Environment International 144 (2020) 105998

Contents lists available at ScienceDirect

Environment International

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

Review article

Long-term exposure to NO_2 and O_3 and all-cause and respiratory mortality: A systematic review and meta-analysis

Peijue Huangfu, Richard Atkinson*

Population Health Research Institute, St George's, University of London, UK

ARTICLE INFO

Keywords:
Nitrogen dioxide
Ozone
Cohort
Mortality
Systematic review
Meta-analysis

ABSTRACT

Background: WHO has published several volumes of Global Air Quality Guidelines to provide guidance on the health risks associated with exposure to outdoor air pollution. As new scientific evidence is generated, air quality guidelines need to be periodically revised and, where necessary, updated.

Objectives: The aims of the study were 1) to summarise the available evidence on the effect of long-term exposure to ozone (O_3) and nitrogen dioxide (NO_2) on mortality; 2) and to assess concentration response functions (CRF), their shape and the minimum level of exposures measured in studies to support WHO's update of the global air quality guidelines.

Data sources: We conducted a systematic literature search of the Medline, Embase and Web of Science databases following a protocol proposed by WHO and applied Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines for reporting our results.

Study eligibility criteria: Cohort studies in human populations (including sub-groups at risk) exposed to long-term concentrations of NO_2 and O_3. Outcomes assessed were all-cause, respiratory, Chronic Obstructive Pulmonary Disease (COPD) and Acute Lower Respiratory Infection (ALRI) mortality.

Study appraisal and synthesis methods: Studies included in the meta-analyses were assessed using a new Risk of Bias instrument developed by a group of experts convened by WHO. Study results are presented in forest plots and quantitative meta-analyses were conducted using random effects models. The certainty of evidence was assessed using a newly developed adaptation of GRADE.

Results: The review identified 2068 studies of which 95 were subject to full-text review with 45 meeting the inclusion criteria. An update in September 2018 identified 159 studies with 1 meeting the inclusion criteria. Of the 46 included studies, 41 reported results for NO_2 and 20 for O_3. The majority of studies were from the USA and Europe with the remainder from Canada, China and Japan. Forty-two studies reported results for all-cause mortality and 22 for respiratory mortality.

Associations for NO_2 and mortality were positive; random-effects summary relative risks (RR) were 1.02 (95% CI: 1.01, 1.04), 1.03 (1.00, 1.05), 1.03 (1.01, 1.04) and 1.06 (1.02, 1.10) per 10 μg/m^3 for all-cause (24 cohorts), respiratory (15 cohorts), COPD (9 cohorts) and ALRI (5 cohorts) mortality respectively. The review identified high levels of heterogeneity for all causes of death except COPD. A small number of studies investigated the shape of the concentration–response relationship and generally found little evidence to reject the assumption of linearity across the concentration range.

Studies of O_3 using annual metrics showed the associations with all-cause and respiratory mortality were 0.97 (0.93, 1.02) and 0.99 (0.89, 1.11) per 10 μg/m^3 respectively. For studies using peak O_3 metrics, the association with all-cause mortality was 1.01 (1.00, 1.02) and for respiratory mortality 1.02 (0.99, 1.05), each per 10 μg/m^3. The review identified high levels of heterogeneity. Few studies investigated the shape of the concentration–response relationship.

Certainty in the associations (adapted GRADE) with mortality was rated low to moderate for each exposure-outcome pair, except for NO_2 and COPD mortality which was rated high.

Limitations: The substantial heterogeneity for most outcomes in the review requires explanation. The evidence base is limited in terms of the geographical spread of the study populations and, for some outcomes, the small number of independent cohorts for meta-analysis precludes meaningful meta-regression to explore causes of heterogeneity. Relatively few studies assessed specifically the shape of the CRF or multi-pollutant models.

Conclusions: The short-comings in the existing literature base makes determining the precise nature (magnitude

* Corresponding author at: Population Health Research Institute, St George's, University of London, Cranmer Terrace, London SW17 0RE, UK.
E-mail address: atkinson@sgul.ac.uk (R. Atkinson).

https://doi.org/10.1016/j.envint.2020.105998
Received 2 March 2020; Received in revised form 16 July 2020; Accepted 16 July 2020
Available online 05 October 2020

and linearity) of the associations challenging. Certainty of evidence assessments were moderate or low for both NO_2 and O_3 for all causes of mortality except for NO_2 and COPD mortality where the certainty of the evidence was judged as high.

1. Introduction

Outdoor air pollution has been a global concern for decades, partially due to economic growth and urbanisation. Air pollution has been recognised as a major environmental hazard to human health and a cause of mortality and morbidity (Burnett et al., 2018; World Health Organization, 2012). Nitrogen dioxide (NO_2) is a toxic gas with both outdoor (e.g. traffic) and indoor (e.g. gas cooking) sources. In outdoor urban environments, NO_2 is derived primarily from the oxidation of nitric oxide (NO) a primary traffic pollutant. Ozone (O_3) is a highly reactive oxidative gas formed by chemical reactions in the atmosphere involving oxides of nitrogen, volatile organic compounds and driven by solar radiation. In urban areas with high traffic density, nitrogen oxides (NO and NO_2) are commonly high and often negatively correlated with O_3 during daylight hours. Evidence suggested that NO_2 and O_3 both detrimentally affect people's health, including respiratory function, hospital admission, and premature death (Nuvolone et al., 2018; Strickland et al., 2010; Malig et al., 2016; Urman et al., 2014).

WHO has previously published Global Air Quality Guidelines (AQGs) to provide guidance to the public and to policy and other decision makers on the health risks associated with exposure to outdoor air pollution (WHO, 2000; WHO, 2005). As new scientific evidence is generated, air quality guidelines need to be periodically revised and, where necessary, updated. The update of the WHO AQGs is a global project coordinated by the WHO Regional Office for Europe's European Centre for Environment and Health (ECEH) in Bonn (Germany), including participation from all WHO Regions and WHO headquarters. In support of this update, systematic reviews of both short- and long-term studies on air pollutants and mortality and morbidity are necessary.

This review focuses upon long-term concentrations of NO_2 and O_3 and all-cause and respiratory mortality studied in epidemiological cohort studies. Previous reviews of NO_2 (Atkinson et al., 2018; Faustini et al., 2014; Hoek et al., 2013; EPA, 2016; WHO, 2013) and O_3 (WHO, 2013; Atkinson et al., 2016; EPA US, 2013) have been undertaken. However, in order to ensure guideline revisions are informed by the latest evidence, a new review was undertaken with formal evaluation of Risk of Bias (RoB) and certainty of evidence (Grading of Recommendations Assessment, Development and Evaluation (GRADE)). For the reviews new adaptations of the RoB and GRADE assessments were developed.

The aims of the study were 1) to conduct an extensive systematic review and meta-analysis on associations between long-term concentrations of NO_2 and O_3 on mortality; and 2) to assess concentration response functions, their shape and the minimum level of exposures measured in studies. The following framework (Appendix Table B1) was used to select the critical health outcome(s) for each pollutant: 1) Evidence on causality for a health outcome based upon the latest determination (causal or likely causal) from US EPA, IARC, Health Canada or other integrated science assessments available; 2) Using the precautionary principle, additional most severe health outcomes other than causal or likely causal (e.g. suggestive causality) were considered for inclusion taking into account contribution to burden of disease (prevalence of disease, disability weight, etc), policy implications, expected increase in exposure to a pollutant in the future, etc.; 3) causality determination superseded severity of a health outcome but, in some cases, two (or more) different health outcomes may be systematically evaluated for the same pollutant (e.g. one with a definite or likely causal link to the pollutant, and another health outcome for which the evidence is suggestive but which is very severe or prevalent in the population). Severity of disease was informed by considerations proposed by the joint European Respiratory Society and American Thoracic Society latest policy statement on health effects from air pollution (fatality, persistence of effect, susceptible groups, and medical/functional significance including loss of autonomy and reduced quality of life) (Thurston et al., 2017).

This systematic review uses the following Population, Exposure, Comparison, Outcome, Study Design (PECOS) statement: in any population, including subgroups of susceptible adults and children (P), what is the health effect of long-term ambient exposure of NO_2 and O_3 (E) per unit increase in $\mu g/m^3$ (C) on all cause, respiratory, Chronic Obstructive Pulmonary Disease (COPD), and Acute Lower Respiratory Infection (ALRI) mortality (O), observed in cohort studies (S)? Additionally, in

Table 1
Inclusion and exclusion criteria for each PECOS domain in relation to long-term exposure and health effects to selected air pollutants.

PECOS	Inclusion	Exclusion
Population	• General human population (including sub-groups at risk: children, pregnant women, elderly, or patients with particular conditions), of all ages, developed and developing areas, both urban and rural. No geographical restrictions. • Study population expose to the pollutant of interest via inhalation through ambient air predominantly	• Study population expose to the pollutant of interest in occupational settings or indoor exposure exclusively
Exposure	• Long-term exposure (order of years) to ambient air O_3 and NO_2 expressed in a concentration unit (ppb and $\mu g/m^3$ respectively).	• Less than one year of data available
Comparator	• Exposure to per concentration increased unit of the air pollutant of interest in the same population	• Increment for hazard ratio not given
Outcome	• Health outcomes selected in relation to long-term exposure include (ICD 10 codes, version 2016 in brackets): **all cause (A00-Z99); respiratory (J00-J99); COPD (J40-47) and ALRI (J12-J18, J20-J28) mortality** [Note: Studies vary in selection of codes.]	• Birth outcomes (due to neonatal exposure of pollutant)
Study	• Human epidemiological studies including: o Prospective and retrospective cohort studies • Published (or accepted for publication i.e. in press) journal articles in any language (abstract in English language), conference abstracts and papers, letters, notes, grey literature. • If suitable articles are identified published in languages not known by the SRT, further assistance will be sought after (members of the GDG or external review team from different regions, colleagues, researcher networks, etc)	• Qualitative studies • Case control studies • No adjustment for socio-economic status (individual or area) • Studies where no original data were analysed • Reviews and methodological papers • Non-human studies (*in vivo, in vitro,* other) • Insufficient information given to standardise hazard ratio and precision (standard error or confidence interval)

SRT: Systematic review team

these studies, what is the lowest concentration that produces a measurable increase in risk?"

2. Methods

2.1. Protocol

The protocol for this review was developed by WHO based largely on standards set by the Cochrane Collaboration and adapted for application to observational studies (Higgins, 2011) and the Preferred Reporting Items for Systematic Review and Meta-analysis Protocols (PRISMA-P) standards (Moher, 2009; Shamseer et al., 2015). The protocol is published in the International Prospective Register of Systematic Reviews (PROSPERO) reference number CRD42018089853.

2.2. Eligibility criteria

The included population comprised general human population (including sub-groups at risk) of all ages, exposed to long-term (i.e. > one year) concentrations (order of years) to ambient NO_2 and O_3 (Table 1). As whole populations are exposed to varying levels of air pollution, the comparison is between subjects in the same population exposed at different concentrations of the pollutant. Outcomes included in the review were mortality from all-causes (A00-Z99); respiratory diseases (J00-J99); COPD (J40-47) and ALRI (J12-J18, J20-J28). We included publication of prospective and retrospective cohort studies, published (or accepted for publication) journal articles in any language, conference abstracts and papers, letters, notes, and grey literature. Cohort studies were selected for the review as they are used in environmental epidemiology to assess associations between long-term (over years) concentrations of pollutants and risk of death.

We excluded 1) studies with exposure of interest in occupational or indoor settings exclusively; 2) studies that explored neonatal exposure and birth outcomes; 3) studies that had less than one year of data available; 4) studies did not report exposure increment for the health effect; 5) qualitative studies; 6) case-control studies (not applicable to the study of mortality in air pollution epidemiology); 7) studies without any adjustment for socio-economic status (either at individual or area level); 8) studies had no original data analysed; 9) reviews and methodological papers; 10) non-human studies (e.g. *in vivo, in vitro*); 11) studies with insufficient information to standardise effect size and precision (Table 1).

2.3. Information sources

To identify articles reporting results of studies matching the PECOS questions the bibliographic databases Medline, Embase and Web of Science were searched without limitation on date. The search strategy included terms related to the study design, pollutant and outcome is documented fully in Appendix Table B2. Results of the three searches were combined and de-duped. In addition, the reference lists of relevant reviews were scanned to identify additional published data matching the PECOS question. All references were downloaded into Endnote reference manager software [Endnote X7.8 Thomson Reuters].

2.4. Study selection

Two authors (PH and RWA) independently screened the titles and abstracts of the studies returned by the systematic searches. Articles that did not meet the prespecified eligibility criteria (Table 1) were identified and excluded.

2.5. Data collection

Data extraction was conducted independently by PH and RWA and compared. Study information collected included citation details (title, authors, date of publication); cohort details (name, country, patient/population group, follow up period(s)); subject characteristics (age at recruitment, sex, occupation); confounders measured; exposure assessment method (e.g. monitor, land use regression model); mean and concentration range of the pollutant (e.g. 5th & 95th percentile or minimum/ maximum or 25th/75th percentile values); outcome assessment (e.g. death records, ICD coding); and details of the risk estimates including exposure unit of measurement, metric description (e.g. annual mean), period of year of exposure assessment (all-year or 'warm/peak season'), and 95% confidence interval (CI) of the risk estimates for relevant outcomes; and details on co-pollutant models.

Where disagreement occurred, it was resolved by discussion. Data extracted from the articles were entered into an Excel spreadsheet. In the absence of complete descriptions of exposure assessment and outcomes, effect estimates, or other important information, individual authors were contacted and the information requested.

2.6. Standardisation of risk estimates

Risk estimates extracted from cohort studies were hazard ratios (HR) and 95% CIs in the units reported in the original studies. For the purpose of this review HRs were considered to be equivalent to relative risks (RR). Where risk estimates were reported in parts per billion (ppb), standard factors were used to convert ppb to $\mu g/m^3$; for NO_2 and O_3 these were 1.88 and 1.96 respectively (Air Information Resource, 2005). RRs (and 95% CIs) were scaled to 10 $\mu g/m^3$ increments by taking the natural logarithm of the risk estimates (and confidence limits) and then standardising to 10 $\mu g/m^3$ by dividing by the original risk increment and multiplying by 10. Standardisation to a common metric is required to enable risk estimates to be combined in a meta-analysis.

2.7. Data synthesis

Some cohorts have been analysed in more than one study (e.g. for different follow-up periods, for more sophisticated air pollution models etc.) or included in a multi-cohort analysis (e.g. the European Study of Cohorts for Air Pollution Effects (ESCAPE) study). We therefore selected only one result from each cohort for inclusion in the meta-analysis. The selection procedure was based upon the following criteria: the cohort using the most recent follow-up period (i.e. more recent studies with longer follow-up, represent more recent exposure status, and with improved exposure measurement to aid the global guidelines update), results from the full cohort rather than a subset, and if results for a cohort were not included in a multi-cohort study.

Meta-analysis was performed using random-effects (RE) models with heterogeneity estimated using restricted maximum likelihood (REML) as implemented in the 'admetan' command in STATA Vn 15 (StataCorp, 2017). Forest plots were produced using the 'admetan' program in STATA. Summary estimates (i.e. RR), 95% CIs, Chi-square statistics, tau^2, I^2 and 80% prediction intervals were reported. Where more than 10 studies were available for analysis, potential small study bias was assessed using the funnel plot and funnel plot asymmetry using Egger's test (Begg and Berlin, 1989; Egger et al., 1997) as implemented in the STATA command 'metabias'. Meta regression was used to study the relationship between study RRs and mean pollutant concentrations ('metareg' in STATA Vn 15) when 10 or more estimates were available.

Cohorts investigating O_3 and mortality may use annual or 'peak' season (e.g. April-September) measures of exposure. Meta-analyses for O_3 were therefore stratified by exposure period.

2.8. Risk of bias evaluation

A new RoB tool was developed by a working group convened by WHO for the assessment of cohort studies in air pollution epidemiology (https://www.euro.who.int/en/health-topics/environment-and-

health/air-quality/publications/2020/risk-of-bias-assessment-instrument-for-systematic-reviews-informing-who-global-air-quality-guidelines-2020). The tool consisted of six domains: confounding, selection bias, exposure assessment, outcome assessment, missing data and selective reporting, each including one to four subdomains. In total, 13 sub-domains (Morgan et al., 2019) were each rated as low, moderate or high risk of bias. If any one sub-domain was rated medium or high RoB then the domain was rated similarly. RoB was applied to each pollutant-outcome pair for studies included in a meta-analysis. Assessment of RoB for the confounding sub-domain "Were all confounders considered adjusted for in the analysis?" was based upon the inclusion in the analysis of critical and potential confounders according to the outcome. For all-cause mortality critical confounders were: age, sex, body mass index (BMI) and an indicator (individual or area) for socio-

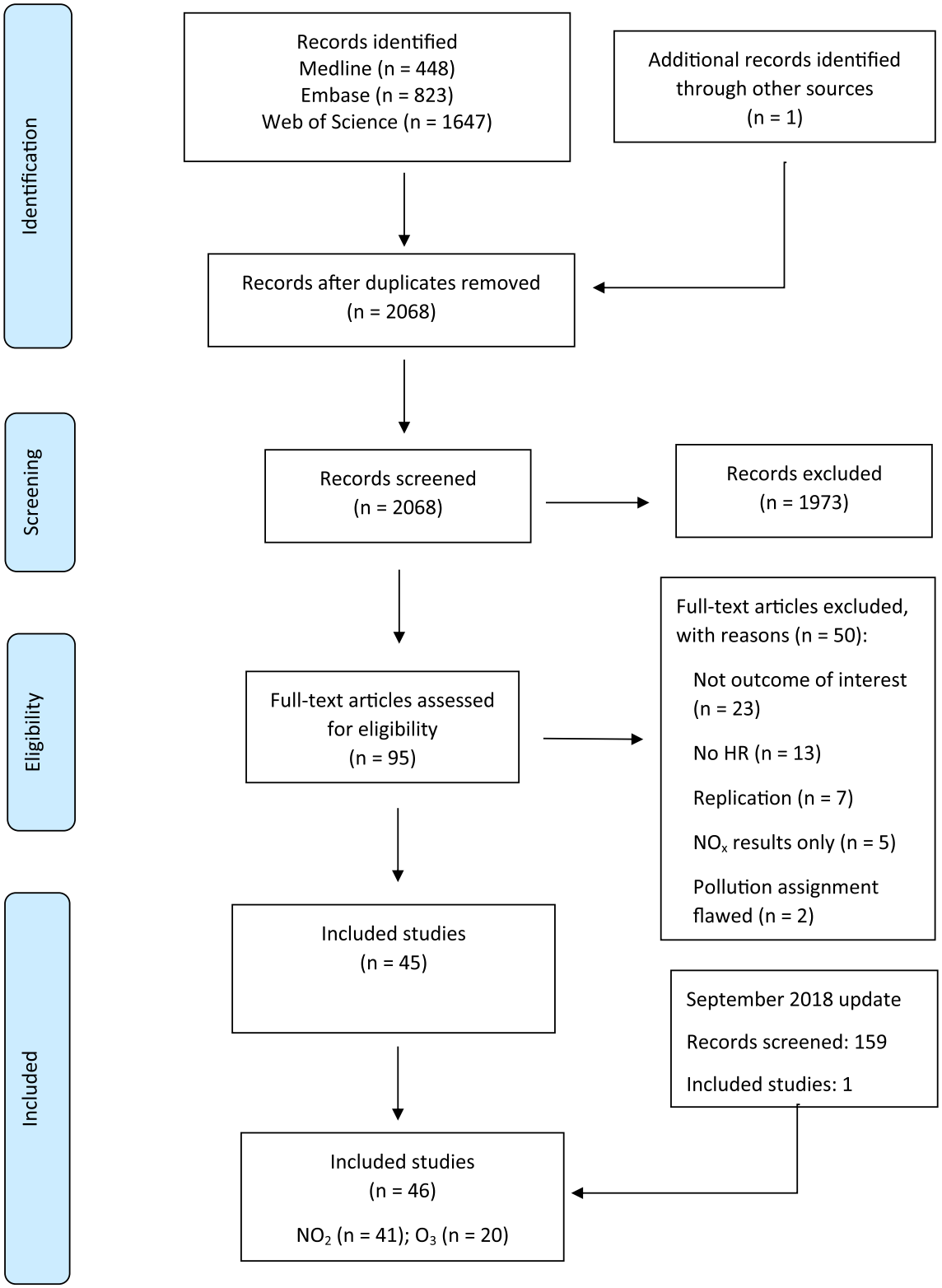

Fig. 1. Flowchart of assessment of studies.

Table 2
Summary of characteristics of studies included in the systematic review – Nitrogen Dioxide.

a) All-cause

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean ($\mu g/m^3$)	Lowest concentration recorded
Abbey 1999 (Abbey et al., 1999)	AHSMOG	General	USA	5,652	639	FM	27–95	Yes	Yes	Indiv	Monitoring station	129.9	NR
Beelen 2014 (Beelen et al., 2014)	ESCAPE	General	Europe	367,251	29,076	FM	All	Yes	Yes	Indiv	Land use regressions	24.9	NR
Bentayeb 2015 (Bentayeb et al., 2015)	Gazel cohort	General	France	20,327	1967	FM	35–50	Yes	Yes	Indiv	Chemistry-transport model	28	NR
Brunekreef 2009 (Brunekreef et al., 2009)	NLCS-AIR	General	Netherlands	120,227	17,674	FM	55–69	No	Yes	Area	Interpolation, land use regression	38	5th (22.0)
Carey 2013 (Carey et al., 2013)	CPRD	General	England	830,429	82,421	FM	40–89	Yes	Yes	Area	Air dispersion model	22.5	5th (4.5)
Cesaroni 2012 (Cesaroni et al., 2012)[i]	Rome longitudinal study	General	Italy	684,204	45,006	FM	45–80	No	No	Indiv	Land use regressions	45.7	Min (18.8)
Cesaroni 2013 (Cesaroni et al., 2013)	Rome longitudinal study	General	Italy	1,265,058	144,441	FM	> =30	No	No	Indiv	Land use regressions	43.6	Min (13.0)
Chen 2016 (Chen et al., 2016)	Four northern Chinese cities	General	China	39,054	1353	FM	23–89	Yes	Yes	Indiv	Land use regressions	40.7	Min (18.0)
Crouse 2015a (Crouse et al., 2015a)	CanCHEC	General	Canada	2,521,525	301,115	FM	25–89	Indirect	Indirect	Indiv	Land use regressions	21.8	Min (0.0)
Crouse 2015b (Crouse et al., 2015b)[i]	CanCHEC	General	Canada	735,590	80,660	FM	25–89	No	No	Indiv	Land use regressions	47.4	6
Desikan 2016 (Desikan et al., 2015)	South London Stroke Register	Patient	UK	1800	729	FM	68.8 (15.8)	No	No	Area	KCLurban model	44.6	25th (41.8)
Filleul 2005 (Filleul et al., 2005)	PAARC	General	France	14,284	2531	FM	25–59	Yes	Yes	Indiv	Monitoring station	36.5	Min (12.0)
Fischer 2015 (Fischer et al., 2015)	DUELS	General	Netherlands	7,218,363	668,206	FM	> =30	No	No	Indiv	Land use regressions	31	5th (19.0)
Gehring 2006 (Gehring et al., 2006)[ii]	German cohort	General	Germany	4752	399	F	50–59	No	Yes	Indiv	GIS monitoring station	39	Min (22.0)
Hart 2011 (Hart et al., 2011)	US trucking industry cohort	General	USA	53,814	4806	M	15.3–84.9	No	No	Indiv	Spatial smoothing and GIS	26.7	5th (8.3)
Hart 2013 (Hart et al., 2013)	Nurses Health Study	General	USA	84,562	11,502	F	30–55	Yes	Yes	Indiv	Spatial smoothing and GIS–based covariates	26.1	5th (8.3)
Hartiala 2016 (Hartiala et al., 2016)	The Cleveland Clinic GeneBank study	Patient	USA	6575	4363	FM	64 (11)	No	Yes	Indiv	Monitoring station	35.9	Min (9.4)
HEI 2000 (Health Effects Institute, 2000)	Six Cities	General	USA	8111	1430	FM	25–74	Yes	Yes	Indiv	Monitoring station	30.3	NR
HEI 2000 (Health Effects Institute, 2000)[i]	ACS CPS-II	General	USA	552,138	38,963	FM	> =30	Yes	Yes	Indiv	Monitoring station	90.1	47.8
Heinrich 2013 (Heinrich et al., 2013)[ii]	German cohort	General	Germany	4752	715	F	50–59	No	Yes	Indiv	GIS Monitoring station	39	Min (20.0)
Hoek 2002 (Hoek et al., 2002)[i]	NLCS-AIR	General	Netherlands	2788	487	FM	55–69	Yes	Yes	Indiv	GIS Monitoring station	36.6	5th (20.3)
Jerrett 2009 (Jerrett et al., 2009)	Toronto respiratory cohort	Patient	Canada	2360	298	FM	60 (49 69)	Yes	Yes	Area	Land use regressions	39.1	NR
Jerrett 2013 (Jerrett et al., 2013)[i]	ACS CPS-II	General	USA	73,711	19,755	FM	> =30	Yes	Yes	Indiv	Land use regressions	23.1	5th (14.9)
Krewski 2009 (Krewski et al., 2009)[i]	ACS CPS-II	General	USA	406,917		FM	> =30	Yes	Yes	Indiv	Monitoring station	52.5	5th (27.4)
Lipfert 2006 (Lipfert et al., 2006)	Washington University-EPRI Veterans	Patient	USA	28,635	5638	M	51 (12)	Yes	Yes	Area	Monitoring station	37.2	5th (16.5)

