
Dolislager 1997	(+)	(++)	(++)	(++)	(+)	NA	(+)	(+)	(+)	(-)	(++)	(++)	(++)	(+)	(-)	(+)	(-)	(++)
Hasunuma 2014	(++)	(+)	(++)	(+)	(++)	(++)	(+)	(+)	(+)	(++)	(++)	(++)	(++)	(-)	(-)	(-)	(+)	(++)
Carillo 2013	(++)	(+)	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)	(+)	(++)	(++)
Davis 2008	(++)	(++)	(++)	(+)	(-)	NA	(++)	(++)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Gallego 2013a	(+)	(-)	(++)	(++)	(-)	NA	(++)	(++)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Gallego 2013b	(+)	(-)	(++)	(++)	(-)	NA	(++)	(++)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Viard 2015	(++)	(++)	(++)	(+)	(++)	NA	(++)	(+)	(+)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Kim 2011	(+)	(+)	(+)	(-)	(+)	(+)	(+)	(+)	(-)	(-)	(-)	(++)	(+)	(-)	(-)	(++)	(+)	(+)
Burr 2004	(+)	(+)	(+)	(+)	NA	(-)	NA	(-)	NR	(++)	(++)	(++)	NA	NA	(-)	(-)	(-)	(+)
Cowie 2012	(+)	(++)	(++)	(++)	(-)	(++)	(++)	(++)	NR	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Gramsch 2013	(+)	(++)	(++)	(+)	(-)	(-)	(++)	(+)	(+)	(+)	(+)	(++)	(+)	(++)	(-)	(++)	(+)	(++)
Peel 2010	(+)	(+)	(-)	(++)	NA	(++)	(-)	(++)	(+)	(++)	(-)	(++)	(+)	(-)	(+)	(++)	(+)	(++)
Titos 2015a	(++)	(++)	(+)	(+)	(+)	(++)	(+)	(++)	(+)	(+)	(+)	(++)	(-)	(-)	(+)	(++)	(+)	(++)
Titos 2015b	(++)	(++)	(+)	(+)	(+)	(+)	(++)	(++)	(+)	(+)	(++)	(++)	(-)	(-)	(+)	(++)	(+)	(++)
Boogaard 2012	(++)	(++)	(+)	(+)	(++)	(+)	(++)	(++)	(+)	(+)	(++)	(++)	(++)	(-)	(+)	(++)	(+)	(+)
Morfeld 2013	(+)	(++)	(-)	(-)	(++)	(+)	(++)	(++)	(++)	(+)	(++)	(+)	(++)	(++)	(++)	(++)	(+)	(+)
Fensterer 2014	(+)	(++)	(+)	(+)	(++)	(+)	(++)	(+)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Morfeld 2014	(++)	(+)	(++)	(+)	(++)	(+)	(++)	(++)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Bel 2013a	(+)	(+)	(++)	(-)	(++)	(-)	(+)	(+)	(-)	(+)	(++)	(++)	(++)	(++)	(++)	(-)	(+)	(+)
Bel 2013b	(+)	(+)	(++)	(-)	(++)	(-)	(+)	(+)	(-)	(+)	(++)	(++)	(++)	(++)	(++)	(-)	(+)	(+)
Dijkema 2008	(+)	(++)	(++)	(+)	(++)	(++)	(++)	(+)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(++)	(++)
Atkinson 2009	(+)	(++)	(+)	(+)	(-)	(+)	(-)	(++)	(+)	(+)	(++)	(++)	(-)	(-)	(-)	(-)	(+)	(+)
Ruprecht 2009	(+)	(+)	(-)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(++)	(+)	(-)	(-)	(-)	(+)	(-)	(+)
Multiple interventions																		
Giovanis 2015	(+)	(+)	(-)	(+)	(++)	(+)	(++)	(-)	(+)	(+)	(++)	(++)	(++)	(++)	(++)	(++)	(+)	(+)
Mullins 2014	(++)	(++)	(+)	(+)	(++)	(+)	(++)	(+)	(+)	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(+)
Zigler 2016	(+)	(+)	(++)	(++)	(+)	(-)	(++)	(++)	(+)	(++)	(++)	(++)	(+)	(++)	(++)	(++)	(++)	(++)

See Appendix 3 for a detailed description of the individual criteria, and Appendix 8 for the support for the individual judgements.

Appendix 8. Modified GATE tool: Support for ratings in 'Risk of bias' assessment of studies

Industrial interventions

Butler 2011

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Eastern United States (EUS)
1.2 Is the eligible population or area representative of the source population or area?	++	20 eastern states participating in the NOx Budget Trading Program
1.3 Do the selected participants or areas represent the eligible population or area?	+	Total number of sites used is 98. Rural CASTNET sites (n = 42 for ambient O3, n = 30 for met-adj O3)

(*Continued*)

Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	EPA route monitors from the states of interest are used for the analyses. Little information is provided about the nature of the monitors (e.g. are the urban monitors urban background monitors or could they be close to streets). Given that ozone is the outcome of interest, however, it is likely that this is not an issue, and no selection bias is present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	Some discussion included about meteorological variables in the method section
2.3 Was the contamination acceptably low?	NA	
2.4 How well were likely confounding factors identified and controlled?	+	Hourly O3 data were meteorologically adjusted to account for variability in meteorological conditions
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	See section 2.2 "O3 and dry-NO3 data" "Data quality for CASTENET is documented in the CASTNET Quality Assurance Project Plan (QAPP) and Quaterly Annual Quality Assurances Reports…Both CASTNET and AWS quality assurance include: measurement uncertainty, precision, bias, accuracy, completeness, detectability, independent audits, and measurement quality checks."
3.2 Were the outcome measurement complete?	++	"Both the ambient and met-adj O3 data have over 98% completeness in terms of site-years"
3.3 Were all important outcomes assessed?	+	Only AQ
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	No comments
3.5 Was follow-up time meaningful?	++	The pre-intervention period was much longer, but a very stable trend was present. Also the 5 years post-intervention were long enough to assess the longer-term impact of the intervention
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Power not discussed, but given the number of sites, estimates were calculated from much data, and power should not be an issue. Also precision around effect estimates (P values reported) from ARIMA analysis indicate that more than sufficient power was present
4.2 Were multiple explanatory variables considered in the analyses?	-	For the time-series analysis, non-adjusted concentrations are used; no further adjustments

(*Continued*)

4.3 Were the analytical methods appropriate?	+	Time points of analysis are not clear (how did they arrive at the datapoints?). Apart from the fact that no variables were adjusted for methods, were appropriate. It is also not clear whether the 2003 assessed step-change was specifically tied to the policy or based on only on the data
4.4 Was the precision of association given or calculable? Is association meaningful?	+	Only effect estimate provided; concentrations and measures of variability not provided
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	No adjustments in the ARIMA time-series modelling, no internal validity concerns, unclear how authors calculated the individual data points
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See Section 1

Deschenes 2012

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Not directly discussed; Industrialized high income countries
1.2 Is the eligible population or area representative of the source population or area?	++	20 Northeastern and Midwestern US where the NBP was implemented, 22 non-adjacent states where it was not Some slight discussion about existing emissions profiles of the regions of interest (e.g. page 6)
1.3 Do the selected participants or areas represent the eligible population or area?	**++ (health)** **+ (AQ)**	Health: Data on mortality was available for all counties, so this outcome is likely very well representative of the eligible population AQ: Unclear to what extent the selected monitors represent that northeast and midwest, because this is not reported in detail. But data from 168 counties were assessed
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	**++ (health)** **+ (AQ)**	Health: data available for all counties should mean that selection bias was not an issue AQ: Criteria of completeness were used to select monitors, but authors did not provide any information about where the selected monitors were located, what types of monitors they were, etc. Thus the exclusion of quite a lot of sites, especially for ozone,

(*Continued*)

		may have introduced bias
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	Not discussed
2.3 Was the contamination acceptably low?	++	"The analysis excludes Alaska, Hawaii, and states adjacent to the NBP participating states, which have ambiguous treatment status given the potential of pollution to cross state borders."
2.4 How well were likely confounding factors identified and controlled?	++	Control for weather-related aspects, as well as for fixed effects related to specific years, counties, state, seasons E.g.: "…county by year fixed effects, which account for all factors common to a county within a year (e.g. local economic activity and the quality of local health care provides)."
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++ **(health)** ++ **(AQ)**	Health: Mortality data from the National Center for Health Statistics should be considered reliable AQ: EPA Air Quality System data should be considered reliable, although not a lot of deals reported on QA/QC procedures
3.2 Were the outcome measurement complete?	++ **(health)** ++ **(AQ)**	Health: This is not explicitly discussed but given that the data come from the National Center for Health Statistics it is likely that they are quite complete AQ: Strict criteria applied for selected monitors. Only those with valid readings for at least 47 weeks in all years 1997 to 2007 were assessed
3.3 Were all important outcomes assessed?	++	Both AQ and health outcomes assessed.
3.4 Was there a similar follow-up time in exposure & comparison groups?	-	One thing to note is that there is more post data than pre data for $PM_{2.5}$ and PM_{10}, since they analyzed 2001 till 2007, instead of starting at 1997, though authors mention that "All regressions limit the sample to a balanced panel of county-season-years." Another issue is with the indicator variable for the policy, which is blurred: they defined Post = 0.5 in 2003 and Post = 1.0 in 2004 through 2007. In addition, they assigned a value of 0.5 in 2003 for all NBP states when the market was operating in 9 of the 20 states (the rest follows in 2004) because they argue that those 11 states may be affected too. That may be the case, but those choices are controversial
3.5 Was follow-up time meaningful?	++	Yes. 10 years of data, included 4 years of data post-intervention
Section 4: Analyses		

(*Continued*)

4.1 Was the study sufficiently powered to detect an effect if one exists?	+	Authors used heavily aggregated observations for both AQ and mortality analyses. If they had not done so, the study would have been better powered
4.2 Were multiple explanatory variables considered in the anlayses?	++	See regression description: sufficient control for weather-related aspects, as well as for fixed effects related to specific years, counties, state, seasons. But the same variables were used for both AQ and health analyses
4.3 Were the analytical methods appropriate?	++	Triple DID analysis, well-controlled, with a range of model specifications (where they try to interpret the regression results as a whole, rather than just one model), and some sensitivity analyses
4.4 Was the precision of association given or calculable? Is association meaningful?	++	SEs and indicators of significance provided for all estimates
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	Reliable data, strong analyses (see 4.3) provide internally valid results. Only the selection of monitors for the AQ analyses may be cause for some concern
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	This is not too well discussed, but these results are likely generalizable to high-income countries

Lin 2013

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Population in the 20 Eastern states and Washington DC where the NOx Budget and Trading Program was implemented
1.2 Is the eligible population or area representative of the source population or area?	++	Population in New York State.
1.3 Do the selected participants or areas represent the eligible population or area?	++	No selection criteria applied for the health outcomes, other than that records were excluded if the patient address was out-of-state. For the air quality outcomes, they do not specify how many monitors are used, and there is not a lot of detail regarding the kriging modelling approach. However, the exposure is at a 12 km grid, which is sufficient for ozone, and they have used regulatory monitoring sites, which are typically population oriented sites

(*Continued*)

Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++ **(Health)** + **(AQ)**	See comments on criteria 1.3.
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Confounders mentioned and explained.
2.3 Was the contamination acceptably low?	++ **(Health)** + **(AQ)**	Yes. All New York State residents were included, and records were excluded if the patient address was out-of-state
2.4 How well were likely confounding factors identified and controlled?	+ **(Health)** - **(AQ)**	Health: the study corrected for most of the typical confounders in time series studies. In addition, they even corrected for $PM_{2.5}$ effects. However, the study does not correct for longer time trends unrelated to the intervention under study. For example, they report that there were populations shifts during the study period for Hispanics, not taken into account in the modelling. They argue that they "did not adjust for a long-term trend, because this would remove the intervention effect, the variable of interest in this study", but in any study covering long time periods (in this case 9 years) one should worry about long-term trends in population and health unrelated to the intervention under study. Especially so, because they report significant and unexplained increases in 'control' admissions AQ: no analyses were done correcting for confounding factors
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	- **(Health)** + **(AQ)**	Health: they have used standard ICD codes, they report unexplained large increases in control admissions (gastrointestinal diseases (009) and non-traffic related accidental injury (E880-E888)) in the post intervention period compared to baseline periods. This leads to questions on the quality of the respiratory hospitalizations data, and the results in general AQ: there is not a lot of information about the ozone measurements and the modelling
3.2 Were the outcome measurement complete?	++	Health: they report a hospitalization coverage of 97%. AQ: not discussed, however as routine regulatory monitoring data were used this is likely not a serious issue
3.3 Were all important outcomes assessed?	++	Both AQ and health.
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	They have one additional summer in the baseline period (1997 to 2000) compared to the post implementation (2004 to 2006) , but unclear whether this is an issue in the health modelling because it seems that they have also used the data from the partial period (2001 to 2003) in the final model, see formula page 7

(*Continued)*

3.5 Was follow-up time meaningful?	++	Yes. No comments.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Health: sufficiently powered for the main analyses (in NYS), lesser so for the region-specific analyses, but still sufficiently powered. For example, they report 142,679 respiratory hospital admissions in the study period AQ: Yes. See Table 1.
4.2 Were multiple explanatory variables considered in the anlayses?	+	Yes, but see comment above about the lack of adjustment for long term
4.3 Were the analytical methods appropriate?	+	Generally appropriate, but see concern above about the lack of adjustment for long-term trends
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Provided.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	Health: strong design, the study corrected for most of the typical confounders in time series studies, however they do not adjust for long-term trends. Although they have used standard ICD codes, they report unexplained large increases in control admissions (gastrointestinal diseases (009) and non-traffic related accidental injury (E880-E888)) in the post-intervention period compared to baseline periods AQ: the study design is not that strong but they have EPA measurements over the whole state and regions
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Pope 2007

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Yes. Source population is the US population exposed to pollution from copper smelters
1.2 Is the eligible population or area representative of the source population or area?	++	Yes. Eligible population is the population in four southwest states (New Mexico, Arizona, Utah, and Nevada)

(*Continued*)

1.3 Do the selected participants or areas represent the eligible population or area?	++	The data, collected from the National Center for Health Statistics' yearly mortality reports of the United States from 1960 to 1975, should well represent the eligible population - it should be noted that state-wide data were used with no inclusion or exclusion criteria
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	No selection criteria were applied. Thus all population living in the four states were included. However, one can think about a scenario that populations closer to the copper smelters would benefit more than populations farther away. However, the study lacks spatial resolution with regard to pollution and mortality data to look into that
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Yes. No comments.
2.3 Was the contamination acceptably low?	++	Yes, the authors explored the use of different populations to control for long-term background trends unrelated to the intervention under study. There is the potential that in the rest of the US, specifically in the bordering states, air quality may also have been somewhat improved because of the nationwide smelter strike, and then using mortality counts in the bordering states to correct for mortality trends may result in overcontrolling and underestimating the mortality effect. This issue has been discussed by the authors as well. The study, however, explored different options, thus a ++ was given
2.4 How well were likely confounding factors identified and controlled?	++	Results were controlled for time trends, mortality trends in other areas of the US, and nationwide mortality counts for influenza/pneumonia, cardiovascular and other respiratory deaths. Multiple options were explored including using total mortality counts from 1) all other US states or 2) the Eastern US states or 3) neighbouring states, or 4) bordering states. No meteo variables were specifically added, but I assume that this time trend with 1 to 3 degrees of freedom is sufficient. Results were presented for models including nationwide mortality counts (thus including the four southeast states) for influenza/pneumonia, and additional models were run correcting for nationwide cardiovascular and other respiratory deaths on top of the influenza/pneumonia. The latter correction is very conservative, with the potential of overcontrolling
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Measures likely reliable, although not directly mentioned. Given that they used total mortality data, there is likely no large potential for error. Note that there was a switch from ICD-7 to ICD-8 in 1968 (right in the middle of the copper strike period). No details are reported regarding possible misclassification error, but given that they only use rather broad areas of disease classes, and only for the purpose of correcting for mortality trends, thus this is likely not an issue
3.2 Were the outcome measurement complete?	**NR**	Not reported.

(*Continued*)

3.3 Were all important outcomes assessed?	+	No air quality indicators were assessed, other than mentioning that there was a 60% decrease in concentrations of suspended sulphate particles. Also no data on transition metals were available. Also only total mortality was assessed, because cause-specific mortality for the four southwest states were not available
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Yes, almost similar follow-up times.
3.5 Was follow-up time meaningful?	++	Yes. They expected almost immediate changes, and a follow up of about 7 years is meaningful for mortality
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Yes, at least to do the analysis for the four southwest states as a region. Of course, power is less in the state-specific analyses, especially the CI are wide in Nevada, but still sufficient
4.2 Were multiple explanatory variables considered in the anlayses?	++	See comment question 2.4.
4.3 Were the analytical methods appropriate?	++	Yes. No comments.
4.4 Was the precision of association given or calculable? Is association meaningful?	+	Although only figures are given, no tables and results only reported in limited form in text
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	Because of the routine data used to assess the impact of the strike on mortality, using appropriate Poisson modelling and relevant covariates, this study can be considered internally valid
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	These results should be generalizable to other areas in the US of high industrial exposure from copper smelters

Sajjadi 2011

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Australian populations impacted by local heavy industries.
1.2 Is the eligible population or area representative of the source population or area?	+	Population in Lower Hunter Region in New South Wales, Australia. The paper included a detailed description of the area, although it was not clear if they

(*Continued)*

		thought the eligible population would be Newcastle (vs control) or the whole Lower Hunter region)
1.3 Do the selected participants or areas represent the eligible population or area?	++	Yes. It seems that no selection is made in the analyses of all respiratory diseases (all ages)
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	It is likely that all residents will be affected by the intervention. However, there is no information about the exact siting of the hospitals, nor where the patients are living
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No. There is hardly any description, and where there is some, this does not seem to be based on a proper theory. For example, in the discussion they list as potential confounders to consider in future studies "BMI, smoking status, rainfall, wind speed, and particularly wind direction", but some of those variables (e.g. smoking) are not typically considered confounders in time series designs. See also box below
2.3 Was the contamination acceptably low?	+	Seen in Figure 1 that the areas are relatively removed from one another geographically; some contamination likely, though this probably did not substantially affect bias
2.4 How well were likely confounding factors identified and controlled?	-	The adjustments made are poorly described. The only information regarding this is "the potential confounding factors of seasonal variation, day of week and public holidays, population, and viral epidemics were included in the model". But it is not clear how they were entered (e.g. as continuous variables, spline?) . In addition, it seems that another important confounder is temperature and perhaps longer-term time trends were not accounted for, although in the data collection section the authors describe that "as other confounders, a combined three-station data set for temperature and relative humidity, were also obtained from the air quality monitoring stations operated by NSW Department of Environment and Climate Change (NSWDECC)". However, those confounders are never mentioned again, and it is not clear whether they have put it into their model
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	-	It seems that the authors have based their selection on ICD codes, but this was not specified in the methods section. Only in the discussion this becomes apparent. Note that the authors report that changes took place from ICD 9 to ICD-10 coding in 1999
3.2 Were the outcome measurement complete?	**NR**	Not reported.
3.3 Were all important outcomes assessed?	++	Air quality and health assessed (see Sajjadi 2012).

(*Continued*)

3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Yes, from 1 January 1996 to 30 June 1999 (3.5 years before closing a major industry) from 1 January 2001 to 30 June 2004 (and 3.5 years after). Thus also exactly in the same period of year
3.5 Was follow-up time meaningful?	+	Yes, but perhaps a little shorter than most other studies of this kind
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	+	Sufficiently powered.
4.2 Were multiple explanatory variables considered in the anlayses?	-	See 2.4.
4.3 Were the analytical methods appropriate?	-	A lot of information is lacking. Even the time points at which the health outcome was analyzed is not apparent. For example, in Figure 3 they refer to monthly estimations. However, when you look at the confounders included, they describe day of week, suggesting that they have used daily counts
4.4 Was the precision of association given or calculable? Is association meaningful?	-	No measure of precision is provided.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	-	Concerns related to the data and the analysis could lead to biased results for this study
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	

Sajjadi 2012

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Source area are areas “impacted by local industrial sources/local heavy industries”
1.2 Is the eligible population or area representative of the source population or area?	+	Yes. Newcastle and Lower Hunter Region in New South Wales, Australia. The paper included a detailed description of the area. Not entirely clear whether the eligible area/population would be Hunter region or Newcastle

(*Continued)*

1.3 Do the selected participants or areas represent the eligible population or area?	+	Three regulatory monitoring sites. The sites were established to provide 'representative regional air quality measurements'. Not much details regarding the exact siting of those sites
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Not much details are given regarding the exact siting of the three sites, and whether they represent regional air quality measurements
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	Hardly any discussion about why certain variables were chosen
2.3 Was the contamination acceptably low?	NA	Not applicable
2.4 How well were likely confounding factors identified and controlled?	+	They controlled for some confounders in the mixed model analysis, but the covariates were poorly specified. For example, it was mentioned that they controlled for • Measurement unit (not further specified, but it seems that they mean months, but not clearly described) • BHP (assume they mean an indicator variable for pre and post closure) • Season (not specified further, e.g. as a categorical variable or a spline) Also it seems that important meteo factors are missing (e.g, temperature)
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	This is hard to tell because not much information is reported. For example, it is not described what the measurement method was, and nothing about QA/QC procedures. The only thing that is known is that they are obtained from a 'regulatory air quality monitor system' and that those are 'EPA' stations, and 'established in 1996 and maintained by New South Wales Department of Environment and Climate Change (NSWDECC).'
3.2 Were the outcome measurement complete?	-	They report missing data in some stations, and that is why they average across the three sites. They considered a daily value as missing when fewer than 12 hourly values were available. If so, they imputed the daily average of the remaining sites. They report that the number of missing days was less than 1%, with the exception of SO2 before the closure (19%); and missing means in this case that all measurements were missing at all three sites
3.3 Were all important outcomes assessed?	++	Health and AQ assessed (see Sajjadi 2011).
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Yes, from 1 January 1996 to 30 June 1999 (3.5 years before closing a major industry) from 1 January 2001 to 30 June 2004 (and 3.5 years after). Thus also exactly in the same period of year
3.5 Was follow-up time meaningful?	++	Yes. No comments.