(*continued on next page*)

Table 2 (*continued*)

a) All-cause

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean ($\mu g/m^3$)	Lowest concentration recorded
Lipfert 2006 (Lipfert et al., 2006)[i]	Washington University-EPRI Veterans	Patient	USA	~15,200	5638	M	51 (12)	Yes	Yes	Area	Monitoring station	38.2	Min (7.3)
Lipsett 2011 (Lipsett et al., 2011)	CTS	General	USA	12,336	4147	F	> =30	Yes	Yes	Area	GIS Monitoring station	63.1	Min (9.9)
Maheswaren 2010 (Maheswaran et al., 2010)[i]	SLSR	Patient	England	3320	1856	FM	70.4 (14.6)	No	Yes	Area	GIS Monitoring station	41	Min (32.2)
Raaschou-Nielsen 2012 (Raaschou-Nielsen et al., 2012)[ii]	DCH	General	Denmark	52,061	5534	FM	50–64	Yes	Yes	Indiv	AirGIS dispersion	16.9	5th (10.5)
Rosenlund 2008 (Rosenlund et al., 2008)	CHD survivors cohort	Patient	Italy	6513	1802	FM	35–84	No	No	Area	Land use regressions	48.5	Min (24.0)
Tonne 2013 (Tonne and Wilkinson, 2013)	MINAP (ACS survivors)	Patient	England & Wales	154,204	39,863	FM	> =25	No	Yes	Area	Dispersion model	18.5	NR
Tonne 2016 (Tonne et al., 2016)[iii]	MINAP (ACS survivors)	Patient	UK	18,138	5129	FM	> =25	No	Yes	Area	Dispersion model	37.1	25th (32.6)
Turner 2016 (Turner et al., 2016)	ACS CPS-II	General	USA	669,046	237,201	FM	> =30	Yes	Yes	Indiv	Land use regressions	21.8	5th (9.6)
Weichenthal 2017 (Weichenthal et al., 2017)	CanCHEC	General	Canada	2,448,500	233,340	FM	25–89	No	No	Indiv	Land use regressions	21.6	5th (6.3)
Yang 2018 (Yang et al., 2018)	Hong Kong elderly	General	China	61,386	NR	FM	> =65	Yes	Yes	Indiv	Land use regressions	104	5th (81.3)
Yorifuji 2010 (Yorifuji et al., 2010)[i]	Shizuoka elderly cohort	General	Japan	12,209	1232	FM	65–84	Yes	Yes	Indiv	Land use regressions	25	5th (1.2)
Yorifuji 2013 (Yorifuji et al., 2013)	Shizuoka elderly cohort	General	Japan	13,412	1663	FM	65–84	Yes	Yes	Indiv	Land use regressions	22	Min (9.4)

b) Respiratory

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean ($\mu g/m^3$)	Lowest concentration recorded
Abbey 1999 (Abbey et al., 1999)	AHSMOG	Population	USA	2278	63	FM	27–95	Yes	Yes	Indiv	Monitoring station	129.9	NR
Brunekreef 2009 (Brunekreef et al., 2009)	NLCS-AIR	Population	Netherlands	120,227	1046	FM	55–69	No	Yes	Area	Interpolation, land use regression	38	5th (22.0)
Carey 2013 (Carey et al., 2013)	CPRD	Population	England	830,429	10,500	FM	40–89	Yes	Yes	Area	Air dispersion model	22.5	Min (4.5)
Cesaroni 2013 (Cesaroni et al., 2013)	Rome longitudinal study	Population	Italy	1,265,058	8825	FM	> =30	No	No	Indiv	Land use regressions	43.6	Min (13.0)
Crouse 2015a (Crouse et al., 2015a)	CanCHEC	Population	Canada	2,521,525	24,900	FM	25–89	Indirect	Indirect	Indiv	Land use regressions	21.8	Min (0.0)
Crouse 2015b (Crouse et al., 2015b)[i]	CanCHEC	Population	Canada	735,590	6450	FM	25–89	No	No	Indiv	Land use regressions	47.4	6
Dimakopoulou 2014 (Dimakopoulou et al., 2014)	ESCAPE	Population	Europe	307,553	1559	FM	All	Yes	Yes	Indiv	Land use regressions	20.4	NR
Fischer 2015 (Fischer et al., 2015)	DUELS	Population	Netherlands	7,218,363	65,132	FM	> =30	No	No	Indiv	Land use regressions	31	5th (19.0)
Hart 2011 (Hart et al., 2011)	US trucking industry cohort	Population	USA	53,814	317	M	15.3–84.9	No	No	Indiv	Spatial smoothing and GIS	26.7	5th (8.3)
Heinrich 2013 (Heinrich et al., 2013)[ii]	German cohort	Population	Germany	4752	34	F	50–59	No	Yes	Indiv	GIS Monitoring station	39	Min (20.0)

(*continued on next page*)

Table 2 (*continued*)

b) Respiratory

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Jerrett 2009 (Jerrett et al., 2009)	Toronto respiratory cohort	Patient	Canada	2360	75	FM	60 (49 69)	Yes	Yes	Area	Land use regressions	39.1	NR
Jerrett 2013 (Jerrett et al., 2013)[i]	ACS CPS-II	Population	USA	73,711	1990	FM	> =30	Yes	Yes	Indiv	Land use regressions	23.1	5th (14.9)
Katanoda 2011 (Katanoda et al., 2011)	3 Japanese Prefectures	Population	Japan	63,520	677	FM	> =40	No	Yes	Indiv	Monitoring stations	32	NR
Lipsett 2011 (Lipsett et al., 2011)	CTS	Population	USA	12,336	404	F	> =30	Yes	Yes	Area	GIS Monitoring station	63.1	Min (9.9)
Turner 2016 (Turner et al., 2016)	ACS CPS-II	Population	USA	669,046	20,484	FM	> =30	Yes	Yes	Indiv	Land use regressions	21.8	5th (9.6)
Weichenthal 2017 (Weichenthal et al., 2017)	CanCHEC	Population	Canada	2,448,500	21,100	FM	25–89	No	No	Indiv	Land use regressions	21.6	5th (6.3)
Yang 2018 (Yang et al., 2018)	Hong Kong elderly	Population	China	61,386	NR	FM	> =65	Yes	Yes	Indiv	Land use regressions	104	5th (81.3)
Yorifuji 2013 (Yorifuji et al., 2013)	Shizuoka elderly cohort	Population	Japan	13,412	281	FM	65–84	Yes	Yes	Indiv	Land use regressions	22	Min (9.4)
Yorifuji 2010 (Yorifuji et al., 2010)[i]	Shizuoka elderly cohort	Population	Japan	12,209	199	FM	65–84	Yes	Yes	Indiv	Land use regressions	25	5th (1.2)

c) COPD

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Carey 2013 (Carey et al., 2013)	CPRD	Population	England	830,429	4104	FM	40–89	Yes	Yes	Area	Air dispersion model	22.5	Min (4.5)
Crouse 2015a (Crouse et al., 2015a)	CanCHEC	Population	Canada	2,521,525	14,170	FM	25–89	Indirect	Indirect	Indiv	Land use regressions	21.8	Min (0.0)
Gan 2013 (Gan et al., 2013)	Vancover	Population	Canada	467,994	541	FM	45–85	No	No	Area	Land use regressions	32.2	Min (15.3)
Hart 2011 (Hart et al., 2011)	US trucking industry cohort	Population	USA	53,814	209	M	15.3–84.9	No	No	Indiv	Spatial smoothing and GIS	26.7	5th (8.3)
Katanoda 2011 (Katanoda et al., 2011)	3 Japanese Prefectures	Population	Japan	63,520	677	FM	> =40	No	Yes	Indiv	Monitoring stations	32	NR
Naess 2007 (Naess et al., 2007)	Oslo Cohort	Population	Norway	143,842	503	FM	51–90	No	No	Indiv	Air dispersion model	39	Min (1.9)
Turner 2016 (Turner et al., 2016)	ACS CPS-II	Population	USA	669,046	9967	FM	> =30	Yes	Yes	Indiv	Land use regressions	21.8	5th (9.6)
Yang 2018 (Yang et al., 2018)	Hong Kong elderly	Population	China	61,386	NR	FM	> =65	Yes	Yes	Indiv	Land use regressions	104	5th (81.3)
Yorifuji 2013 (Yorifuji et al., 2013)	Shizuoka elderly cohort	Population	Japan	13,412	50	FM	65–84	Yes	Yes	Indiv	Land use regressions	22	Min (9.4)
Yorifuji 2010 (Yorifuji et al., 2010)[i]	Shizuoka elderly cohort	Population	Japan	12,209	35	FM	65–84	Yes	Yes	Indiv	Land use regressions	25	5th (1.2)

d) ALRI

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Carey 2013 (Carey et al., 2013)	CPRD	Population	England	830,429	4065	FM	40–89	Yes	Yes	Area	Air dispersion model	22.5	Min (4.5)
Katanoda 2011 (Katanoda et al., 2011)	3 Japanese Prefectures	Population	Japan	63,520	677	FM	> =40	No	Yes	Indiv	Monitoring stations	32	NR
Turner 2016 (Turner et al., 2016)	ACS CPS-II	Population	USA	669,046	6599	FM	> =30	Yes	Yes	Indiv	Land use regressions	21.8	5th (9.6)
Yang 2018 (Yang et al., 2018)	Hong Kong elderly	Population	China	61,386	NR	FM	> =65	Yes	Yes	Indiv	Land use regressions	104	5th (81.3)

(*continued on next page*)

Table 2 (*continued*)

d) ALRI

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure		
								BMI	Smoking	SES	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Yorifuji 2013 (Yorifuji et al., 2013)	Shizuoka elderly cohort	Population	Japan	13,412	159	FM	65–84	Yes	Yes	Indiv	Land use regressions	22	Min (9.4)
Yorifuji 2010 (Yorifuji et al., 2010)[i]	Shizuoka elderly cohort	Population	Japan	12,209	35	FM	65–84	Yes	Yes	Indiv	Land use regressions	25	5th (1.2)

Abbreviation: BMI – body mass index; SES – socio-economic status; NR – not reported.
Confounding adjustment: for BMI and smoking, if it was adjusted (yes), direct or indirect adjustment were recorded; for SES, all studies were adjusted for, therefore adjustment at area or individual level were recorded
[i] Excluded from analysis due to more recent cohort follow-up was available.
[ii] Excluded from analysis due to cohorts was included in ESCAPE study.
[iii] Excluded from analysis due to cohort was a subset of Tonne 2013.

economic status (SES). For respiratory outcomes critical confounders included age, sex, smoking and SES. Potential critical confounders included: year of enrolment, ethnicity, diet, physical activity, marital status, and smoking/BMI according to inclusion as critical confounder.

2.9. Additional analyses

Pre-specified sub-group analyses were performed where sufficient numbers of studies were available for meaningful analysis (i.e. a minimum of five studies in each subgroup). Sub-groups were defined by: 1) cohorts comprised of patient group versus general population cohorts; 2) cohorts able to control for individual measures of BMI, smoking and SES; 3) WHO region (Region of the Americas (AMR); European Region (EUR); Western Pacific Region (WPR)); and 4) by low/high RoB. Sensitivity analyses were conducted excluding high RoB studies (where sub-group analysis was not performed).

2.10. Certainty of evidence assessment

Certainty of evidence for each pollutant / outcome pair was assessed using a modified GRADE adapted following discussions of a working group composed of methodologists and GDG members, under the oversight of the WHO Secretariat (see Appendix A for WHO guidance in detail). We briefly describe the approach here.

The GRADE instrument is comprised of eight domains. In each domain the starting level of certainty in the evidence was 'moderate'. In five domains: limitations in studies; indirectness; inconsistency; imprecision; and publication bias the certainty of evidence could be downgraded following assessment of the evidence. In three domains: large effect size; plausible confounding towards null; and dose–response relationship the certainty in the evidence could be upgraded. The overall certainty assessment of the body of evidence was then determined by adding together the downgrades and upgrades across domains. An overall rating of high meaning that further research is very unlikely to change the confidence in the estimate of the effect; moderate that further research is likely to have an important impact on the confidence in the estimate of the effect; low, that further research is very likely to have an important impact on the confidence in the estimate of the effect; or very low, meaning that the estimate of the effect is very uncertain. Some domains of this tool were evaluated using results of the RoB, heterogeneity, sensitivity, and publication bias analyses, which were previously described in the methodology.

A brief outline of each domain is given below:

Domain 1, limitation in studies, incorporated assessment of RoB, with certainty of evidence downgraded only if meta-analysis of studies of low RoB differed from meta-analysis of all studies. Hence, the presence of small studies with high RoB but limited influence on the meta-analysis was not a reason to downgrade.

Domain 2, indirectness, related to how well the PECO in the studies in the meta-analysis reflected the original PECO;

Domain 3, inconsistency domain, addressed heterogeneity using an 80% prediction interval. The evidence certainty was downgraded if substantial heterogeneity was present as indicated by the 80% PI including 1 and twice the width of the 95% CI;

Domain 4, imprecision, was evaluated using sample size calculations rather than the confidence interval for the pooled estimate since in environmental health there are no clinical decision thresholds involved;

Domain 5, small study bias, assessment was based upon a funnel plot and Eggers test used to assess funnel plot asymmetry. The evidence certainty was downgraded only if there was clear indication of bias/asymmetry;

Domain 6, effect size. Potential upgrades to certainty of evidence related to effect size was assessed using the E-value calculated with increments of 40 μg/m^3 and 30 μg/m^3 for NO_2 and O_3 respectively. (VanderWeele and Ding, 2017) E-values were not calculated when the summary RR was below 1;

Domain 7, statistically significant RR after adjustment for plausible confounding. As the omission of potential confounders could alter the RR in either direction no upgrading was considered.

Domain 8, evidence of a dose–response relationship. A RR with lower 95% CI above 1 together with evidence from studies that examined specifically the shape of the concentration response function was considered sufficient evidence to upgrade certainty for this domain; else no upgrade was applied.

2.11. Deviations from protocol

The following deviations from the published protocol were implemented:

1. STATA program 'admetan' used instead of 'metan' in order to implement estimation of between study heterogeneity using restricted maximum likelihood. This was required as it is acknowledged that the method of D&L underestimates tau^2. (Veroniki et al., 2016)
2. O_3 studies assign estimated concentrations for annual and 'peak' periods. As O_3 is a seasonal pollutant it is not appropriate to combine study results for the different exposure windows, hence all analyses were stratified by annual and warm season exposures.

3. Results

3.1. Search strategy

The search strategies were applied in January 2018 and returned 2918 studies. One further study not captured by the searches was identified from another review. (Atkinson et al., 2018) After combining the search results and removal of duplicates, 2068 studies remained for screening via title/abstract. The searches were re-run on 11th September 2018 to identify new studies published during the review process. After removal of duplicates, this update identified a further 159 studies for screening of titles/abstracts. The results of the search strategy and the screening process are documented in the PRISMA flow diagram (Fig. 1).

3.2. Study selection

Of the 2068 studies identified in the initial search, 1973 were excluded after title and abstract screening. The remaining studies (n = 95) were subject to full-text assessment. Fifty studies did not meet the inclusion criteria (hence 45 studies were included in the review (Fig. 1).

Of the 159 studies identified at the review update in September 2018, one study was eligible for inclusion in the review. Hence, a total of 46 studies were included in the review. Tables 2 and 3 show the included studies by exposure and outcome.

3.3. Description of excluded studies

Fifty studies did not meet the inclusion criteria and were excluded. The reasons for exclusion were: 23 studies did not include the outcome of interest; 13 did not report results that can be converted into RR or HR; seven replicated results from other papers; five reported results for NO_x; and two studies were excluded because the assignment of pollution concentrations were related to length of follow-up. References for the excluded studies are listed in Appendix Table B3.

3.4. Evaluation of included studies

Of the 46 included studies, 12 studies assessed cohorts recruited from patient groups as opposed to the general population (Tables 2 and 3). Forty-one studies reported risk estimates for NO_2 and 20 for O_3, 15 studies reported estimates for both pollutants. About half of the studies were from the USA (n = 15) and Canada (n = 7), 19 studies from Europe (i.e. UK (n = 5), Netherlands (n = 3), Italy (n = 3), France (n = 2), Germany (n = 2), Denmark (n = 1), Norway (n = 1), and multiple European study populations (n = 2)), and with remainder from China (n = 2) and Japan (n = 3). Forty-two studies reported risk estimates for all-cause mortality and 22 for respiratory mortality. All cohorts assigned air pollution concentrations to cohort subjects retrospectively. Cohort sample size varied from 1800 to 60,000,000. A number of cohorts were analysed in more than one study, varying by length of follow-up, number of events, and methods used to estimate pollution concentrations. A small number of studies used a sub-group of subjects taken from a cohort analysed and reported elsewhere or reported a meta-analysis of a number of individual cohorts, some of which were published separately. Studies investigating O_3 used annual concentrations and/or peak season concentrations as the exposure metric. Studies included used various methods in exposure assessment, including local monitoring networks, atmospheric dispersion models, and land use regression model. Outcome (mortality) ascertainment methods were similar among studies including national death records, insurance records, and hospital records.

3.5. Risk of bias

RoB grading for each domain for studies meta-analysed are given in Tables 4–6. For the confounding domain most studies were graded moderate or high RoB; for the selection bias domain most were graded low with only a small number assessed as high/moderate. For all other domains, RoB was graded as low for all studies. Details of RoB assessment of individual pollutant-outcome pairs are provided in Appendix C Supplementary file.

3.6. Conflict of interest

The majority of studies either did not publish a conflict of interest statement or declared no conflict. Only a very small number of authors declared grant income or additional income and none constituted a conflict that warranted sensitivity analyses.

3.7. Meta-analyses

3.7.1. Nitrogen dioxide

3.7.1.1. All-cause mortality. Thirty-six studies reported results for NO_2 and all-cause mortality (Table 2a). One study reported results for two separate cohorts. (Health Effects Institute, 2000) Thirteen results were excluded from meta-analysis as results from more recent publications were available, available for full cohorts, rather than samples, or included in the ESCAPE study (see Table 2a for more details). Individual study estimates, weights, RE (95% CI) summary estimate, model statistics and 80% prediction interval are shown in Fig. 2. A 10 μg/m^3 increase in NO_2 was associated with a RR of 1.02 (95% CI: 1.01, 1.04) for mortality from all-causes. Heterogeneity indicated by I^2 was very high (96.9%). No evidence of small study bias/funnel plot asymmetry was found (Egger's test, P = 0.61, see Fig. B1). The E-value was 1.38.

A slightly larger, more precisely estimated summary RR was observed in general population versus patient cohorts; 1.02 (1.01, 1.04) and 1.01 (0.98, 1.04) per 10 μg/m^3 respectively (Appendix Fig. B2). Meta-analysis stratified by cohorts that controlled for individual measures of BMI, smoking and SES versus those that did not, reported RR of 1.03 (1.00, 1.05) and 1.03 (1.03, 1.04) respectively (Appendix Fig. B3). Stratification by WHO region is shown in Appendix Fig. B4. Meta-regression including study mean NO_2 concentration indicated a negative relationship, (-0.00042 (standard error 0.00028) change in ln(RR) per unit increase in study mean NO_2 concentration. Stratification by RoB for the confounding domain (high versus moderate/low) is shown in Appendix Fig. B5. Exclusion of the five studies (Table 4a) assessed as

Table 3
Summary of characteristics of studies included in the systematic review – Ozone.

a) All-cause

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure			
								BMI	Smoking	SES	Period	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Abbey 1999 (Abbey et al., 1999)	AHSMOG	General	USA	6338	1628	FM	58.5	Yes	Yes	Indiv	Annual	Monitoring station	51.2	NR
Bentayeb 2015 (Bentayeb et al., 2015)	Gazel cohort	General	France	20,327	1967	FM	43.7	Yes	Yes	Indiv	Peak	Chemistry-transport model	96	NR
Cakmak 2018 (Cakmak et al., 2018)	CANCHEC	General	Canada	2,291,250	522,305	FM	25–90	Indirect	Indirect	Indiv	Peak	Interpolation	76.8	Min (0)
Cakmak 2016 (Cakmak et al., 2016)[i]	CANCHEC	General	Canada	2,415,505	NR	FM	> =25	No	No	Indiv	Peak	Interpolation	60	Min (48.2)
Carey 2013 (Carey et al., 2013)	CPRD	General	UK	824,654	83,103	FM	40–89	Yes	Yes	Area	Annual	Air dispersion	51.7	Min (44.5)
Crouse 2015a (Crouse et al., 2015a)[i]	CANCHEC	General	Canada	2,521,525	301,115	FM	25–90	Indirect	Indirect	Indiv	Peak	Interpolation	77.6	Min (21)
Desikan 2016 (Desikan et al., 2015)	SLSR	Patient	UK	1800	729	FM	68.8 (15.8)	No	No	Area	Annual	KCLurban	36.7	25th (34.4)
Di 2018 (Di et al., 2017)	MCBS	Patient	USA	60,925,443	22,567,924	FM	70.1	Indirect	Indirect	Area	Peak	Monitoring stations Prediction model	90.7	5th (71.1)
Jerrett 2009 (Jerrett et al., 2009)[i]	ACS CPS II	General	USA	448,850	118,777	FM	56.6	Yes	Yes	Indiv	Annual	Monitoring stations	133.3	NR
Jerrett 2013 (Jerrett et al., 2013)[i]	ACS CPS II	General	USA	73,711	19,755	FM	57.4(10.6)	Yes	Yes	Indiv	Annual	Monitoring station, inverse distance weighting	98.7	5th (56.5)
HEI 2000 (Health Effects Institute, 2000)	Six Cities	General	USA	8111	1430	FM	49.7	Yes	Yes	Indiv	Annual	Monitoring stations	42.3	NR
HEI 2000 (Health Effects Institute, 2000)[i]	ACS CPS II	General	USA	552,138	38,963	FM	58.5	Yes	Yes	Indiv	Annual	Monitoring stations	54.4	NR
Krewski 2009 (Krewski et al., 2009)[i]	ACS CPS II	General	USA	531,826	128,954	FM	58.5	Yes	Yes	Indiv	Peak	Monitoring stations	44.9	5th (29.5)
Lipfert 2006 (Lipfert et al., 2006)	WU-EPRI	Patient	USA	28,635	5638	M	51 (12)	Yes	Yes	Area	Peak	Interpolation	NR	NR
Lipfert 2006 (Lipfert et al., 2006)[ii]	WU-EPRI	Patient	USA	NR	NR	M	51(12)	Yes	Yes	Area	Annual & Peak	Interpolation	101.5	Min (80.6)
Lipsett 2011 (Lipsett et al., 2011)	CTS	General	USA	124,614	7381	F	> =30	Yes	Yes	Area	Annual & Peak	Interpolation Monitoring stations	94.3	Min (49.8)
Rush 2017 (Rush et al., 2017)	NIS	Patient	USA	93,950	30,155	FM	> =18	Yes	No	Area	Annual	Monitoring stations	NR	NR
Smith 2009 (Smith et al., 2009)[i]	ACS CPS II	General	USA	352,242	NR	FM	NR	Yes	Yes	Indiv	Peak	Monitoring stations	NR	NR
Tonne 2016 (Tonne et al., 2016)	MINAP	Patient	UK	18,138	5129	FM	68 (14)	No	Yes	Area	Annual	KCLurban	40.3	25th (37.8)
Turner 2016 (Turner et al., 2016)	ACS CPS II	General	USA	669,046	237,201	FM	> =30	Yes	Yes	Indiv	Annual & Peak	Hierarchical Bayesian space–time model	74.9	5th (61)
Weichenthal 2017 (Weichenthal et al., 2017)	CANCHEC	General	Canada	2,448,500	233,340	FM	25–89	No	No	Indiv	Peak	Interpolation	75	5th (54.1)

(*continued on next page*)

Table 3 (*continued*)

b) Respiratory

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure			
								BMI	Smoking	SES	Period	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Abbey 1999 (Abbey et al., 1999)	AHSMOG	General	USA	6338	135	FM	58.5	Yes	Yes	Indiv	Annual	Monitoring stations	51.2	NR
Carey 2013 (Carey et al., 2013)	CPRD	General	UK	824,654	10,583	FM	40–89	Yes	Yes	Area	Annual	Air dispersion	51.7	Min (44.5)
Crouse 2015a (Crouse et al., 2015a)	CANCHEC	General	Canada	2,521,525	24,900	FM	25–89	Indirect	Indirect	Indiv	Peak	Interpolation	77.6	Min (21)
Jerrett 2009 (Jerrett et al., 2009)[i]	ACS CPS II	General	USA	448,850	9891	FM	56.6	Yes	Yes	Indiv	Annual	Monitoring stations	133.3	NR
Jerrett 2013 (Jerrett et al., 2013)[i]	ACS CPS II	General	USA	73,711	1990	FM	57.4(10.6)	Yes	Yes	Indiv	Annual	Monitoring station, inverse distance weighting	98.7	5th (56.5)
Lipsett 2011 (Lipsett et al., 2011)	CTS	General	USA	101,784	702	F	> =30	Yes	Yes	Area	Annual & Peak	Interpolation Monitoring stations	94.3	Min (49.8)
Smith 2009 (Smith et al., 2009)[i]	ACS CPS II	General	USA	352,242	NR	FM	NR	Yes	Yes	Indiv	Peak	Monitoring stations	NR	NR
Turner 2016 (Turner et al., 2016)	ACS CPS II	General	USA	669,046	20,484	FM	> =30	Yes	Yes	Indiv	Annual & Peak	Hierarchical Bayesian space–time model	74.9	5th (61)
Weichenthal 2017 (Weichenthal et al., 2017)	CANCHEC	General	Canada	2,448,500	21,100	FM	25–89	No	No	Indiv	Peak	Interpolation	75	5th (54.1)

c) COPD

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure			
								BMI	Smoking	SES	Period	Exposure assessmentExposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Cakmak 2018 (Cakmak et al., 2018)	CANCHEC	General	Canada	2,291,250	16,470	FM	25–90	Indirect	Indirect	Indiv	Peak	Interpolation	76.8	Min (0)
Carey 2013 (Carey et al., 2013)	CPRD	General	UK	824,654	4083	FM	40–89	Yes	Yes	Area	Annual	Air dispersion	51.7	Min (44.5)
Crouse 2015a (Crouse et al., 2015a)[i]	CANCHEC	General	Canada	2,521,525	14,170	FM	25–89	Indirect	Indirect	Indiv	Peak	Interpolation	77.6	Min (21)
Turner 2016 (Turner et al., 2016)	ACS CPS II	General	USA	669,046	9967	FM	> =30	Yes	Yes	Indiv	Annual & Peak	Hierarchical Bayesian space–time model	74.9	5th (61)

d) ALRI

Author year	Cohort	Study population	Country	Sample size	N (death)	Sex	Age	Confounding adjustment			Exposure			
								BMI	SmokingSmoking	SES	Period	Exposure assessment	Mean (μg/m^3)	Lowest concentration recorded
Carey 2013 (Carey et al., 2013)	CPRD	General	UK	824,654	4042	FM	40–89	Yes	Yes	Area	Annual	Air dispersion	51.7	Min (44.5)
Turner 2016 (Turner et al., 2016)	ACS CPS II	General	USA	669,046	6599	FM	> =30	Yes	Yes	Indiv	Annual & Peak	Hierarchical Bayesian space–time model	74.9	5th (61)

Abbreviation: BMI – body mass index; SES – socio-economic status; NR – not reported.

Confounding adjustment: for BMI and smoking, if it was adjusted (yes), direct or indirect adjustment were recorded; for SES, all studies were adjusted for, therefore adjustment at area or individual level were recorded

[i] Excluded from analysis due to more recent cohort follow-up was available.

[ii] Results for peak season analysis not included, longer follow-up study available.

Table 4
RoB assessment for studies included in meta-analysis – NO_2. RoB Domains: CO – confounding; SB – selection bias; EA – exposure assessment; OM = outcome measurement; MD – missing data; SR – selective reporting.

a) All-cause

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Fischer	2015	DUELS	mod	low	low	low	low	low
Chen	2016	Four northern Chinese cities	mod	low	low	low	low	low
Bentayeb	2015	Gazel cohort	mod	mod	low	low	low	low
Desikan	2016	South London Stroke Register	high	high	low	low	low	low
Beelen	2014	ESCAPE	mod	low	low	low	low	low
Tonne	2013	MINAP (ACS survivors)	mod	low	low	low	low	low
Cesaroni	2013	Rome longitudinal study	high	low	low	low	low	low
Carey	2013	CPRD	mod	low	low	low	low	low
Hart	2013	Nurses Health Study	low	low	low	low	low	low
Lipsett	2011	CTS	low	low	low	low	low	low
Hart	2011	US trucking industry cohort	mod	low	low	low	low	low
Brunekreef	2009	NLCS-AIR	high	low	low	low	low	low
Jerrett	2009	Toronto respiratory cohort	mod	low	low	low	low	low
Rosenlund	2008	CHD survivors cohort	high	low	low	low	low	low
Lipfert	2006	Washington University-EPRI Veterans	mod	low	low	low	low	low
Abbey	1999	AHSMOG	mod	low	low	low	low	low
Weichenthal	2017	CanCHEC	high	low	low	low	low	low
Hartiala	2016	The Cleveland Clinic GeneBank study	mod	low	low	low	low	low
Turner	2016	ACS CPS-II	mod	low	low	low	low	low
Yorifuji	2013	Shizuoka elderly cohort	mod	mod	low	low	low	low
Filleul	2005	PAARC	mod	low	low	low	low	low
HEI	2000	Six Cities	mod	low	low	low	low	low
Yang	2018	Hong Kong elderly	mod	low	low	low	low	low
Crouse	2015	CanCHEC	mod	low	low	low	low	low

b) Respiratory

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Fischer	2015	DUELS	mod	low	low	low	low	low
Dimakopoulou	2014	ESCAPE	mod	low	low	low	low	low
Cesaroni	2013	Rome longitudinal study	mod	low	low	low	low	low
Carey	2013	CPRD	mod	low	low	low	low	low
Katanoda	2011	3 Japanease Prefectures	mod	low	low	low	low	low
Lipsett	2011	CTS	low	low	low	low	low	low
Hart	2011	US trucking industry cohort	high	low	low	low	low	low
Brunekreef	2009	NLCS-AIR	mod	low	low	low	low	low
Jerrett	2009	Toronto respiratory cohort	mod	low	low	low	low	low
Abbey	1999	AHSMOG	mod	low	low	low	low	low
Weichenthal	2017	CanCHEC	high	low	low	low	low	low
Turner	2016	ACS CPS-II	mod	low	low	low	low	low
Yorifuji	2013	Shizuoka elderly cohort	mod	mod	low	low	low	low
Yang	2018	Hong Kong elderly	mod	low	low	low	low	low
Crouse	2015	CanCHEC	mod	low	low	low	low	low

c) COPD

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Carey	2013	CPRD	mod	low	low	low	low	low
Katanoda	2011	3 Japanese Prefectures	mod	low	low	low	low	low
Hart	2011	US trucking industry cohort	high	low	low	low	low	low
Naess	2007	Oslo Cohort	mod	low	low	low	low	low
Turner	2016	ACS CPS-II	mod	low	low	low	low	low
Crouse	2015	CanCHEC	mod	low	low	low	low	low
Gan	2013	Vancover	high	low	low	low	low	low
Yorifuji	2013	Shizuoka elderly cohort	mod	mod	low	low	low	low
Yang	2018	Hong Kong elderly	mod	low	low	low	low	low

d) ALRI

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Carey	2013	CPRD	mod	low	low	low	low	low
Katanoda	2011	3 Japanese Prefectures	mod	low	low	low	low	low
Turner	2016	ACS CPS-II	mod	low	low	low	low	low
Yorifuji	2013	Shizuoka elderly cohort	mod	mod	low	low	low	low
Yang	2018	Hong Kong elderly	mod	low	low	low	low	low

Table 5
RoB assessment for studies included in meta-analysis – O_3 annual average concentrations. RoB Domains: CO – confounding; SB – selection bias; EA – exposure assessment; OM = outcome measurement; MD – missing data; SR – selective reporting.

a) All-cause

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Rush	2017	NIS	high	low	mod	low	low	low
Tonne	2016	MINAP	high	low	low	low	low	low
Desikan	2016	SLSR	high	high	low	low	low	low
Carey	2013	CPRD	mod	low	low	low	low	low
Lipsett	2011	CTS	low	low	low	low	low	low
Abbey	1999	AHSMOG	mod	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low
Lipfert	2006	WU-EPRI	mod	low	low	low	low	low
Krewski	2000	Six Cities	mod	low	low	low	low	low

b) Respiratory

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Carey	2013	CPRD	mod	low	low	low	low	low
Lipsett	2011	CTS	low	low	low	low	low	low
Abbey	1999	AHSMOG	mod	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low

c) COPD

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Carey	2013	CPRD	mod	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low

d) ALRI

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Carey	2013	CPRD	mod	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low

high RoB for the confounding domain gave a summary RR for the remaining 19 studies of 1.03 (1.01, 1.04) per 10 μg/m^3. Of the 24 studies included in the meta-analysis, 20 included both male and female participants; two included males only, and another two with females only (Table 2a). Hence sub-group analysis by sex was not undertaken. Age of subjects at cohort entry varied substantially between studies but all studies adjusted for age in the analyses.

3.7.1.2. Respiratory mortality. Nineteen studies reported results for NO_2 and respiratory mortality. Four results were excluded from meta-analysis as results from more recent publications were available, or included in the ESCAPE study (Table 2b). Individual study estimates, weights, RE (95% CI) summary estimate, model statistics and 80% prediction interval are shown in Fig. 3. A 10 μg/m^3 increase in NO_2 was associated with a RR of 1.03 (1.01, 1.05) for mortality from respiratory disease. Heterogeneity indicated by I^2 was high (82.9%). No evidence of small study bias/funnel plot asymmetry was found (Egger's test, P = 0.22, Appendix Fig. B6). The E-value was 1.5.

One of 15 studies reported results from a patient group (Table 2b). Stratification by confounding adjustment (Appendix Fig. B7) suggested a difference between studies that controlled for individual measures of key confounders (1.02 (0.99, 1.05)) compared to those that did not (1.04 (1.02, 1.07)). Appendix Fig. B8 presents the results stratified by WHO region and clearly illustrates differences between WHO regions – summary RR for Eur and AMR were 1.04 (1.00, 1.07) and 1.02 (1.00, 1.05) compared to 1.07 (0.98, 1.17) per 10 μg/m^3 for WPR region. Meta-regression including study mean NO_2 concentration indicated a negative relationship −0.00046 (standard error 0.00020) change in ln (RR) per unit increase in study mean NO_2 concentration. Exclusion of the two studies assessed as high RoB for the confounding domain

Table 6
RoB assessment for studies included in meta-analysis – O_3 peak concentrations. RoB Domains: CO – confounding; SB – selection bias; EA – exposure assessment; OM = outcome measurement; MD – missing data; SR – selective reporting.

a) All-cause

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Cakmak	2018	CANCHEC	mod	low	low	low	low	low
Di	2018	MCBS	mod	low	low	low	low	low
Bentayeb	2015	Gazel cohort	mod	mod	low	low	low	low
Lipsett	2011	CTS	low	low	low	low	low	low
Lipfert	2006	WU-EPRI	mod	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low
Weichenthal	2017	CANCHEC	high	low	low	low	low	low

b) Respiratory

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Lipsett	2011	CTS	low	low	low	low	low	low
Weichenthal	2017	CANCHEC	high	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low
Crouse	2015	CANCHEC	mod	low	low	low	low	low

c) COPD

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Cakmak	2018	CANCHEC	mod	low	low	low	low	low
Turner	2016	ACS CPS II	mod	low	low	low	low	low

d) ALRI

Author	Year	Cohort	CO	SB	EA	OM	MD	SR
Turner	2016	ACS CPS II	mod	low	low	low	low	low

(Table 4b) gave a summary RR for the remaining 13 studies of 1.03 (1.01, 1.05) per 10 μg/m^3 (results not shown).

3.7.1.3. COPD. Ten studies reported results for NO_2 and COPD mortality with a single study excluded from the meta-analysis as a more recent publication was available (Fig. 4, Table 2c). A 10 μg/m^3 increase in NO_2 was associated with a RR of 1.03 (1.01, 1.04) for COPD mortality. Heterogeneity indicated by I^2 was low (22.7%) and the E-value was 1.5. Because of the small number of studies, no sub-group analyses were undertaken. Exclusion of the two studies assessed as high RoB for the confounding domain (Table 4c) gave a summary RR for the remaining seven studies of 1.03 (1.01, 1.05) per 10 μg/m^3 (results not shown).

3.7.1.4. Acute lower respiratory infection. Six studies reported results for NO_2 and ALRI mortality with a single study excluded from the meta-analysis as a more recent publication was available (Fig. 5, Table 2d). A 10 μg/m^3 increase in NO_2 was associated with a RR of 1.06 (1.02, 1.10) for ALRI mortality. Heterogeneity indicated by I^2 was 81.3%. Because of the small number of studies, no sub-group analyses were undertaken. RoB was low/moderate for all domains (Table 4d). The E-value was 1.8.

3.7.1.5. Minimum concentrations recorded. For NO_2 and all-cause mortality, 18 out of 24 studies included in the meta-analysis reported details of the range of NO_2 concentrations in the studies (Table 2). Metrics reported included minimum (n = 9) (Crouse et al., 2015a; Lipsett et al., 2011; Hartiala et al., 2016; Carey et al., 2013; Filleul et al., 2005; Rosenlund et al., 2008; Cesaroni et al., 2013; Chen et al., 2016; Yorifuji et al., 2013); 5 th percentile (n = 8) (Weichenthal et al., 2017; Turner et al., 2016; Hart et al., 2011; Hart et al., 2013; Lipfert et al., 2006; Fischer et al., 2015; Brunekreef et al., 2009; Yang et al., 2018)and 25th percentile (n = 1) (Desikan et al., 2015) values of the distribution of NO_2 concentrations; values ranged from 4.5 μg/m^3

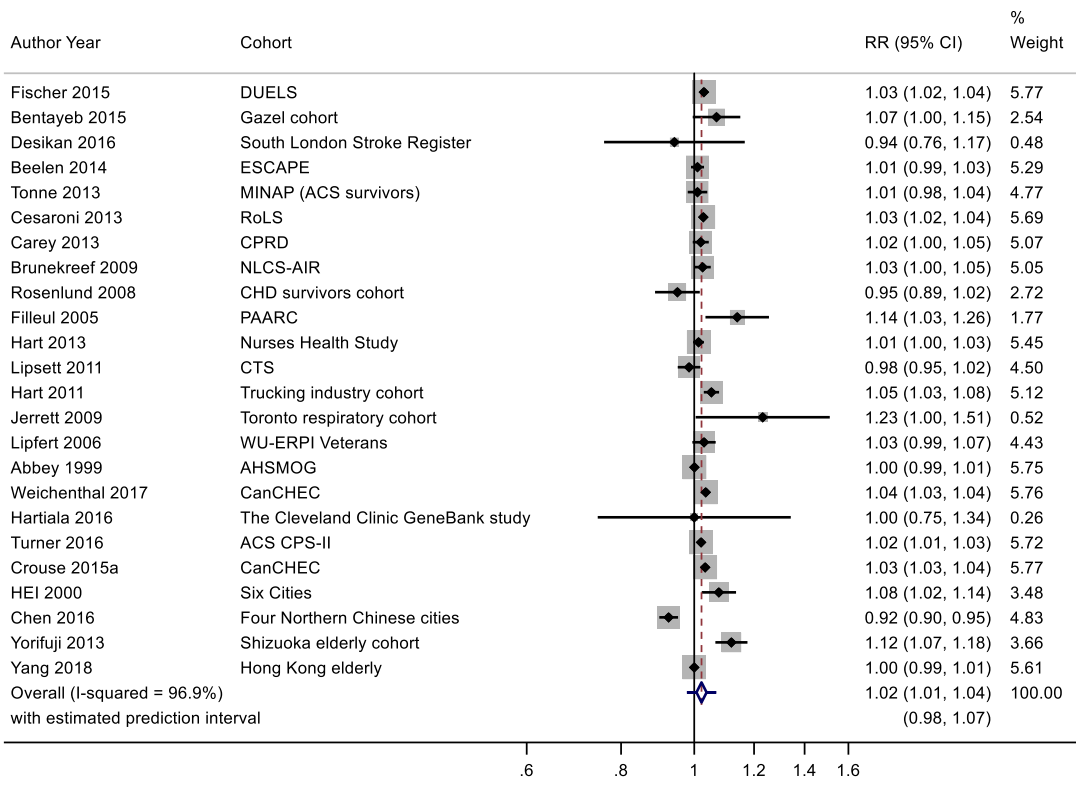

Fig. 2. NO_2 and all-cause mortality.

(Carey et al., 2013) to 81.3 μg/m^3 (Chen et al., 2016). For respiratory mortality 11 out of 15 studies included in the meta-analysis reported details of low NO_2 concentrations in the studies (five minimum (Crouse et al., 2015a; Lipsett et al., 2011; Carey et al., 2013; Cesaroni et al., 2013; Yorifuji et al., 2013) and six 5th percentile (Weichenthal et al., 2017; Turner et al., 2016; Hart et al., 2011; Fischer et al., 2015; Brunekreef et al., 2009; Yang et al., 2018) values, ranging from 4.5 μg/m^3 to 81.3 μg/m^3. Eight of the nine studies of COPD reporting low concentrations, five (Crouse et al., 2015a; Carey et al., 2013; Yorifuji et al., 2013; Gan et al., 2013; Naess et al., 2007) were for minimum

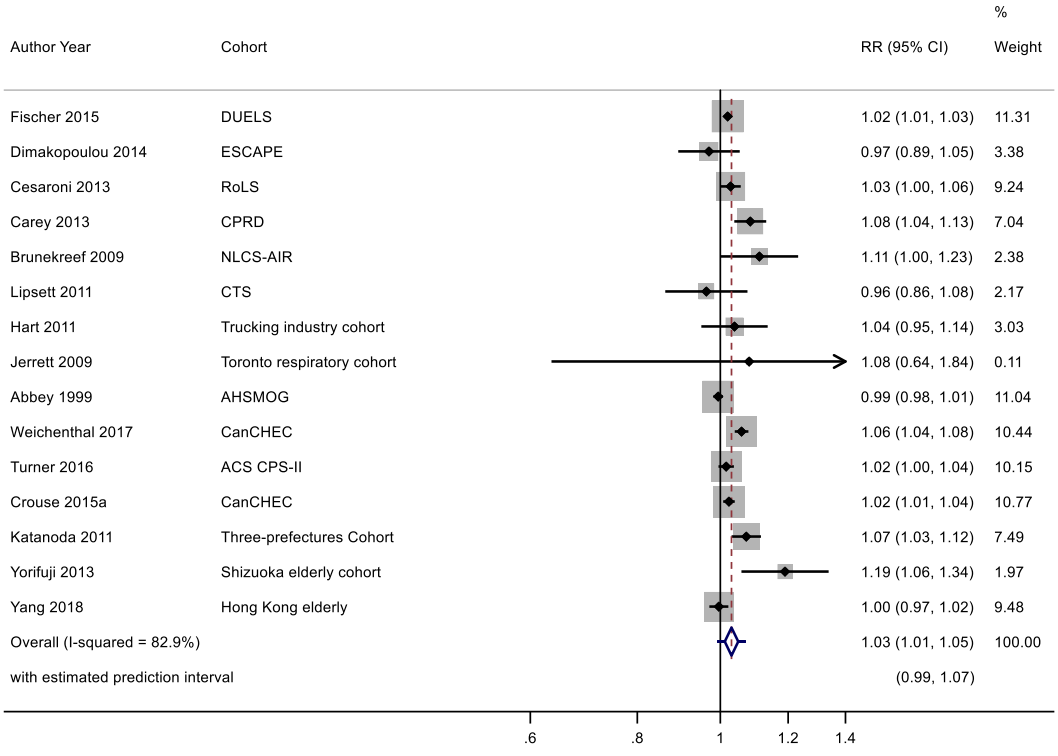

Fig. 3. NO_2 and respiratory mortality.

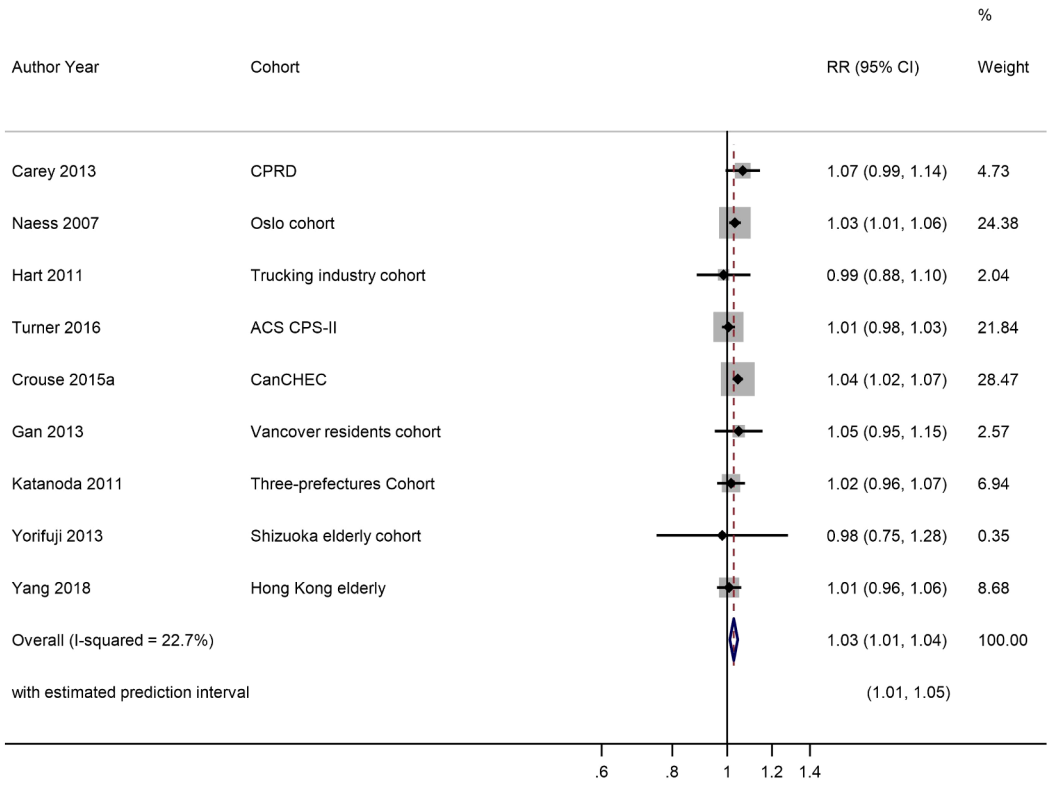

Fig. 4. NO_2 and COPD mortality.

concentrations and three (Turner et al., 2016; Hart et al., 2011; Yang et al., 2018; Yorifuji et al., 2010) for 5th percentile values. The lowest reported concentration was 0 μg/m^3 (Crouse et al., 2015a). Two (Carey et al., 2013; Yorifuji et al., 2013) of the four studies of ALRI mortality reported minimum NO_2 concentrations and two (Turner et al., 2016; Yang et al., 2018; Yorifuji et al., 2010) reported concentrations for the 5th percentile with values ranging from 4.5 μg/m^3 to 81.3 μg/m^3.