(*Continued*)

Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Yes. No power calculation mentioned, but based on precision seems that the study was sufficiently powered
4.2 Were multiple explanatory variables considered in the anlayses?	++	Yes, as described above.
4.3 Were the analytical methods appropriate?	+	A lot of information is lacking but adjusted analyses are conducted
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Yes, they report SDs and P values. Note: no range is given for the estimated percentage change (Table 3)
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	-	Models are not well described, covariates not clear, perhaps missing important confounders (e.g. no control for temperature), and they did not exclude outlier values
5.2 Are the results generalisable to the source population (i.e. externally valid)?	+	Yes, but not much details regarding the exact siting of those sites, therefore not scored a ++

Saaroni 2010

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Local urban environments polluted by power stations.
1.2 Is the eligible population or area representative of the source population or area?	++	Local areas of Tel Aviv.
1.3 Do the selected participants or areas represent the eligible population or area?	-	Only one monitor was selected to represent the urban area of Tel Aviv, and it is not at all clear to what extent this area is representative of the rest of the city
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	The selection of one monitor to measure ambient air changes likely introduced bias into the study
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No explanatory variables.

(*Continued*)

2.3 Was the contamination acceptably low?	++	As they were upwind of the power station, it is unlikely that the two reference sites measuring PM were substantially affected by the intervention
2.4 How well were likely confounding factors identified and controlled?	-	None assessed.
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	No comments.
3.2 Were the outcome measurement complete?	++	The fraction of data missing was relatively small, 7% to 12.5%
3.3 Were all important outcomes assessed?	+	A range of pollutants were assessed, but no further impact of the intervention
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	No comments.
3.5 Was follow-up time meaningful?	+	Four months in two years is not long enough to completely rule out seasonal or other variations
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	No power calculation mentioned; but given the amount of data, power should not have been an issue
4.2 Were multiple explanatory variables considered in the anlayses?	-	None included.
4.3 Were the analytical methods appropriate?	+	Simple t-tests were applied before and after. Confounding factors could have been considered
4.4 Was the precision of association given or calculable? Is association meaningful?	+	Only approximate P values were given
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	-	Concerns with selection bias and aspects of the analysis may have compromised the internal validity of the study
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	Too little information is provided about the one intervention sampling site, and the rest of the city, to know how well generalizable these data would be to the rest of Tel Aviv or other metropolitan areas

Tanaka 2015

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Developing countries - well described, examples from Chile and India are cited as relevant. "The findings in this study accordingly present relevant estimates for the effect of environmental regulations in developing countries implementing similar policies on coal in the power industry"
1.2 Is the eligible population or area representative of the source population or area?	++	Likely somewhat representative of developing countries. The economic growth experienced in China makes it perhaps somewhat different from other developing countries. "As China's economy continued to grow at unprecedented rates for the last several decades..."
1.3 Do the selected participants or areas represent the eligible population or area?	++	Yes, likely representative of China. "…we draw the IMR data from the Chinese Disease Surveillance Points (DSP) system that collected birth and death registrations for 145 nationally representative sites from 1991 through 2000." "In total, 61 of 145 DSP sites are in the TCZ prefectures and thus comprise the treatment group, and 84 sites are in the non-TCZ prefectures, forming the control group."
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	All 145 sites were analyzed. The prefectures designated as TCZ were those violating nationally mandated pollution levels, while the remaining non-TCZ were thus less polluted. Authors have tried to include variables in the analyses to adjust for any differences in sites, but some remaining bias could be present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Lots of information dispersed throughout the introduction about various potential confounders (birth characteristics, parental attributes, socioeconomic status, unobserved characteristics - all on p 91)
2.3 Was the contamination acceptably low?	+	Probably yes, although the TCZ policy may also have reduced pollution beyond TCZ cities, especially when non-TCZ cities are located near TCZ cities, either directly through the policy effect on even non-TCZ cities or indirectly through reducing pollution that travels to non-TCZ cities. There is no data reporting the actual distance between the two. However contamination is probably low because of the "large amount of high-sulfur coal and SO2 emissions was produced in the TCZ cities (about 90%)"

(*Continued*)

2.4 How well were likely confounding factors identified and controlled?	++	The author controlled rigorously for many important confounding variables, such as DSP sites fixed effects, birth and parental characteristics (share of male, birth shares in respective month, birth order, mother's age, mother with high school degree or more), DSP sites characteristics (number of births, total population, rainfall) and DSP-site-specific time trends
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Routinely collected government data, not reported, but assumedly mostly reliable. "The micro-level data on infant mortality come from the Chinese Disease Surveillance Points (DSP) system. The DSP covers 145 sites, primarily at the county-level, established on the representative sample of the national population…Overall, the original data record approximately 500,000 deaths (for all ages) and 1,000,000 births from 1991 through 2000, from which the dataset we obtained was aggregated to the DSP site by year level."
3.2 Were the outcome measurement complete?	+	Not explicitly reported, but likely that the records are mostly complete
3.3 Were all important outcomes assessed?	+	"Due to lack of reliable pollution data in our study area…" Only health outcomes.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	No comments.
3.5 Was follow-up time meaningful?	++	7 years pre-intervention ; 3 years post-intervention.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	+	Power not explicitly discussed, however for the most part Table 3 estimates show that this is not a problem. Authors' "preferred" model, however, results in a large standard error, and the association of interest is not significant for this model
4.2 Were multiple explanatory variables considered in the anlayses?	++	Four models including a range of individual and district characteristics
4.3 Were the analytical methods appropriate?	++	Yes, appropriate. The difference in difference is a strong method to estimate causal effects, mimicking a randomized controlled trial. The author went to great lengths to investigate alternative assumptions, model specifications and key assumptions of the model
4.4 Was the precision of association given or calculable? Is association meaningful?	++	SDs and significant P values included for the main variables

(*Continued)*

Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	Strong analyses with reliable data and a long study period.
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Residential interventions

Allen 2009

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	The source area can be considered those areas of North America and northern Europe where "residential wood combustion (RWC) is a common heating method and a major source of air pollution in many locations"
1.2 Is the eligible population or area representative of the source population or area?	++	Two communities (Telka and Smithers) in British Columbia were the selected area; from the following text it is clear that heating is common and thus a major source of air pollution, meaning the selected area is representative to the eligible area: "Specifically, this study was conducted in two communities in the Bulkley Valley and Lakes District (BVLD) of [British Columbia], a region in which 7200 of 11,500 homes heat with wood, and 4200 (58%) of the wood-burning appliances are non-EPA-certified."
1.3 Do the selected participants or areas represent the eligible population or area?	++	"Outdoor equipment was placed in a secure location near the home (in the yard or on a deck or patio), and not directly adjacent to trees, sheds, or other large objects."
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Monitoring sites were placed directly near relevant homes - for the purposes, as all houses were monitored, of this simple before and after comparison this should not introduce selection bias "Outdoor equipment was placed in a secure location near the home (in the yard or on a deck or patio), and not directly adjacent to trees, sheds, or other large objects." Simple control: selection not described; if this is influenced by the stove change-out at the study homes, this could bias results. "In addition to measurements collected as part of this study, tapered element oscillating microbalance (TEOM) PM2.5 data at centrally located pollution monitoring stations in both Smithers and Telkwa were obtained from the

(*Continued*)

		[British Columbia] Ministry of Environment.”
2.2 Was the selection of explanatory variables based on sound theoretical basis?	NA	Not applicable.
2.3 Was the contamination acceptably low?	-	No information provided, it is possible that levels at the control site were influenced by the stove changeout at study homes
2.4 How well were likely confounding factors identified and controlled?	-	Simple median changes reported.
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	“During each 6-day monitoring period, PM2.5 samples were collected onto Teflon filters during two consecutive 3-day samples using single-stage Harvard Impactors (Air Diagnostics and Engineering, Harrison, ME) and 10-lpm pumps (Leland Legacy, SKC Inc., Eighty Four, PA).” “Outdoor equipment was placed in a secure location near the home (in the yard or on a deck or patio), and not directly adjacent to trees, sheds, or other large objects.”
3.2 Were the outcome measurement complete?	++	Not reported, but given measuring technique described above, likely that missing data did not lead to bias
3.3 Were all important outcomes assessed?	-	Only $PM_{2.5}$ assessed.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	“Each study home was monitored during one 6-day monitoring period prior to the stove exchange and one 6-day monitoring period after the stove exchange”; (this applies to central site measurements as well)
3.5 Was follow-up time meaningful?	-	6 days is enough to potentially detect an immediate effect, but not to assess long-term effects of the changes in stoves, or to rule out spurious trends either before or after the introduction of improved stoves
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	As significant median changes were seen at both outdoor and central site monitors, the study seems to be sufficiently powered
4.2 Were multiple explanatory variables considered in the anlayses?	-	None included.
4.3 Were the analytical methods appropriate?	+	Median changes assessed using t-test at outdoor and central site monitors. This is a very basic method and a more sophisticated method of comparing the two site types would have been much more appropriate

(*Continued*)

4.4 Was the precision of association given or calculable? Is association meaningful?	+	P value given; no other measure of precision.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	-	Findings are based on a very short follow-up and a very basic analysis. Additionally, no information is provided on the "control" central site, therefore the risk of contamination is unclear
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Likely that the reported results are relevant for other communities relying on wood stoves to heat in North America and northern Europe

Aung 2016

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Developing regions in Africa and Asia where biomass combustion is highly prevalent. "Household biomass combustion is a major contributor of BC emissions; in Africa and Asia, the sector is thought to account for 70% of the region's BC emissions"
1.2 Is the eligible population or area representative of the source population or area?	++	The source area is a small village (Hire Waddarkal) in India where biomass burning is nearly universally practised. "The study site was in Koppal District of northern Karnataka, India. Most households (99%) in this region burn biomass fuels…"
1.3 Do the selected participants or areas represent the eligible population or area?	-	With one village centre monitor and one upwind monitor, it cannot be said that (this component) of the study is well representative of the village
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	A site which is "predominantly upwind" was chosen as comparison. Wind direction is, however, not actually assessed. Also very little description of the sites
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	Some covariate discussion, but only with regard to the indoor air pollution
2.3 Was the contamination acceptably low?	+	Not assessed, but as wind direction was not addressed it is likely that some contamination was possible

(*Continued)*

2.4 How well were likely confounding factors identified and controlled?	++	They have collected meteo variables. A weather station (model PWS 1000 TB, Zephyr Instruments, East Granby, CT) was placed in the centre of the village next to the community measurement location and recorded temperature, relative humidity, atmospheric pressure, wind speed, and wind direction every 30 min
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Reliable measurement procedures, with QA/QC reported
3.2 Were the outcome measurement complete?	-	Quite a lot of missing data, especially at the upwind site due to "unstable flow rate and negative measurement of filter mass". Authors report that this was mainly due to "negative filter masses [...] presumably a result of low ambient concentrations and low pump flow rates, particularly in the post-intervention season, because of the reduced sampling duty cycle."
3.3 Were all important outcomes assessed?	+	Only AQ outcomes
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	Outdoor samples were roughly collected in the same season: Pre-intervention (2 September to 28 September 2011), and the post-intervention (14 July to 4 August 2012). Time duration of the measurements differ between the pre and post period, namely 24 hr and 22 hr, but they have developed an approach to adjust for a shorter sampling period using Dustrak and microaethalometers. Also only two measurements make up the post-intervention time
3.5 Was follow-up time meaningful?	+	Short follow-up, not clear to what extent meteorology may be influencing concentrations in the short term
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	No, very small pilot study, for example, looking at the number of samples available in Table S7 for $PM_{2.5}$, there are only 2 samples for the post-intervention period for the upwind site as compared to 8 samples for the pre intervention period
4.2 Were multiple explanatory variables considered in the anlayses?	-	None considered.
4.3 Were the analytical methods appropriate?	-	Means and SDs reported. The applied statistical test assesses whether the "order" of the measurements remains the same for two groups of measurements. It seems very likely that the upwind (thus not influenced by the village centre and heating) monitor should be expected to be lower, even if the intervention was to lead to improved air quality. Also, with no consideration of any potential confounders, like meteorology, this analysis was not very appropriate
4.4 Was the precision of association given or calculable? Is association meaningful?	-	No direct group comparison reported.

(*Continued*)

Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	-	The internal validity of this study suffers from the nature of the two monitors (i.e. control a “predominantly” upwind monitor), from the analysis methods and from the lack of consideration of potential confounders
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Dockery 2013

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Source population is the population in Irish cities.
1.2 Is the eligible population or area representative of the source population or area?	++	Yes. The 12 Irish cities and towns affected by the bans. These are well described under Study Area, p 4
1.3 Do the selected participants or areas represent the eligible population or area?	++	Yes. No selection is made for mortality because death registry data is used. Note that in the 1990 Dublin ban, city residents were included, whereas in the 1995 and 1998 ban, all population in the county were selected. (+) For the Air quality outcomes, they have 1 to 6 sites per city, but they do not specify whether those sites are representative for the population. However, it is fair to make this assumption because they use regulatory sites, and typically those are population-oriented sites
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	For the mortality analysis no selection made, as registry data used, which should not have introduced any bias. (+) The selection of the air quality monitoring sites were likely representative for the respective cities, but this is not described, thus bias is still possible
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No rationale for the selection of explanatory variables provided
2.3 Was the contamination acceptably low?	++	Cities compared to the coastal and midland counties; geographically removed from one another, unlikely that contamination biased results
2.4 How well were likely confounding factors identified and controlled?	++	The ITS controlled for all the typical confounders in time series studies (influenza, temperature and a season smooth of the standardized mortality rates

(*Continued*)

		in a reference population unaffected by the bans). They compared the current results with the earlier study investigating the Dublin ban (Clancy 2002), and investigated the differences in data and methods running several sensitivity analyses. They also did simulation analyses (Appendix G) how to best control for long-term trends. (-) No confounding factors considered in the AQ UBA analysis
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	They have used death registry data. No specific issues reported other than with the coding of the exact residence; there have been some errors reported in whether a person was living in a specific city or its county. To avoid misclassification, they have chosen to analyze the 1995 and 1998 ban at the county level instead of the city level. For hospital admissions, however, some concerns existed with regard to underreporting pre-1995. All admission values from before this time were adjusted
3.2 Were the outcome measurement complete?	++	Yes. They have used death registry data. It is required to register deaths within 3 months of the date of death. Written permission from the Registrar General is required to register deaths more than one year after the date of death. Approximately 400 deaths are registered late in Ireland each year. For hospital admissions, they have documented underreporting issues before 1995, but have developed an approach to correct for it using hospital admission data on digestive disorders
3.3 Were all important outcomes assessed?	++	By measuring health outcomes (and a basic uncontrolled analysis for AQ) a complete picture of the effect of the coal ban on the air quality and the related effect on human health is attained
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	In the 24 year analysis period, the pre-intervention times were generally longer than the post-intervention time (except in Dublin). As both periods were sufficiently long to detect effects, this should not have introduced bias. The same applies to the analysis of hospital admissions
3.5 Was follow-up time meaningful?	++	They expected immediate changes, and a follow-up of a few years (5 years is follow-up in study) is meaningful for mortality
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	No power analysis calculated, but they assessed a large amount of data; precision around estimates shows that power was likely not a problem
4.2 Were multiple explanatory variables considered in the anlayses?	++	The study controlled for all the typical confounders in time series studies (influenza, temperature and a season smooth of the standardized mortality rates in a reference population unaffected by the bans). They compared the current results with the earlier study investigating the Dublin ban (Clancy 2002), and investigated the differences in data and methods running several sensitivity analyses. They also did simulation analyses (Ap-

(*Continued*)

		pendix G) how to best control for long-term trends
4.3 Were the analytical methods appropriate?	++	Well controlled, well adjusted, long analyses.
4.4 Was the precision of association given or calculable? Is association meaningful?	++	No comments.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	Strong analysis based on complete routine data.
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Johnston 2013

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Tasmania: climate and population of the state of Tasmania (and for climate the differences when compared to Australia) given in the Methods: Setting
1.2 Is the eligible population or area representative of the source population or area?	+	Launceston: "The impact on air quality was particularly severe in Launceston, which is in a river valley where both topographical and metrological conditions limit atmospheric dispersion of air pollution." i.e. due to the geography of Launceston, it is likely more affected by air pollution than the rest of Tasmania
1.3 Do the selected participants or areas represent the eligible population or area?	++	City-wide data from the Australian Bureau of Statistics should be representative for Launceston
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	As for 1.3, using mortality data coming from the Australian Bureau of Statistics for the city of Launceston as the selected participants should not introduce bias into the study
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	From "Potential Confounders" in the discussion: "We included smooth daily mortality data from all of Tasmania to adjust for secular trends because the entire state has similar distributions of health outcomes, socioeconomic status, and demographic structure. The changing prevalence of population risk factors through time, such as smoking and diabetes, is likely to have been similar."

(*Continued*)

2.3 Was the contamination acceptably low?	**NA**	"We are not explicitly using Hobart as a control site, but even so, contamination is unlikely due to the geographic separation."
2.4 How well were likely confounding factors identified and controlled?	++	The ITS mortality analysis controlled for the effects of meteorology, epidemics of respiratory infections, and secular trends in daily mortality in Tasmania
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	"Mortality data were obtained from Austrlian Bureau of Statistics. These data undergo considerable auditing for quality before being released for publication. "
3.2 Were the outcome measurement complete?	++	Mortality data for Australia likely very complete
3.3 Were all important outcomes assessed?	++	Study offers a clear picture of the effect of the intervention on air quality, and the associated effect on health
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Included data from 1994 to 2000 (pre-intervention) and 2001 to 2007 (post-intervention)
3.5 Was follow-up time meaningful?	++	6.5 years before and after intervention allows sufficient time for short- and long-term trends both in mortality and air quality in effectiveness to be assessed
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	No power calculation mentioned in text for the ITS regression analysis - the effect direction for the total population and females favoured the intervention, very large confidence intervals may suggest that the study may not be sufficiently powered
4.2 Were multiple explanatory variables considered in the anlayses?	++	The ITS mortality analysis controlled for the effects of meteorology, epidemics of respiratory infections, and secular trends in daily mortality in Tasmania
4.3 Were the analytical methods appropriate?	++	ITS analysis, controlled for secular trends as well as other potential confounders, was performed appropriately
4.4 Was the precision of association given or calculable? Is association meaningful?	+	CIs provided for the % change estimates from the ITS regression
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	+	ITS analysis of Launceston well performed, the only major concern relates to whether the study is statistically powered to detect a meaningful effect
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Although the geography of Launceston makes it somewhat unique in how air pollution can be dispersed, the findings of the mortality analysis are likely still generalizable to the rest of Tasmania

Yap 2015

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	
1.2 Is the eligible population or area representative of the source population or area?	++	
1.3 Do the selected participants or areas represent the eligible population or area?	**+ (Health)** **+ (AQ)**	Health: Hospital admissions from the California Office of Statewide Health Planning and Development are likely well representative to the area, although this is not described, and they use only people 45 years and older, without providing a rationale AQ: Monitor locations and characteristics not described. These were routine monitors but there is no information regarding how many and where. The only text implies that the monitors are likely picking up background concentrations: "We used the available ambient PM2.5 monitoring data from central outdoor monitoring stations and assumed that an average of the ambient PM2.5 measurements was representative of the complex spatial and temporal pattern of exposures over a large area."
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	**++ (Health)** **+ (AQ)**	Health: No likely selection bias AQ: Site selection (and site characteristics) not described, see above; likely background concentrations across the study period
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No explanation of why certain variables were considered as confounding variables
2.3 Was the contamination acceptably low?	++	Assessment across entire SJVAB.
2.4 How well were likely confounding factors identified and controlled?	++	Health: examined, in addition to the variables listed above, socioeconomic variables, such as percentages of poverty, unemployment, and low education using census data AQ: Meteorological variables (temperature, dew point, wind speed) assessed, and included in models where deemed relevant
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	**++ (Health)** **+ (AQ)**	Health: Hospital admission based on the ICD codes reliable. AQ: no discussion of how data were measured.