3.7.1.6. Shape of the concentration response function. Naess (Naess et al., 2007) assessed the relationship between NO_2 concentrations and all-cause and COPD mortality stratified by age groups (51–70 and 71–90 years). The authors reported that in younger subjects the risk of death from all-causes started to increase from 40 μg/m^3 whereas in the oldest age group the relationship was linear across the concentration range (2–73 μg/m^3). Rosenlund (Rosenlund et al.,

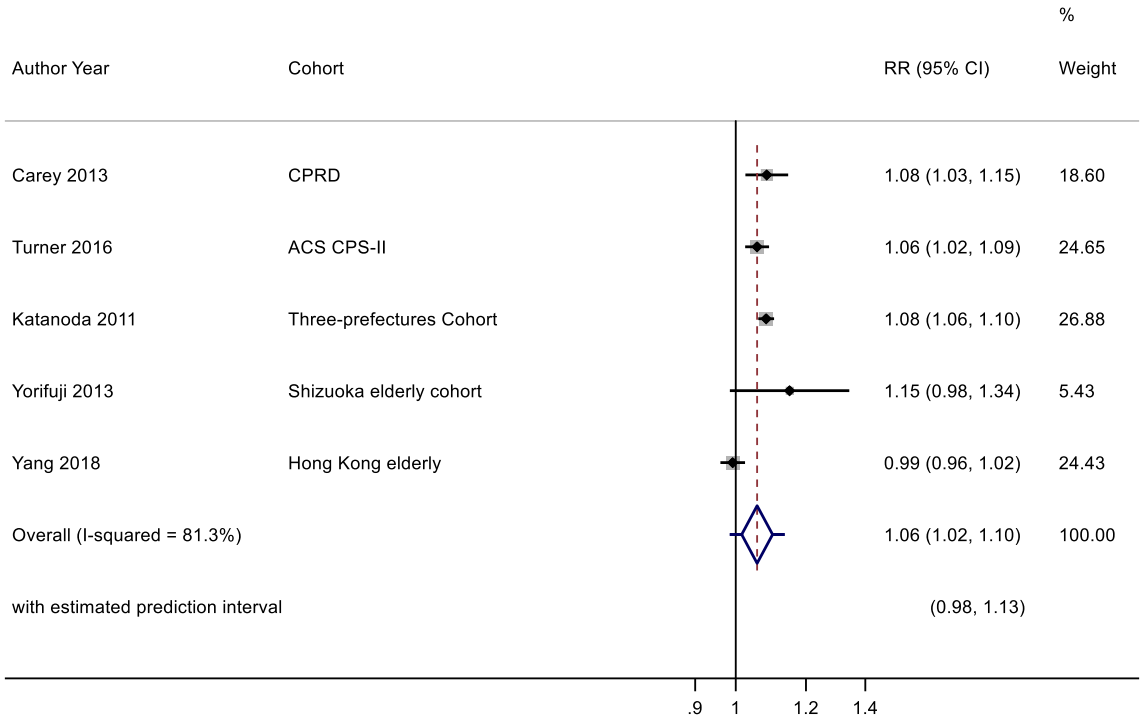

Fig. 5. NO_2 and ALRI mortality. Cochran's Q: Chi-square = 22.0, df = 4, P = 0.000. tau2 = 0.0014.

2008) investigated mortality within 28 days of first coronary events. Risk estimates stratified by quintile of NO_2 concentration indicated that there was no evidence of nonlinearity, although the risk in the 2nd quintile was close to 1 and the risk in the top quintile was lower than in the 3rd and 4th quintiles. (Raaschou-Nielsen et al., 2012) investigated the exposure–response relationship between log_2NO_2 and all-cause mortality using spline functions. They found no evidence to reject a linear relationship across the concentration range (5th-95th percentile values: 11.6–29.5 μg/m^3). Analysis of a 20% sample from the Rome longitudinal cohort by Cesaroni (Cesaroni et al., 2013) using natural splines showed no evidence of deviation from linearity for all-cause mortality and NO_2 (minimum concentration approximately 20 μg/m^3). Fischer (Fischer et al., 2015) assessed the shape of the concentration–response relationship for all-cause and respiratory mortality using natural splines and tested deviation from linearity using the likelihood ratio test. They found no evidence of deviation for linearity for either causes of death for NO_2 concentrations to approximately 10 μg/m^3 (5th percentile 19 μg/m^3). Naess et al. (2007), Gan et al. (2013), Gan et al. (2013) evaluated the concentration response relationship using natural cubic spline models and reported 'no discernible exposure–response trends' for NO_2 and COPD mortality. None of the studies of ALRI mortality assessed the shape of the concentration–response function.

3.7.1.7. Co-pollutant adjustment. Studies reporting results for NO_2 and all-cause, respiratory and COPD/ALRI from multipollutant models are shown respectively in Appendix Figs. B9-B11. A range of co-pollutants were investigated including Black Carbon (Yang et al., 2018; Gan et al., 2013), particles with a median diameter of < 2.5 μm ($PM_{2.5}$) (Crouse et al., 2015a; Cesaroni et al., 2013; Turner et al., 2016; Yang et al., 2018; Jerrett et al., 2013; Beelen et al., 2014), sulphur dioxide (SO_2) (Carey et al., 2013; Chen et al., 2016; Hart et al., 2011) and O_3 (Crouse et al., 2015a; Carey et al., 2013; Turner et al., 2016; Jerrett et al., 2013). In some studies associations between NO_2 and mortality was attenuated upon adjustment for co-pollutants (Carey et al., 2013; Turner et al., 2016; Fischer et al., 2015; Beelen et al., 2014) but not in others.

3.7.1.8. Certainty of evidence assessment. Tables 7–10 present the certainty of evidence assessments for all-cause, respiratory, COPD and ALRI mortality respectively. For NO_2 and mortality we assessed the certainty of evidence from single pollutant models to be moderate for all-causes (mean RR = 1.02 per 10 μ/m^3), moderate for respiratory (mean RR 1.03 per 10 μ/m^3); high for COPD (mean RR = 1.03 per 10 μ/m^3); and moderate for ALRI (mean RR = 1.06 per 10 μ/m^3).

3.7.2. Ozone

3.7.2.1. All year concentrations

3.7.2.1.1. All- cause mortality. Twelve studies reported results for all-year O_3 exposure and all-cause mortality (Table 3a). We selected the most recent study results for meta-analyses, therefore three studies (Health Effects Institute, 2000; Jerrett et al., 2013; Jerrett et al., 2009) were excluded and nine studies (Health Effects Institute, 2000; Lipsett et al., 2011; Carey et al., 2013; Turner et al., 2016; Desikan ET AL., 2015; Abbey et al., 1999; Lipfert et al., 2006; Tonne et al., 2016; Rush et al., 2017)were included for main analysis (note: one study (Health Effects Institute, 2000) included two cohorts, results from one cohort was included, the other was excluded) (Fig. 6). Pooled results showed no significant association between increased O_3 exposure and all-cause mortality, 0.97 (0.93, 1.02) per 10 μg/m^3 with large heterogeneity (I^2 = 98.7%). Publication bias was not assessed due to small number of included studies. Exclusion of the three studies (Table 5) with high RoB did not materially alter the summary risk (results not shown).

3.7.2.1.2. Respiratory mortality. Six studies (Lipsett et al., 2011; Carey et al., 2013; Turner et al., 2016; Jerrett et al., 2013; Jerrett et al., 2009; Abbey et al., 1999) reported the all-year O_3 exposure and respiratory mortality, while four studies (Lipsett et al., 2011; Carey et al., 2013; Turner et al., 2016; Abbey et al., 1999) with most recent study results were included in the pooled analysis (Table 3b, Fig. 7). No significant association was found between increased O_3 exposure and respiratory mortality, 0.99 (0.89, 1.11) per 10 μg/m^3.

3.7.2.1.3. COPD. Only two studies reported the association between annual O_3 exposure and COPD mortality (Table 3c). Turner 2016 (Turner et al., 2016) showed that increased O_3 exposure was associated with higher risk of COPD mortality, 1.07 (1.04, 1.10) per 10 μg/m^3, while Carey 2013 (Carey et al., 2013) found no significant association between O_3 exposure and risk of COPD mortality.

3.7.2.1.4. Acute lower respiratory Infection. Two studies reported O_3 exposure and risk of ALRI mortality (Table 3d). Turner 2016 (Turner et al., 2016) showed that increased O_3 exposure was associated with a higher risk of mortality, 1.07 (1.04, 1.11) per 10 μg/m^3 while Carey 2013 (Carey et al., 2013) found the association was in the opposite

Table 7
GRADE assessment – NO_2 and all-cause mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	24 included studies. Risk of bias moderate because although not all studies adjusted for all confounders, exclusion of high risk of bias studies did not reduce the summary RR (Appendix Fig. B5).	No downgrading
Indirectness	All studies included the desired population, exposures and outcomes	No downgrading
Inconsistency	The 80% prediction interval included 1 & > twice CI (Fig. 2). High level of heterogeneity in general population studies. Studies controlling for individual measures of BMI, smoking, SES (Appendix Fig. B3) gave slightly higher, less precise summary RR. Exclusion of patient cohorts (6) did not change summary RR & CI (Appendix Fig. B2).	Downgrade one level
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	According to the funnel plot and Egger's test (P < 0.1), there were no sign of publication bias/funnel plot asymmetry.	No downgrading
Large Effect Size	Summary RR = 1.02. Precision reduced for cohorts with all individual confounder adjustment but not summary estimate. Insufficient information on unmeasured potential confounders available.	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected.	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies. 5 studies investigated the shape of the dose response relationship with no evidence to suggest non-linear. 95% CI for linear RR excluded 1.	Upgrade one level
GRADE conclusion	Downgrade one level and upgrade one level	MODERATE CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 1.02 PER 10μ/m^3

Table 8
GRADE assessment – NO_2 and respiratory mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	15 included studies. Risk of bias moderate because although not all studies adjusted for all confounders, exclusion of high risk of bias studies did not alter summary RR.	No downgrading
Indirectness	All studies included the desired population, exposures and outcomes	No downgrading
Inconsistency	The 80% prediction interval included 1; PI = 2 × CI (Fig. 3). Studies controlling for individual measures of BMI, smoking, SES gave lower summary RR and CI included 1 (Appendix Fig. B7). Exclusion of single patient cohort did not change summary RR & CI. High level of heterogeneity in general population studies	Downgrade one level
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	According to the funnel plot little evidence of publication bias	No downgrading
Large Effect Size	Summary RR = 1.03Insufficient information on unmeasured potential confounders available	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies, 95% CI for linear RR excluded 1. No evidence to confirm shape of the dose response relationship.	Upgrade one level
GRADE conclusion	No downgrade and no upgrade	MODERATE CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 1.03 PER 10μ/m^3

direction, 0.84 (0.73, 0.97) per 10 μg/m^3.

3.7.2.2. Peak exposures

3.7.2.2.1. All- cause mortality. Twelve studies (Crouse et al., 2015a; Lipsett et al., 2011; Weichenthal et al., 2017; Turner et al., 2016; Lipfert et al., 2006; Lipfert et al., 2006; Bentayeb et al., 2015; Cakmak et al., 2018; Di et al., 2017; Krewski et al., 2009; Smith et al., 2009; Cakmak et al., 2016) reported the association between warm season O_3 exposure and all-cause mortality. Seven studies (Lipsett et al., 2011; Weichenthal et al., 2017; Turner et al., 2016; Lipfert et al., 2006; Bentayeb et al., 2015; Cakmak et al., 2018; Di et al., 2017) with most recent cohort results were included for pooled analysis (Table 3a). Meta-analysis result showed that a 10 μg/m^3 increase in O_3 exposure was associated with a RR of 1.01 (1.00, 1.02) per 10 μg/m^3 for all-cause mortality although the heterogeneity was high among studies (I^2 = 98%) (Fig. 8). Exclusion of the single study judged high RoB for confounding domain (Table 6) did not change the summary RR and CI (results not shown). The E-value was 1.25.

3.7.2.2.2. Respiratory mortality. Five studies (Crouse et al., 2015a; Lipsett et al., 2011; Weichenthal et al., 2017; Turner et al., 2016; Smith et al., 2009) reported the association between warm season O_3 exposure and respiratory mortality, and four studies (Crouse et al., 2015a; Lipsett et al., 2011; Weichenthal et al., 2017; Turner et al., 2016) with most recent results were included in the pooled analysis (Table 3b). Meta-analysis showed that increased O_3 exposure was associated with an increased risk of respiratory mortality, 1.02 (0.99, 1.05) per 10 μg/m^3 (Fig. 9). Exclusion of the single study judged high RoB for confounding domain (Table 6) did not change the summary RR and CI (results not shown). The E-value was 1.38.

3.7.2.2.3. COPD. Only two studies reported the warm season O_3 exposure with COPD mortality (Table 3c). Turner 2016 (Turner et al., 2016) showed that increased O_3 exposure was associated with higher risk of COPD mortality, while Cakmak 2018 (Cakmak et al., 2018) found no significant association between O_3 exposure and COPD mortality.

3.7.2.2.4. Acute lower respiratory infection. Only a single study reported results for O_3 peak exposure and ALRI mortality (Table 3d). Turner (Turner et al., 2016) found that a 10 ppb increase in O_3 was associated with a RR of 1.10 (1.03, 1.18).

3.7.2.3. Minimum concentrations recorded.

Minimum O_3 concentrations were recorded in 6 of 21 studies; 5th percentile in 5 of 21; 25th percentile in 1 of 21; and not recorded in 9 of 21 studies (Table 3). The lowest minimum and 5th percentile concentrations values in annual exposure studies were 44 μg/m^3 and 57 μg/m^3 respectively. In 'peak' season studies, the corresponding values were 21 μg/m^3 and 30 μg/m^3.

3.7.2.4. Shape of the concentration response function.

A small number of studies examined the shape of the concentration response relationships for O_3 and mortality. Turner 2016 (Turner et al., 2016) analysing the American Cancer Society (ACS) cohort reported evidence that a threshold model (35 ppb) offered an improved fit over the linear

Table 9
GRADE assessment – NO_2 and COPD mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	9 included studies. Risk of bias moderate because although not all studies adjusted for all confounders, exclusion of 2 high risk of bias studies did not alter summary RR.	No downgrading
Indirectness	All studies included the desired population, exposures and outcomes	No downgrading
Inconsistency	The 80% prediction interval did not include 1 (Fig. 4)	No downgrading
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	No analysis of publication bias – too few studies (n = 9)	No downgrading
Large Effect Size	Summary RR = 1.02 Insufficient information on unmeasured potential confounders available	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies, 95% CI for linear RR excluded 1. 2 studies investigated the shape of the dose response relationship with no evidence to suggest non-linear	Upgrade one level
GRADE conclusion	No downgrade and upgrade one level	HIGH CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 1.03 PER 10μ/m^3

Table 10
GRADE assessment – NO_2 and ALRI mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	5 included studies. Risk of bias moderate for all studies, not all studies adjusted for all confounders.	No downgrading
Indirectness	All studies included the desired population, exposures and outcomes	No downgrading
Inconsistency	The 80% prediction interval included 1 but the PI was not > 2 × CI (Fig. 5). Substantial heterogeneity amongst small number of studies.	Downgrade one level
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	No analysis of publication bias – too few studies	No downgrading
Large Effect Size	Summary RR = 1.02 Insufficient information on unmeasured potential confounders available	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected	No upgrading
Dose-response relation	No information on shape. 95% CI for linear RR excluded 1.	Upgrade one level
GRADE conclusion	No downgrade and no upgrade	MODERATE CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 1.06 PER 10μ/m^3

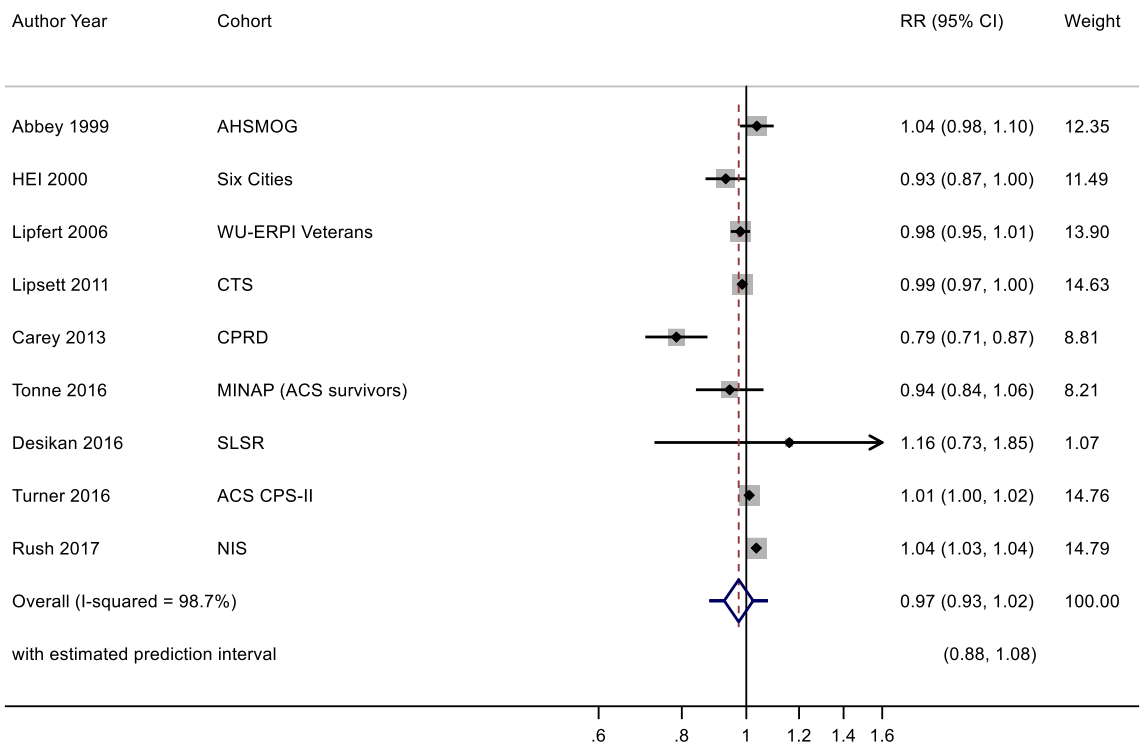

Fig. 6. O_3 annual exposure and all-cause mortality. Cochran's Q: Chi-square = 98.7, df = 8, P < 0.001. tau2 = 0.004.

model for annual O_3 concentrations and both respiratory and cardiovascular mortality (Rush et al., 2017). Using thin-plate–spline models, Di (Di et al., 2017) reported a relationship between O_3 and all-cause mortality that was almost linear, with no signal of a threshold down to 30 ppb.

3.7.2.5. Co-pollutant adjustment. A small number of studies reported results for O_3 and mortality from multipollutant models (Appendix Figs. B12 and B13). A range of co-pollutants were investigated including Black Carbon, $PM_{2.5}$, SO_2 and NO_2 in two- and three-pollutant models. Associations between O_3 and mortality were attenuated upon adjustment for co-pollutants in some studies but not in others and no discernible pattern between unadjusted and adjusted studies was observed.

3.7.2.6. Certainty of evidence assessment. Certainty of evidence assessments were completed for studies using annual and peak O_3 concentrations and all-cause and respiratory mortality (Tables 11–14). Too few studies were available for COPD and ALRI mortality for GRADE assessment. For studies reporting annual O_3 metrics we assessed the certainty of the evidence from single pollutant models to be low for all-cause mortality (mean RR 0.97 per 10μ/m^3); and low for respiratory mortality (mean RR 0.99 per 10μ/m^3). For peak O_3 exposures we assessed the certainty of evidence from single pollutant models to be moderate for all-cause mortality (mean RR 1.01 per 10μ/m^3) and low for respiratory mortality (mean RR 1.02 per 10μ/m^3).

3.8. New studies published after final search

We rapidly reviewed studies (n = 5) that published since our last search (Appendix Table B4). Two studies were conducted in the USA, NIH-AARP (Lim et al., 2019) and Medicare beneficiaries (Kazemiparkouhi et al., 2019); two studies conducted in Europe, including Danish Diet, Cancer and Health (Hvidtfeldt et al., 2019) cohort

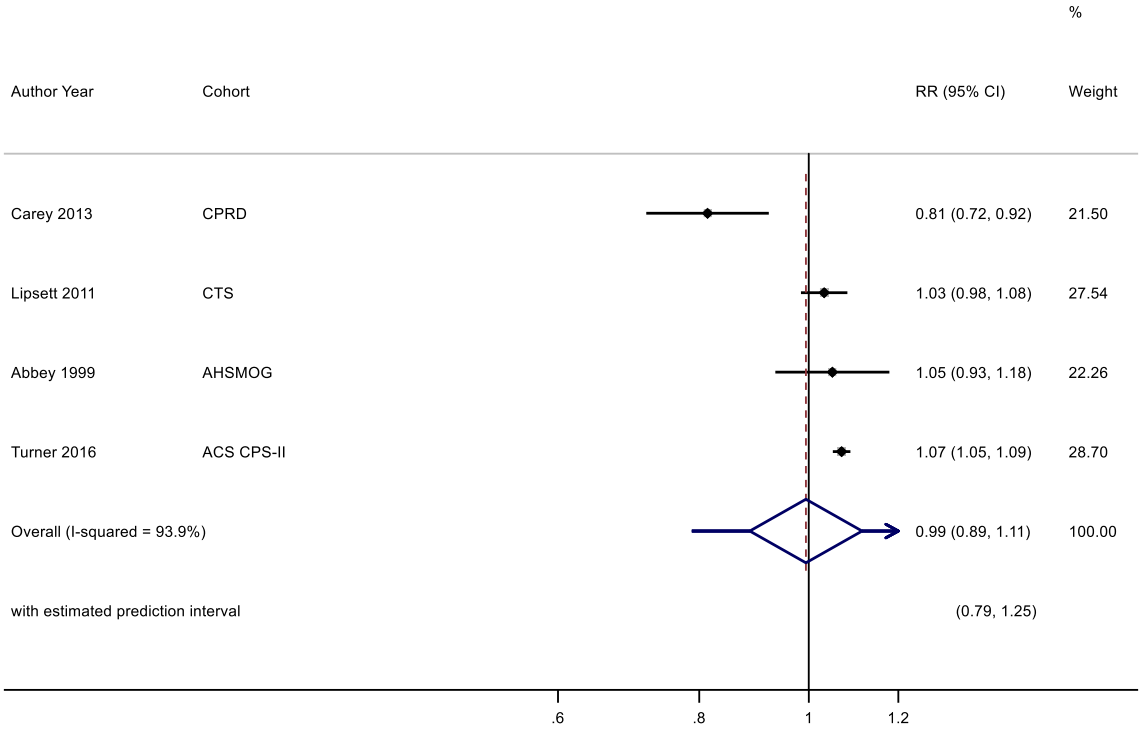

Fig. 7. O_3 annual exposure and respiratory mortality. Cochran's Q: Chi-square = 19.4, df = 3, P < 0.001. tau2 = 0.012.

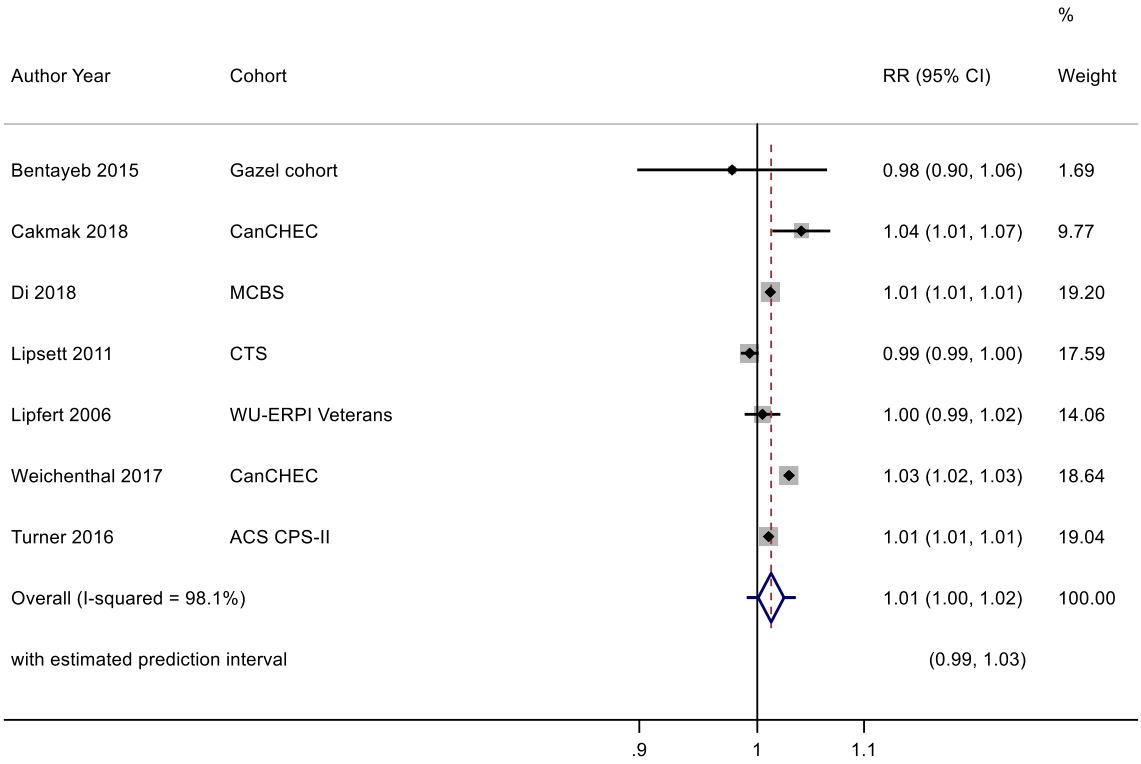

Fig. 8. O_3 peak exposure and all-cause mortality. Cochran's Q: Chi-square = 78.48, df = 6, P < 0.001. tau2 = 0.0002.

and Dutch National Health Survey cohort (Klompmaker et al., 2020) follow-up; and the "45 and up" cohort based in Australia (Hanigan et al., 2019). All studies were conducted among general population rather than patient cohort.

Among the new studies, four explored the association between NO_2 and all-cause mortality, two reported results consistent with our pooled analyses (Lim et al., 2019; Hvidtfeldt et al., 2019); an Australian study reported an association in the same direction (Hanigan et al., 2019), while another Dutch study showed no clear association which may due to relatively shorter period of follow up (Klompmaker et al., 2020). Three studies investigated the impact of NO_2 exposure on respiratory mortality: two showed a consistent direction of association (Lim et al., 2019; Hvidtfeldt et al., 2019) while no clear association was found in the Dutch cohort (Klompmaker et al., 2020). Lim 2019 also found

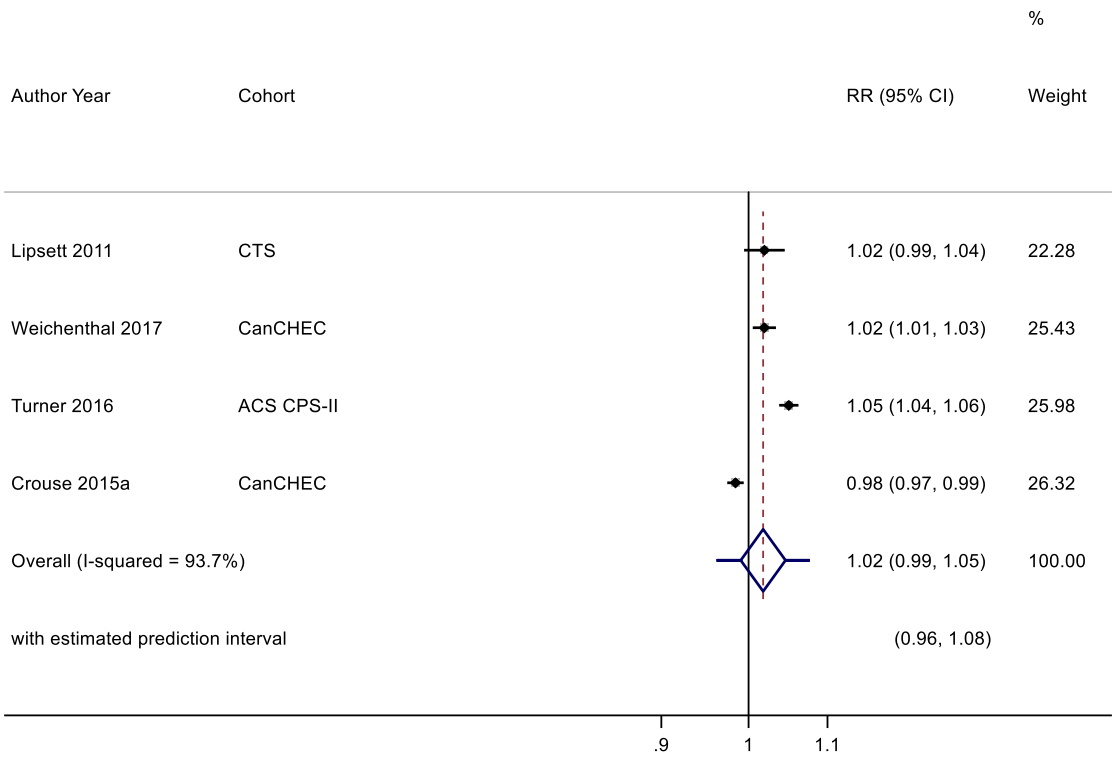

Fig. 9. O_3 peak exposure and respiratory mortality.

adverse associations for NO_2 concentrations and ALRI mortality, while the association was less clear for COPD mortality (Lim et al., 2019).

Lim et al. (Lim et al., 2019) and Hvidtfeldt et al. (Hvidtfeldt et al., 2019) both found no clear association between annual O_3 concentrations and all-cause mortality, which was consistent with our summary estimates; while Hvidtfeldt et al found an adverse association between annual O_3 concentrations and respiratory, COPD, but not ALRI mortality. Kazemiparkouhi et al. (Kazemiparkouhi et al., 2019) found warm season O_3 exposure increased the risk of all-cause, respiratory, and COPD mortality among Medicare beneficiaries. Lim et al showed consistent results for respiratory and COPD mortality, but adverse associations with all-cause and ALRI mortality which were of borderline statistical significance (Lim et al., 2019). In summary, most of the newly published studies reported similar effect estimates compared to our summary estimates, therefore our pooled estimates is unlikely to be altered by the small number of newly published studies.

4. Discussion

4.1. Summary of evidence and comparison with existing literature

4.1.1. Nitrogen dioxide

The review identified 41 articles reporting results for NO_2 and mortality. Associations with mortality were positive; RR (95% CI) were 1.02 (1.01, 1.04); 1.03 (1.01, 1.05); 1.03 (1.01, 1.04); and 1.06 (1.02, 1.10) per 10 μg/m^3 for all-cause, respiratory, COPD and ALRI mortality respectively. The review identified high levels of heterogeneity, as indicated by the I^2 statistic, together with a wide variation between studies in the magnitude and precision of the associations for most pollutant/outcome pairs.

Reviews published in 2013 (Hoek et al., 2013); 2014 (Faustini et al., 2014) and 2018 (Atkinson et al., 2018) have assessed the growing literature on NO_2 and mortality. The evidence base continues to be dominated by studies from North America and Europe. Furthermore, a number of the more recent studies included re-analyses of existing

Table 11
GRADE assessment – O_3 annual exposure and all-cause mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	9 included studies. 3 studies with a total weight of 28% in the meta-analysis had high risk of bias. Excluding these studies did not change significantly the summary RR (text).	No downgrading
Indirectness	1 study with study sample of stroke patients based in London. However, it was a small study and only carried 1% weight	No downgrading
Inconsistency	The 80% prediction interval included 1 & PI > 2 × CI (Fig. 6).	Downgrade one level
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	No analysis of publication bias – too few studies (n = 9)	No downgrading
Large Effect Size	Summary RR = 0.97	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies. 95% CI for linear RR included 1. None of the studies reported the dose–response relationship	No upgrading
GRADE conclusion	Downgrade one level and no upgrade	LOW CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 0.97 PER 10μ/m^3

Table 12
GRADE assessment – O_3 annual exposure and respiratory mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	Only 4 studies; all rated low or moderate risk of bias	No downgrading
Indirectness	All studies included the desired population, exposures and outcomes	No downgrading
Inconsistency	The 80% prediction interval included 1 & PI > 2 × CI (Fig. 7). Substantial heterogeneity amongst small number of studies.	Downgrade one level
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	No analysis of publication bias – too few studies (n = 4)	No downgrading
Large Effect Size	Summary RR = 0.99	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected.	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies. 95% CI for linear RR included 1. None of the studies reported dose–response relationship.	No upgrading
GRADE conclusion	Downgrade one level and no upgrade	LOW CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 0.99 PER 10μ/m^3

cohorts. The summary RR for all-cause mortality from this review is broadly comparable to previous reviews; Faustini et al. (Faustini et al., 2014) assessed 12 studies that also included results for particulate matter and reported a summary HR (per 10 μg/m^3) of 1.04 (1.02, 1.06) ; Hoek et al. (Hoek et al., 2013) assessed 11 cohorts, summary HR = 1.06 (1.04, 1.08); and Atkinson et al. (Atkinson et al., 2018) 23 cohorts with a summary HR = 1.02 (1.01, 1.03). For respiratory mortality, Atkinson et al. (Atkinson et al., 2018) reported a RE summary HR of 1.03 (1.01, 1.05) per 10 μg/m^3 increment in NO_2 based upon 13 studies - the addition of two further studies for this review did not materially alter the summary estimate. Similarly, as few additional studies reporting results for COPD were available, the results from this review and Atkinson et al. (Atkinson et al., 2018) were very similar.

4.1.2. Ozone

The review identified 20 articles reporting results for O_3 and mortality. The majority of the evidence came from cohorts in North America and Europe. A number of cohorts were analysed more than once hence reducing the number of independent estimates available for meta-analyses. Studies also differed in the O_3 metric used; in some studies, O_3 concentrations were calculated for 'peak' or warm season months only, whereas others used annual metrics. Combining studies using annual and peak O_3 metrics was not considered appropriate because the lowest O_3 concentrations are unlikely to occur during 'peak' O_3 months; and secondly correlations between O_3 and other pollutants are known to vary by 'season'.

The associations between annual O_3 and mortality were 0.97 (0.93, 1.02) and 0.99 (0.89, 1.11) per 10 μg/m^3 for all-cause and respiratory mortality respectively. The review identified high levels of heterogeneity, as indicated by the I^2 statistic. Few studies investigated the shape of the concentration–response relationship.

Reviews of the health effects of long-term exposure to O_3 are limited. Early reviews have provided narrative assessments of the cohort literature as part of more comprehensive assessments of the epidemiological and toxicological literature (WHO, 2013; EPA US, 2013). A quantitative review in 2016 (Atkinson et al., 2016) found a limited number of studies for synthesis: no evidence of associations between long-term annual O_3 concentrations and all-cause and respiratory mortality were found, a result confirmed in this review. The 2016 review and this review using updated analyses from the ACS and CanCHEC cohorts differed in their findings for peak season concentrations of O_3 and all-cause and respiratory mortality.

4.2. Heterogeneity

Heterogeneity is an indicator of the extent to which variation between study estimates is too great to be explained by chance. Large variations in study sample sizes/number of events (as for most outcomes included in this review) can lead to an artificially high I^2 statistic, a measure of heterogeneity (IntHout et al., 2016). The I^2 statistic does not provide information about the range of the size of the estimates in a meta-analysis; for this purpose the forest plots and prediction intervals are more informative (Borenstein et al., 2017). One consequence of the high levels of heterogeneity and variation in the size of study estimates found in the evidence assembled for this review is that a random effects model is preferable to a fixed effects model for the meta-

Table 13
GRADE assessment – O_3 peak exposure and all-cause mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	7 included studies. 1 study with high risk of bias- exclusion did not change summary RR (text).	No downgrading
Indirectness	1 study might have introduced some selection bias due to the volunteering sample chosen. However, it was only weighted at < 2% among all studies.	No downgrading
Inconsistency	The 80% prediction interval included 1; PI = 2 × CI (Fig. 8)	No downgrading
Imprecision	The number of person years in the included studies was greater than 940 000.	No downgrading
Publication Bias	No analysis of publication bias – too few studies (n = 6)	No downgrading
Large Effect Size	Summary RR = 1.01. All critical confounders were adjusted for. Insufficient information on unmeasured potential confounders available	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected.	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies. 95% CI for linear RR included 1. 1 study investigated the shape of the dose response relationship with no evidence to suggest non-linear.	No upgrading
GRADE conclusion	No downgrade and no upgrade	MODERATE CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 1.01 PER 10μ/m^3

Table 14
GRADE assessment – O_3 peak exposure and respiratory mortality.

Domain	Judgement	Down/Up Grade
Limitations in studies	4 included studies. 1 study high risk of bias. Exclusion did not alter significantly the RR and CI (text).	No downgrading
Indirectness	All studies included the desired population, exposures and outcomes	No downgrading
Inconsistency	The 80% prediction interval included 1; PI = 2 × CI (Fig. 9). Substantial heterogeneity amongst small number of studies.	Downgrade one level
Imprecision	The number of person years in the included studies was greater than 940 000	No downgrading
Publication Bias	No analysis of publication bias – too few studies (n = 3)	No downgrading
Large Effect Size	Summary RR = 1.02. Insufficient information on unmeasured potential confounders available	No upgrading
Plausible confounding towards null	Confounding direction unknown but precision may be affected.	No upgrading
Dose-response relation	A linear dose–response relationship was assumed in all studies. 95% CI for linear RR included 1. 1 study investigate the dose–response relationship. No evidence to confirm shape of the dose response relationship for Ozone exposure	No upgrading
GRADE conclusion	No downgrade and no upgrade	LOW CERTAINTY EVIDENCE MEAN RR UNADUSTED FOR CO-POLLUTANTS EQUALS 1.02 PER 10μ/m^3

analysis. This random effect model assumes a distribution of true *population* associations; that is the magnitude of the association between pollutant and mortality in one study population is different to another study population. However, as NO_2 and O_3 are gases, and therefore the same compositions in all study locations, it could be argued that a random effects model is not appropriate. The variation between the observed associations may arise because of differences between study characteristics (e.g. study population, adjustment for confounders, spatial resolution of the pollution models, co-pollutants or analytical methods employed) and should be (mostly) explainable using meta-regression techniques provided the appropriate data and sufficient numbers of studies are available. None of the factors available in this review explained fully the observed heterogeneity. Further investigation is needed therefore to inform the appropriate interpretation of the evidence in this review (Egger et al., 1998).

4.3. Concentration response functions

Only a small number of studies investigated the shape of the concentration–response function. In general, these studies found limited evidence to reject the assumption of linearity. A statistically significant linear regression coefficient (log-scale) is indicative of an important concentration response function but does not inform of the actual shape of the relationship, the existence of a threshold or whether the association is present for a given pollutant concentration range. The majority of included studies did not indicate whether the shape of the relationship was investigated prior to fitting a linear model. The minimum or 5th percentile values of the distributions of the pollutant concentrations provide an indication of the lower concentrations included in the set of observations to which a linear model was fitted. This is not the same as saying the concentration–response relationship is linear down to these concentrations. A linear model will fit a linear relationship between two variables irrespective of whether not the exposure variable and the outcome are linearly related. Furthermore, data are often sparse at the extremes of the pollutant distributions and therefore the corresponding prediction intervals wide. Caution is therefore required when interpreting the results from linear models in relation to the range of observed concentrations.

Even for studies that only reported results from linear models, the distribution of the pollutant concentrations is not always reported. In such cases authors were contacted by email and asked to provide the relevant data with variable response. Hence, in a sizeable proportion of the studies, the 5th percentile/minimum pollutant concentrations are missing. Moreover, studies from geographical regions with high levels of pollutants comprised a small proportion in our review, but with vast majority of studies from countries with low-middle range of pollutant concentration. These gaps make interpretation of the evidence more difficult. Given the importance of both the shape of the concentration response function and the range of observed concentrations to the achievement of the review objective, a strategy for evidential judgement is required. For example, should guideline recommendations be based upon only those studies with complete data and have specifically set out to evaluate the shape of the concentration response function?

4.4. Multi-pollutant models

Assessment of the impact of co-pollutants on the associations between NO_2 and O_3 and mortality was limited by the small number of studies reporting results from two-pollutant models and the high correlation between pollutants in some studies. The difficulties in interpreting coefficients in multi-pollutant models are well recognised (Greenbaum and Shaikh, 2010; Dominici et al., 2010) and include high correlation between pollutants (limiting the ability of two-pollutant models to separate out associations) leading to unstable parameter estimation; differential measurement error between pollutants which can lead to the 'transfer' of an association from the less well measured (but true) pollutant to the better measured (but incorrect) pollutant; and finally analysts rarely assess interactions between pollutants which is necessary to interpret correctly model main effects. Some investigators have proposed methods for dealing with correlated predictors, for example composite hazard ratios for more than one pollutant. An assessment of results from multi-pollutant models should consider changes in risk estimates from single and multi-pollutant pollutants for each pollutant jointly, not individually (Dominici et al., 2010). The relatively small numbers of studies reporting results from multi-pollutant models limits out assessment of these issues. Caution should be exercised therefore when interpreting results from single pollutant models as reported associations may reflect pollutant mixtures rather than individual pollutants *per se*.

4.5. Strengths and limitations

This review uses a comprehensive search strategy applied to three databases and updated to include more recent publications. It also includes a narrative assessment of the evidence for the shape of the concentration response function for both NO_2 and O_3 and consideration of results from multi-pollutant models. A key strength of the review involved the application of new RoB and GRADE tools developed specifically for application in environmental epidemiology.

In common with many reviews of cohort studies of outdoor air pollutants and mortality the evidence base can be limited both in terms of the number of independent cohorts and their geographical spread. These limitations may restrict the applicability of the review findings worldwide. The number of available cohorts also precludes meaningful

meta-regression to explore causes of heterogeneity. Whilst sub-group analysis, even *a priori* sub-group analysis, is useful to explore differences between studies, it is a univariate procedure and does not rule out the possibility of group differences arising due to other confounding factors.

This review of associations between NO_2 and O_3 and mortality in epidemiological cohort studies provides evidence for the assessment of the strength of evidence for associations only. It has focused on results form single pollutant models. The question of the independence of these associations from other pollutants requires careful consideration. A separate causal determination is required to proceed to quantification of health impacts.

Hazard ratios from cohort studies are typically small. The choice of studies for meta-analyses can have a relatively large impact on the summary HRs. Because of the ubiquitous nature of exposure to outdoor air pollution, small HRs derived from reviews of this kind, can have a substantial estimated health impact because of the large populations exposed. The review protocol including decisions on study selection relating to confounder adjustment, patient vs general population cohorts, spatial resolution of the air pollution models will have a major bearing on the included studies, the meta-analytical estimates and consequently health impact assessments. For these reasons the prediction intervals provide useful and important information regarding the range of the risk estimates in the studied populations.

The RoB tool was discriminatory for a small number of studies in the confounding domain only and all but a very few studies were rated low for other domains. One possible explanation for this lack of discrimination is that the included studies are of a high quality and at low risk of bias. Another is that the RoB tool is not sensitive enough to assess the risk of potential biases in the literature. For example, the bias assessment in the confounding domain relied upon the inclusion/exclusion of a set of critical and potential confounders in the studies. The list of potential confounders was long and lead to most studies being rated as moderate risk of bias. The RoB tool used in this review was recently developed for environmental epidemiology, therefore we cannot rule out the potential misapplication of the new tool at this stage. However, the new RoB tool does provide a framework to assess bias in the included literature systematically. Future development of the tool could further improve its capacity to recognise the flaws in study design and report which potentially attenuate the effect observed.

There remains no widely accepted GRADE-like system to assess evidence in observational studies in environmental health. The adapted GRADE framework used in this review was less strict than the standard GRADE. The modified framework, which was developed by methodologists and experts in environmental epidemiology, is a step towards achieving a robust set of criteria for evidence evaluation. However, it is not without its difficulties. The development of the tool was a lengthy process with differences between group members relating to the detail and application of the tool. The use of the E-value (VanderWeele and Ding, 2017), derived from the summary risk ratio in a meta-analysis, to assess the size of associations reported in cohort studies remains problematical because some cohorts control for most critical and potential confounders. Determining a potential unmeasured confounder is therefore not straightforward, nor is finding the relevant literature with which to assess the degree of potential confounding. Furthermore, the E-value is not without its critics (Ioannidis et al., 2019). Another challenging aspect of the tool is the assessment of small study bias in the presence of heterogeneity (Peters et al., 2010). In this respect we have been circumspect in the application of this specific GRADE criterion, interpreting both the funnel plot and the result of the Eggers test with caution. The criterion that the 80% prediction interval is twice the confidence interval and contains unity was also new. Where the body of evidence is heterogeneous and a wide range in the magnitude of associations is observed, this criterion downgrades the certainty of evidence on the basis that there may exist one (or more) populations in which an adverse association is not found. There may well be positive associations in other populations however. During the development of the tool there was discussion on whether this rule should be strictly applied or whether a more flexible application was appropriate. We therefore applied downgrading in this domain only when the 80% PI contained unity and did not downgrade when the ratio was marginal (the CI are approximately symmetrical about the point estimate on the relative risk scale when the risks are small). Exclusion of the few, high RoB studies did not lead to major changes in the summary risk ratios and hence did not indicate a downgrade for the domain, though this is perhaps not surprising given the small risks reported in many air pollution cohort studies and the relative imprecision of some studies.

5. Conclusion

This review of cohort studies found positive associations between long-term concentrations of NO_2 and mortality and limited evidence for O_3 and mortality. However, there was very high levels of heterogeneity between study estimates giving rise to 80% prediction intervals that included the null for most pollutant-outcome pairs with insufficient studies to explore reasons using meta-regression. Relatively few studies reported results from multi-pollutant models.

For NO_2 and mortality we assessed the certainty of evidence (adapted GRADE) from single pollutant models to be moderate for all-causes (mean RR = 1.02 per 10 μ/m^3), moderate for respiratory (mean RR 1.03 per 10 μ/m^3); high for COPD (mean RR = 1.03 per 10 μ/m^3; and moderate for ALRI (mean RR = 1.06 per 10 μ/m^3). For studies reporting annual O_3 metrics we assessed the certainty of the evidence from single pollutant models to be low for all-cause mortality (man RR = 0.97 per 10 μ/m^3); and low for respiratory mortality (mean RR = 0.99 per 10 μ/m^3). For peak O_3 exposures we assessed the certainty of evidence from single pollutant models to be moderate for all-cause mortality (mean RR = 1.01 per 10 μ/m^3) and low for respiratory mortality (mean RR = 1.02 per 10 μ/m^3).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This systematic review has been funded by the World Health Organization Regional Office for Europe, supported by the European Commission (DG Environment), Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (Germany), Federal Ministry of Health (Germany), Government of the Republic of Korea, Federal Office for the Environment (Switzerland) and United States Environmental Protection Agency, and delivered as part of the evidence base that informs the ongoing development of WHO global air quality guidelines. All rights in the work, including ownership of the original work and copyright thereof, are vested in WHO. The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions or the stated policy of the World Health Organization. We thank Jos Verbeek, Román Pérez-Velasco, Hanna Yang, Dorota Jarosinska, Rebecca Morgan, the rest of the Systematic Review Team and the Guideline Development Group for their contributions to the conduct of the systematic review.

Appendix A. WHO GRADE guidance

Approach to assessing the certainty of evidence from systematic reviews informing WHO global air quality guidelines

By: The WHO global air quality guidelines Working group on certainty of evidence assessment

Acknowledgements

This supplementary material consists of an approach to assessing the certainty of evidence from systematic reviews of epidemiologic studies of air quality and health, based on the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.

The approach was developed by external methodologist Jos Verbeek (Cochrane Work), with inputs from the WHO Global Air Quality Guidelines Working Group on Certainty of Evidence Assessment, convened by the WHO European Centre for Environment and Health (WHO Regional Office for Europe) in the context of the forthcoming WHO global air quality guidelines. The Working Group was composed of the Guideline Development Group members: Aaron Cohen (Health Effects Institute), Bert Brunekreef (Utrecht University), Francesco Forastiere (King's College London), Nino Künzli (Swiss Tropical and Public Health Institute), and external methodologist: Rebecca Morgan (McMaster University); and, from the staff of the WHO Regional Office for Europe: Román Pérez-Velasco, Hanna Yang and Dorota Jarosińska. Additional comments were provided at different stages by external methodologist Eva Rehfuess (Cochrane Public Health Europe) and GDG members Michal Krzyzanowski (King's College London), and Jonathan Samet (Colorado School of Public Health).

The WHO Regional Office for Europe acknowledges funding and in-kind contributions from the European Commission (Directorate-General for Environment); the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety; the German Federal Ministry of Health; the Government of the Republic of Korea; the Swiss Federal Office for the Environment; and the United States Environmental Protection Agency.

Background

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) has been developed to standardize the approach to judging the certainty of the effects of interventions (Schunemann et al., 2013). As such, the approach is currently the basis for evidence review in support of WHO Guidelines (World Health Organization, 2014).

The main value of the system is that the comparability of the judgements increases when all assessors consider the same arguments underpinning their certainty in a similar manner. That is how the factors for downgrading and upgrading the certainty have been developed: to guide expert judgement. Behind each down- and upgrading factor in GRADE, there is a rationale for its importance and guidance for elaborating good reasons for downgrading or not downgrading. These ideas are well explained in the *GRADE Handbook* (Schunemann et al., 2013). Most of the reasoning in this framework can be equally well used for observational studies of exposure as for randomized studies of interventions (Morgan et al., 2016). However, at some points there is a need for elaboration or clarification on how to use the GRADE criteria for observational studies of exposure.