(*Continued*)

3.2 Were the outcome measurement complete?	**+ (Health)** **+ (AQ)**	Health: not reported; likely that some values were missing, but likely not sufficient to bias estimates too greatly AQ: not reported.
3.3 Were all important outcomes assessed?	++	$PM_{2.5}$, coarse particles and hospital admissions.
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	It is not described whether exactly the same PM monitors were used before and after
3.5 Was follow-up time meaningful?	+	Yes, probably but potentially there is a timing issue since the rule was adopted already in 1992, though enforcement of the rule did not begin until the 2003 amendment. In the analyses they have compared 2000 to 2002 as before and November 2003 to 2006 as post. There is no information regarding the enforcement, and there were 15 wintertime days in the pre period during which residential wood burning was banned
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	**+ (Health)** **++ (AQ)**	Health: for the adults 45 to 64 years, it can be seen in table 3 that the study may not have been sufficiently powered to show potentially relevant effects AQ: See Table 2 and Figure 1; AQ analysis sufficiently powered to show assessed effects
4.2 Were multiple explanatory variables considered in the anlayses?	++	For both AQ and health analyses, a model selection process was described, where variables were tested for relevance before being included in final model
4.3 Were the analytical methods appropriate?	-	Regression-based approach controlling for relevant covariates is a valid approach. No consideration for pre-existing time trends is considered, which would have been more appropriate. For the health outcomes, age and influenza episodes were not considered. In addition,the variable no days (with a no-burn day defined as a day when air quality was forecast to reach an air quality index of at least 150 (approximately 65 µg/m3 of $PM_{2.5}$) and wood burning was therefore banned) are questionable; the authors should have perhaps added an interaction term between the no-days variable and the rule variable because the rule only seem to apply when air quality was forecast to be poor
4.4 Was the precision of association given or calculable? Is association meaningful?	++	No comments
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	**- (Health)** **+ (AQ)**	Health: No consideration of pre-existing trends in outcomes, potentially important confounders not considered, and ambiguity

(*Continued*)

		with the intervention timing may have led to bias AQ: Regression-based approach controlling for relevant covariates; the intervention timing may, however have led to some bias
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Vehicular interventions

Burr 2004

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Severely congested cities (in the UK, in Europe, in high-income countries??)
1.2 Is the eligible population or area representative of the source population or area?	+	Congested areas in northern Wales.
1.3 Do the selected participants or areas represent the eligible population or area?	+	The two sites are likely somewhat representative of northern Wales, but the limited number of sites and the lack of descriptions limits the certainty. For health: 165 at congested area, and 283 in uncongested area
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	An uncongested site somewhat separated from the intervention congested street was chosen. Authors write that they chose a site close by so that the houses and other relevant characteristics would not differ. Unclear to what extent the investigators actually assessed the appropriateness of this as a control site
2.2 Was the selection of explanatory variables based on sound theoretical basis?	NA	None included.
2.3 Was the contamination acceptably low?	-	Uncongested control area separated from the intervention area for which a bypass was opened by only 20 metres. Unlikely that contamination did not occur
2.4 How well were likely confounding factors identified and controlled?	NA	None included.
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	-	Self-reported and self-assessed health outcomes, asking for recall over the last year. Survey is non-validated

(*Continued*)

3.2 Were the outcome measurement complete?	+	"Many of the subjects who participated in the the first phase had moved away by this time [follow-up]" In the congested streets group 386 at baseline and 165 at follow-up. In the uncongested streets group 425 at baseline 283 and at follow-up
3.3 Were all important outcomes assessed?	++	Health and AQ
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	"After the by-pass opened…fixed site pollutant measurements were repeated for the same periods of time at the same seasons as before, using the same methodology."
3.5 Was follow-up time meaningful?	++	Given that short-term effects are being assessed, approximately 9 months is sufficient (although data for a longer period would help rule out any spurious trends seen only in one given year/season)
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	Large loss to follow-up; possible lack of power in calculated precision mentioned in the discussion as limitation
4.2 Were multiple explanatory variables considered in the anlayses?	-	None considered.
4.3 Were the analytical methods appropriate?	+	Net improvement in each group, and the difference in net improvement assessed for symptom prevalence. Change in variability assessed within groups for peak flow, but not between groups. These analyses are informative, but more clinically relevant endpoints, and especially for symptoms, a more structure analysis would have been more appropriate
4.4 Was the precision of association given or calculable? Is association meaningful?	-	Association not meaningful.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	-	Because of loss to follow-up, self-reported outcomes with a high potential for bias, and a lack of between-group comparison, these results are not very internally valid
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Dolislager 1997

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Metropolitan areas in NAAQS CO non-attainment areas.
1.2 Is the eligible population or area representative of the source population or area?	++	Metropolitan areas in NAAQS CO non-attainment areas in California
1.3 Do the selected participants or areas represent the eligible population or area?	++	The 16 assessed non-attainment areas studied here represent all of the NA areas in California
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	As only non-attainment metropolitan areas were assessed, and as all of those were included, selection bias should not be an issue for the study
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	The use of NOx as the main "explanatory" factor (in controlling for meteorology) is well explained
2.3 Was the contamination acceptably low?	NA	Not applicable
2.4 How well were likely confounding factors identified and controlled?	+	NOx is used to correct for metrological factors
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Not quite clear how reliable the conversion to actual changes in CO concentrations is
3.2 Were the outcome measurement complete?	+	No mention of completeness of data provided, though given that continuously collected pollutant data were used, it is likely that outcome incompleteness would not have led to bias. Some observations were excluded as outliers, but this "was generally less than ten out of 234 possible during three winters."
3.3 Were all important outcomes assessed?	-	Health outcomes were not assessed, nor were changes in concentrations over the intervention and non-intervention time periods
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	The pre-intervention time period was 7 years, while the post-intervention period was only 3 years, but this imbalance in time periods is unlikely to lead to bias
3.5 Was follow-up time meaningful?	++	Follow-up time was sufficient for assessing effectiveness.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	This is not explicitly stated, but given the amount of data that were analyzed, it is likely that the power was sufficient

(*Continued*)

4.2 Were multiple explanatory variables considered in the anlayses?	+	Multiple explanatory variables not included, but NOx was used as a proxy for multiple explanatory variables
4.3 Were the analytical methods appropriate?	-	Analytical methods not optimal for obtaining the intervention effect
4.4 Was the precision of association given or calculable? Is association meaningful?	+	
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	-	The analysis methods may have introduced bias into measurements
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

El-Zein 2007

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Source population are populations in "rapidly urbanized developing countries"
1.2 Is the eligible population or area representative of the source population or area?	+	Children under 17 years admitted to the emergency room of selected hospitals. No other eligibility criteria reported. Not sure whether children under 17 are representative of the source population
1.3 Do the selected participants or areas represent the eligible population or area?	+	"Emergency admissions for respiratory illnesses of children under 17 years of age were selected from 5 (1419 beds) out of 8 (1902 beds) eligible hospitals (= 75%). Accredited hospitals (Class A or B as per the Ministry of Public Health classification) with 50 or more hospitals beds and 24-hour emergency services were considered eligible. No reasons given why 3 hospitals declined to participate - and how they differ in terms of characteristics with the other hospitals"
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	Poorly described. No details listed regarding the actual location of the hospitals, or population characteristics of the patients admitted. Also no info reported where the patients are residing. It is likely that populations (and hospitals) closer to traffic may be impacted more by the intervention than populations further away - no info is provided on this

(*Continued)*

2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No. No discussion.
2.3 Was the contamination acceptably low?	NA	No control group.
2.4 How well were likely confounding factors identified and controlled?	+	Yes. They adjusted for temperature, humidity and rainfall. In addition, they repeated the analysis excluding the months January and February (the typical flu months) since no data on flu was available. In addition, they argue that the study "is restricted to a well-defined age group with limited confounding exposures (e.g., no or minimal smoking, no occupational hazards." No table of baseline characteristics of patients admitted to the ER. New and recurrent admissions counted, not able to report on socio-economic status
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	They have used hospital data from "accredited" hospitals, thus assume that data is reliable. However, it was not reported whether for example International Classification of Diseases (ICD) codes were used for the diagnosis
3.2 Were the outcome measurement complete?	+	They dropped the variable "access to private health insurance from the analyses, used as a proxy for socioeconomic status, because it was found to be poorly recorded at the hospitals". Thus, perhaps we can assume that there were no problems with completeness for the other variables, including the outcome
3.3 Were all important outcomes assessed?	+	No air quality indicators were assessed.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Yes. No comments.
3.5 Was follow-up time meaningful?	++	Yes. No comments.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	Not convinced when looking at Table 1 and 2, but this is partly due to the choice to analyse monthly averages instead of daily averages No sample size calculation.
4.2 Were multiple explanatory variables considered in the anlayses?	++	Yes. No additional comments.
4.3 Were the analytical methods appropriate?	-	Regression analysis was based on at most 20 data points - 10 before and 10 after, because they have used monthly averages instead of daily averages (although collected)
4.4 Was the precision of association given or calculable? Is association meaningful?	-	No variances reported.
Section 5: Summary		

(*Continued*)

5.1 Are the study results internally valid (i. e unbiased)?	-	See concerns with analyses techniques.
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Hasunuma 2014

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Japanese urban areas and children living in urban areas. These aspects are not described in great detail, but the excerpt below provides enough to justify the study in this population. "In Japan, the observed increase has slowed recently, and the prevalence rate of atopic dermatitis has decreased slightly, for unknown reasons. There have been reports of alleviation from respiratory disorders and symptoms in children associated with improvements in the air quality, but knowledge of changes in allergic disorders is limited."
1.2 Is the eligible population or area representative of the source population or area?	+	The eligible population is 3-year-old children in 40 areas in Japan across the nation
1.3 Do the selected participants or areas represent the eligible population or area?	++	The selected participants and areas are from all areas across Japan where data was collected in 3-year-olds from 1997 to 2009 (28 survey areas). These are likely representative of all of Japan (See Fig. 1)
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Intervention and control sites were determined based on where the PM-enforcement law had been enacted. This is not explicitly discussed, but this was likely decided upon based on whether PM levels were of concern (i.e. urban areas with heavy traffic and high exposure to pollutant concentrations). However, not a lot of detail on the assessed areas. Also see baseline differences in Table 2
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	"The following items were considered as confounding factors because they had been reported to be significantly related to the prevalence of asthma and wheezing among the items monitored in the Environmental Health Surveillance: Maternal smoking; allergic predispositions of the child or parent(s); use of nursery school in daytime; presence of a pet animal; and feeding method for 3 months after birth."

(*Continued*)

2.3 Was the contamination acceptably low?	++	Regarding geographical contamination for health outcomes, residents that had not lived in the area of study for at least one year were excluded from the analysis, thus also likely not an issue
2.4 How well were likely confounding factors identified and controlled?	+	A fairly extensive list of covariates was identified, but a few important ones may have been forgotten
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	"The surveys were performed using a self-administered questionnaire. The reliability of the results is considered to be increased by adding the diagnoses by the physicians and objective indices, but this is difficult to implement in a large-scale survey. However, the questionnaire used for the surveys was prepared on the basis of the ATS-DLD questionnaire and was sufficiently validated."
3.2 Were the outcome measurement complete?	+	85% response rate across time periods.
3.3 Were all important outcomes assessed?	++	No comments.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	No comments.
3.5 Was follow-up time meaningful?	++	Four years before and after is sufficiently long to detect and effect and check for longer-term trends perhaps not related to the intervention
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Not discussed, however the estimates and measures of variance suggest that a lack of power was not an issue
4.2 Were multiple explanatory variables considered in the anlayses?	-	In the t-tests relevant for the review, no explanatory variables considered in the analysis
4.3 Were the analytical methods appropriate?	-	t-tests in this case, of course, can detect mean differences, but an analysis including adjustment and consideration of time would have been much more appropriate
4.4 Was the precision of association given or calculable? Is association meaningful?	-	Variance measures are missing.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	+	Analysis and self-reported outcomes are concerns to internal validity

(*Continued*)

5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Likely well generalizable to the general Japanese population

Yorifuji 2016

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Japanese cities with heavy pollution from traffic sources.
1.2 Is the eligible population or area representative of the source population or area?	++	Tokyo is the source area - discussed as heavily trafficked area where traffic policies are very relevant. “Particularly in Tokyo, about 60% of trucks use diesel engines, and diesel vehicles are among the largest contributors to emission of nitrogen dioxide (NO2) and particulate matter (PM) ”
1.3 Do the selected participants or areas represent the eligible population or area?	++	The mortality data are taken for the whole of the two cities. These data are likely very representative.
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	City-wide mortality data from Tokyo and Osaka should be relatively unbiased
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No information given justifying the inclusion of selected covariables (other than rates in Osaka - which is the control site)
2.3 Was the contamination acceptably low?	++	Unlikely that contamination was a large concern for either AQ or health outcomes - although given that trucks are replaced in some cases, it could be that trans-city transport, for example, could have been affected
2.4 How well were likely confounding factors identified and controlled?	++	Authors considered day of the week and public holidays, daily number of influenza patients, temperature, relative humidity
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	The Ministry of Health, Labor, and Welfare in Japan provided electronic data on all deaths in Tokyo’s 23 wards and Osaka.. Likely high quality data, well-controlled data, but this is not described
3.2 Were the outcome measurement complete?	++	Hard to judge, no mention of data completion, but can probably be from the description it seems that the Ministry collects data on “all death”. Additionally, data for all days were available for the analysis (see table 1)

(*Continued*)

3.3 Were all important outcomes assessed?	++	AQ outcomes also assessed descriptively.
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	Yes, though there are some timing issues with the studies as to when the policies were implemented, though the authors mentioned a 7-year grace period for new vehicles to meet the obligation. Especially the use of Osaka as a reference population to account for background trends debatable for the last period (October 2009 to September 2012) because at that time similar policies were also implemented in Osaka
3.5 Was follow-up time meaningful?	++	Yes. No comments.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	No discussion of power, but data from approximately 4400 days used in the health analysis. Also see the relatively narrow confidence intervals
4.2 Were multiple explanatory variables considered in the anlayses?	++	Time-series analysis adjusted for Osaka mortality, same-day temperature, same-day relative humidity, number of influenza patients, public holiday and day of the week
4.3 Were the analytical methods appropriate?	++	Time-series analysis adjusted for mortality trends in a control site and several other relevant covariates, also alternative ways of "controlling" were performed as sensitivity analyses. An actual analysis of AQ levels would have been more informative than simply reporting descriptive statistics
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Confidence intervals for all estimates provided in Tables 3 and 4. None provided for AQ measures, but these were also not analyzed
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	Data included for a long period of time, analyzed as a time-series analysis with adjustment for a control site. Comprehensive list of confounders. Various sensitivity analyses checking methods and included data. Some concerns with Osaka as appropriate control for Tokyo, but given it is the second largest city, there would be no better choice within Japan. There is an issue with the final time period in the analysis, as a similar intervention was introduced in Osaka in 2009. Also slight concern about contamination, if the Tokyo intervention could have had larger geographical implications
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	These results are likely quite gerenalizable to other large cities

Carillo 2013

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Cities in developing countries. "In developing countries…the health effects of the air pollution generated by the growing numbers of vehicles on the road. These vehicles are often "dirtier" than vehicles in developed countries…"
1.2 Is the eligible population or area representative of the source population or area?	+	City of Quito is not very well described. The pollution character of the city is well described in Introduction. "A 2007 emissions inventory for the MDQ indicates that vehicles subject to PyP accounted for 57.7% of CO emissions, 4.4% of SO2 emissions, 18.9% of NOX emissions and 5.1% of PM10 emissions "
1.3 Do the selected participants or areas represent the eligible population or area?	++	See Figure 1: Assessed monitors spread across the city limits, incorporating both the restricted zone and the non-restricted zone
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	Careful explanation behind the selection of intervention and control sites. Given the discussed nature of CO as a very local pollutant, it is likely that selection bias was not introduced because of the site selection
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Yes. Based on existing literature: "The set of variables is based on the pollution meteorology literature and past studies of driving restrictions. They are temperature, relative humidity, precipitation, an indicator variable that takes on a value of 1 for hours in which there is precipitation, solar radiation, atmospheric pressure and wind speed interacted with one of eight dummy variables capturing the eight principal wind directions."
2.3 Was the contamination acceptably low?	+	As described in the study, some contamination may have been possible "This proximity suggests that the stations are suited to being in a control group, but the proximity also raises the possibility that traffic flows in the vicinity of these stations are reduced by PyP, to the extent that traffic into the restricted zone originates or passes through these areas. In addition to these negative traffic spillovers from PyP, positive spillovers are possible, that is, PyP could result in increased traffic flows outside the restricted zone as a result of drivers' avoidance behaviour." Additionally, the temporal control is discussed by authors: "A concern with the use of off-peak hours as a control is the possibility that the policy has induced traffic to shift from peak hours to off-peak hours, though

(*Continued*)

		they provide some evidence that not a lot of traffic shifting occurred.”
2.4 How well were likely confounding factors identified and controlled?	++	Comprehensive list of potentially important variables: station, hour-of-week heterogeneity, time fixed effects and meteorological factors
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Routinely monitored CO data; describe some quality assurance steps such as a US EPA audit of the monitoring system, which concluded that the measurements were of “good quality”
3.2 Were the outcome measurement complete?	++	Authors mention the relative completeness of the data: “We encountered relatively few missing observations in the time series that we worked with.” Also see table 1 (about 5% max) in time series.
3.3 Were all important outcomes assessed?	+	Only CO
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	No comments.
3.5 Was follow-up time meaningful?	++	No comments.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Not discussed, but hourly data over a 5-year period were analyzed, which should have provided sufficient power. Additionally, the variance around all effect estimates were very precise
4.2 Were multiple explanatory variables considered in the anlayses?	++	See Results tables; a range of model specifications including a number of fixed effect terms and covariables of interest
4.3 Were the analytical methods appropriate?	++	Difference-in-difference-in-differences, using both a geographic control as well as a temporal control is a strong methodology to estimate causal effects. In addition, they have explored many modelling choices and assumptions
4.4 Was the precision of association given or calculable? Is association meaningful?	+	Concentrations not provided, and some uncertainty about what is reported in the results tables (Table 7)
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	Difference-in-difference-in-differencess, using both a geographic control as well as a temporal control is a strong methodology to show changes in pre- and post-intervention concentrations

(*Continued*)

5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Davis 2008

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	"...the analysis has implications for air quality and transportation policies throughout the urban developing world." (p 41)
1.2 Is the eligible population or area representative of the source population or area?	++	The study should be representative for the whole of Mexico City
1.3 Do the selected participants or areas represent the eligible population or area?	++	"Air quality in Mexico City is recorded by the Automated Environmental Monitoring Network maintained by the city environmental agency. Established in 1986, the network consists of monitoring stations distributed throughout Mexico City." (p 41)
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Concentrations from monitors across the entire city are used to assess the effectiveness of the intervention. Authors do not discuss how likely it is to assume that the intervention will be effective across such a large geographic area, or discuss the possibility of varying effects dependent on traffic, etc
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No theoretical basis provided.
2.3 Was the contamination acceptably low?	NA	Not applicable.
2.4 How well were likely confounding factors identified and controlled?	++	"The vector of covariates includes indicator variables for month of the year, day of the week, and hour of the day as well as interactions between weekends and hour of the day. In addition, xi, includes weather variables including current and 1-hour lags of quartics in temperature, humidity, and wind speed." (p 48)
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Routine monitoring process likely ensure reliable data.
3.2 Were the outcome measurement complete?	+	No measure of outcome completeness given. As these are routinely collected pollutant data, it is likely that some observations are missing, but not to the extent to cause bias

(*Continued)*

3.3 Were all important outcomes assessed?	+	Neither health outcomes nor a primary AQ outcome included, but a wide range of secondary pollutant outcomes
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Symmetrical time window around the intervention considered
3.5 Was follow-up time meaningful?	++	3-year follow-up sufficient for assessing effectiveness and likely sufficient for protecting against spurious temporal trends
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	No sample size calculation described. With over 200,000 observations per pollutant (Table 1), it is very likely that analysis is sufficiently powered
4.2 Were multiple explanatory variables considered in the anlayses?	++	Yes. No comments.
4.3 Were the analytical methods appropriate?	++	Appropriate analysis for assessing changes related to the intervention, adjusted for important covariates
4.4 Was the precision of association given or calculable? Is association meaningful?	++	No comments.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	Bias introduced through the study methods are unlikely to change the conclusions of the study
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Results are likely relevant for heavily polluted urban areas throughout the developing world

Gallego 2013a

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Source area is Latin American cities with serious air pollution and congestion problems
1.2 Is the eligible population or area representative of the source population or area?	-	Eligible areas are Mexico City and Santiago - they are not well-described
1.3 Do the selected participants or areas represent the eligible population or area?	++	Data from 15 monitoring stations in Mexico City and 7 in Santiago were used. Maps of the locations of the monitors, stratified for income category are provided in online supplement

(*Continued*)

Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	All monitors for the two respective cities were used. From the online supplemental material it can be seen that monitors were spread across the cities, near enough to roads that the intervention should make an impact - it is not likely that substantial selection bias is present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No rationale for explanatory variables included.
2.3 Was the contamination acceptably low?	**na**	Not applicable.
2.4 How well were likely confounding factors identified and controlled?	++	Extensive list of weather and economic covariates considered in analyses
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	CO measured by state-wide monitoring networks in each case - likely that these are well audited and reliable
3.2 Were the outcome measurement complete?	+	"The average failure rate of the network [in Mexico City] is about 31% and roughly constant over time and across days of the week and hours of the day" "Failure rates [in Santiago] are much smaller than in Mexico City (9.4% on average) but there are different patterns before and after TS...we will see below that this measurement change hardly affect our estimations" Some imputation was used in sensitivity analyses and did not make a substantial difference in estimations
3.3 Were all important outcomes assessed?	+	Only CO used, but authors justify this well as a good proxy to assess whether the policies resulted in less traffic during peak hours
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Symmetric window around intervention assessed.
3.5 Was follow-up time meaningful?	++	Follow-up facilitated the impact of the intervention on concentrations in the short term and long term
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Analyses in both cities are fed with over 33,000 observations - likely that any null finding does not stem from too little power
4.2 Were multiple explanatory variables considered in the anlayses?	++	Expansive list of weather and economic-related variables included
4.3 Were the analytical methods appropriate?	++	Strong analysis with several model specifications.