Although different groups have adapted the approach for environmental exposures in recent years, no consensus has emerged among experts yet. Unlike some previous efforts, the aim of this work is not assessing the strength of evidence for causal inference by considering all the relevant streams of research (Woodruff and Sutton, 2011), but to rate how certain one is that the 'true' estimate of the epidemiological association between an air pollutant and an adverse health effect lies within a particular range (Hultcrantz et al., 2017). Consistent with the standard GRADE framework, the certainty of the effect estimate is graded as high, moderate, low or very low. The ratings are subsequently used to select and underpin concentration − response functions in the process of deriving guideline exposure levels.

The current approach was designed specifically to assess the certainty of the evidence from the systematic reviews commissioned by WHO to inform the update of global air quality guidelines (AQGs). Its development benefitted from previous experiences in applying GRADE in the field of occupational and environmental health, as well as specific expertise in air pollution epidemiology. The approach was extensively discussed in two Guideline Development Group meetings, pilot tested by the members of the Systematic Review Team and improved iteratively according to the feedback received.

The Working Group accepted to start the rating of the certainty of the evidence for observational studies at moderate certainty evidence and not at high certainty, because of the risk of unmeasured confounding in observational studies. The certainty of the evidence from this level can then be downgraded or upgraded, based on the criteria per GRADE domain. The GRADE domains and the criteria considered when judging the certainty of the evidence are elaborated below.

Reasons for downgrading

Limitations in studies: Downgrade one or two levels

For risk of bias in studies, there should be serious concern about bias in the studies that have the most weight in the meta-analysis to rate down the certainty of the total body of evidence with one level. If there are very serious concerns, the certainty can be downgraded with two levels.

This is a judgement and there are no clear pre-set cut-off points (WHO, 2020). A judgement is based on the number of studies and the impact they have in the meta-analysis, as well as the seriousness of the risk of bias in these studies. One small study with very serious risk of bias but hardly an influence on the meta-analysis should not be a reason to downgrade, but two big studies with a considerable weight in the meta-analysis should.

If the sensitivity analysis for risk of bias shows a considerable impact on the effect size, the conclusions could be based on the studies at low risk of bias only. In that case, there is no reason to downgrade because the body of evidence on which the conclusions are based is considered to be at low risk of bias only.

Indirectness: Downgrade one or two levels

The assessors should consider the extent to which the Population, Exposure, Comparator, Outcome(s), Study Design (PECOS) of the studies in the meta-analysis reflects the original PECOS question formulated at the beginning of the systematic review process (Guyatt et al., 2011).

If there are considerable differences between the elements of the PECOS in the body of evidence compared to the original question, then the certainty of the body of evidence should be rated down with one level. This would, for example, be the case if the evidence consists of studies of

occupational exposure instead of exposure in the general population.

Inconsistency: Downgrade one or two levels

Inconsistency among studies means that there is a considerable difference in effect size between studies. For example, if there are studies in the body of evidence that show a harmful effect and also studies that show a preventive effect, this indicates serious inconsistency or heterogeneity.

Usually there is more heterogeneity in observational than in experimental studies, because more factors can influence the effect size. Therefore, it is important to try to explain the heterogeneity. The first step should be to consider the factors that are listed for subgroup analyses in the protocol, as those that are most likely to be moderators of effect sizes. Another source of heterogeneity can be variation in risk of bias. This may explain part of the heterogeneity, and evaluation of only studies at low risk of bias should then decrease the heterogeneity. The difference in effect sizes between the subgroups should be tested for statistical significance. A rule of thumb to be used is to check if the confidence intervals of the subgroup pooled effect sizes do not overlap.

Ideally, a meta-regression should be conducted including all moderators of the effect size, to find out how much heterogeneity remains after allowing for previously established reasons for heterogeneity. In practice, it is unlikely that all studies in a systematic review will have the necessary information to do a complete meta-regression including all previously documented reasons for heterogeneity. This could then be done on subsets of studies having the relevant information.

Heterogeneity is often measured with the I^2 statistic which varies between 0 and 100%, where 0% would indicate no heterogeneity and 100% large heterogeneity. Because the I^2 statistic is a relative measure, it is difficult to make a judgement about the absolute amount of heterogeneity. As a result, the use of the prediction interval has been suggested (IntHout et al., 2016; Borenstein, 2019).

The prediction interval provides an estimate of the distribution of the true effect sizes. To prevent overstating heterogeneity in observational studies, an 80% interval, and not the usual 95% interval, was chosen. For an 80% prediction interval, the true effect size for 80% of all populations would fall in this interval. This tells if the effect is consistent or if it varies substantially. It also tells if the effect is harmful in all populations, or if there is no effect in some populations or maybe even a preventive effect.

To make a judgement about the amount of heterogeneity that cannot be explained and that would be a reason for concern and a reason for downgrading, the following approach is proposed.

If the 80% prediction interval for a specific meta-analysis of relative risks is of the same size as the confidence interval, this indicates that there is no more variation in effect sizes than the statistical uncertainty. Then there is no reason for concern about heterogeneity.

However, if the prediction interval is considerably wider than the confidence interval (e.g., double the size) and overlaps with 1, there is reason for concern about heterogeneity. The effect sizes of the studies vary so much that with different samples of studies the conclusions of the meta-analysis could be substantially different. For example, an alternative conclusion could be that there would be no risk. In this case, the certainty of the body of evidence would be downgraded with one level.

Assessors need to provide a rationale for downgrading or not downgrading by explicitly addressing all of the issues mentioned above. This includes an assessment of how much of the heterogeneity can be explained.

Imprecision: Downgrade one or two levels

Precision of the pooled effect size is another domain to be judged for downgrading. If there are only a few participants and the confidence interval around the pooled effect size is wide, one is less inclined to believe that the results reflect the true effects. If there is considerable imprecision, there is a reason to downgrade.

The cut-offs for downgrading because of imprecision given by the standard GRADE approach are applicable to clinical decision-making. Since in environmental health there are no clinical decision thresholds involved, only the second criterion of optimal information size can be applied to air pollution and health studies.

Therefore, the proposed approach consists of calculating the number of participants needed for a single study that can measure the relative risk of interest with sufficient precision (Rothman and Greenland, 2018). If the number of participants in the meta-analysis is considerably lower than the number that would be needed for an adequately powered study, the certainty of the evidence is rated down. This is a relatively conservative approach, and implies that the information size of the meta-analysis would need to be larger than the single study because heterogeneity has to be taken into account.

A method of calculating the sample size needed for a study with a specific relative risk and confidence interval was recently proposed by Rothman and Greenland (Ostro et al., 2010). As guidance, the calculation of the sample size needed to be able to assess a relative risk for mortality of 1.05 per 10 $\mu g/m^3$ increase of $PM_{2.5}$ with a confidence interval with a width of 0.09 (1.01–1.10) is provided below.

The event rate of mortality would be 0.0116 per person-year as in (Ostro et al., 2010); (Guyatt et al., 2011) . This would lead to a number of about 940,000 person-years in the meta-analysis, containing sufficient information to assess the relative risk of interest with sufficient precision.

The event rate in the example above was observed over a five-year follow-up period in a cohort of female public school teachers aged around 54 years on average at baseline. As the confidence interval of the relative risk depends also and strongly on the event rate, the calculated number of about 940,000 person years should be viewed as indicative. It could be considerably smaller in older populations with higher event rates, and considerably larger in populations with lower event rates.

Separate calculations are needed for short-term studies which do not deal with person years but with numbers of daily events.

Publication bias: Downgrade one level

Publication bias is assessed by a funnel plot and Egger's test. If the funnel plot upon visual inspection shows that small studies with non-harmful effects are missing, this would be an indication of publication bias. This means that small (imprecise) studies that have a relative risk smaller than 1 are missing. If there is no indication for these missing studies in the funnel plot, there is no use for the Egger's test, because significance will result from other factors causing heterogeneity (Borenstein, 2019). The Egger's test would just be used to confirm suspected publication bias detected from the funnel plot.

It is important to note that the Egger's test can easily produce statistical significance for other reasons than publication bias in case of heterogeneity. Members of the Working Group noted that the Egger's test should not be used in case of heterogeneity, and that funnel plots should only include the studies included in the meta-analysis. Then, assessors should examine if small imprecise studies are missing in the funnel plots.

Other approaches to assessing reporting bias, such as a subgroup analysis of multi-centre studies compared to single city studies in case of

evidence based on time series studies, an analysis of differences in effect estimates from earlier versus later studies, and a comparison to published results of attempts to quantify the magnitude of reporting bias, may help make a judgement.

Reasons for upgrading

The majority of the Working Group decided to recommend that upgrades for reasons of large effect size, all plausible confounding moving the relative risk estimate towards the null, and concentration − response gradient should be addressed independently from the results of applying the downgrading factors. Domains would be treated equally and independently, thus, leading to upgrading, downgrading or not changing the evidence level. A downgrade for any reason would not necessarily preclude upgrading for another reason.

Large magnitude of effect size: Upgrade one level

The standard GRADE approach proposes upgrading the certainty of the evidence in observational studies if the pooled effect size is large or very large, so that 'the study design that is more prone to bias is unlikely to explain all of the apparent benefit or harm'. The cut-off point for a large effect size for harm is a relative risk > 2, while for a very large effect size is a relative risk > 5. These numbers are somehow arbitrary, and are not in the order of magnitude of the many relative risks reported in environmental health.

Instead of taking a certain value of the relative risk as the cut-off point, it is reasonable to judge whether confounding could have easily influenced the pooled effect size found in the meta-analysis. To this end, the application of the E-value approach is helpful (VanderWeele and Ding, 2017; Haneuse et al., 2019; Ioannidis et al., 2019; VanderWeele et al., 2019). This statistic is based on an assessment of how easily unmeasured confounders could explain away the relationship found between the exposure and the health outcome. It is based on the mathematical calculation of how large the effect of a confounder should be to explain away the relative risk that has been found in a study. With 'explain away', it is meant that such a confounder would reduce the relative risk that resulted from the observations in the study to 1. This effect (or E-value) is a function of the relative risk that has been found in a study or in a meta-analysis and is calculated as follows: E-value = RR + sqrt {RR * (RR – 1)}. The idea behind it is very similar to the 'large effect' concept in the standard GRADE framework but does not use absolute cut-offs for large effect sizes.

The judgement is then to ascertain if an unmeasured confounder could easily have an association with the exposure and the outcome with relative risks as large as or larger than the E-value. It is important to note that this is always the *covariate-adjusted* association between the unmeasured confounder and the outcome, and also the *covariate-adjusted* association between the unmeasured confounder and exposure to air pollution. If such a confounder could realistically have such strong relationships with both exposure and outcome, then unmeasured confounding could explain away the observed pooled relative risk. If one judges that it would be very unlikely that an unmeasured confounder would attain a relative risk as high as the E-value, then one can conclude that unmeasured confounding is unlikely to explain away the relative risk that has been observed. In that case, the certainty of the evidence can be upgraded because of a large effect size.

It is important to note that a major part of the judgement is what a realistic value for the relative risk of the unmeasured confounder could possibly be. Preferably, this should be based on what is known about strong confounders for the association at hand. For the association air pollution– mortality, smoking would be an obvious choice about which much information is available concerning its relationship with all-cause and cause-specific mortality. However, the residual association between smoking and air pollution is highly variable across published studies, and calculations of E-values should report the covariate-adjusted associations with both air pollution and the outcome. The same logic applies to short-term studies where the covariate-adjusted associations between the confounder and the exposure (and the confounder–outcome) is relevant.

All plausible confounding shifts the relative risk towards the null: Upgrade one level

Another proposed reason for upgrading is if all plausible confounding would shift the relative risk towards the null and still there would be a significant relative risk. This requires considerable judgement of possible confounders.

In most air quality and health studies, there would be a long list of possible confounders that would shift the relative risk in both directions. However, if one can reasonably argue that all confounding would have reduced the relative risk towards 1, then this will be a reason to upgrade the certainty of the evidence with one level.

Concentration − response gradient: Upgrade one level

The standard GRADE proposes upgrading the certainty of the evidence if there is a concentration − response relationship between exposure and adverse health outcomes.

This domain is readily applicable to air quality and health studies. If there is an increase in risk with increasing exposure, either linearly or non-linearly, the certainty of the evidence would be upgraded with one level.

Acknowledgement

This systematic review has been funded by the World Health Organization Regional Office for Europe, supported by the European Commission (DG Environment), Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (Germany), Federal Ministry of Health (Germany), Government of the Republic of Korea, Federal Office for the Environment (Switzerland) and United States Environmental Protection Agency, and delivered as part of the evidence base that informs the ongoing development of WHO global air quality guidelines. All rights in the work, including ownership of the original work and copyright thereof, are vested in WHO. The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions or the stated policy of the World Health Organization. We thank Jos Verbeek, Román Pérez-Velasco, Hanna Yang, Dorota Jarosinska, Rebecca Morgan, the rest of the Systematic Review Team and the Guideline Development Group for their contributions to the conduct of the systematic review.

Appendix B

Tables B1–B4.

Table B1
Health outcomes selected for the update of the AQGs in relation to long-term exposure to ambient air pollutants.

Long-term exposure			
Pollutants	**Health outcome(s) used in latest WHO AQGs (2006)**	**Health outcomes selectedfor updating WHO AQGs**	**Justification for health outcome selection**
O_3	No long-term guideline provided	• **All-cause mortality** • **Respiratory mortality**	**CAUSALITY DETERMINATION** • US EPA total mortality (suggestive, 2013) • HC respiratory mortality (suggestive, 2013) **SUPPORTING CONSIDERATIONS** • Severity of health outcome, burden of disease • Precautionary principle from expected increase of this pollutant due to climate change (policy implications and end-user perspectives). **OTHER RELEVANT CAUSAL DETERMINATIONS***(to be described in guidelines background chapter)* • US EPA respiratory effects (likely, 2013) • HC respiratory effects (suggestive, 2013)
NO_2	Respiratory effects in children	• **All-cause mortality** • **Respiratory mortality**	**CAUSALITY DETERMINATION** • US EPA (suggestive for total mortality, 2016) • HC (suggestive for total mortality, 2016) **SUPPORTING CONSIDERATIONS** • Severity of health outcome, burden of disease • Recent studies show associations with respiratory mortality, consistent with likely causality for respiratory effects (see other causal determinations below). • The causal determination of US EPA for mortality is suggestive, in light of the limited number of studies properly addressing confounding by other transport-related air pollutants. • The causal determination of US EPA of "likely causal" for respiratory effects (see other causal determinations below) takes into account respiratory mortality. • Also, studies on asthma incidence (mainly from USA) considered for respiratory effects are observed in children with specific genetic profiles, which may not be applying globally (differences in genetic makeup). **OTHER RELEVANT CAUSAL DETERMINATIONS***(to be described in guidelines background chapter)* • US EPA respiratory effects (likely, 2016) • HC respiratory effects (likely, 2016)

HC: Health Canada science assessments, US EPA: United States Environmental Protection Agency Integrated Science Assessments (ISA), COHb: carboxyhaemoglobin, ED: Emergency Department visits; HA: Hospital Admissions, IHD: Ischaemic Heart Disease; COPD: chronic obstructive pulmonary disease ALRI: acute lower respiratory infections; CV: cardiovascular admissions, IHD: Ischaemic Heart Disease; COPD: chronic obstructive pulmonary disease ALRI: acute lower respiratory infections; CV: cardiovascular

Table B2
Search strategy.

MEDLINE (Search timeline: 1946–15.Jan.2018)		
#11	#8 and #9 and #10	448
#10	#5 or #6 or #7	55,425
#9	#3 or #4	1,068,043
#8	#1 or #2	1,111,602
#7	("Nitrogen Dioxide" or NO2 or ozone or O3).tw.	27,723
#6	(Nitrogen Dioxide or ozone).nm.	17,689
#5	(Nitrogen Dioxide or ozone or air pollution).sh.	43,385
#4	(cohort or Cox or hazard* or prospective).tw.	978,195
#3	cohort studies.sh.	245,055
#2	(mortality or death).tw.	1,092,011
#1	(mortality or death).sh.	59,284
EMBASE (1980–15.Jan.2018)		
#10	#7 and #8 and #9	823
#9	#5 or #6	95,646
#8	#3 or #4	2,020,425
#7	#1 or #2	1,760,699
#6	("Nitrogen Dioxide" or NO2 or ozone or O3).tw.	44,586
#5	(Nitrogen Dioxide or Ozone or air pollution).sh.	77,983
#4	(cohort or Cox or hazard*).tw.	942,137
#3	(cohort analysis or follow up).sh.	1,488,257
#2	(mortality or death).tw.	1,508,949
#1	(mortality or death).sh.	860,304
Web of Science 1970–11.Jan.2018		
#4	#3 AND #2 AND #1	1,647
#3	TS = ("nitrogen dioxide") OR TS = (NO2) OR TS = (ozone) OR TS = (O3) OR TS = ("air pollution")	175,687
#2	TS = (cohort) OR TS = (cox) OR TS = (hazard*) OR TS = (prospective)	1,198,902
#1	TOPIC: (mortality) OR TOPIC: (death)	1,398,001

Table B3
Excluded studies (with reasons).

No quantitative HR provided

1. Peng Z, Liu C, Xu B, Kan H, Wang W. Long-term exposure to ambient air pollution and mortality in a Chinese tuberculosis cohort. Science of the Total Environment. 2017;580:1483–8.
2. Lin JH, Yen TH, Weng CH, Huang WH. Environmental NO2 level is associated with 2-year mortality in patients undergoing peritoneal dialysis. Medicine (United States). 2015;94(1):e368.
3. Andersen ZJ, de Nazelle A, Mendez MA, Garcia-Aymerich J, Hertel O, Tjonneland A, et al. A study of the combined effects of physical activity and air pollution on mortality in elderly urban residents: The Danish Diet, Cancer, and Health Cohort. Environmental Health Perspectives. 2015;123(6):557–63.
4. Villeneuve PJ, Jerrett M, Su J, Burnett RT, Chen H, Brook J, et al. A cohort study of intra-urban variations in volatile organic compounds and mortality, Toronto, Canada. Environmental Pollution. 2013;183:30–9.
5. Vedal S, Campen MJ, McDonald JD, Larson TV, Sampson PD, Sheppard L, et al. National Particle Component Toxicity (NPACT) initiative report on cardiovascular effects. Research report (Health Effects Institute). 2013(178):5–8.
6. Gan WQ, Davies HW, Koehoorn M, Brauer M. Association of long-term exposure to community noise and traffic-related air pollution with coronary heart disease mortality. American Journal of Epidemiology. 2012;175(9):898–906.
7. Nawrot TS, Vos R, Jacobs L, Verleden SE, Wauters S, Mertens V, et al. The impact of traffic air pollution on bronchiolitis obliterans syndrome and mortality after lung transplantation. Thorax. 2011;66(9):748–54.
8. Krewski D, Burnett RT, Goldberg M, Hoover K, Siemiatycki J, Abrahamowicz M, et al. Reanalysis of the Harvard Six Cities Study, Part II: Sensitivity analysis. Inhalation Toxicology. 2005;17(7–8):343–53.
9. Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama-Journal of the American Medical Association. 2002;287(9):1132–41.
10. Lipfert FW, Perry Jr HM, Miller JP, Baty JD, Wyzga RE, Carmody SE. The Washington University-EPRI Veterans' Cohort Mortality Study: preliminary results. Inhalation toxicology. 2000;12 Suppl 4:41–73.
11. Abbey DE, Lebowitz MD, Mills PK, Petersen FF, Beeson WL, Burchette RJ. LONG-TERM AMBIENT CONCENTRATIONS OF PARTICULATES AND OXIDANTS AND DEVELOPMENT OF CHRONIC DISEASE IN A COHORT OF NONSMOKING CALIFORNIA RESIDENTS. Inhalation Toxicology. 1995;7(1):19–34.
12. Dockery DW, Pope ICA, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association between air pollution and mortality in six U.S. cities. New England Journal of Medicine. 1993;329(24):1753–9.
13. Abbey DE, Colome SD, Mills PK, Burchette R, Beeson WL, Tian Y. Chronic disease associated with long-term concentrations of nitrogen dioxide. Journal of exposure analysis and environmental epidemiology. 1993;3(2):181–202.

Results replicated elsewhere

1. Beelen R, Hoek G, Raaschou-Nielsen O, Stafoggia M, Andersen ZJ, Weinmayr G, et al. Natural-cause mortality and long-term exposure to particle components: An Analysis of 19 European cohorts within the multi-center ESCAPE project. Environmental Health Perspectives. 2015;123(6):525–33.
2. Wang M, Beelen R, Stafoggia M, Raaschou-Nielsen O, Andersen ZJ, Hoffmann B, et al. Long-term exposure to elemental constituents of particulate matter and cardiovascular mortality in 19 European cohorts: Results from the ESCAPE and TRANSPHORM projects. Environment International. 2014;66:97–106.
3. Gan WQ, Tamburic L, Davies HW, Demers PA, Koehoorn M, Brauer M. Changes in residential proximity to road traffic and the risk of death from coronary heart disease. Epidemiology. 2010;21(5):642–9.
4. Beelen R, Hoek G, van den Brandt PA, Goldbohm RA, Fischer P, Schouten LJ, et al. Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). Environmental Health Perspectives. 2008;116(2):196–202.
5. Krewski D, Burnett R, Jerrett M, Pope CA, Rainham D, Calle E, et al. Mortality and long-term exposure to ambient air pollution: ongoing analyses based on the American Cancer Society cohort. Journal of toxicology and environmental health. 2005;Part A. 68(13–14):1093–109.
6. Jerrett M, Burnett RT, Ma RJ, Pope CA, Krewski D, Newbold KB, et al. Spatial analysis of air pollution and mortality in Los Angeles. Epidemiology. 2005;16(6):727–36.
7. Krewski D, Burnett RT, Goldberg MS, Hoover BK, Siemiatycki J, Jerrett M, et al. Overview of the Reanalysis of the Harvard Six Cities Study and American Cancer Society study of particulate air pollution and mortality. Journal of Toxicology and Environmental Health - Part A. 2003;66(16–19):1507–51.

HR for NO_X not NO_2

1. Cohen G, Levy I, Yuval, Kark JD, Levin N, Broday DM, et al. Long-term exposure to traffic-related air pollution and cancer among survivors of myocardial infarction: A 20-year follow-up study. European Journal of Preventive Cardiology. 2017;24(1):92–102.
2. Stockfelt L, Andersson EM, Molnar P, Rosengren A, Wilhelmsen L, Sallsten G, et al. Long term effects of residential NOx exposure on total and cause-specific mortality and incidence of myocardial infarction in a Swedish cohort. Environmental Research. 2015;142:197–206.
3. Cao J, Yang C, Li J, Chen R, Chen B, Gu D, et al. Association between long-term exposure to outdoor air pollution and mortality in China: A cohort study. Journal of Hazardous Materials. 2011;186(2–3):1594–600.
4. Lipfert FW, Wyzga RE, Baty JD, Miller JP. Air pollution and survival within the Washington University-EPRI veterans cohort: Risks based on modeled estimates of ambient levels of hazardous and criteria air pollutants. Journal of the Air and Waste Management Association. 2009;59(4):473–89.
5. Nafstad P, Haheim LL, Wisloff T, Gram F, Oftedal B, Holme I, et al. Urban air pollution and mortality in a cohort of Norwegian men. Environmental Health Perspectives. 2004;112(5):610–5.

Other causes of death

1. Cappellari M, Turcato G, Zannoni M, Forlivesi S, Maccagnani A, Bonora A, et al. Association between short- and medium-term air pollution exposure and risk of mortality after intravenous thrombolysis for stroke. Journal of Thrombosis and Thrombolysis. 2017:1–7.
2. Tseng E, Ho WC, Lin MH, Cheng TJ, Chen PC, Lin HH. Chronic exposure to particulate matter and risk of cardiovascular mortality: cohort study from Taiwan. BMC public health. 2015;15:936.
3. Sorensen M, Luhdorf P, Ketzel M, Andersen ZJ, Tjonneland A, Overvad K, et al. Combined effects of road traffic noise and ambient air pollution in relation to risk for stroke? Environmental Research. 2014;133:49–55.
4. Bhinder S, Chen H, Sato M, Copes R, Evans GJ, Chow CW, et al. Air pollution and the development of posttransplant chronic lung allograft dysfunction. American Journal of Transplantation. 2014;14(12):2749–57.
5. Beelen R, Stafoggia M, Raaschou-Nielsen O, Andersen ZJ, Xun WW, Katsouyanni K, et al. Long-term Exposure to Air Pollution and Cardiovascular Mortality An Analysis of 22 European Cohorts. Epidemiology. 2014;25(3):368–78.
6. Raaschou-Nielsen O, Sorensen M, Ketzel M, Hertel O, Loft S, Tjonneland A, et al. Long-term exposure to traffic-related air pollution and diabetes-associated mortality: a cohort study. Diabetologia. 2013;56(1):36–46.
7. Chen H, Goldberg MS, Burnett RT, Jerrett M, Wheeler AJ, Villeneuve PJ. Long-term exposure to traffic-related air pollution and cardiovascular mortality. Epidemiology. 2013;24(1):35–43.
8. Zanobetti A, O'Neill MS, Gronlund CJ, Schwartz JD. Summer temperature variability and long-term survival among elderly people with chronic disease. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(17):6608–13.
9. Andersen ZJ, Kristiansen LC, Andersen KK, Olsen TS, Hvidberg M, Jensen SS, et al. Stroke and long-term exposure to outdoor air pollution from nitrogen dioxide: A cohort study. Stroke. 2012;43(2):320–5.
10. Zhang PF, Dong GH, Sun BJ, Zhang LW, Chen X, Ma NN, et al. Long-Term Exposure to Ambient Air Pollution and Mortality Due to Cardiovascular Disease and Cerebrovascular Disease in Shenyang, China. Plos One. 2011;6(6).
11. Zanobetti A, Schwartz J. Ozone and survival in four cohorts with potentially predisposing diseases. American Journal of Respiratory and Critical Care Medicine. 2011;184(7):836–41.
12. Spencer-Hwang R, Knutsen SF, Soret S, Ghamsary M, Beeson WL, Oda K, et al. Ambient air pollutants and risk of fatal coronary heart disease among kidney transplant recipients. American Journal of Kidney Diseases. 2011;58(4):608–16.

(*continued on next page*)

Table B3 (*continued*)

13. Gan WQ, Koehoorn M, Davies HW, Demers PA, Tamburic L, Brauer M. Long-term exposure to traffic-related air pollution and the risk of coronary heart disease hospitalization and mortality. Environmental Health Perspectives. 2011;119(4):501–7.
14. McKean-Cowdin R, Calle EE, Peters JM, Henley J, Hannan L, Thurston GD, et al. Ambient air pollution and brain cancer mortality. Cancer Causes & Control. 2009;20(9):1645–51.
15. Beelen R, Hoek G, Houthuijs D, Van Den Brandt PA, Goldbohm RA, Fischer P, et al. The joint association of air pollution and noise from road traffic with cardiovascular mortality in a cohort study. Occupational and Environmental Medicine. 2009;66(4):243–50.
16. Schikowski T, Sugiri D, Ranft U, Gehring U, Heinrich J, Wichmann HE, et al. Does respiratory health contribute to the effects of long-term air pollution exposure on cardiovascular mortality? Respiratory research. 2007;8:20.
17. Chen LH, Knutsen SF, Shavlik D, Beeson WL, Petersen F, Ghamsary M, et al. The association between fatal coronary heart disease and ambient particulate air pollution: Are females at greater risk? Environmental Health Perspectives. 2005;113(12):1723–9.
18. Goss CH, Newsom SA, Schildcrout JS, Sheppard L, Kaufman JD. Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. American Journal of Respiratory and Critical Care Medicine. 2004;169(7):816–21.
19. Lipfert FW, Perry Jr HM, Miller JP, Baty JD, Wyzga RE, Carmody SE. Air pollution, blood pressure, and their long-term associations with mortality. Inhalation Toxicology. 2003;15(5):493–512.
20. McDonnell WF, Nishino-Ishikawa N, Petersen FF, Chen LH, Abbey DE. Relationships of mortality with the fine and coarse fractions of long-term ambient PM10 concentrations in nonsmokers. Journal of Exposure Analysis and Environmental Epidemiology. 2000;10(5):427–36.
21. Mills PK, Abbey D, Beeson WL, Petersen F. Ambient air pollution and cancer in California Seventh-day Adventists. Archives of Environmental Health. 1991;46(5):271–80.
22. Abbey DE, Mills PK, Petersen FF, Beeson WL. Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California Seventh-Day Adventists. Environmental Health Perspectives. 1991;94:43–50.
23. Dong GH, Zhang PF, Sun BJ, Zhang LW, Chen X, Ma NN, et al. Long-Term Exposure to Ambient Air Pollution and Respiratory Disease Mortality in Shenyang, China: A 12-Year Population-Based Retrospective Cohort Study. Respiration. 2012;84(5):360–8.

Pollution assignment flawed

1. Kim H, Kim J, Kim S, Kang SH, Kim HJ, Kim H, et al. Cardiovascular Effects of Long-Term Exposure to Air Pollution: A Population-Based Study With 900 845 Person-Years of Follow-up. Journal of the American Heart Association. 2017;6(11).
2. Eckel SP, Cockburn M, Shu YH, Deng H, Lurmann FW, Liu L, et al. Air pollution affects lung cancer survival. Thorax. 2016;71(10):891–8.

Table B4
New studies published after last search.

Exposure	Study	Country	Cohort	Study population	Follow up period	Outcome	HR (95%CI)
NO_2	Hanigan 2019	Australia	45 and Up	General population	2007–2015	All cause	1.03 (0.98, 1.07)
	Hvidtfeldt 2019	Denmark	Danish Diet, Cancer and Health	General population	1993–2015	All cause	1.05 (1.01, 1.09)
						Respiratory	1.03 (0.97, 1.09)
	Klompmaker 2020	Netherlands	Dutch National Health Survey	General population	2013–2017	All cause	0.99 (0.97, 1.01)
						Respiratory	0.98 (0.91, 1.06)
	Lim 2019	US	NIH-AARP	General population	1995–2011	All cause	1.04 (1.02, 1.05)
						Respiratory	1.04 (0.99, 1.08)
						COPD	1.03 (0.98, 1.08)
						ALRI	1.22 (1.11, 1.35)
O_3(Annual)	Hvidtfeldt 2019	Denmark	Danish Diet, Cancer and Health	General population	1993–2015	All cause	0.95 (0.91, 1.00)
						Respiratory	0.97 (0.89, 1.05)
	Lim 2019	US	NIH-AARP	General population	1995–2011	All cause	0.99 (0.98, 1.00)
						Respiratory	1.04 (1.00, 1.09)
						COPD	1.09 (1.03, 1.15)
						ALRI	1.00 (0.90, 1.11)
O_3(Warm season)	Lim 2019	US	NIH-AARP	General population	1995–2011	All cause	1.00 (0.99, 1.01)
						Respiratory	1.04 (1.02, 1.06)
						COPD	1.05 (1.02, 1.08)
						ALRI	1.05 (0.99, 1.10)
	Kazemiparkouhi 2019	US	Medicare beneficiaries	General population	2000–2008	All cause	1.013 (1.012, 1.014)
						Respiratory	1.036 (1.032, 1.039)
						COPD	1.065 (1.060, 1.069)

Figs. B1–B13.

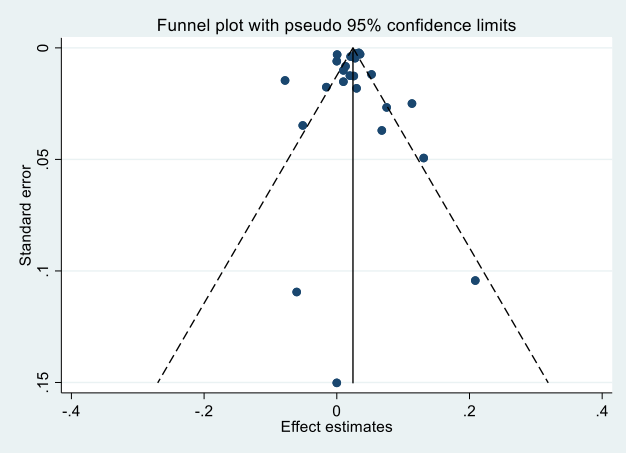

Fig. B1. NO_2 and all-cause mortality – funnel plot.

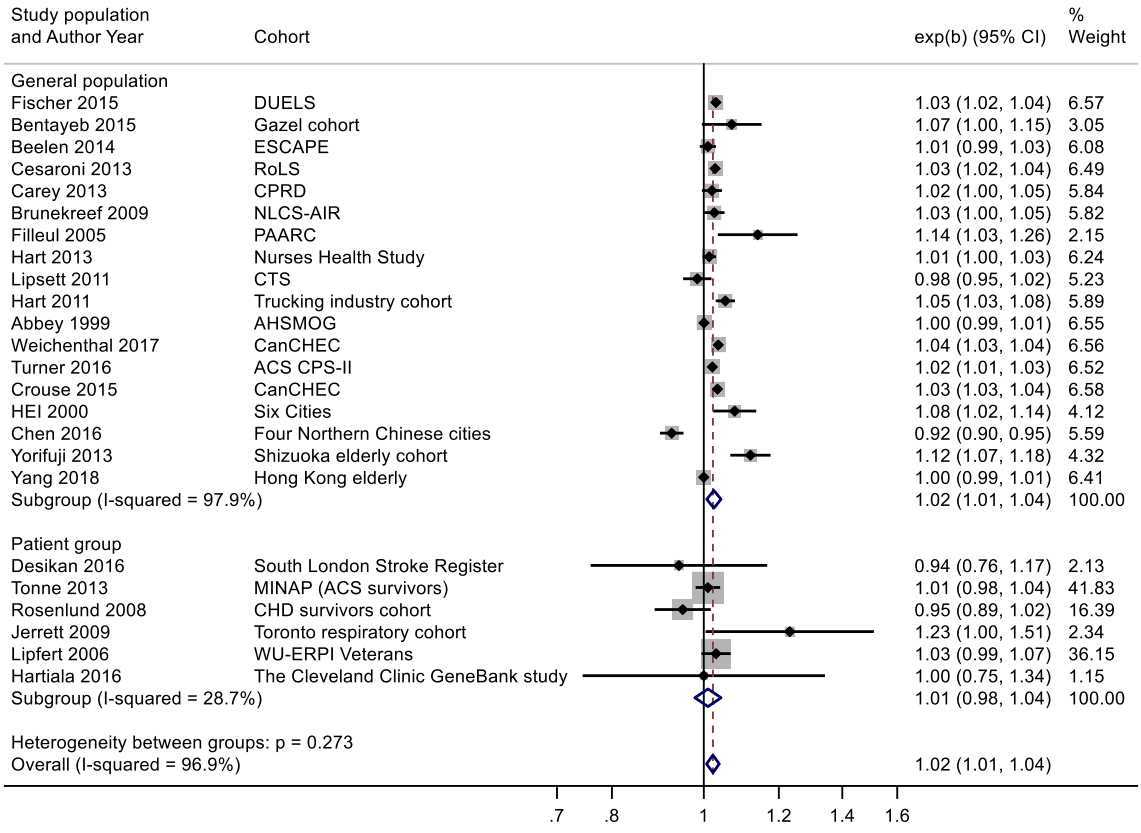

Fig. B2. NO_2 and all-cause mortality – stratification by patient and population groups.

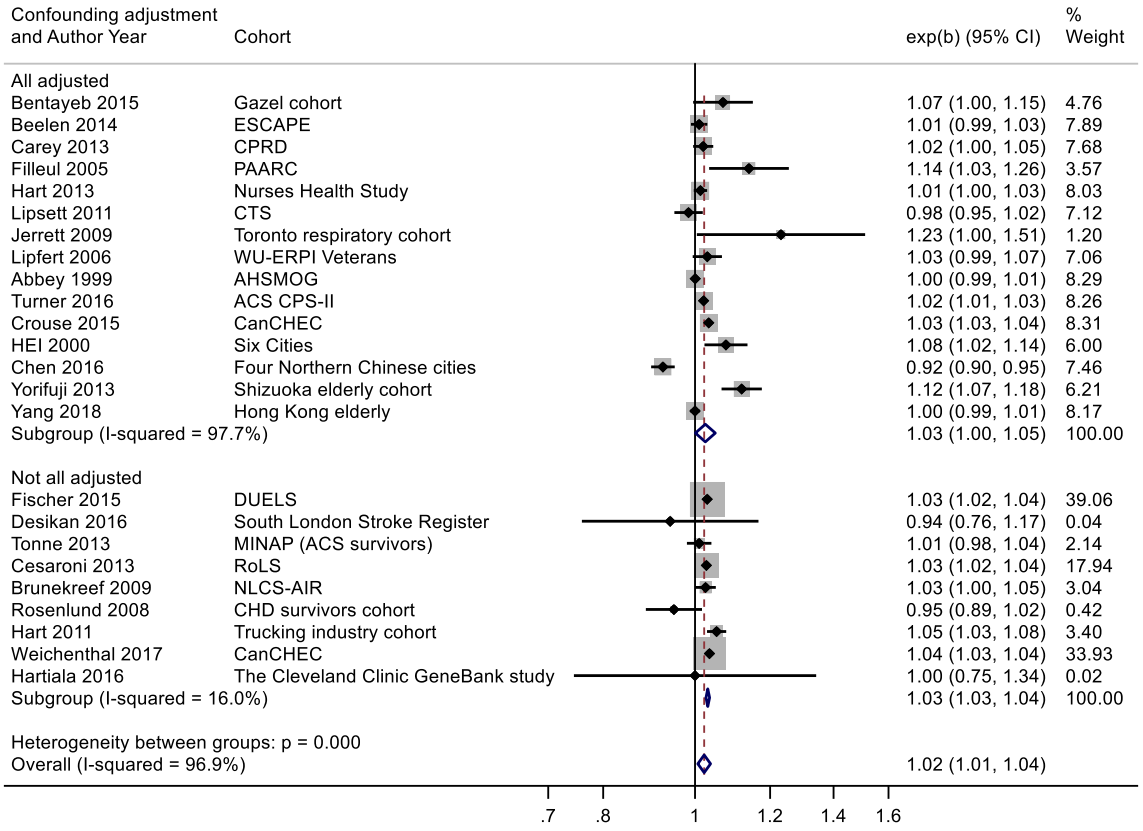

Fig. B3. NO_2 and all-cause mortality – stratification by individual vs area-level confounder control.

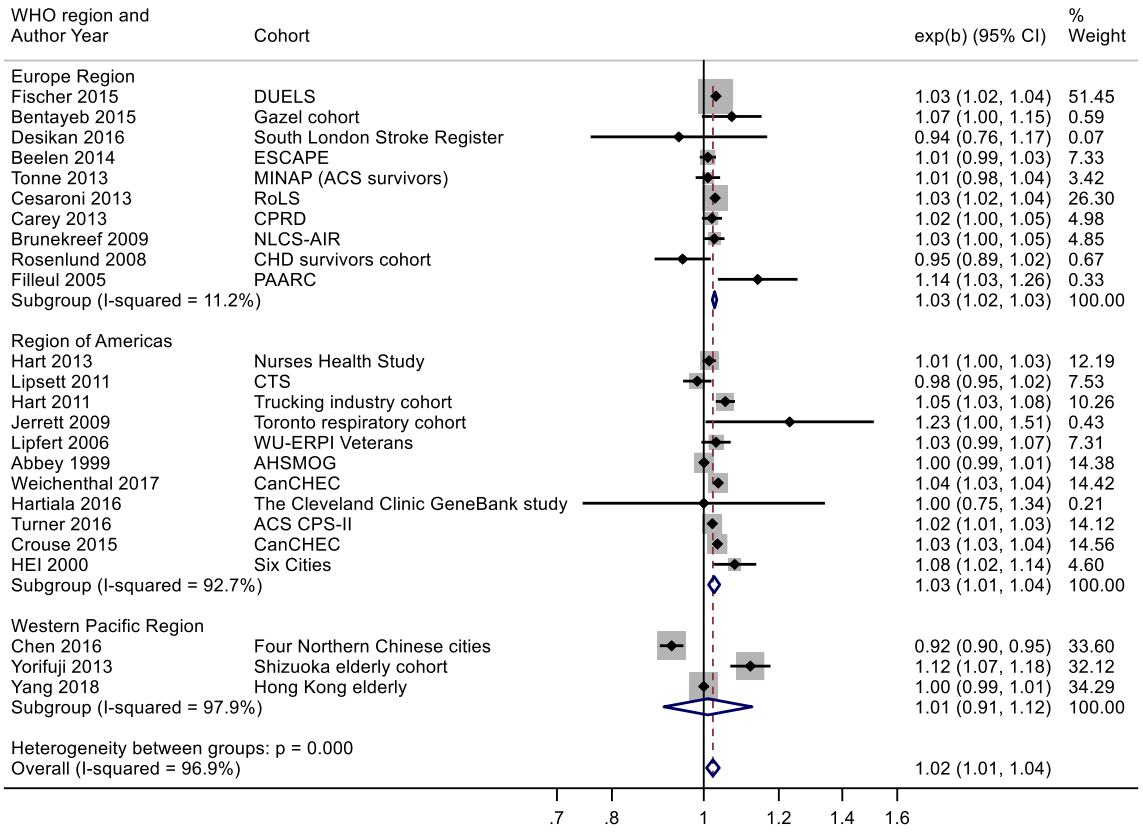

Fig. B4. NO_2 and all-cause mortality – stratification by WHO region.

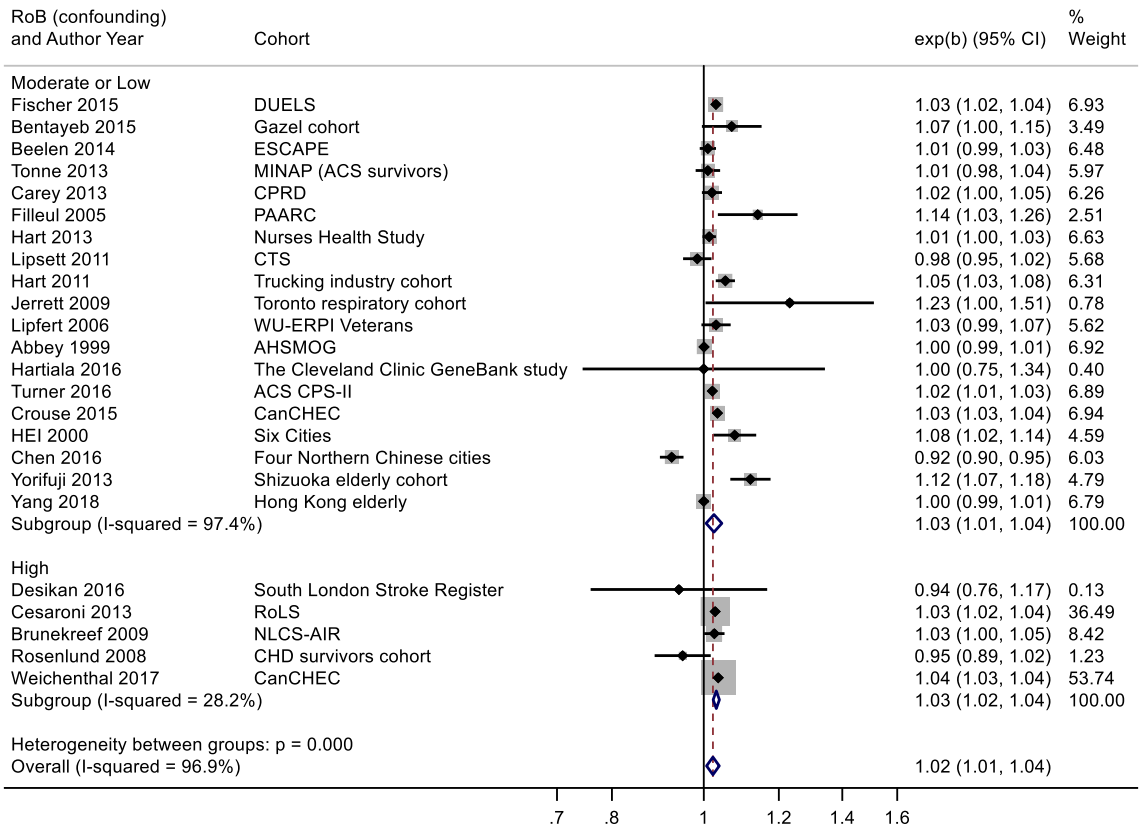

Fig. B5. NO_2 and all-cause mortality – stratification by risk of bias for confounding domain.

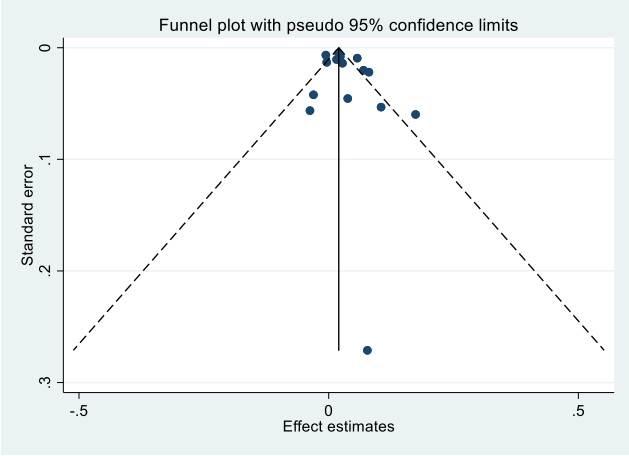

Fig. B6. NO_2 and respiratory mortality – funnel plot.

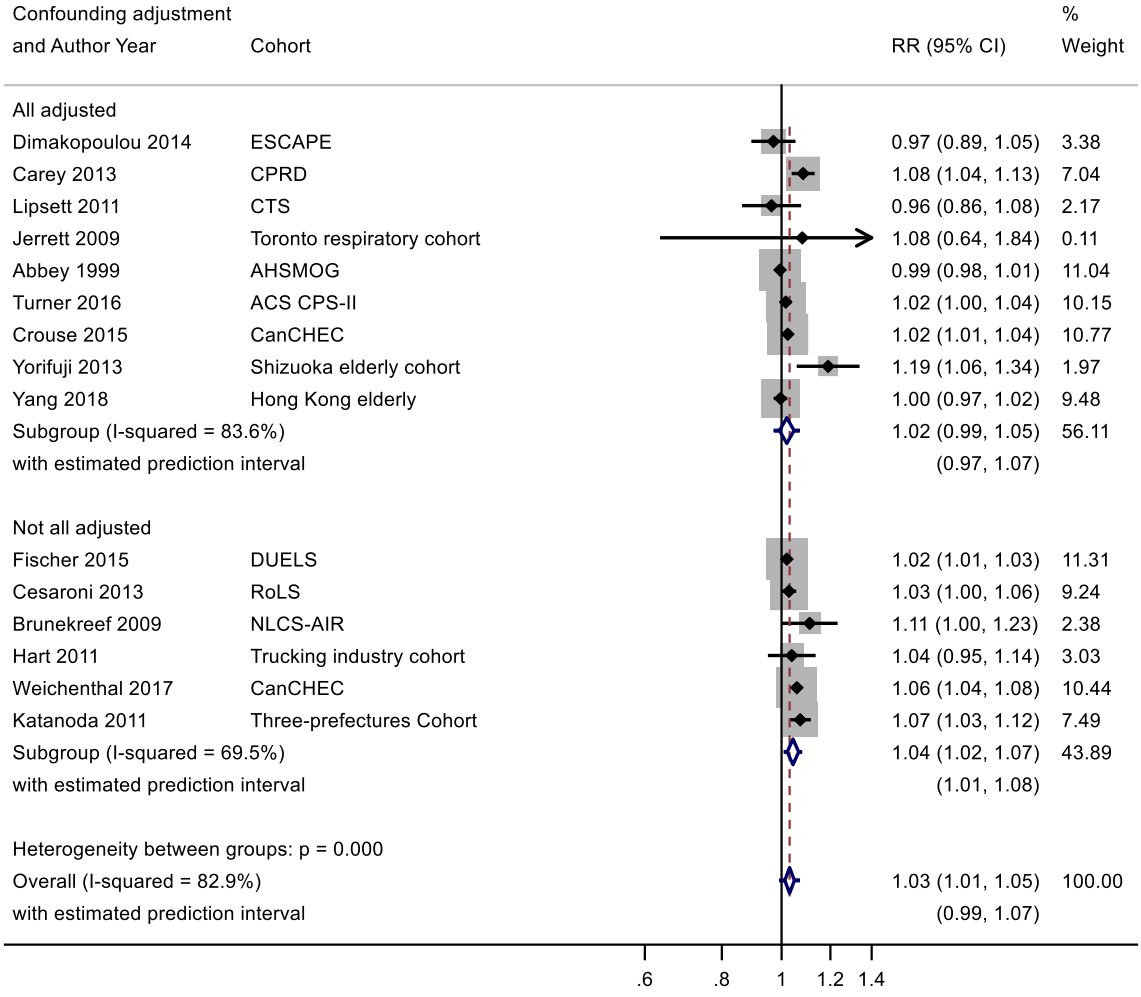

Fig. B7. NO_2 and respiratory mortality, stratification by confounder control.