(*Continued*)

4.4 Was the precision of association given or calculable? Is association meaningful?	++	SEs and approximate P values provided.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	Sophisticated analysis methods well-controlled for known confounders and tested in several sensitivity analyses
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Results are likely generalizable to large heavily polluted Latin American cities

Gallego 2013b

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Source area is Latin American cities with serious air pollution and congestion problems
1.2 Is the eligible population or area representative of the source population or area?	-	Eligible areas are Mexico City and Santiago - they are not well described
1.3 Do the selected participants or areas represent the eligible population or area?	++	Data from 15 monitoring stations in Mexico City and 7 in Santiago were used. Maps of the locations of the monitors, stratified for income category are provided in online supplement
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	All monitors for the two respective cities were used. From the online supplemental material it can be seen that monitors were spread across the cities, near enough to roads that the intervention should make an impact - it is not likely that substantial selection bias is present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No rationale for explanatory variables included.
2.3 Was the contamination acceptably low?	NA	Not applicable.
2.4 How well were likely confounding factors identified and controlled?	++	Extensive list of weather and economic covariates considered in analyses
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	CO measured by state-wide monitoring networks in each case - likely that these are well audited and reliable

(*Continued*)

3.2 Were the outcome measurement complete?	+	"The average failure rate of the network [in Mexico City] is about 31% and roughly constant over time and across days of the week and hours of the day" "Failure rates [in Santiago] are much smaller than in Mexico City (9.4% on average) but there are different patterns before and after TS...we will see below that this measurement change hardly affect our estimations" Some imputation was used in sensitivity analyses and did not make a substantial difference in estimations
3.3 Were all important outcomes assessed?	+	Only CO used, but authors justify this well as a good proxy to assess whether the policies resulted in less traffic during peak hours
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Symmetric window around intervention assessed.
3.5 Was follow-up time meaningful?	++	Follow-up facilitated the impact of the intervention on concentrations in the short term and long term
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Analyses in both cities are fed with over 33,000 observations - likely that any null finding does not stem from too little power
4.2 Were multiple explanatory variables considered in the anlayses?	++	Expansive list of weather and economic-related variables included
4.3 Were the analytical methods appropriate?	++	Strong analysis with several model specifications.
4.4 Was the precision of association given or calculable? Is association meaningful?	++	SEs and approximate P values provided.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	Sophisticated analysis methods well controlled for known confounders and tested in several sensitivity analyses
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Results are likely generalizable to large heavily polluted Latin American cities

Viard 2015

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Cities with heavy pollution due to automobiles - not well described; Santiago, Mexico City, Bogota, San Jose, La Paz, Athens, Barcelona, Amsterdam, Tokyo, Honduras, and several Italian cities listed in Footnote 2
1.2 Is the eligible population or area representative of the source population or area?	++	Beijing "Jiang (2006) reports that approximately 53% of Bejing's PM10 is attributable to motor vehicles"
1.3 Do the selected participants or areas represent the eligible population or area?	++	The aggregate API from which PM_{10} concentrations were based, is measured at multiple sites all across Beijing. Likely representative for the city as a whole, as this is the goal of the API
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	It seems that all available monitor stations, a total of 27/28 in Bejing, were used, and that they were pretty well spread out, though details are lacking regarding characteristics of the sites, thus hard to answer
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	"Our pollution regressions include daily weather variables known to affect particulate matter (EPA, 2010)…Higher wind speeds can remove particulates but also import them from neighboring areas…We include daily hours of sunshine to control for atmospheric solar radiation, which creates ozone and particulate mater. Humidity can interact with pollutants to create secondary ones. Rain can interact with existing pollutants to create secondary ones but can also wash particles... Daily maximum surface temperature has an intermediate effect on particulate matter..."
2.3 Was the contamination acceptably low?	**na**	Not applicable.
2.4 How well were likely confounding factors identified and controlled?	++	Comprehensive list of weather-related and other potentially important confounders
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	PM_{10} was derived using API values, and this comes with some uncertainty. Also exact sampling methods for PM_{10} were not described in the paper, though they used governmental sites, thus standard methods and QA/QC procedures can be assumed
3.2 Were the outcome measurement complete?	+	Because PM_{10} was derived using API, there was some data missing for 143 days when the API was below 50 and the maximal pollutant unknown, 29 days when the worst pollutant was other than PM_{10}, and 7 days when the API was above 50 but the pollutant identity is missing; making the total number of observations available for PM_{10} analysis 917 (compared to 1096 for API analyses, which is 84%)

(*Continued)*

3.3 Were all important outcomes assessed?	+	Only PM_{10}
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	Post period was shorter than pre period for the one-day policies. Also their aggregated PM_{10} measure was based on a network that differs a bit over the years - both in composition and the number of sites (five stations are dropped in 2008 and 2009, four additional ones added)
3.5 Was follow-up time meaningful?	++	Approximately 1.5 years pre- and post-intervention (although the 'evolution' of the intervention after the initial point may have caused this to introduce some bias)
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Not discussed, but effect estimates and precision shown in Table 2 show that power was not an issue in the models
4.2 Were multiple explanatory variables considered in the anlayses?	++	See Table 2.
4.3 Were the analytical methods appropriate?	++	Method allows for the assessment of the intervention effect (of several different "intervention" stages), while still checking for underlying time trends
4.4 Was the precision of association given or calculable? Is association meaningful?	++	See Table 2.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	Other than concerns with the conversion from API to PM10, the study seems to be highly internally valid
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Kim 2011

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	No information given; urban areas in Korea
1.2 Is the eligible population or area representative of the source population or area?	+	Urban areas in 7 major cities. It seems that they included all major cities. Not much detail is given however

(*Continued*)

1.3 Do the selected participants or areas represent the eligible population or area?	+	Many different stations in metropolitan areas, but no information where exactly. The 16 road sites were selected out of the available 30 sites because they had long-term coverage. There is no detailed description about the representativeness but it is likely that they represent the near road environment rather than the greater urban environment. There is no description about the representativeness of the background sites either
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	Choosing the existing ones that are available for time frame. But no information where the control sites are, why these are chosen (except for one example, but also this example is not ideal). Furthermore, they selected only road sites, thus probably not representative of urban areas
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	There is some discussion about confounding or other factors potentially influencing the effect of the intervention on air quality, for example related to the Asian Dust (AD) effect and seasonal change (table for season and citations included)
2.3 Was the contamination acceptably low?	+	Probably, but since a few interventions were implemented nationwide, the selected regional background sites may be impacted as well, and no information is given about the exact siting of those sites
2.4 How well were likely confounding factors identified and controlled?	+	they looked at some confounders, such as seasonal data and time trend (MK test and season variation)
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	They report minimum detectable sensitivity and precision (less than 1%) of the instrument as reported by the manufacturer, but no specific QA/QC procedures and results are reported
3.2 Were the outcome measurement complete?	-	They do not explicitly report except that for Period 1 MK test was not possible to conduct due to missing data. Also looking at Fig. 2 missing data points for Period 1 are visible
3.3 Were all important outcomes assessed?	-	PM_{10} is not a good indicator to investigate interventions targeting to reduce diesel emissions. Also no health outcomes assessed
3.4 Was there a similar follow-up time in exposure & comparison groups?	-	Before period was much shorter (1998 to 2000), then after period because main intervention implemented started in June 2000 (until 2007 (although gradually introduced))
3.5 Was follow-up time meaningful?	++	Yes, they were interested in the long-term effects of those interventions, thus a follow-up of about 7 years is meaningful
Section 4: Analyses		

(*Continued*)

4.1 Was the study sufficiently powered to detect an effect if one exists?	+	No information provided, but power should not be an issue given the amount of data assessed
4.2 Were multiple explanatory variables considered in the anlayses?	-	Not included in analysis of effect estimates relevant for the review (t-test)
4.3 Were the analytical methods appropriate?	-	Likely t-test, do not mention method explicitly.
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Yes. P values and SDs.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	There are some methodological concerns, but the long study period and the number of monitoring sites can be seen as strengths of the study
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Cowie 2012

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Individuals in large Australian city exposed to air pollution from traffic sources
1.2 Is the eligible population or area representative of the source population or area?	++	Yes. The area representative of the source area is the "Lane Cove Tunnel (LCT) study area, approximately a 5*10 km area, incorporating motorways and other major and local roads". The opening of this road traffic tunnel in March 2007 presented the opportunity to study the effect of a local traffic intervention on air quality in the vicinity of the tunnel and the bypassed main road
1.3 Do the selected participants or areas represent the eligible population or area?	++	Yes, they have established 4 fixed site monitors "as part of the planning conditions for the construction of the tunnel representative of community exposures in background locations"
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	The monitors were carefully selected, see also question 1.3. Meaningful selection bias not likely
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No theoretical basis.

(*Continued*)

2.3 Was the contamination acceptably low?	++	Yes, they adjusted for changes in regional air pollution levels. It is very likely that the regional background sites were not influenced by the intervention under study, given that they were at least 6 km away from the study area
2.4 How well were likely confounding factors identified and controlled?	++	Analyses were adjusted for local weather conditions (daily changes in temperature, wind direction weighted by speed, and wind speed). In addition, they adjusted for changes in regional air pollution levels. Also they accounted for autocorrelation of daily values
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Yes. They used standardized methods at the fixed sites. In addition, they report that "independent audits of the equipment, processes, and reporting were conducted twice per year." For the passive monitoring campaign, QA/QC results were reported, and show good agreements with standardized methods at the fixed sites
3.2 Were the outcome measurement complete?	**nr**	This was not reported. Note: they only report that $PM_{2.5}$ was only measured at one of the three control sites that were used to adjust for changes in regional air pollution levels
3.3 Were all important outcomes assessed?	+	No health outcomes are reported.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Yes. Data were collected before the tunnel opened, from 25 March 2006 to 24 March 2007 (year 1), and after the tunnel opened, from 25 March 2007 to 24 March 2008, (year 2) and from 25 March 2008 to 24 March 2009 (year 3) . Thus exactly the same time period before and after
3.5 Was follow-up time meaningful?	++	A follow-up of two years after the opening of the tunnel is sufficient to assess the effect, and to assess whether the effect is sustainable over time
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Although no formal power calculation was presented, the study seems sufficiently powered. For example a decrease of 0.73 ppb in NO2 reached statistical significance (see table 1, last column)
4.2 Were multiple explanatory variables considered in the anlayses?	++	Analysis was adjusted for local weather conditions by including as covariates daily changes in temperature, wind direction weighted by speed, and wind speed. In addition, analysis was adjusted for changes in regional air quality
4.3 Were the analytical methods appropriate?	++	Yes. No comments.
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Yes, SDs are reported, 95 % Cis and P values.
Section 5: Summary		

(*Continued*)

5.1 Are the study results internally valid (i. e unbiased)?	++	Reliable data, well-selected sites and a strong analysis point to strong internal validity for this study
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Gramsch 2013

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Unclear what the source area should be considered - authors talk almost solely of Santiago in both the introduction and the discussion sections (except for one sentence citing the Beijing Olympic Games in the discussion). But it can be assumed that the source population could be any large city in Latin America heavily impacted by traffic
1.2 Is the eligible population or area representative of the source population or area?	++	Santiago, Chile - well described e.g. “The number of trips in a working day in Santiago is 16.3 million, from which 10.1 million are done in vehicles.” “According to emission estimates for Santiago, traffic is the largest source of air pollution (Dictuc, 2007) accounting for 37% of the PM10 emissions, 35% of the PM2.5 emissions and 90% of CO emissions”
1.3 Do the selected participants or areas represent the eligible population or area?	++	See p155-156 and Figure 1 - a lot of information is provided on the individual monitoring sites. These should be well representative of the city
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Selection was based on whether an intervention effect was expected or not. Authors describe that the Usach, Alameda and Departamental sites were “directly influenced by the Tranantiago project”, while the E. Yanez site “was not influenced…because it had no circulation of public transportation before or after Transantiago”. Authors also claim that the predominant wind direction likely helped to avoid contamination at E. Yanez. Although with only 1 control site there is still a risk of selection bias
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No theoretical basis.

(*Continued*)

2.3 Was the contamination acceptably low?	-	Authors address the risk of contamination: "This site is about 150 m south of Alameda, however, it is unlikely that contamination from this street influences E. Yañez site because the predominant wind direction is south-west" But, given that the policy restructured Santiago's entire public transportation system, it is likely that it impacted air pollution at a larger scale and as such the control may be influenced by the policy
2.4 How well were likely confounding factors identified and controlled?	++	They describe meteo conditions before and after. They document that wind conditions are similar for the two periods (though no data shown for 2007). In addition, they compare air pollution data from an urban background site (Parque O'Higgins) to argue that the differences between these two years are related mostly to meteorological conditions. For example, they describe that in year 2005, the period with cold fronts and rain coming from the south started in May and lasted until July. In the year 2007, the cold fronts started in June and lasted until July
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	BC was measured with instruments that measure black carbon using an optical method, built at the University of Santiago, thus not a standard device such as the aethalometer. Corrections have been made to transform the measured absorption coefficient to BC concentrations, using co-located measurements. Though this is common, they report large changes in the absorption coefficient from 2005 to 2007 , which the authors attribute to changes in the chemical composition over time. That may be one possibility. But they also report that different instruments were used to measure BC, the instrument used in 2007 was of better quality than the one used in 2005
3.2 Were the outcome measurement complete?	+	Large differences in the number of observation between the two time periods, as well as across different sites. "The measurements in Eliodoro Yañez in 2007 had considerable more errors than the other stations. There were many electricity failures in this station resulting in loss of data."
3.3 Were all important outcomes assessed?	+	Only BC, with some assessment of CO and PM_{10} as well (not included due to study design)
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	All monitoring sites had different numbers of observations and measurements were not taken simultaneously, and the control site drew from the least amount of observations, which could potentially lead to bias
3.5 Was follow-up time meaningful?	++	No comments.
Section 4: Analyses		

(*Continued*)

4.1 Was the study sufficiently powered to detect an effect if one exists?	+	No discussion of power - but for Alameda, for example, the non-significant differences observed could be underpowered
4.2 Were multiple explanatory variables considered in the anlayses?	++	Regression models controlled for time of day, wind speed, relative humidity, wind direction and temperature
4.3 Were the analytical methods appropriate?	-	Some of the predictor variables may correlate highly. Analyses conducted for each site separately, and most importantly no statistical tests are provided to test whether changes are different from the intervention sites compared to the control site
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Tables 2 and 4 and Figure 9 all include relevant measures of precision
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	Some concerns with the internal validity, including potential selection bias, missing observations, and suboptimal analyses lead to some concerns with internal validity
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Well-generalizable to Santiago as a whole.

Peel 2010

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	US population in major cities.
1.2 Is the eligible population or area representative of the source population or area?	+	Metropolitan area of Atlanta.
1.3 Do the selected participants or areas represent the eligible population or area?	**++ (health)** **- (AQ)**	Health: hospital data should be well-representative of Atlanta AQ: Data based on limited number of monitors.
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	**+ (health)** **++ (AQ)**	Health: only 12 hospitals responded and provided data. AQ: all available monitoring sites used.
2.2 Was the selection of explanatory variables based on sound theoretical basis?	**- (health)** **NA (AQ)**	Health: No explanation provided. AQ: Not applicable.

(*Continued*)

2.3 Was the contamination acceptably low?	**NA (Health)** **++ (AQ)**	Health: Not applicable. AQ: no contamination based on the various geographical controls assessed
2.4 How well were likely confounding factors identified and controlled?	**++ (Health)** **- (AQ)**	Health: Numerous potential confounders included in the analysis AQ: None included.
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Routine data, as well as hospital data and electronic records assessed
3.2 Were the outcome measurement complete?	**++ (health)** **+ (AQ)**	Health: unlikely that missing hospital record data led to substantial bias AQ: Some issues with missing data: "CO values were missing from site B for 10 of 17 days within the Olympic Games period; therefore, we excluded this site from further analyses. Data from other sites were complete during the Olympic Games period and nearly complete during the Olympic Games baseline periods (the other site for CO, site A, was missing 2 of 73 days; site Cwas missing 1 day for NO2; site D was missing 2 days for NO2; site C was missing 1 day for O3; all other sites had data for all 73 days)"
3.3 Were all important outcomes assessed?	++	Both a range of AQ outcomes and emergency department visits were assessed
3.4 Was there a similar follow-up time in exposure & comparison groups?	-	No; the Olympic period was much shorter than the pre- and post-Olympic periods
3.5 Was follow-up time meaningful?	++	10 years measured summer periods including time period of Olympic period and 4 weeks before and 4 weeks after this period
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	+	Given the amount of data analyzed it is unlikely that power was an issue
4.2 Were multiple explanatory variables considered in the anlayses?	**++ (health)** **- (AQ)**	Health: day-of-week, daily minimum temperature (lag 1), daily average dew point temperature (lag 1), linear, quadratic, and cubic terms for day-of-summer, an indicator variable for 1996 (compared with all other years), and an interaction term between the year indicator and the Olympic period indicator AQ: None included in the analysis.

(*Continued*)

4.3 Were the analytical methods appropriate?	**++ (health)** **+ (AQ)**	Health: Poisson GLMs adjusted for potential confounders and secular trend AQ: GLM but unadjusted models shown only before-after comparison within each site
4.4 Was the precision of association given or calculable? Is association meaningful?	++	See detailed manuscript tables.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	**++ (health)** **+ (AQ)**	Strong design for both health and AQ analyses; lack of potentially confounding factors for AQ analysis perhaps of slight concern
5.2 Are the results generalisable to the source population (i.e externally valid)?	**++ (health)** **+ (AQ)**	See section 1 above.

Titos 2015a

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Medium-sized European cities.
1.2 Is the eligible population or area representative of the source population or area?	++	Cities of Ljubljana and Granada heavily affected by vehicular traffic (well described)
1.3 Do the selected participants or areas represent the eligible population or area?	+	3 sites per city selected: 2 intervention sites which can be considered traffic sites and 1 urban background site
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	The use of urban background sites as control sites is at least debatable. It would have been better to compare the intervention sites with similar traffic sites - without the intervention. Also, the small number of sites cannot exclude the possibility that there is not bias present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	Limited mention of the effect of temperature, and how resulting wood smoke pollution (p 21 under Measurements), and how this could bias observed effects
2.3 Was the contamination acceptably low?	++	Very local effect expected, does not appear that the intervention influenced urban background

(*Continued)*

2.4 How well were likely confounding factors identified and controlled?	+	Ljubljana: temperature “corrected” for using source apportionment
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Two different aethalometers were used for BC, and authors describe in detail how they correct the results for systematic differences, thus no remaining large issue
3.2 Were the outcome measurement complete?	+	Not mentioned, and some data were excluded due to weather or other external factors such as road construction, but from time series it would appear that there is not substantial missing data. Also no BC measurements before the intervention at Palacio de Congresos
3.3 Were all important outcomes assessed?	+	BC only (for our review), however authors argue that this is a key indicator for monitoring changes in concentrations due to changes in traffic
3.4 Was there a similar follow-up time in exposure & comparison groups?	+	Pre-intervention measurements from summer and post-intervention measurements from winter; this is a concern that is “corrected” for by using source apportionment to only include BC from traffic
3.5 Was follow-up time meaningful?	++	Follow-up sufficient to detect a meaningful effect.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	Study underpowered. Large SDs are reported, see Table 3, e.g. a quite large 14% reduction in BC at CEAMA site does not reach statistical significance
4.2 Were multiple explanatory variables considered in the anlayses?	-	No correction for important confounders, other than simple subtractions of background concentrations
4.3 Were the analytical methods appropriate?	+	Simple t tests conducted. For example, no statistical test conducted to see whether the changes observed at the control site were different than at the intervention sites. No correction for important confounders, other than simple substractions of background concentrations
4.4 Was the precision of association given or calculable? Is association meaningful?	++	No comments
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	Concerns with the selection of sites, the reliability of the data and the analysis methods
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Well-generalizable to mid-sized European cities.

Titos 2015b

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Medium-sized European cities.
1.2 Is the eligible population or area representative of the source population or area?	++	Cities of Ljubljana and Granada heavily affected by vehicular traffic (well described)
1.3 Do the selected participants or areas represent the eligible population or area?	+	3 sites per city selected: 2 intervention sites which can be considered traffic sites and 1 urban background site
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	The use of urban background sites as control sites is at least debatable. It would have been better to compare the intervention sites with similar traffic sites - without the intervention. Also, the small number of sites cannot exclude the possibility that there is not bias present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	Limited mention of the effect of temperature, and how resulting wood smoke pollution (p 21 under Measurements), and how this could bias observed effects
2.3 Was the contamination acceptably low?		Implementation of a new public transportation system is expected to have wider impacts, with urban background levels potentially being affected as well
2.4 How well were likely confounding factors identified and controlled?	+	
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Meteo variables were described for the before and after period in the text: wind speed and direction; temperature; relative humidity
3.2 Were the outcome measurement complete?	++	Two different aethalometers were used for BC, and authors describe in detail how they corrected the results for systematic differences, thus no remaining large issue
3.3 Were all important outcomes assessed?	+	BC only (for our review), however authors argue that this is a key indicator for monitoring changes in concentrations due to changes in traffic
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	No comment.
3.5 Was follow-up time meaningful?	++	Follow-up sufficient to detect a meaningful effect.
Section 4: Analyses		

(*Continued*)

4.1 Was the study sufficiently powered to detect an effect if one exists?	-	Study underpowered. Large SDs are reported, see Table 3, e.g. a quite large 14% reduction in BC at CEAMA site does not reach statistical significance
4.2 Were multiple explanatory variables considered in the anlayses?	-	No correction for important confounders, other than simple substractions of background concentrations
4.3 Were the analytical methods appropriate?	+	Simple t tests conducted. For example, no statistical test conducted to see whether the changes observed at the control site were different than at the intervention sites. No correction for important confounders, other than simple substractions of background concentrations
4.4 Was the precision of association given or calculable? Is association meaningful?	++	No comment.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	Concerns with the selection of sites, the reliability of the data and the analysis methods
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Well-generalizable to mid-sized European cities.

Boogaard 2012

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	European cities (introduction European standard and cities implementing LEZs mentioned)
1.2 Is the eligible population or area representative of the source population or area?	+	Not very much information given about the areas but enforcement mentioned and geographical information given
1.3 Do the selected participants or areas represent the eligible population or area?	+	1 or 2 monitoring sites per intervention city and 1 monitoring site per control site
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Range of monitors assessed, some selection bias could be present
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Provided.
2.3 Was the contamination acceptably low?	+	“suburban background locations (likely not affected by LEZ).”

(*Continued*)

2.4 How well were likely confounding factors identified and controlled?	++	Weather data measured, temporal trend.
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Likely that routinely monitored data are reliable.
3.2 Were the outcome measurement complete?	+	Mixed - they had to exclude some measurements due to unexpected road reparations, they also had to exclude PM_{10} and $PM_{2.5}$ concentrations from the first sampling week of the 2010 sampling period because a problem occurred in the pre-weighing of the filters
3.3 Were all important outcomes assessed?	+	AQ outcomes only.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Same follow-up.
3.5 Was follow-up time meaningful?	++	2 years sufficient for assessing effect.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Differences due to LEZ are too small to be detected but this is not due to limited power of study
4.2 Were multiple explanatory variables considered in the anlayses?	-	They also applied a regression analysis including wind speed but results are not shown
4.3 Were the analytical methods appropriate?	+	t-test comparison of means.
4.4 Was the precision of association given or calculable? Is association meaningful?	++	
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	+	Slight concerns with the selection of sites, as well as the analysis
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Morfeld 2013

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Not directly described, but they talk about German cities implementing LEZs, therefore they could be the source population
1.2 Is the eligible population or area representative of the source population or area?	++	Munich
1.3 Do the selected participants or areas represent the eligible population or area?	-	The authors describe in discussion that the chosen 5 monitoring stations do not represent the Munich population very well
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	The authors say that the monitoring stations were the ones available. For the reference stations, they do not give any explanations why they took Johanneskirchen as the reference station included in analysis and not the other station. They correct for baseline data in regression model
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	They give references for meteorological data as important confounding variables
2.3 Was the contamination acceptably low?	+	Monitor located outside of LEZ could still be influenced by the LEZ
2.4 How well were likely confounding factors identified and controlled?	++	Meteorological data, baseline data, days of LKW traffic excluded
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Measurements of PM_{10} through two difference techniques.
3.2 Were the outcome measurement complete?	++	Regulatory monitoring data likely complete.
3.3 Were all important outcomes assessed?	+	only PM_{10}
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Same follow-up.
3.5 Was follow-up time meaningful?	+	The authors do not report anything about intervention implementation/fidelity. It seems no exploratory analysis was done or information collected whether the LEZ was being obliged and e.g. incentives or penalties given in case of breach. Therefore it is difficult to judge whether follow-up time was meaningful, taking only one year after implementation of the regulation
Section 4: Analyses		

(*Continued*)

4.1 Was the study sufficiently powered to detect an effect if one exists?	++	No comment.
4.2 Were multiple explanatory variables considered in the anlayses?	++	Meteorological and baseline data were included as variables. Furthermore days with LKW traffic were excluded from analysis; indirectly also for time trend was adjusted by including a reference station and comparing the reference and index stations on the exact date/time with each other and calculating their difference to evaluate an intervention effect
4.3 Were the analytical methods appropriate?	++	New statistical method developed to compare pre-post pollutant data from an intervention and reference station
4.4 Was the precision of association given or calculable? Is association meaningful?	++	CIs and SEs given.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	+	Slight concerns with the selection of sites and potential contamination
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	See section 1 above.

Fensterer 2014

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	European cities
1.2 Is the eligible population or area representative of the source population or area?	++	Munich - not discussed in detail but likely generalizable to many other European cities
1.3 Do the selected participants or areas represent the eligible population or area?	+	Streetside, urban background and regional background should ensure that the selected areas are representative of Munich. However number of sites limited
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Though the selection of a control site outside the LEZ is appropriate, they have corrected both intervention sites (the street site and the urban background site) using the same control site, which was characterized as a regional background site (which seems more an urban background site). Perhaps it would have been even better to have at least an additional control site at a street location outside the LEZ to correct the street intervention site

(*Continued*)

2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	See detailed reasons for including various covariates (wind direction, seasonal variation etc.) and excluding others (temperature and precipitation) on pp 5098 and 5099
2.3 Was the contamination acceptably low?	+	"The measurements at the reference station represented the regional background pollution level, which was mostly not affected by the measures."
2.4 How well were likely confounding factors identified and controlled?	++	Comprehensive list of variables considered (both those included in the regression analysis, as well as those that are not important due to the inclusion of the reference group)
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Not reported, but routinely collected PM_{10} data by the Bavarian Environment Agency are likely reliable
3.2 Were the outcome measurement complete?	+	Data missing for the street site for the summer post period, because the site was closed from 1 July 2010 to 30 September 2010
3.3 Were all important outcomes assessed?	+	Only PM_{10}, but this was the focus of the study because it is the pollutant monitored to track air quality guidelines in Europe. However, PM_{10} is not a good indicator to evaluate a traffic policy such as the LEZ
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Same follow-up.
3.5 Was follow-up time meaningful?	++	No comment.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	No discussion of power, but calculated estimates show that power was not an issue
4.2 Were multiple explanatory variables considered in the anlayses?	++	PM_{10} reference values; wind direction; public holiday (discuss that temperature and precipitation are not important, as they are implicitly included in the reference station values)
4.3 Were the analytical methods appropriate?	++	Yes, very extensive analyses. Perhaps a bit too many variables in the model (e.g. 4 interaction terms), but it seems that they had enough data to allow that, so no real concerns
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Provided and meaningful.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	No concerns regarding the internal validity of the study.

(*Continued*)

5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Study likely well generalizable to European cities.

Morfeld 2014

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	European cities with elevated levels of traffic-related pollutants (PM and NO2) "Values were and are in excess: about 69% of all stations near to traffic showed annual averages higher than 40 mg/m3 in Germany. This non-compliance is not restricted to Germany but the European limit value for NO2 is exceeded in many European cities"
1.2 Is the eligible population or area representative of the source population or area?	+	German LEZs "…as many as eligible" Authors describes how the LEZs across Europe are quite heterogeneous, but these should still nevertheless be somewhat generalizable for those across Europe
1.3 Do the selected participants or areas represent the eligible population or area?	++	17 of the 34 active LEZs at the time of the study were included based on the study inclusion criteria
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Little information is given about the location or characteristics of the index and reference monitors (other than that they are inside or outside of the respective LEZs). The following sensitivity analysis, however, does suggest that any bias based on the "type" of station (i.e. whether it was background, industry, traffic) actually leads to a conservative bias. "The NO2 analysis was based on 192 comparisons of index vs reference stations, among them were 31 index stations characterized as "background", one characterized as "industry" and 160 as "traffic" stations. We performed a sensitivity analysis by restricting the evaluation to the stations close to traffic. The additive linear type 2 model estimated an effect of -1.73 ug/m3 at all index stations. When the analysis only accounted for the traffic stations we got a slightly more pronounced LEZ effect estimate of -2. ug/m3."
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Yes. Selected parameters for statistical models all based off of cited literature. See p 13
2.3 Was the contamination acceptably low?	+	Supplemental Figures S1-S14 show the geographical locations of the various LEZs, as well as index and reference stations. As the "intervention effect" is not

(*Continued*)

		constrained to the borders as seen on these maps, any reference stations close to the LEZ borders could introduce the potential for contamination. In cities such as Karlsruhe, Munich, Frankfurt am Main and Berlin there are stations where such contamination may have been relevant. However, on the aggregate analysis level, it is unlikely that this made a huge difference, and any bias would have likely led to more conservative effect estimates
2.4 How well were likely confounding factors identified and controlled?	++	Models 1 and 2 represent a well controlled, and even more extended model
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	"The original NO2 and NO measurements were performed by the Environmental State Institutions in Germany (Landesumweltämter)."
3.2 Were the outcome measurement complete?	+	NR - however given that this are regulatory data collected by the State Institutions, quality control and assurance processes are likely, and it is unlikely that substantial amounts of data were missing
3.3 Were all important outcomes assessed?	+	AQ outcomes
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Identical measurement procedures at index and reference monitors
3.5 Was follow-up time meaningful?	++	No further comments.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	See effect estimates and confidence intervals (even for small effects, very tight and significant confidence intervals were calculated)
4.2 Were multiple explanatory variables considered in the anlayses?	++	See above.
4.3 Were the analytical methods appropriate?	++	Regression of matched intervention and reference stations pre- and post-intervention
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Provided. Questionable whether the small effect estimates are relevant
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	Most aspects that could have led to bias (selection of monitors; contamination) would have led to more conservative effect estimates). Difference-in-difference with measurements coming from 364 days previous is very sound methodologically. Additionally, multiple analysis (linear vs. log-linear; continuous data vs. continuous and diffuse sampler data; model 1 vs. model 2) showed mostly consistent estimates across outcomes

(*Continued*)

5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Bel 2013a

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Metropolitan residential zones, possibly in Europe - not well described
1.2 Is the eligible population or area representative of the source population or area?	+	Barcelona - not well described.
1.3 Do the selected participants or areas represent the eligible population or area?	++	They have used 15 air quality monitoring stations from government sites. “Barcelona metropolitan area has one of the densest networks of such stations in Europe”
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	They have used 15 air quality monitoring sites, but there are no further details regarding characteristics of the sites, exact location, and whether they are located in the treatment zone (zones with an 80 km/h speed limit or zones with a variable speed limit) or in the control zone (zones with neither an 80 km/h speed limit nor a variable speed limit), and if in the control zone at what distance to the intervention zone. It is impossible to assess the comparability of the intervention sites and the control sites since no data is provided, thus hard to say something about potential selection bias
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	All included methods discussed in detail in the Methods section “The explanatory variables selected (see Table 1 for these and their main descriptive statistics) aim to capture the variability in pollutant sources and the transport, sedimentation and/or reaction of the pollutants.”
2.3 Was the contamination acceptably low?	-	Given that all monitors are in Barcelona, it is likely that some contamination exists
2.4 How well were likely confounding factors identified and controlled?	+	The authors adjusted for important confounder variables (See also question 4.2). There is no descriptive comparison of confounder variables before and after the policies, and also not for the intervention and control group separately, and the lack of description of certain included variables in the methods (year dummy variables)
Section 3: Outcomes		

(*Continued)*

3.1 Were the outcome measures and procedures reliable?	+	Assume this is the case since they used government sites, but exact measurement methods and QA/QC procedures are not documented
3.2 Were the outcome measurement complete?	-	A lot of missing data reported for PM_{10}, with only ~30% (626/1826) of the data available per site. The authors mention that they sampled PM_{10} "manually on a daily basis, which means few measurements are available for weekends and holidays". Note the sample size in Tables 3 and 4 which is for PM_{10} only ~20% of that of NOx. There seems to be a mismatch between the observations available as reported in Table 1, and the final sample size in Tables 3-4. This is unclear
3.3 Were all important outcomes assessed?	+	Only AQ outcomes.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	More data available in post period compared to pre policy for variable speed intervention, but unlikely this is a serious concern
3.5 Was follow-up time meaningful?	++	Two years before and three after are sufficient for detecting the effect of interest
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Unclear how many observations were used in the analysis and measure of precision not provided, but the size of the P values would suggest that power was not an issue for the study
4.2 Were multiple explanatory variables considered in the anlayses?	++	As shown in tables 3 and 4, potentially important covariables included in the models - traffic, temperature, relative humidity, precipitation, wind speed, atmospheric pressure, and years, though the last is not clearly listed
4.3 Were the analytical methods appropriate?	++	Difference-in-difference adjusted for autocorellation is a strong method for estimating causal effects. Basic analysis assumptions (i.e. that the pre-intervention trends were similar among treatment and control sites) also tested. In addition, they used overall PM_{10} levels instead of the traffic contribution only. This is less an issue for NOx, which is a much better indicator of traffic-related air pollution. Note: as some sites already had a speed limit of 80 km/h at the beginning of the study, these had a value of 1 for the whole study period (thus not technically a full CBA). We are not sure to what extent this would affect the analysis
4.4 Was the precision of association given or calculable? Is association meaningful?	-	P values provided, but measures of variability for the effects of interest are not provided
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	Strong analysis, but the nature of the intervention and control sites is somewhat questionable. First of all, there is no information characterizing the monitors, or how many and which ones belonged to the intervention and control groups; secondly, we have no information on the location of the monitors; it is likely

(*Continued*)

		given that all were geographically close, that contamination may have been an issue here and that the effect estimates would have been impacted. Also the use of the indicator variable (where some sites were 1 for the whole period) is somewhat questionable
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	Once again the lack of information on the monitors and their location limits the generalizable of the results

Bel 2013b

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Metropolitan residential zones, possibly in Europe - not well described
1.2 Is the eligible population or area representative of the source population or area?	+	Barcelona - not well described
1.3 Do the selected participants or areas represent the eligible population or area?	++	They have used 15 air quality monitoring stations from government sites. "Barcelona metropolitan area has one of the densest networks of such stations in Europe"
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	-	They have used 15 air quality monitoring sites, but there are no further details regarding characteristics of the sites, exact location, and whether they are located in the treatment zone (zones with an 80 km/h speed limit or zones with a variable speed limit) or in the control zone (zones with neither an 80 km/h speed limit nor a variable speed limit), and if in the control zone at what distance to the intervention zone. It is impossible to assess the comparability of the intervention sites and the control sites since no data is provided, thus hard to say something about potential selection bias
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	All included methods discussed in detail in the methods section "The explanatory variables selected (see Table 1 for these and their main descriptive statistics) aim to capture the variability in pollutant sources and the transport, sedimentation and/or reaction of the pollutants."
2.3 Was the contamination acceptably low?	-	Given that all monitors are in Barcelona, it is likely that some contamination exists
2.4 How well were likely confounding factors identified and controlled?	+	The authors adjusted for important confounder variables (See also question 4.2). There is no descriptive comparison of confounder variables before and after the policies, and also not for the intervention and control group separately,

(*Continued*)

		and the lack of description of certain included variables in the methods (year dummy variables)
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Assume this is the case since they used government sites, but exact measurement methods and QA/QC procedures are not documented
3.2 Were the outcome measurement complete?	-	A lot of missing data reported for PM_{10}, with only ~30% (626/1826) of the data available per site. The authors mention that they sampled PM_{10} "manually on a daily basis, which means few measurements are available for weekends and holidays". Note the sample size in Tables 3 and 4 which is for PM_{10} only ~20% of that of NOx. There seems to be a mismatch between the observations available as reported in Table 1, and the final sample size in Tables 3-4. This is unclear
3.3 Were all important outcomes assessed?	+	Only AQ outcomes.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	More data available in post period compared to pre policy for variable speed intervention, but unlikely this is a serious concern
3.5 Was follow-up time meaningful?	++	Two years before and three after are sufficient for detecting the effect of interest
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Unclear how many observations were used in the analysis and measure of precision not provided, but the size of the P values would suggest that power was not an issue for the study
4.2 Were multiple explanatory variables considered in the anlayses?	++	As shown in tables 3 and 4, potentially important covariables included in the models - traffic, temperature, relative humidity, precipitation, wind speed, atmospheric pressure, and years, though the last is not clearly listed
4.3 Were the analytical methods appropriate?	++	Difference-in-difference adjusted for autocorellation is a strong method for estimating causal effects. Basic analysis assumptions (i.e. that the pre-intervention trends were similar among treatment and control sites) also tested. In addition, they used overall PM_{10} levels instead of the traffic contribution only. This is less an issue for NOx, which is a much better indicator of traffic-related air pollution. Note: as some sites already had a speed limit of 80 km/h at the beginning of the study, these had a value of 1 for the whole study period (thus not technically a full CBA). We are not sure to what extent this would affect the analysis
4.4 Was the precision of association given or calculable? Is association meaningful?	-	P values provided, but measures of variability for the effects of interest are not provided
Section 5: Summary		

(*Continued*)

5.1 Are the study results internally valid (i.e unbiased)?	+	Strong analysis, but the nature of the intervention and control sites is somewhat questionable. First of all, there is no information characterizing the monitors, or how many and which ones belonged to the intervention and control groups; secondly, we have no information on the location of the monitors; it is likely given that all were geographically close, that contamination may have been an issue here and that the effect estimates would have been impacted. Also the use of the indicator variable (where some sites were 1 for the whole period) is somewhat questionable
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	Once again the lack of information on the monitors and their location limits the generalizability of the results

Dijkema 2008

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Urban areas and populations affected by traffic pollution in Europe. Not much detail given, however
1.2 Is the eligible population or area representative of the source population or area?	++	Yes. An urban area in Amsterdam, the Netherlands impacted by an urban highway. Specifically in the introduction, they describe that "approximately 40,500 people live within close proximity that is within 500 m of the road section where the intervention was taken". The area is fairly well described
1.3 Do the selected participants or areas represent the eligible population or area?	++	They have selected two road sites located on the same highway, one affected by the intervention (A10W) at 6.7 m distance to highway, and one chosen as a control site (A10S) at 8 m distance to the highway. In addition, data on urban background concentrations (BN, BC, BW) are available from at least two urban background monitoring stations. They use the latter data to derive a 'traffic contribution' concentration
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	The intervention section is located on the western side, where there are apartment buildings < 20 m from the road and thus resembling a street canyon which is different from the control side where no buildings are present next to the road. If they use those two sites to represent the area affected (< 500 m), then some selection bias is likely because areas and populations closer by would be more affected than areas further away from the highway
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Yes. Authors referenced, for example, other published articles when explaining why they subtract urban background concentrations from roadside concentrations to derived 'traffic contribution' concentrations

(*Continued*)

2.3 Was the contamination acceptably low?	++	Yes. It is unlikely that the intervention impacted the control site. One possible scenario would be if the intervention impacted urban background sites in any substantial way and then subtracting those levels would lead to underestimations of the intervention, but it is unlikely that this substantially impacted results. Different areas on the highway, with the speed limit only applying to the western section. Risk of contamination therefore low
2.4 How well were likely confounding factors identified and controlled?	++	They controlled for daily traffic flow, congestion and wind direction. In addition they used 'traffic contribution ' concentrations for the analyses, instead of roadside concentration as a way to control for 'factors other than local sources of air pollution such as meteorology factors and long range air pollution'. They provide correlation coefficients between the urban background sites as well as between the roadside monitors and argue because these were high (> 0.70) that "meteorology and other long range atmospheric processes affect the concentrations over the whole city in a similar way"
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Generally yes, although no details are given other than "the Amsterdam Air Quality Monitoring Network complies with the accreditation criteria."
3.2 Were the outcome measurement complete?	+	Not specifically described, but given in Table 1 and 2, showing almost complete data for PM_{10}, but quite some missing data for PM1 at the control site (A10S) in the post year (232/335) with ~30% missing. Also note that data on urban background concentrations were typically available from two of the three indicated sites; PM1 was available from one site urban background site only
3.3 Were all important outcomes assessed?	+	No health outcomes were assessed.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Yes, they compared exactly one year before and one year after the intervention
3.5 Was follow-up time meaningful?	++	Yes. No comments.
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Yes, though not so much for NOx , a secondary outcome for our review, given the high variability
4.2 Were multiple explanatory variables considered in the anlayses?	++	Yes. No further comments.
4.3 Were the analytical methods appropriate?	++	Yes, although one comment is that they did not describe which method they used to test whether changes at the two sites were significant different from each other

(*Continued)*

4.4 Was the precision of association given or calculable? Is association meaningful?	++	Yes. No comments.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	No serious internal validity concerns for the study.
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	See section 1 above.

Atkinson 2009

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	No discussion of the source area was provided; heavily trafficked metropolitan areas in Northern Europe "This study provides important pointers for study design and data requirements for the evaluation of similar schemes in terms of air quality"; "...this is the first evaluation of the effects of a permanent traffic management scheme on pollution levels in a major city. With road pricing schemes being considered in the UK and elsewhere in the world this study provides...."
1.2 Is the eligible population or area representative of the source population or area?	++	Area of London affected by the CCS is the study area (as well as the area not affected, the control area). The CCS is in "the centre of the city - an area covering approximately 22km2 or 1.4% of the Great London Area"
1.3 Do the selected participants or areas represent the eligible population or area?	+	Roadside monitors were used to assess the main study question: 1 intervention monitor and 7 control monitors. It is unclear how well the 1 roadside intervention monitor is representative of the whole CCS area, but multiple monitors may have been more appropriate For background monitors (3 intervention; 7) a similar situation is observed
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	Intervention and control monitor sites were selected based on geographic location - CCZ sites were within the CCZ, and control sites were at least 8 km removed. The use of only 1 CCZ roadside site could have potentially biased the results Additionally, exclusion of monitors where completeness criteria were not met

(*Continued*)

		may have also led to bias, if these monitors were somehow different than those not excluded
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	None considered.
2.3 Was the contamination acceptably low?	+	Unclear to what extent the intervention may have influenced pollutant concentrations 8 km removed. A secondary analysis assessing the change in concentrations moving away from the CCZ through the boundary zone and control zone did not offer solid clarification, as no clear pattern emerged among pollutants
2.4 How well were likely confounding factors identified and controlled?	-	Authors hoped to exclude all seasonal and temporal variations simply by including 2 years pre and 2 years post intervention; this is likely not sufficient to adjust for potential confounders
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	Data extracted from "the London Air Quality Network (LAQN) database" are likely well calibrated and quality controlled
3.2 Were the outcome measurement complete?	+	"Completeness criteria applied to the calculation of the daily average values (75% of hourly observations available) and to the selection of sites for analysis (daily average values available for at least 75% of days in the four year period) ."
3.3 Were all important outcomes assessed?	+	A variety of AQ outcomes were assessed.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	2 years both pre- and post-intervention for both intervention and control zones represented similar follow-up times for the time with and without intervention
3.5 Was follow-up time meaningful?	++	2 years both pre- and post-intervention were likely sufficient both to recognize an effect, and to assess whether an effect would be sustained
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	
4.2 Were multiple explanatory variables considered in the anlayses?	-	
4.3 Were the analytical methods appropriate?	-	Comparing values before and after the intervention at intervention and control monitors assesses changes, but no analysis is performed and more consideration into confounders would have allowed for a much stronger analysis
4.4 Was the precision of association given or calculable? Is association meaningful?	-	Precision not provided for the % change of the various pollutants, which is the parameter of most interest for monitoring the effect of the intervention

(*Continued*)

Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	+	Aspects of the study were well designed, but concerns, especially with regard to the analysis, limit the study's internal validity
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	With 1 intervention roadside and 3 intervention background monitors, it is likely that these results are not completely generalizable to either the whole CCZ zone or other large metropolitan areas

Ruprecht 2009

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	Source population: Not clear from description in paper. Presumably cities in Italy or Northern Italy
1.2 Is the eligible population or area representative of the source population or area?	+	Eligible population: Milan historic city centre; appropriate description. Not clear to what extent the eligible population (i.e. city centre area of Milan) is representative of other Italian city centres or cities although some degree of transferability likely
1.3 Do the selected participants or areas represent the eligible population or area?	-	Selected population: monitoring stations in and outside of Ecopass zone; no description of how these were selected and whether these are representative of the intervention and control areas
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	No description as to selection of monitoring stations inside and outside Ecopass zone; only one monitoring station per site selected; no significant baseline differences between the two monitoring stations
2.2 Was the selection of explanatory variables based on sound theoretical basis?	-	No explanatory factors described, assessed or controlled.
2.3 Was the contamination acceptably low?	-	Contamination likely, as described in Discussion of article.
2.4 How well were likely confounding factors identified and controlled?	-	No confounding factors described, assessed or controlled.
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	PM_{10} represents an objective measure; no description of quality of air pollution monitoring by ARPA at the two monitoring sites

(*Continued*)

3.2 Were the outcome measurement complete?	+	No reporting on completeness of monitoring data but presumably reasonably complete
3.3 Were all important outcomes assessed?	+	Only PM_{10}, no health outcomes; they tried also to measure PM1, 2.5 and 10 with different measurement techniques, but we cannot use the data
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Follow-up time is identical (mean for two months pre- and post-intervention) for both groups
3.5 Was follow-up time meaningful?	+	Relatively short term (i.e. two months before and after intervention)
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	-	No power analyses reported; single monitoring station problematic
4.2 Were multiple explanatory variables considered in the anlayses?	-	No, only simple Student t-tests.
4.3 Were the analytical methods appropriate?	-	No adjustment for potential confounders, no time series analyses (which would have been more powerful)
4.4 Was the precision of association given or calculable? Is association meaningful?	+	Standard deviations provided.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	-	Problematic because of (i) poor description of selection of monitoring sites with only two sites selected, (ii) no statistical analysis conducted
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	Very little detail provided on setting (city of Milan), selection of monitoring sites and intervention (Ecopass zone); some transferability to other Italian cities but difficult to judge given poor reporting

Multiple interventions

Giovanis 2015

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	In their literature review, authors cite studies almost exclusively from urban areas in the US (San Francisco, Chicago, Atlanta)

(*Continued*)

1.2 Is the eligible population or area representative of the source population or area?	+	North Carolina.
1.3 Do the selected participants or areas represent the eligible population or area?	-	No information regarding monitor characteristics. Unclear how representative these are of the areas they represent
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	+	"One of the reasons for choosing the treated and non-treated counties is that all of them are considered as "non-attainment areas". Additionally, these counties share common demographic and economic characteristics." Although, once again, not clear what monitors were actually selected - and no information provided about the baseline differences between sites, no matching, etc
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	"The weather data used in the estimates are the average daily temperature, wind speed, wind direction and solar radiation. A negative association between wind speed and actual ozone levels is expected, while a positive relationship between temperature, solar radiation and observed ozone concentrations is anticipated. "
2.3 Was the contamination acceptably low?	+	As some of the relevant counties shared borders (see Map 1) some contamination may have been possible, and as ozone is a regional pollutant some contamination may have been present
2.4 How well were likely confounding factors identified and controlled?	++	"The model controls for the day of the week, month, year, counties, ozone regions and weather conditions, such as temperature, wind speed, wind direction, and solar radiation."
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	-	Not reported.
3.2 Were the outcome measurement complete?	+	Not reported, but pretty complete, when looking at the number of observations in Table 3
3.3 Were all important outcomes assessed?	+	Although there is a small section on health outcomes, this is not considered for the review because important information is lacking
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	2000 to 2010 (2006 intervention point), the ozone forecast period of May to September
3.5 Was follow-up time meaningful?	++	No further comment.
Section 4: Analyses		

(*Continued)*

4.1 Was the study sufficiently powered to detect an effect if one exists?	++	Statistical power not discussed. Standard errors and significance levels (35,463 observations) reported in Table 5 (p 31) suggest that power was not an issue
4.2 Were multiple explanatory variables considered in the anlayses?	++	See above.
4.3 Were the analytical methods appropriate?	++	DiD estimator (treatment*program), controlled for a range of potential confounders, allows for a regression-based assessment of the Clean Air Works Program effect. Key assumptions are checked. The use of quadruple differences is possibly questionable - could have influenced the estimate of interest here. No information about the baseline variables within the intervention and control communities, no matching, etc
4.4 Was the precision of association given or calculable? Is association meaningful?	++	No further comment.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	+	Analysis is very good, yet some serious concerns stem from the data that was used in the analysis: what types of sites were selected, and what data were used. Also see the few concerns about the analysis above
5.2 Are the results generalisable to the source population (i.e externally valid)?	+	Dependent upon what data were used, which is not described.

Mullins 2014

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Description of similar problems and interventions in large cities both industrial countries (Paris) and developing countries (Dehli, Beijing) found in the introduction (p. 1108)
1.2 Is the eligible population or area representative of the source population or area?	++	Santiago Chile - very thoroughly described throughout the introduction - large city particularly susceptible to high air-pollution levels
1.3 Do the selected participants or areas represent the eligible population or area?	**++ (health)** **+ (AQ)**	Health: City-wide death statistics likely representative of the city of Santiago AQ: Not clear to what extent the 3 assessed monitors are representative of the city. In the data section, it is mentioned that “placements intended to capture traditional hotspots and provide observations on representative pollution levels”, but this refers to all 9 monitors, not just the 3 included. The expansion of the net-

(*Continued*)

		work, however, implies that the 3 that were already in place were not sufficient
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	**++ (health)** **+ (AQ)**	Health: selection bias should not lead to bias in this city-wide selection AQ: not clear whether monitors represent hotspot or rather background concentrations. Values are aggregated across sites, so this will not necessarily bias results, it is just not possible to fully interpret the results
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Well described, e.g.: "As weather conditions are expected to covary with many of the outcomes of interest in this study, observational weather controls are of critical importance."
2.3 Was the contamination acceptably low?	+	Probably, but not sure, since they included some pre-PPDA Episodes as part of the control days. The policy of identifying and announcing Episodes was technically established in the early 1990s, yet they treat 1997 as the first year the intervention starts, because of the desire to keep the matching pool as large as possible. However, the authors do provide evidence suggesting that the policy was not vigorously implemented until much later, and provide some arguments to justify their modelling choices
2.4 How well were likely confounding factors identified and controlled?	**+ (health)** **++ (AQ)**	Health: Only considered confounders for PM_{10} because matching procedure was based only on confounders for PM_{10}, though they matched on baseline mortality in a sensitivity analysis. Important confounders such as influenza episodes may have been missed AQ: The study rigorously controlled for important confounders (mean PM, temperature, average wind speed, and precipitation, day of the week, and month) using matching procedures, and using an additional approach to control for remaining confounding (of the variables included in the matching procedure). In addition, they explored whether the results are robust to the addition of more meteorological covariates in the matching process, and inclusion of multiple lags
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Routinely collected PM_{10} data by the Chilean Ministry of the Environment; routinely collected mortality data from the Chilean Ministry of Health's Department of Statistics and Health Information

(*Continued*)

3.2 Were the outcome measurement complete?	+	Health: No info reported, but likely reasonably complete. AQ: "Due to the centrality of PM10 levels in our examination, days for which PM10 data are not available from any of these three stations are omitted from our analysis. This criterion leads us to omit 185 days in the pre-PPDA period and 17 days in the post-PPDA period (all in 1997) from the matching analysis."
3.3 Were all important outcomes assessed?	++	Both air pollution and health assessed.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Follow-up 5 days after announcement, both when there was an announcement and when there was no announcement
3.5 Was follow-up time meaningful?	++	Very short term changes expected based on the intervention - 5 days appears to have been appropriate for assessing these changes
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	+	Although they had twenty year of data available (for PM_{10}), in the end they identified 34 treatment days and 100 (PM_{10}) and 85 (mortality) control days for the analysis, which is not an awful lot. One reason for this is that they excluded any events within 5 days of another event, which reduced the number of events by two-thirds. Sensitivity analyses A10 shows with increased numbers of observations did increase power. Also SD are sometimes larger, especially for mortality
4.2 Were multiple explanatory variables considered in the anlayses?	++	Yes
4.3 Were the analytical methods appropriate?	++	Difference-in-Difference assessed directly and in the form of a regression with further control for potential confounders. DiD is a strong method for estimating causal effects, mimicking a randomized controlled trial. Authors went to great lengths to investigate alternative assumptions, model specifications and key assumptions of the model
4.4 Was the precision of association given or calculable? Is association meaningful?	++	See tables 2 and 3.
Section 5: Summary		
5.1 Are the study results internally valid (i.e unbiased)?	++	No internal validity concerns for this study.
5.2 Are the results generalisable to the source population (i.e externally valid)?	**++ (health)** **+ (AQ)**	Health: routinely monitored data across the metropolitan are likely generalizable AQ: likely genrealizable but lack of reporting regarding monitoring sites is a concern

Zigler 2016

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	+	US counties (all were subjected to attainment and non-attainment designation through the CAA)
1.2 Is the eligible population or area representative of the source population or area?	+	The western United States was chosen for the study ”because virtually all initial nonattainment designations for PM10 occurred in this part of the country“
1.3 Do the selected participants or areas represent the eligible population or area?	++	See Table 4. Yes, likely representative of entire US Western region. ”For our analysis, data were considered at the monitor level, that is for each monitoring location we have a specific location (latitude and longitude), measures of ambient pollution, demographic characteristics of th county containing the monitor, and aggregated health information on all Medicare beneficiaries residing within a 6-mile radius. The initial data set contained the 547 monitoring locations...“
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	The 268 non-attainment areas are obviously different from the 279 attainment areas, because designation was based on pollutant levels. However, the propensity score methods applied, as well as further adjustment, should have ensured that similar groups were compared. ”The obvious threat to validity of the decision to estimate causal effects of the nonattainment designations by comparing outcomes with attainment areas is that the designations were decidedly not randomly assigned and thus attainment areas share important differences with nonattainment areas…required careful confounding adjustment.”
2.2 Was the selection of explanatory variables based on sound theoretical basis?	+	It is clear from the description of the methods for building propensity scores, that authors feel the aspects listed in Table 1 “constitute (or are proxies for) all factors that could confound comparisons between attainment and nonattainment areas.”
2.3 Was the contamination acceptably low?	-	Clear from Figure 5 that in many areas attainment and non-attainment areas were geographically close to one another. It is likely that the air quality of non-attainment areas influenced that of attainment areas and vice versa. Decreases at non-attainment areas due to the intervention could potentially have decreased pollution at attainment areas, which would have neutralized any observable intervention effect

(*Continued*)

2.4 How well were likely confounding factors identified and controlled?	++	Table 1 shows an extensive list of demographic aspects that may have influenced associations. Only one meteorological aspect was included, this is likely not the only such relevant aspect
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	++	US EPA monitoring data and Medicare health data.
3.2 Were the outcome measurement complete?	+	"284 monitoring locations (131 in nonattainment areas) had missing PM10 measurements in 1990... Average ambient PM10 concentrations for 1999-2001 were missing for 157 monitoring locations (70 in nonattainment areas)" These were imputed with procedures described on page 18. It does not seem that data were missing differentially at either time between groups
3.3 Were all important outcomes assessed?	++	PM_{10}, hospitalization and mortality.
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	1999 to 2001 for both groups.
3.5 Was follow-up time meaningful?	++	Yes, long-term changes could be assessed so long after the attainment designation status
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	+	Power not discussed, but health data from 3 million Medicare recipients is very unlikely underpowered. PM_{10} data is less clear, but with daily measurements this is likely well powered as well
4.2 Were multiple explanatory variables considered in the anlayses?	++	Through both propensity scores and direct adjustment, to handle any residual confounding
4.3 Were the analytical methods appropriate?	++	See 4.2
4.4 Was the precision of association given or calculable? Is association meaningful?	++	Precision provided for all estimates.
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	++	The methods are valid, especially the adjusted causal analysis, which uses propensity scores (and a pruned dataset) to create similar groups for comparison. One concern is the long data gap between 1990 and 1999-2001, which represent the pre- and post-intervention time frames
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	Results should be generalizable for the western region of the USA

Li 2011

Criteria	Rating	Support for rating
Section 1: Population (external validity)		
1.1 Is the source population or source area well described?	++	Mega cities (and their populations) in China, with high levels of pollution and an increasing trend of asthma
1.2 Is the eligible population or area representative of the source population or area?	++	City of Beijing (and its population): representative of the source area and population
1.3 Do the selected participants or areas represent the eligible population or area?	++	Data on outpatient visits for asthma were obtained from the database of the asthma registry of the Institute of Respiratory Medicine, Beijing Chaoyang Hospital. It covers adult residents (mean age: 51.1 years) of urban areas of Beijing
Section 2: Method of selection of exposure (or comparison) group		
2.1 Selection of exposure (and comparison) group. How was selection bias minimised?	++	From the text (see above) it would seem that the asthma data from the asthma registry should be representative for the whole city. We assume that the data are collected from all Beijing and the hospital is only the place of data gathering (not the only place where asthma cases are collected)
2.2 Was the selection of explanatory variables based on sound theoretical basis?	++	Influence of meteorology well cited, and the use of the Plam index well explained
2.3 Was the contamination acceptably low?	**na**	Not applicable
2.4 How well were likely confounding factors identified and controlled?	+	In the time-series regression, covariates included day of the week, mean temperature and humidity - other potential confounders (seasonality and time trends) were not included because of the short study period. Missing are other health trends, medical covariables such as 'flu epidemics etc. that could influence asthma rates
Section 3: Outcomes		
3.1 Were the outcome measures and procedures reliable?	+	Data on outpatient visits for asthma taken from the registry were likely reliable. However, no information is actually given about how the data is retrieved
3.2 Were the outcome measurement complete?	+	For both AQ and asthma outcomes, data are likely relatively complete, but no information is given
3.3 Were all important outcomes assessed?	++	Assessment of air quality and health outcomes allows a relatively complete picture of the intervention effects
3.4 Was there a similar follow-up time in exposure & comparison groups?	++	Three time periods are assessed and are approximately equal.

(*Continued*)

3.5 Was follow-up time meaningful?	++	This is a short-term intervention, and an immediate effect can be seen in the short follow-up time, although this short time does not allow for the valid assessment of potential confounders
Section 4: Analyses		
4.1 Was the study sufficiently powered to detect an effect if one exists?	+	No mention of a power calculation, but effect precision suggests that study is sufficiently powered at least for the Olympic period (wide CIs for pre-Olympic period could suggest lack of power). Although the authors claim: "The special nature of the Olympic Games, the relatively short intervention period and limited statistical power, and the limited number of air pollution monitoring sites and medical data make firm conclusions difficult."
4.2 Were multiple explanatory variables considered in the anlayses?	++	Sufficient control for potential confounders.
4.3 Were the analytical methods appropriate?	+	ITS analysis appropriate for assessing effect of intervention, but lack of trend assessment and potentially relevant other confounders
4.4 Was the precision of association given or calculable? Is association meaningful?	++	The RRs for adjusted and unadjusted analyses were given with confidence intervals
Section 5: Summary		
5.1 Are the study results internally valid (i. e unbiased)?	+	Appropriate data, slight concerns with appropriate outcome measures and the ITS analysis
5.2 Are the results generalisable to the source population (i.e externally valid)?	++	These results are likely generalizable to other heavily polluted Chinese mega cities

Appendix 9. Data and effect measurements from included main studies

Primary health outcomes

Intervention category	Study ID	Pre-intervention outcome level (intervention)	Pre-intervention outcome level (control)	Post-intervention outcome level (intervention)	Post-intervention outcome level (control)	Effect estimate and measure of precision	P value	Narrative interpretation
All-cause mortality								

(*Continued*)

Industrial sources	Deschênes 2012	NR	NR	NR	NR	DiD Estimator (SE): −1.557 (0.813)	> 0.05	No observed change in total mortality associated with the intervention
	Pope 2007	NR	NR	NR	NR	Adjusted % change (95% CI): −2.5 (−7.2 to −4.1)	NR	Significant decrease in mortality observed after the implementation of the intervention
	Tanaka 2015	NR	NR	NR	NR	DiD Estimator (SE): −3.287 (2.128)	> 0.05	No observed change in infant mortality associated with the intervention, although a slight borderline significant decrease was observed, and authors discuss in depth why the model may have inflated standard errors, thus leading to an insignificant result
Residential sources	Clancy 2002 (per 1000 person years)	9.41	NR	8.65	NR	Adjusted % change (95% CI): −5.7 (−4.0 to −1.1)	< 0.0001	Significant decrease in mortality observed after the implementation of the

(*Continued*)

								inter-vention; this decrease not seen in other causes of mortality
	Dockery 2013a (per 1000 person years)	9.87	9.88	8.2	7.84	Adjusted % change (95% CI): int: −1.0 (−6.0 to 4.4); con: −2.7 (−7.7 to 2.7)	int: 0.72; con: 0.32	No observed change in total mortality associated with the intervention
	Dockery 2013b (per 1000 person years)	9.7	9.44	7.07	7.41	Adjusted % change (95% CI): int: −4.4 (−9.6 to 1.0); con: −3.6 (−8.8 to 2.0)	int: 0.11; con: 0.20	No observed change in total mortality associated with the intervention
	Dockery 2013c (per 1000 person years)	9.47	9.22	7.47	7.07	Adjusted % change (95% CI): int: 0.2 (−3.1 to 3.6); con: −0.2 (−6.7 to 6.8)	int: 0.90; con: 0.96	No observed change in total mortality associated with the intervention
	Johnston 2013 (per 1000 person years)	Annual: 8.57; Winter: 9.2	Annual: 8.25 ; Winter 9.52	Annual: 7.42; Winter: 8.08	Annual: 7.22; Winter: 8.12	Adjusted % change (95% CI): Annual: int: −2.7 (−8.7 to 3.7); con: 1.4 (−3.0 to 6.0); Winter:	Annual: int: 0.40; con: 0.54; Winter: int: 0.73; con: 0.64	No observed change in total mortality associated with the intervention

(Continued)

						int: −2. 2 (−14.1 to 11.3); con: −2.0 (−10.2 to 6. 9)		
Vehicular sources	Yorifuji 2011 (per 1000 person years)	NR	NR	NR	NR	Adjusted % change (95% CI): −0.13 (−1. 99 to 1.77)	0.893	No observed change in total mortality associated with the intervention
	Yorifuji 2016 - Diesel standards (per 1000 person years)	7.52	8.72	7.22	8.44	Adjusted % change (95% CI): −0.61 (−1. 3 to 0.056)	NR	No observed change in total mortality associated with the intervention
	Yorifuji 2016 - Tightening of standards (per 1000 person years)	7.22	8.44	6.87	8.14	Adjusted % change (95% CI): −2.1 (−2.8 to −1.4)	NR	Significant decrease in mortality observed after the implementation of the intervention
Multiple sources	Mullins 2014	64.64	64.64	63.9	67.6	DiD Estimator (SE): −3.611 (2. 48)	> 0.05	No observed change in total mortality associated with the intervention on the day of the intervention. 3 days after the intervention a significant decrease in mortality is seen

(*Continued*)

	Zigler 2016 (per 1000 person years)	62.51	62.58	62.5	62.6	Causal effect (95% posterior interval) : −1.08 (−3.27 to 0.99)	NR	No observed change in total mortality associated with the intervention
Cardiovascular mortality								
Industrial sources	Deschênes 2012 (cardiovascular + respiratory	NR	NR	NR	NR	DiD Estimator (SE): −0.547 (0.0675)	> 0.05	No observed change in total mortality associated with the intervention
Residential sources	Clancy 2002 (per 1000 person years)	4.37	NR	3.78	NR	Adjusted % change (95% CI): −10.3 (−12.6 to −8.0)	< 0.0001	Significant decrease in mortality observed after the implementation of the intervention; this decrease not seen in other causes of mortality
	Dockery 2013a (per 1000 person years)	4.55	5.45	3.39	3.62	Adjusted % change (95% CI): int: 0.1 (−8.5 to 9.5); con: −1.8 (−10.0 to 7.2)	int: 0.98; con: 0.68	No observed change in overall cardiovascular mortality associated with the intervention, nor in the individual sub-categories
	Dockery 2013b (per 1000 person	5.00	5.05	3.41	3.26	Adjusted % change (95% CI): int: −3.7	int: 0.42; con: 0.47	No observed change in cardiovascu-

(*Continued*)

	years)					(−12.2 to 5.6); con: −3.4 (−12.0 to 6.1)		lar mortality associated with the intervention
	Dockery 2013c (per 1000 person years)	4.68	4.84	3.07	3.00	Adjusted % change (95% CI): Intervention: −1.1 (−6.1 to 4.1); Control: −3.1 (−12.6 to 7.3)	int: 0.67; con: 0.54	No observed change in overall cardiovascular mortality associated with the intervention, nor in most individual subcategories. A greater decrease in cerebrovascular mortality was observed at intervention sites than at control sites
	Johnston 2013 (per 1000 person years)	Annual: 3.88; Winter: 4.52	Annual: 3.58; Winter: 4.16	Annual: 2.74; Winter: 2.96	Annual: 2.68; Winter: 2.96	Adjusted % change (95% CI): Annual: int: −4.9 (−15.5 to 7.0); con: 0.9 (−7.1 to 9.6); Winter: int: −19.6 (−36.3 to 1.5); con: −7.0 (−20.8 to 9.2)	Annual: int: 0.40; con: 0.83; Winter: int: 0.06; con: 0.38	No observed change in cardiovascular mortality associated with the intervention

(*Continued*)

Vehicular sources	Yorifuji 2011 (per 1000 person years)	NR	NR	NR	NR	Adjusted % change (95% CI): 1.27 (−2.11 to 4.78)	0.466	No observed change in circulatory or IHD mortality associated with the intervention; a significant decrease in cerebrovascular mortality was observed
	Yorifuji 2016 - Diesel standards (per 1000 person years)	2.34	2.48	2.16	2.34	Adjusted % change (95% CI) : −1.9 (−3. 3 to −0.60)	NR	Significant decrease in CVD mortality observed after the implementation of the intervention
	Yorifuji 2016 - Tightening of standards (per 1000 person years)	2.16	2.34	1.96	2.2	Adjusted % change (95% CI) : −5.9 (−7. 2 to −4.6)	NR	Significant decrease in both CVD and IHD mortality observed after the implementation of the intervention
Respiratory mortality								
Residential sources	Clancy 2002 (per 1000 person years)	1.38	NR	1.16	NR	Adjusted % change (95% CI): −15. 5 (−19.1 to −11.6)	< 0.0001	Significant decrease in mortality observed after the implementation of the intervention; this decrease not

(*Continued*)

								seen in other causes of mortality
	Dockery 2013a (per 1000 person years)	1.46	1.37	1.23	1.26	Adjusted % change (95% CI): int: −16.8 (−24.4 to −8.4); con: −2.3 (−11.5 to 7.9)	Intervention: 0.0002; Control: 0.65	Significant decrease in overall respiratory mortality seen at intervention areas, while not at control areas
	Dockery 2013b (per 1000 person years)	1.35	1.34	1.14	1.25	Adjusted % change (95% CI): int: −9.3 (−18.2 to 0.7); con: −1.4 (−10.9 to 9.1)	int: 0.067; con: 0.78	No observed change in overall respiratory mortality associated with the intervention
	Dockery 2013c (per 1000 person years)	1.49	1.34	1.26	1.19	Adjusted % change (95% CI): int: −2.6 (−8.1; 3.4); con: 1.4 (−10.2; 14.5)	int: 0.39; con: 0.82	No observed change in overall respiratory mortality associated with the intervention not at control sites
	Johnston 2013 (per 1000 person years)	Annual: 0.86; Winter: 1.16	Annual: 0.76; Winter: 1.0	Annual: 0.64; Winter: 0.76	Annual: 0.64; Winter: 0.88	Adjusted % change (95% CI): Annual: int: −8.5 (−23.2 to 9.0); con: 4.8 (−7.4 to 18.6); Winter: int: −27.9	Annual: int: 0.32; con: 0.50; Winter: int: 0.07; con: 0.60	No observed change in respiratory mortality associated with the intervention. A non-significant decrease, however, was observed in in-

(*Continued*)

						(−49.5 to 3. 1); con: 8. 0 (−16.9 to 40.4)		tervention areas, while an non-significant increase was seen in control areas
Vehicular sources	Yorifuji 2011 (per 1000 person years)	NR	NR	NR	NR	Adjusted % change (95% CI): 3.02 (−0.16 to 6.29)	0.063	No observed change in respiratory mortality associated with the intervention, although a slight borderline significant increase was observed
	Yorifuji 2016 - Diesel standards (per 1000 person years)	1.09	1.36	1.07	1.37	Adjusted % change (95% CI) : −6.0 (−8. 1 to −3.9)	NR	Significant decrease in respiratory mortality observed after the implementation of the intervention
	Yorifuji 2016 - Tightening of standards (per 1000 person years)	1.07	1.37	1.02	1.35	Adjusted % change (95% CI): −10.0 (−12 to −8.1)	NR	Significant decrease in respiratory disease observed after the implementation of the intervention

Primary AQ outcomes

Intervention category	Study ID	Pre-intervention concentration (intervention)	Pre-intervention concentration (control)	Post-intervention concentration (intervention)	Post-intervention concentration (control)	Effect estimate and measure of precision	p-value	Narrative interpretation
PM_{10} (ug/m^3)- mean (SD)								
Industrial sources	Deschênes 2012	NR	NR	NR	NR	DiD estimator (SE): −0.896 (1.018); % change: −3.0%;	> 0.05	No observed change in mean concentration associated with the intervention
	Saaroni 2010	47.9	36.8	42	48.3	NR	< 0.05	Concentrations at the intervention site were significantly lower after the intervention than at control sites
	Sajjadi 2012	18.2 (8.4)	NA	20.9 (11.2)	NA	Mean change: 13.2%	0.021	Significant increase in mean concentration observed after the intervention
Vehicular sources	Atkinson 2009	streetside: 41; background: 35.6	streetside: 30.6; background: 23.5	streetside: 43.3; background: 30.1	streetside: 31.4; background: 23.3	Mean change: Streetside int: 5.6% Background INT: −15.4%; Streetside con: 2.5%; Background con: −0.8%	NR	Increases at streetside monitors were observed at both intervention and control sites. A large decrease was seen at background intervention sites

(*Continued*)

	Bel 2013a	47.7	48.8	37.8	35.9	DiD estimator: 2.594 % change: 5.4%	< 0.05	An increase in concentrations was seen after the implementation of the intervention
	Bel 2013b	38.9	NR	32.8	NR	DiD estimator: −6.196; % change: −14.7%	< 0.01	A significant decrease in concentrations was observed after the implementation of the intervention
	Boogaard 2012	streetside: 28.1; background: 25.1	22.4	streetside: 25.0; background: 21.2	19	Mean change: Streetside int: −3.1; Background int: −4.0; Suburban con: −3.3	Streetside int. vs. Suburban con: > 0.05; Background int vs. Suburban con.: > 0.05	Similar decreases in concentrations observed at all monitors. When comparing changes at intervention monitors with those at control monitors, no differences were observed
	Burr 2004	35.2	11.6	27.2	8.2	Mean change: int: −22.7%; con: −28.9%	NR	Concentrations decreased at both the congested and uncongested streets between the pre-

(*Continued*)

								and post-intervention time. This change was to a slightly greater extent at the control site (uncongested street), but no statistical analysis was performed
	Cowie 2012	17.6 (6.9)	NA	Year 1: 15.2 (6.2); Year 2: 15.9 (6.4)	NA	Adjusted mean change: Year 1: −0.38 (−1.51 to 0.75); Year 2: −0.67 (−1.40 to 0.07)	Year 1: > 0.05; Year 2: > 0.05	No observed change in mean concentration associated with the intervention, after adjustment for local meteorology and regional background
	Dijkema 2008	29.72 (range: 12.60 to 85.50)	25.20 (range: 6.60 to 80.40)	27.55 (range: 11.60 to 59.20)	24.21 (range: 9.20 to 54.30)	Adjusted mean change (95% CI): int: −2.20 (−2.98 to −1.43); con: −0.97 (−1.68 to −0.25)	< 0.05 (data not shown)	Decreases in concentrations at both intervention and control sites observed. Decrease at intervention site statistically greater than at control site
	Fensterer 2014	Streetside summer: 27.2 (14.3); Background summer: 21.	Summer: 19.3 (12.2); Winter: 24.3 (21.6)	Streetside summer: 23.4 (14.5); Background summer: 20.	Summer: 18.9 (12.3); Winter: 24.5 (20.8)	Adjusted mean change (95% CI): Streetside	Streetside summer: < 0.001; Background summer: <	Decreases in concentrations at both streetside and back-

(*Continued*)

		3 (12.9); Streetside winter 30.8 (21.6); Background winter: 28.3 (23.6)		8 (15.3); Streetside winter 30.2 (23.6); Background winter: 27.6 (22.0)		summer: −19. 63% (−22. 75 to −16. 52%); Background summer: −5. 73% (−7.71 to −3.74%) ; Streetside winter: −6. 80% (−10. 14 to −3. 47%); Background winter: −3. 18% (−5.24 to −1.11%) All seasons: Streetside: 13% Background: 4. 5%	0.001; Streetside winter: < 0. 001; Background winter: 0.003 All seasons: Streetside: < 0.001; Background: < 0. 001	ground intervention sites, both in summer and in winter, after control for concentration at a reference station
	Kim 2011	61.3 (10.3)	54.4 (14.3)	70.3 (19.4)	51.9 (15.4)	Mean change: int: 14.7%; con: −4.7%	int: 0.01; con: 0.6	Increase in concentration observed when taking all intervention sites into account. Slight decrease was associated with no significant change at control sites
	Morfeld 2013	33.87	24.64	38.98	30.52	Adjusted mean change (95% CI):	0.326	Concentrations increased at both inter-

(Continued)

						0.4% (−0.4% to 1.1%)		vention and control sites. After adjusting for changes at the control sites, no change associated with the intervention was seen at intervention sites
	Peel 2010	37.6 (14.2)	Surrounding states 1: 42.2 (19.2); Surrounding states 2: 37.6 (14.9)	31.2 (10.4)	Surrounding states 1: 35.3 (12.9); Surrounding states 2: 32.6 (13.4)	NR	int: 0.239; Surrounding states 1 con: 0.432; Surrounding states 2 con: 0.479	No observed change in mean concentration associated with the intervention at any sites
	Ruprecht 2009	71.2 (32.6)	74.8 (38.4)	67.3 (36.4)	70.9 (38.3)	Pre-, post-int concentration ratio: int: 0.9517 con: 0.9504	NR	Similar decreases in concentrations observed at all monitors. When comparing changes at intervention monitors with those at control monitors, no differences were observed
	Viard 2015	NR	NR	NR	NR	Adjusted mean change (SE): Even-odd policy: −31% (0.1090);	Even-odd policy: < 0.01; One-day policy: < 0.01	Significant decrease in concentration observed after the implemen-

(*Continued*)

						One-day pol-icy: −27% (0.0681)		tation of the intervention
Multiple sources	Mullins 2014	133	133	105	130	DiD estimator (SE): −22.53 (4.99)	< 0.01	Significant decrease in concentrations at intervention sites the day after the intervention, compared to the change at the control sites
	Zigler 2016	40.4	27	31.6	21.6	Causal estimate (95% posterior interval): −1.17 (−7.33 to 4.00)	> 0.05	No observed change in air quality due to intervention
$PM_{2.5}$ (ug/m^3)- mean (SD)								
Industrial sources	Deschênes 2012	NR	NR	NR	NR	DiD estimator (SE): −0.382 (0.278); % change: −2.3%	> 0.05	No observed change in mean concentration associated with the intervention
Residential sources	Allen 2009	18.5	10	10.5	7	Median change: int: −2.7; con: - 3.4	int: 0.04; con: 0.03	Similar decreases in concentrations observed at both intervention and control homes
	Aung 2016	23 (15)	4 (3.1)	29 (23)	5 (0.5)	Mean difference (95% CI): Pre-int: 13	Pre-intervention: < 0.05 Post-inter-	Concentration increased at both inter-

(*Continued*)

						(8 to 24); Post-int: 18 (−1 to 62)	vention: > 0. 05	vention and control sites after the implementation of the intervention. No observed change associated with the intervention
	Yap 2015	30.76 (22. 88)	NA	26.10 (16. 56)	NA	Mean change (95% CI): −3.79 (−2. 25 to −4.5)	< 0.05	Decrease in concentration observed after the implementation of the intervention
Vehicular sources	Boogaard 2012	streetside: 16.8; background: 14.7	13.8	streetside: 11.8; background: 10.8	11.1	Mean change: Streetside int: −5.1; Background int: −3.9; Suburban con: −2.7	Streetside int. vs. Suburban con: < 0.05 ; Background int. vs. Suburban con:> 0.05	Decreases in concentrations were observed at all sites. The change at streetside intervention sites were, however, significantly greater than at suburban control sites. This difference was not present when comparing background intervention sites with suburban control sites

(*Continued*)

	Burr 2004	21.2	6.7	16.2	4.9	Mean change: int: −23.5%; con: −26.6%	NR	Concentrations decreased at both the congested and uncongested streets between the pre- and post-intervention time. This change was to a slightly greater extent at the control site (uncongested street), but no statistical analysis was performed
	Cowie 2012	5.8 (3.5)	NA	Year 1: 4.9 (4.3); Year 2: 5.1 (4.7)	NA	Adjusted mean change (95% CI): Year 1: −0.16 (−0.57 to 0.26); Year 2: 0.17 (−0.23 to 0.56)	Year 1: > 0.05; Year 2: > 0.05	No observed change in mean concentration associated with the intervention, after adjustment for local meteorology and regional background
	Yorifuji 2016 (diesel standards)	24.4 (12.6)	22.7 (11.0)	21.0 (11.0)	19.9 (9.3)	Mean change: int: −3.4%; con: −2.9%	NR	Similar decreases in concentrations observed at both intervention and control sites

(*Continued*)

	Yorifuji 2016 (tightening of standards)	21.0 (11.0)	19.9 (9.3)	18.0 (9.0)	19.1 (10.7)	Mean change: int: −6.5%; con: −3.6%	NR	Similar decreases in concentrations observed at both intervention and control sites, difference is slightly larger at intervention sites
Coarse PM- mean (ug/m^3) - (SD)								
Residential sources	Yap 2015	19.02 (16.91)	NA	14.63 (12.09)	NA	Mean change (95% CI): −1.61 (−2.25 to −1.25)	< 0.05	Decrease in concentration observed after the implementation of the intervention
Combustion-related PM (black smoke) (ug/m^3)- mean (SD)								
Vehicular sources	Dijkema 2008	23.83 (range: 0.43 to 104.06)	20.12 (range: 0.33 to 93.24)	19.41 (range: 0.89 to 92.51)	15.82 (range: 0.63 to 53.93)	Mean change (95% CI): int: −3.57 (−5.65 to −1.50); con: −2.43 (−3.80 to −1.05)	NR	Decrease in concentrations observed at both intervention and control sites
Combustion-related PM (black carbon) (ug/m^3)- mean (SD)								
Residential sources	Aung 2016	3.3 (2.1)	0.3 (0.3)	3.2 (2.2)	1.2 (0.9)	Mean difference (95% CI): Pre-int: 2.7 (1.4 to 3.9); Post-int: 1.6 (0.5 to 2.9)	Pre-intervention: < 0.05; Post-intervention: > 0.05	Concentration increased at both intervention and control sites after the implemen-

(*Continued*)

								tation of the intervention. This increase was greater at the intervention site
Vehicular sources	Gramsch 2013	7.91 (5.69)	5.05 (2.87)	8.29 (5.78)	5.93 (3.81)	Mean change: int: 4.8% con: 17.4%	int: 0.028; con: < 0.01	Slight significant increases observed at both intervention and control sites
	Titos 2015a	5.6 (8.1)	2.5 (4.9)	1.6 (5.9)	2.4 (6.3)	Mean change: int: −72%; con: 6%	int: < 0.01; con: > 0.05	Statisticfally significant decrease observed at intervention sites, slight increase observed at control sites
	Titos 2015b	3.8 (2.7)	1.4 (0.9)	2.5 (1.6)	1.2 (1.0)	Mean change: int: −37%; con : −14%	int: < 0.01; con: > 0.05	Statisticfally significant decrease observed at intervention sites, slight decrease observed at control sites
Combustion-related PM (soot) (ug/m^3)- mean (sd)								
Vehicular sources	Boogaard 2012	Streetside: 2.93; Background: 1.61	1.48	Streetside: 2.89; Background: 1.48	1.27	Mean change: Streetside int: −0.04; Background int: −0.13; Suburban con: −0.11	Streetside int vs. Suburban con: > 0.05; Background int vs. Suburban con: > 0.05	Similar decreases in concentrations observed at all monitors. When comparing changes at intervention moni-

(*Continued*)

								tors with those at control monitors, no differences were observed

Secondary health outcomes

Intervention category	Study ID	Pre-intervention outcome level (intervention)	Pre-intervention outcome level (control)	Post-intervention outcome level (intervention)	Post-intervention outcome level (control)	Effect estimate and measure of precision	p-value	Narrative interpretation
Cardiovascular hospitalizations								
Residential sources	Dockery 2013b (per 1000 persons years)	14.25	NA	13.49	NA	Adjusted % change (95% CI): −3.6 (−9.8 to 2.9)	0.27	No observed change in cardiovascular hospitalizations associated with the intervention
	Dockery 2013c (per 1000 persons years)	Limmerick: 16.45; Louth: 15.86; Wexford: 11.09; Wicklow: 8.88	NA	Limmerick: 12.16; Louth: 15.13; Wexford: 12.13; Wicklow: 9.02	NA	Adjusted % change (95% CI): −3.2 (−5.7 to −0.6)	0.016	An overall significant decrease cardiovascular hospitalizations observed; heterogeneous effects were, however, seen across counties
	Yap 2015 (per 1000 persons years)	Ages 45 to 64: 41; Ages >65: 152.2	NA	Ages 45 to 64: 39.9; Ages > 65: 81.1	NA	Adjusted relative risk (95% CI): Ages 45 to	NR	For the oldest age group (>65) significant

(Continued)

						64: 0.97 (0. 90 to 1.05); Ages >65: 0. 93 (0.89 to 0.97)		decreases in CVD hospital-izations observed. For the younger age group, no change was observed
Vehicular sources	Peel 2010	NR	NR	NR	NR	Adjusted relative risk (95% CI): 0.996 (0.829 to 1. 195)	NR	For total cardiovas-cular disease hospital-izations no change was observed
Multiple sources	Zigler 2016 (per 1000 persons years)	92.09	83.74	92.1	83.7	Causal effect (95% Posterior interval) : 1.44 (−4.64 to 6.16)	NR	No observed change in cardiovascular hospitalizations associated with the intervention
Respiratory hospitalizations								
Industrial sources	Lin 2013	NR	NR	NR	NR	Adjusted % change (95% CI): −0.15 (−9. 83 to 10.55)	NR	No observed change in respiratory hospitaliza-tions associated with the intervention
	Sajjadi 2011 (per 100000 population)	Respiratory diease: 3.91; COPD (65+): 2. 671; Asthma (< 15): 2.199	Respiratory diease: 3.81; COPD (65+): 3. 243; Asthma (< 15): 1.652	Respiratory diease: 3.34; COPD (65+): 3. 656; Asthma (< 15): 1.450	Respiratory diease: 3.41; COPD (65+): 4. 264; Asthma (< 15): 1.048	Adjusted % change: int: Respiratory diease: NR; COPD (65+): 36.9; Asthma (< 15): −34.1;	int: Respiratory diease: NR; COPD (65+): < 0. 0001; Asthma (< 15): 0.0031; con:	Across all indica-tors, similar changes observed at inter-vention and control sites; significant decreases in

(*Continued*)

						con: Respiratory diease: NR; COPD (65+): 31.5; Asthma (< 15): −36.6	Respiratory diease: NR; COPD (65+): 0. 0003; Asthma (< 15): 0.0008	overall respiratory disease admissions, all-ages asthma, and age < 15 asthma, a significant increase in age +65 COPD
Residential sources	Dockery 2013b (per 1000 persons years)	17.31	NA	17.19	NA	Adjusted % change (95% CI): 3. 6 (−2.5 to 10)	0.25	No observed change in respiratory hospitalizations associated with the intervention
	Dockery 2013c (per 1000 persons years)	Limmerick: 22.80; Louth: 15. 21; Wexford: 15.87; Wicklow: 9. 52	NA	Limmerick: 18.67; Louth: 14. 18; Wexford: 15.25; Wicklow: 8. 55	NA	Adjusted % change (95% CI): −8.5 (−10. 5 to −6.2)	< 0.0001	An overall significant decrease cardiovascular hospitalizations observed; mostly consistent decreases seen across counties
	Yap 2015 (per 1000 persons years)	COPD (45 to 64): 7.2; COPD (>65): 23.7	NA	COPD (45 to 64): 6.5; COPD (> 65): 13.7	NA	Adjusted relative risk (95% CI): COPD (45 to 64): 0.90 (0.78 to 1. 05); COPD (> 65): 0.93 (0. 83 to 1.04)	NR	No observed change in COPD hospitalizations associated with the intervention
Vehicular sources	El-Zein 2007 (daily admissions)	Two year follow-up: 617; One	NR	Two year follow-up: 817; One	NR	Regression coefficient: Two year follow-	Two year follow-up: 0.32 One	For shorter-term follow-up, a significant

(*Continued*)

		year follow-up: 925		year follow-up: 591		up: 0.128; One year follow-up: −0.165	year follow-up: 0.04	decrease in overall respiratory disease admissions was observed. For the longer-term follow-up, no change associated with the intervention was observed
	Peel 2010	NR	NR	NR	NR	Adjusted relative risk (95% CI): 1.012 (0.920 to 1.113)	NR	For total respiratory disease hospitalizations, no change was observed, however, heterogeneous effects were seen across subcategories, and a significant increase in COPD admissions was observed
Multiple sources	Li 2011 (daily admissions)	12.5	NA	Partial int: 16.5; Full int: 7.3	NA	Adjusted relative risk (95% CI): Partial int: 1.24 (0.93 to 1.76); Full int: 0.50 (0.47 to 0.55)	Partial int: > 0.05; Full int: < 0.01	A significant decrease in asthma outpatient visits per day observed during the full intervention period. This decrease was not seen in the period

(Continued)

								in which the intervention was partially implemented
	Zigler 2016 (per 1000 person years)	28.41	28.39	28.4	28.4	Causal effect (95% Posterior interval) : −1.47 (−3.86 to 0.70)	NR	No observed change in respiratory hospitalizations associated with the intervention
Respiratory effects								
Vehicular sources	Burr 2004 (symptoms)	Wheeze: 33.9; Winter cough: 15.6; Plegm: 12.2; Rhinitis: 37.3	Wheeze: 32.5; Winter cough: 20.1; Plegm: 15.5; Rhinitis: 38.3	NR	NR	Net improvement (95% CI): Wheeze: −6.5 (−14.9 to 2.0); Winter cough: 1.5 (−6.2 to 9.3); Plegm: 0 (−7.6 to 7.6); Rhinitis: 5.4 (−3.1 to 15.0)	Wheeze: > 0.05; Winter cough: > 0.05; Plegm: > 0.05; Rhinitis: > 0.05	No significant changes with regard to the health outcomes were observed after implementation of the intervention
	Hasunuma 2014 (asthma symptoms)	3.40	3.67	2.81	3.55	Mean change (95% CI): int: −0.59 (−0.88 to −0.31); con: −0.13 (−0.46 to 0.20)	int: < 0.05; con: > 0.05	Decreases in asthma symptoms seen at both intervention and control sites

Secondary AQ outcomes

Intervention category	Study ID	Pre-intervention concentration (intervention)	Pre-intervention concentration (control)	Post-intervention concentration (intervention)	Post-intervention concentration (control)	Effect estimate and measure of precision	p-value	Narrative interpretation
NO_x (ppb) - mean (SD)								
Vehicular sources	Atkinson 2009	streetside: 107.6; background: 33.8	streetside: 74.4; background: 21.6	streetside: 102.2; background: 31.6	streetside: 71.8; background: 20.4	Mean change: Streetside int: −5%; Background int: −6.4; Streetside con: −4.4%; Background con: −5%	NR	Similar decreases seen at streetside and background monitors at both intervention and control sites
	Bel 2013a	82.3	74.7	63.9	69.4	DiD Estimator: 1.887; % change: 1.7%	< 0.01	An increase in concentrations was seen after the implementation of the intervention
	Bel 2013b	60.5	NR	59.2	NR	DiD Estimator: −10.462; % change: −16%	< 0.01	A significant decrease in concentrations was observed after the implementation of the intervention
	Boogaard 2012	streetside: 81.8; background: 47.7	38.3	streetside: 74.3; background: 40	32.3	Mean change: Streetside int: −7.5; Background int: −7.7; Suburban con: −6.1	Streetside int vs. suburban con: > 0.05; Background int vs suburban con:> 0.05	Similar decreases in concentrations observed at all monitors. When comparing changes at interven-

(*Continued*)

								tion monitors with those at control monitors, no differences were observed
Cowie 2012	25.3 (18.6)	NA	Year 1: 21 (13.9); Year 2: 20.5 (13.4)	NA	Adjusted mean change (95% CI): Year 1: −2.24 (−4.59 to 0.11); Year 2: −2.06 (−4.73 to 0.61)	Year 1: > 0.05; Year 2: > 0.05	No observed change in mean concentration associated with the intervention, after adjustment for local meteorology and regional background
Davis 2008	NR	NR	NR	NR	Adjusted mean change (SE): 17.3% (3.3%)	NR	An increase in concentrations was observed after the implementation of the intervention
Dijkema 2008	90.00 (range: 8.80 to 334.40)	68.65 (range: 8.00 to 322.40)	83.99 (range: 8.80 to 218.40)	61.60 (range: 4.80 to 179.20)	Mean change (95% CI): int: −2.13 (−7.25 to 3.00); con: −1.87 (−5.68 to 1.94)	> 0.05	No significant changes in concentrations observed at intervention or control sites
Morfeld 2014	49.479	34.153	46.373	31.025	Adjusted mean change (95% CI): −1.74 (−2.334 to 1.	< 0.001	Small yet significant decrease in concentrations observed at in-

(*Continued*)

						145)		tervention sites
NO_2								
Industrial sources	Deschênes 2012	NR	NR	NR	NR	DiD estimator (SE): −1.210 (0.397); % Change: −7.2%	< 0.01	Significant 7.2% decrease in mean concentration seen after the implementation of the intervention
	Sajjadi 2012	0.92 (0.39)	NA	0.90 (0.39)	NA	Mean change: −3.3%	> 0.05	No observed change in mean concentration associated with the intervention
Vehicular sources	Atkinson 2009	streetside: 42.1; background: 19.8	streetside: 27.9; background: 13.8	streetside: 43; background: 21	streetside: 71.8; background: 13.4	Mean change: Streetside int: 2.1%; Background int: 7.1%; Streetside con: 3.7%; Background con: −2.3%	NR	Increase seen at all sites, except for background control sites, where a slight decrease was observed
	Boogaard 2012	streetside: 47.2; background: 32	25.8	streetside: 45.7; background: 28.6	21.2	Mean change: Streetside int: −1.5; Background int: −3.4; Suburban con: −4.5	Streetside int vs suburban con: < 0.05; Background int vs. suburban con: > 0.05	Decreases in concentrations were observed at all sites. The change at suburban control sites were, however, significantly greater than at streetside

(*Continued*)

								intervention sites. This difference was not present when comparing background intervention sites with suburban control sites
	Cowie 2012	12.6 (4.8)	NA	Year 1: 11.5 (4.0); Year 2: 11.1 (4.0)	NA	Adjusted mean change (95% CI): Year 1: −0.34 (−0.72 to 0.05); Year 2: −0.36 (−0.91 to 0.19)	Year 1: > 0.05; Year 2: > 0.05	No observed change in mean concentration associated with the intervention, after adjustment for local meteorology and regional background
	Davis 2008	NR	NA	NR	NA	Adjusted mean change (SE): 8.9% (3.4%)	NR	An increase in concentrations was observed after the implementation of the intervention
	Hasunuma 2014	26.9	14.8	20.6	11.6	Mean change (95% CI): int: −6.04 (−7.10 to −4.99); con: −3.20 (−4.42 to 1.98)	int: < 0.01 con: > 0.05	Significant decrease in concentration observed at the intervention sites, while slight, non-significant change observed at the control sites

(*Continued*)

	Kim 2011	44.3 (6.3)	5.33 (1.38)	43.8 (5.77)	5.86 (1.50)	Mean change: int: -1.13%; con: 1.0%	int: 0.78; con: 0.35	No observed change in mean concentration either intervention or control sites
	Morfeld 2014	51.959	26.383	50.831	26.17	Adjusted mean change (95% CI): −1.12 (−1.137 to −0.087)	< 0.001	Small yet significant decrease in concentrations observed at intervention sites
	Peel 2010	int 1: 49.1 (15.9); int 2: 36.2 (13.3)	Immediate area: 5.23 (2.54); Surrounding states 1: 35.0 (15.0); Surrounding states 2: 39.0 (12.0)	int 1: 43.7 (8.17); int 2: 31.2 (9.89)	Immediate area: 5.18 (4.43); Surrounding states 1: 30.0 (9.0); Surrounding states 2: 36.0 (8.0)	NR	int 1: 0.450; int 2: 0.397; Immediate area con: 1.0; Surrounding states con 1: 0.367; Surrounding states con 2: 0.523	Slight decreases observed at all intervention and control sites
	Yorifuji 2016 (diesel standards)	30.9 (11.7)	29.7 (11.2)	28.0 (10.7)	28.2 (10.0)	Mean change: int: −2.8%; con: −1.4%	NR	Similar decreases in concentrations observed at both intervention and control sites
	Yorifuji 2016 (tightening of standards)	28.0 (10.7)	28.2 (10.0)	24.3 (10.0)	25.0 (9.9)	Mean change: int: −6.6%; con: −4.7%	NR	Similar decreases in concentrations observed at both intervention and control sites, difference is

(*Continued*)

								slightly larger at intervention sites
NO								
Vehicular sources	Atkinson 2009	streetside: 63.9; background: 13.0	streetside: 44.7; background: 6.7	streetside: 57.8; background: 8.9	streetside: 40.6; background: 6.3	Mean change: Streetside int: −9.5%; Background int: −31; Streetside con: −9.4%; Background con: −6.6%	NR	Similar decreases seen at streetside and background monitors at both intervention and control sites, the largest being at background intervention sites
	Morfeld 2014	49.479	34.153	46.373	31.025	Adjusted mean change (95% CI): −1.128 (−1.555 to −0.702)	< 0.001	Small yet significant decrease in concentrations observed at intervention sites
SO_2								
Industrial sources	Deschênes 2012	NR	NR	NR	NR	DiD estimator (SE): 0.097 (0.183); % change: 2.1%	> 0.05	No observed change in mean concentration associated with the intervention
	Sajjadi 2012	0.29 (0.26)	NA	0.18 (0.14)	NA	Mean change: −40.5%	< 0.0001	Significant decrease in mean concentration after the implementation of the intervention

(Continued)

Vehicular sources	Davis 2008	NR	NA	NNR	NA	Adjusted mean change (SE): −9.2% (7.6%)	NR	A slight decrease in concentration observed after the implementation of the intervention. With such a large SE, however, this is not likely significant
	Peel 2010	int 1: 13.7 (11.0); int 2: 13.4 (14.8)	Immediate area: 16.9 (27.3); Surrounding area 1: 11.0 (14.1); Surrounding area 2: 20.8 (20.4)	int 1: 14.8 (11.8); int 2: 18.3 (13.5):	Immediate area: 7.2 (7.25); Surrounding area 1: 8.18 (9.02); Surrounding area 2: 24.9 (36.8)	NR	int 1: 0.941; int 2: 0.613; Immediate area con: 0.185; Surrounding area 1 con: 0.662; Surrounding area 2 con: 0.855	Very slight, non-significant increases observed at intervention sites. A mix of very slight increases and decreases observed at control sites
O_3								
Industrial sources	Butler 2011	55	NA	50	NA	Mean change: 5.0	< 0.0001	Significant reduction in concentration after the implementation of the intervention
	Deschênes 2012	NR	NR	NR	NR	DiD estimator (SE): −2.965 (0.747); % change: −5.8%	< 0.01	Significant reduction in concentration at the intervention relative to the control site,

(*Continued*)

								after the implementation of the intervention
	Lin 2013	NR	NA	NR	NA	Adjusted mean change (95% CI): −2.47% (−3.22 to −1.72)	< 0.05	Significant reduction in concentration after the implementation of the intervention
Multiple sources	Giovanis 2015	54.344 (17.244)	52.250 (16.627)	51.936 (14.476)	51.110 (13.951)	DiD estimator (SE): −1.268 (0.3887)	< 0.01	Significant reduction in concentration at the intervention relative to the control site, after the implementation of the intervention
Vehicular sources	Atkinson 2009	12.4	17.8	16.9	20.1	Mean change: Background int: −35.7%; Background con: −11.9%	NR	Decreases in concentrations observed at all monitors, though the difference was much greater at intervention sites. No statistical tests were performed, and no measure of variance was provided
	Davis 2008	NR	NA	NR	NA	Adjusted mean change (SE):	NR	An increase in concentrations was

(Continued)

						28% (5.4%)		observed after the implementation of the intervention
	[Friedman 2001]	81.3	con 1: 66.2; con 2: 61.2; con 3: 64.1	58.6	con 1: 58.8; con 2: 50.5; con 3: 52.2	Mean change: int: −27.9%; con 1: −11.1%; con 2: −17.5%; con 3: −18.5%	int: < 0.01; con 1: 0.11; con 2: 0.003; con 3: 0.01	Significant decreases observed at all intervention sites, as well as at all but one control sites
	[Peel 2010]	int 1: 76.3 (20.3); int 2: 68.5 (21.4)	Immediate area con: 71.8 (16.4); Surrounding area 1 con: 50.3 (19.7); Surrounding area 2 con: 59.5 (9.97); Surrounding area 3 con: 60.5 (12.1); surrounding states 1 con: 70.0 (26.0); Surrounding states 2 con: 49.0 (20.0); Surrounding states 3 con: 84.0 (22.0); Surrounding states 4 con: 77.1 (13.9)	Intervention 1: 53.6 (17.0); Intervention 2 45.9 (16.2):	Immediate area con: 52.4 (12.7); Surrounding area 1 con: 35.5 (7.28); Surrounding area 2 con: 49.4 (6.97); Surrounding area 3 con: 45.4 (8.17); surrounding states 1 con: 44.0 (21.0); Surrounding states 2 con: 40.0 (8.0); Surrounding states 3 con: 70.0 (14.0); Surrounding states 4 con: 62.9 (15.7)	NR	int 1: < 0.001; int 2: < 0.001; Immediate area con: < 0.001; Surrounding area 1 con: 0.004; Surrounding area 2 con: 0.001; Surrounding area 3 con: < 0.001; surrounding states 1 con: < 0.001; Surrounding states 2 con: 0.114; Surrounding states 3 con: 0.034; Surrounding states 4 con: 0.035	Significant decreases observed at all intervention sites, as well as at all but one control sites
CO								

(*Continued*)

Industrial sources	Deschênes 2012	NR	NR	NR	NR	DiD estimator (SE): −0.042 (0.035); % change: −8.1%	> 0.05	No observed change in mean concentration associated with the intervention
Vehicular sources	Atkinson 2009	0.4	0.32	0.3	0.3	Mean change: Background int: −19%; Background con: −3.8%	NR	Decreases in concentrations observed at all monitors, though the difference was much greater at intervention sites. No statistical tests were performed, and no measure of variance was provided
	Carrillo 2016	NR	NR	NR	NR	DiDiD estimator (SE): −0.0890 (0.0175); % change: −9%	< 0.001	Significant decrease in concentrations observed at peak hours at intervention sites relative to control sites due to the intervention
	Davis 2008	NR	NA	NR	NA	Adjusted mean change (SE): 31% (4.8%)	NR	An increase in concentrations was observed after the implementation of the intervention

(*Continued*)

	Dolislager 1997	NR	NA	3.5 (0.4)	NA	Adjusted mean change (SE): 8% (2%)	NR	A decrease in concentrations was observed after the implementation of the intervention
	Gallego 2013a	NR	NR	NR	NR	Mean change (SE): Immediate: −13% (5%); Long-term: 11.3% (8.1%)	Immediate: < 0.05; Long-term: 0.12	A decrease in concentrations was observed immediately following the intervention. However, the long-term effect showed an increase in concentrations over time
	Gallego 2013b	NR	NR	NR	NR	Mean change (SE): Immediate: −5.9% (9.8%); Long-term: 26.8% (7.1%)	Immediate: > 0.1; Long-term: < 0.01	A slight decrease in concentrations observed immediately following the intervention. However, the long-term effect shown a significant increase in concentrations over time
	Peel 2010	2.26 (1.38)	Immediate area con: 0.28 (0.10); Surround-	1.55 (0.43)	Immediate area con: 0.22 (0.09); Surround-	NR	int: 0.053; Immediate area con: 0.355;	Slight decreases observed at the intervention

(*Continued)*

			ing states 1 con: 2.03 (1.33); Surrounding states 2 con: 1.07 (0.52); Surrounding states 3 con: 1.70 (0.74)		ing states 1 con: 1.57 (1.26); Surrounding states 2 con: 1.06 (0.53); Surrounding states 3 con: 1.81 (0.71)		Surrounding states 1 con: 0.466; Surrounding states 2 con: 0.999; Surrounding states 3 con: 0.867	site, as well as the immediate area and one surrounding states

CONTRIBUTIONS OF AUTHORS

JB drafted the initial protocol, which was finalized after several rounds of input from all coauthors. RT and JB planned and conducted all search-related activities. JB, HB, SP, AR, AvE, LP and ER contributed to the data collection and analysis stages of the review. JB performed the initial GRADE assessment; ER reviewed all GRADE ratings, and supported JB in finalizing the GRADE assessment. JB drafted the initial manuscript, which was finalized after several rounds of input from HB, SP, AR, AvE, LP and ER. JB coordinated the work at all stages of review conduct, and HB provided substantial support in relation to the coordination.

DECLARATIONS OF INTEREST

Hanna Boogaard and Annemoon van Erp are employed by the Health Effects Institute, an independent non-profit organization supported by the US Environmental Protection Agency and world-wide automotive manufacturers. The views expressed in this review are those of the authors and do not necessarily reflect the views of the Health Effects Institute or its sponsors.

As several studies published by the Health Effects Institute were included in this review, we ensured that the data collection and risk of bias assessment of these studies were completed by other authors.

Jacob Burns: none known

Stephani Polus: none known

Lisa Pfadenhauer: none known

Anke Rohwer: none known

Ruth Turley: none known

Eva Rehfuess: none known

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We endeavoured to apply the a priori defined methods, as outlined in the published review protocol (Burns 2014), however we decided that certain changes to the methods were necessary. These changes are outlined in the following.

In listing the study designs to be included at the protocol stage, we did not foresee the need to specify the controlled ITS study design as distinct from the uncontrolled ITS study design. After identifying the relevant evidence base, however, we decided to follow the cITS-EPOC study design classification for labelling those studies applying an ITS design and analysis, and assessing data from one or multiple control sites.

In the protocol, we planned a single-reviewer title and abstract screening to remove any clearly irrelevant evidence. Given that only very few studies at this stage appeared to be clearly irrelevant, this step was not performed, and we instead followed a more rigorous duplicate title and abstract screening.

We planned to extract aspects related to intervention complexity using the Methodological Investigation of Cochrane Reviews of Complex Interventions (MICCI). This tool, now called the intervention Complexity Assessment Tool for Systematic Reviews (iCat˙SR), underwent substantial further development, and was only recently published (Lewin 2017). Thus we were unable to use it in the review.

Based on substantial differences in reporting and study quality, we made the post hoc decision to further classify included studies into main studies (cITS-EPOC, ITS-EPOC, CBA-EPOC, and CBA studies) and supporting studies (UBA studies and those not providing a relevant analytical comparison). This decision was extensively discussed among the author team and with the Editors of Cochrane Public Health. Only main studies contributed to the evidence synthesis of effects through harvest plots and narrative synthesis and were used to develop 'Summary of findings' tables.

The protocol details the use of both the Cochrane-EPOC 'Risk of bias' tool and the modified GATE tool to assess risk of bias. In piloting both tools, we felt that the modified GATE tool much better captured the risk of bias from the included studies, very few of which were designed and conducted as classical 'clinical' trials. We therefore only used the modified GATE tool for the assessment of risk of bias of included studies.

With regard to the application of the modified GATE tool for assessing risk of bias, we had initially planned to use the version of the tool developed for quantitative intervention studies. During piloting, however, we found that the version of the tool developed for correlation studies allowed for a much more appropriate assessment of the study designs covered by our review. We thus used the modified GATE tool for correlation studies for the risk of bias assessment of all included studies.

We had planned to plot intervention effects for PM_{10} and $PM_{2.5}$ reductions against WHO air quality guidelines to explore to what extent specific interventions may help in reaching these targets. Given the lack of homogeneous data fit for this purpose, this was not done.