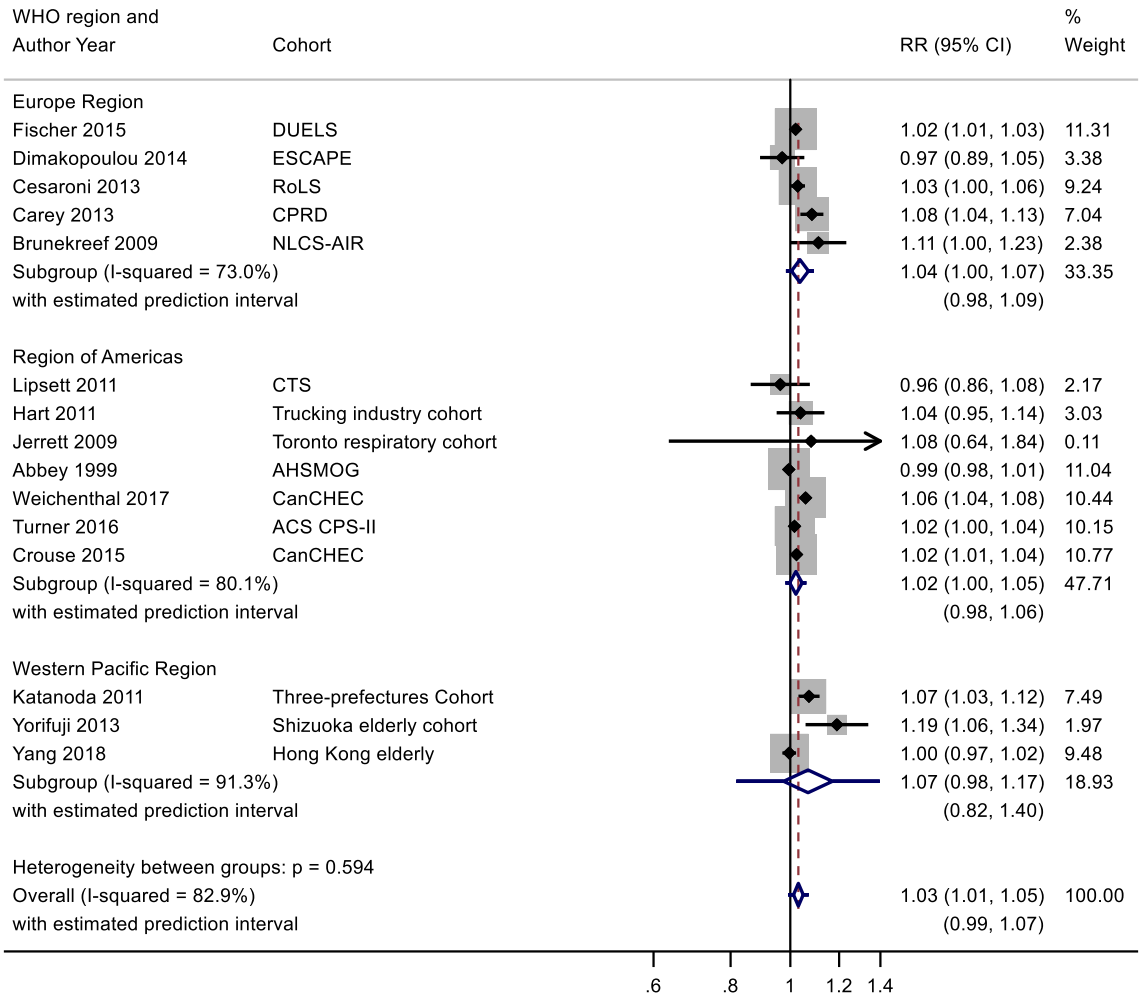

Fig. B8. NO_2 and respiratory mortality, stratification by WHO region.

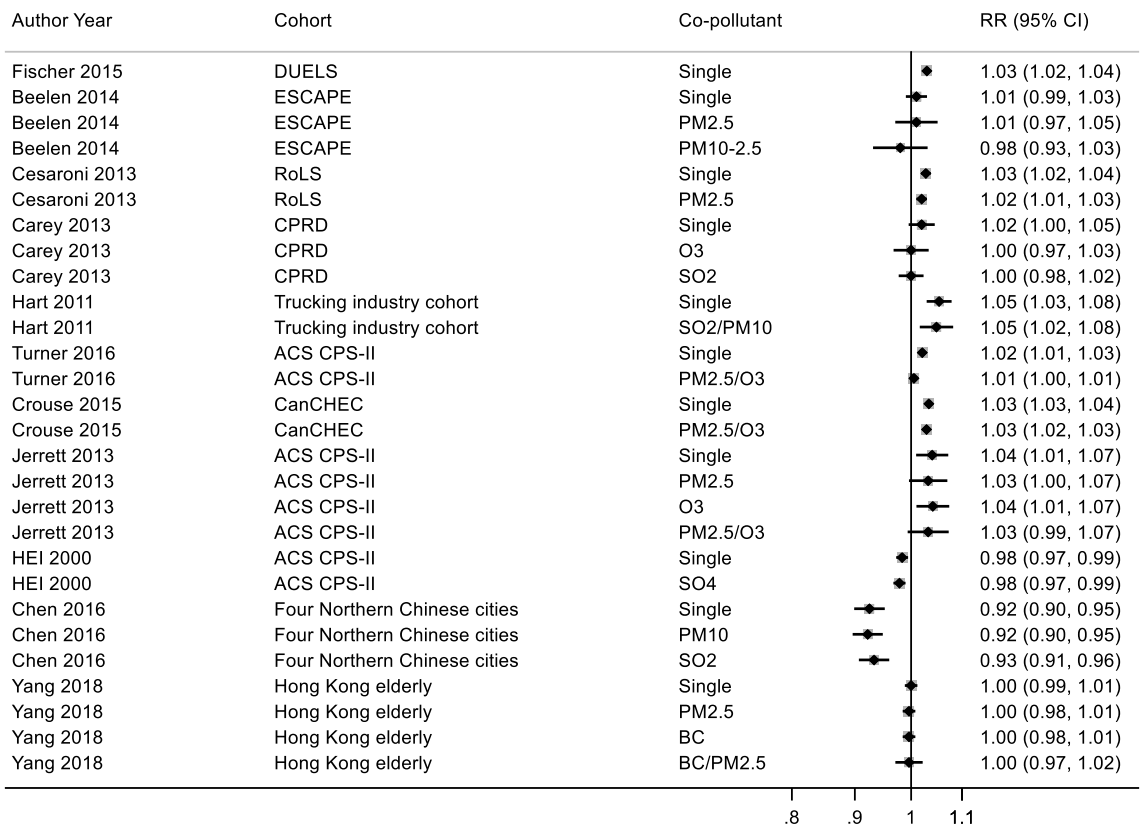

Fig. B9. NO_2 and all-cause mortality – multi-pollutant models.

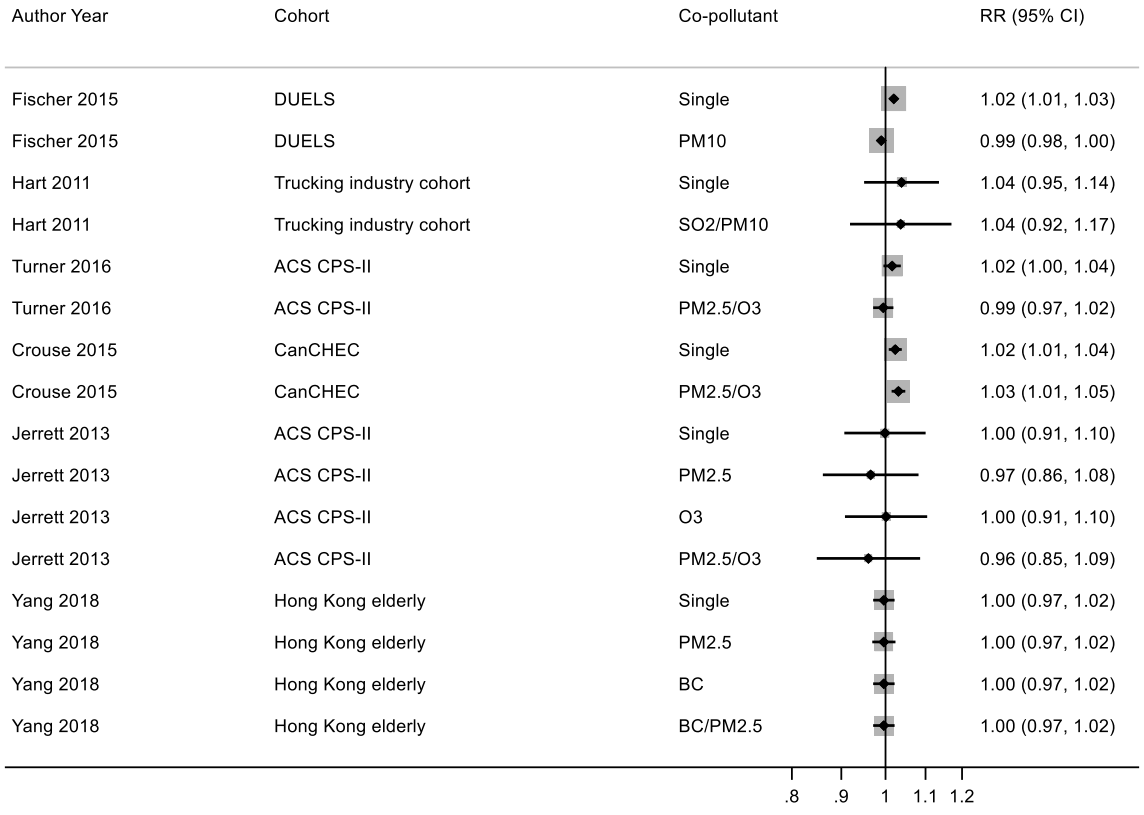

Fig. B10. NO_2 and respiratory mortality – multi-pollutant models.

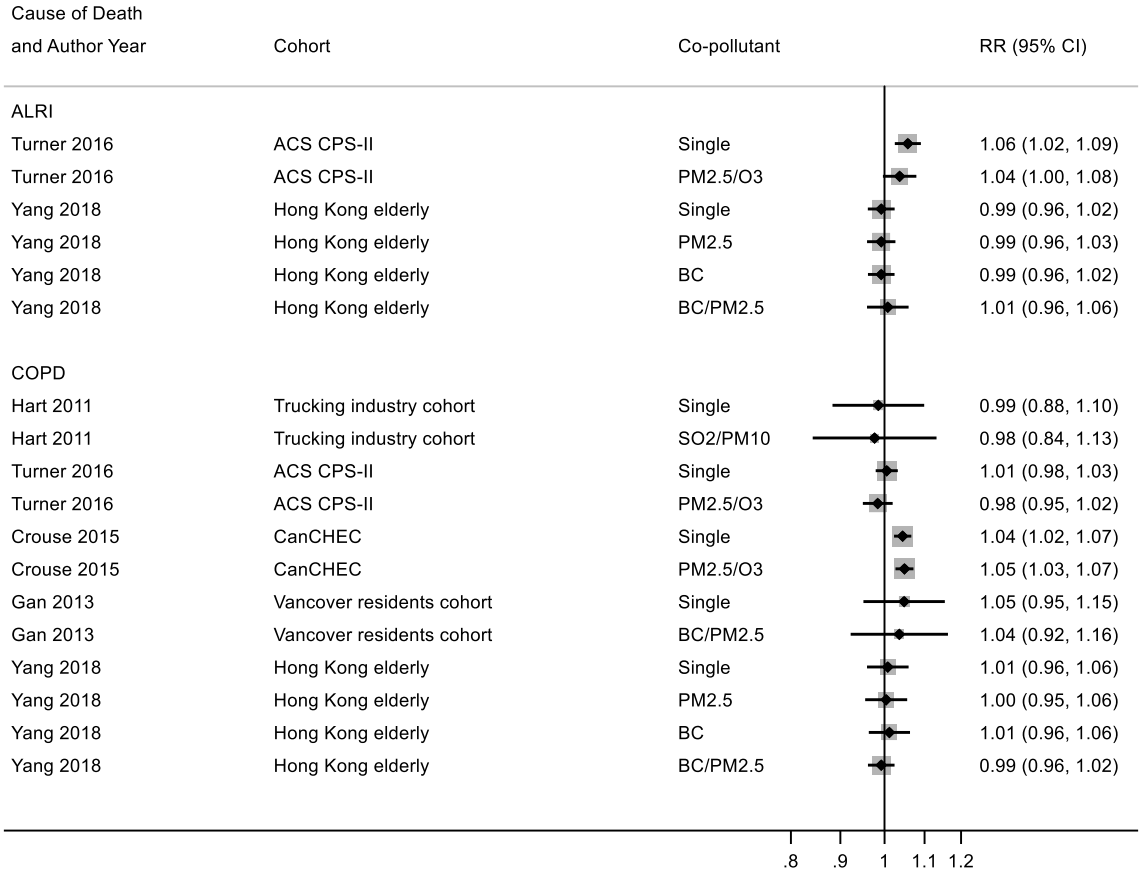

Fig. B11. NO_2 and COPD & ALRI mortality – multi-pollutant models.

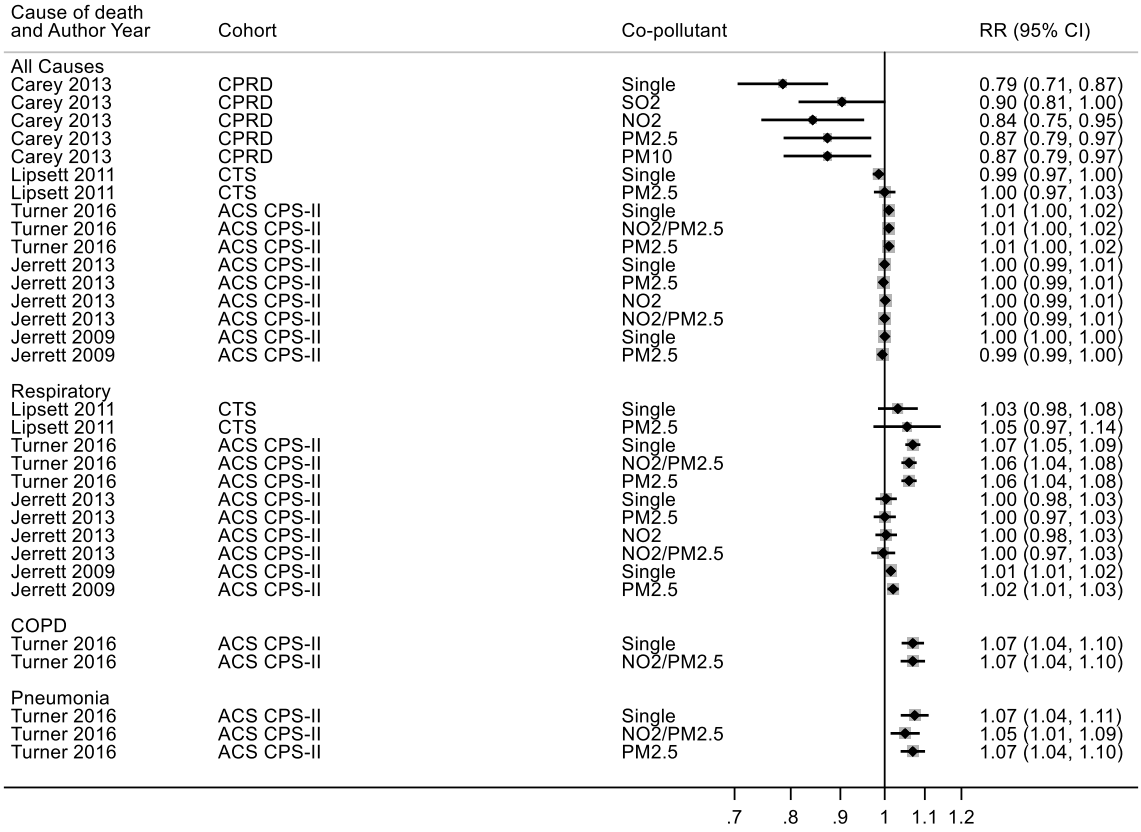

Fig. B12. O_3 annual exposure and mortality – multi-pollutant models.

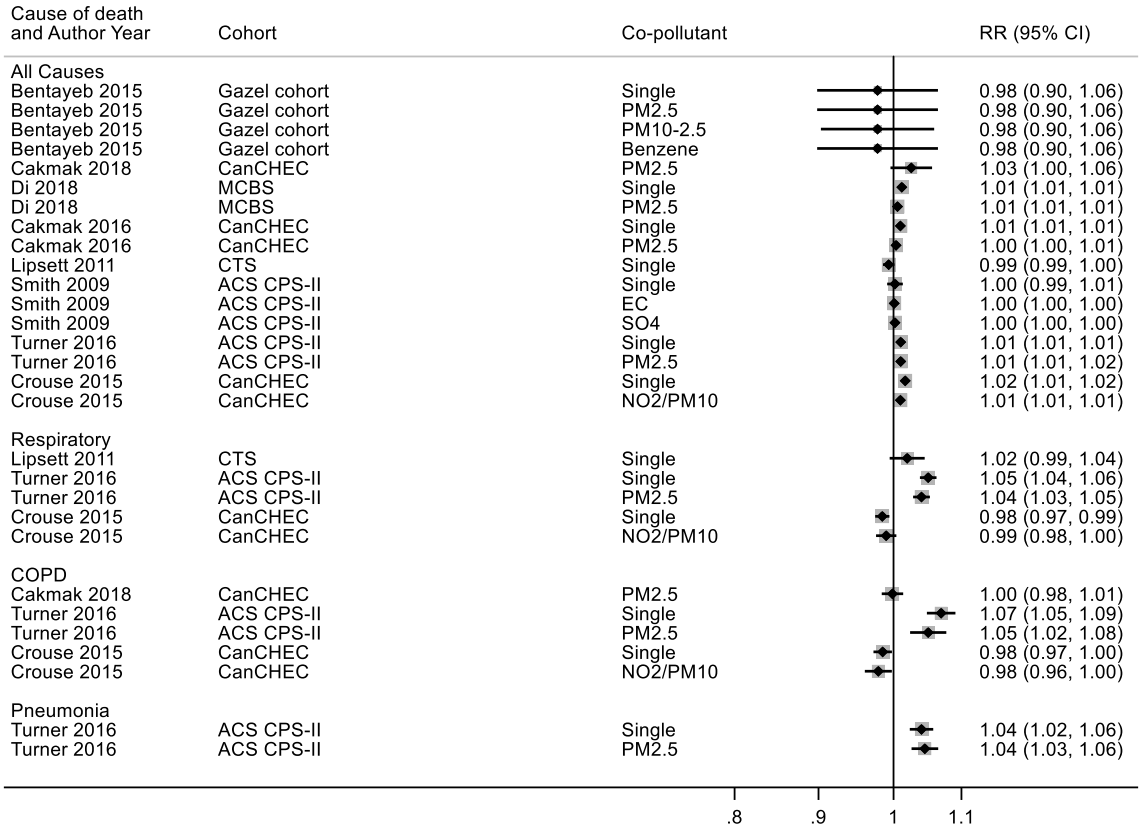

Fig. B13. O_3 peak exposure and mortality – multi-pollutant models.

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2020.105998.

References

Abbey, D.E., Nishino, N., McDonnell, W.F., Burchette, R.J., Knutsen, S.F., Beeson, W.L., et al., 1999. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. Am. J. Respir. Crit. Care Med. 159 (2), 373–382.

UK Air Information Resource DfEFRA; 2005.

Atkinson, R.W., Butland, B.K., Dimitroulopoulou, C., Heal, M.R., Stedman, J.R., Carslaw, N., et al., 2016. Long-term exposure to ambient ozone and mortality: a quantitative systematic review and meta-analysis of evidence from cohort studies. BMJ Open. 6 (2), e009493.

Atkinson, R.W., Butland, B.K., Anderson, H.R., Maynard, R.L., 2018. Long-term concentrations of nitrogen dioxide and mortality: a meta-analysis of cohort studies. Epidemiology 29 (4), 460–472.

Beelen, R., Raaschou-Nielsen, O., Stafoggia, M., Andersen, Z.J., Weinmayr, G., Hoffmann, B., et al., 2014. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. The Lancet. 383 (9919), 785–795.

Begg, C.B., Berlin, J.A., 1989. Publication bias and dissemination of clinical research. J. Natl. Can. Inst. 81 (2), 107–115.

Bentayeb, M., Wagner, V., Stempfelet, M., Zins, M., Goldberg, M., Pascal, M., et al., 2015. Association between long-term exposure to air pollution and mortality in France: A 25-year follow-up study. Environ. Int. 85, 5–14.

Borenstein, M., 2019. Common mistakes in meta-analysis and how to avoid them. Biostat, Inc, Englewood, NJ.

Borenstein, M., Higgins, J.P., Hedges, L.V., Rothstein, H.R., 2017. Basics of meta-analysis: I(2) is not an absolute measure of heterogeneity. Res. Synthesis Methods 8 (1), 5–18.

Borenstein, M., Higgins, J.P., Hedges, L.V., Rothstein, H.R., 2017. Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. Res Synth Methods. 8 (1), 5–18. https://doi.org/10.1002/jrsm.1230.

Brunekreef, B., Beelen, R., Hoek, G., Schouten, L., Bausch-Goldbohm, S., Fischer, P., et al. Effects of long-term exposure to traffic-related air pollution on respiratory and cardiovascular mortality in the Netherlands: the NLCS-AIR study. Research report (Health Effects Institute). 2009(139):5–71; discussion 3-7189.

Burnett, R., Chen, H., Szyszkowicz, M., Fann, N., Hubbell, B., Pope 3rd, C.A., et al., 2018. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc. Natl. Acad. Sci. USA 115 (38), 9592–9597.

Cakmak, S., Hebbern, C., Vanos, J., Crouse, D.L., Burnett, R., 2016. Ozone exposure and cardiovascular-related mortality in the Canadian Census Health and Environment Cohort (CANCHEC) by spatial synoptic classification zone. Environ. Pollut. 214, 589–599.

Cakmak, S., Hebbern, C., Pinault, L., Lavigne, E., Vanos, J., Crouse, D.L., et al., 2018. Associations between long-term $PM_{2.5}$ and ozone exposure and mortality in the Canadian Census Health and Environment Cohort (CANCHEC), by spatial synoptic classification zone. Environ. Int. 111, 200–211.

Carey, I.M., Atkinson, R.W., Kent, A.J., Van Staa, T., Cook, D.G., Anderson, H.R., 2013. Mortality associations with long-term exposure to outdoor air pollution in a national English cohort. Am. J. Respir. Crit. Care Med. 187 (11), 1226–1233.

Cesaroni, G., Badaloni, C., Gariazzo, C., Stafoggia, M., Sozzi, R., Davoli, M., et al., 2013. Long-term exposure to urban air pollution and mortality in a cohort of more than a million adults in Rome. Environ. Health Perspect. 121 (3), 324–331.

Cesaroni, G., Porta, D., Badaloni, C., Stafoggia, M., Eeftens, M., Meliefste, K., et al., 2012. Nitrogen dioxide levels estimated from land use regression models several years apart and association with mortality in a large cohort study. Environmental Health: A Global Access Science Source. 2012;11 (1) (no pagination)(48).

Chen, X., Zhang, L.W., Huang, J.J., Song, F.J., Zhang, L.P., Qian, Z.M., et al., 2016. Long-term exposure to urban air pollution and lung cancer mortality: a 12-year cohort study in Northern China. Sci. Total Environ. 571, 855–861.

Crouse, D.L., Peters, P.A., Hystad, P., Brook, J.R., van Donkelaar, A., Martin, R.V., et al., 2015a. Ambient PM2.5, O-3, and NO2 exposures and associations with mortality over 16 years of follow-up in the Canadian census health and environment COhort (CanCHEC). Environ. Health Perspect. 123 (11), 1180–1186.

Crouse, D.L., Peters, P.A., Villeneuve, P.J., Proux, M.O., Shin, H.H., Goldberg, M.S., et al., 2015b. Within- and between-city contrasts in nitrogen dioxide and mortality in 10 Canadian cities; a subset of the Canadian Census Health and Environment Cohort (CanCHEC). J. Eposure Sci. Environ. Epidemiol. 25 (5), 482–489.

Desikan A, Crichton S, Hoang U, Barratt B, Kelly FJ, Wolfe C. The effect of exhaust and non-exhaust related components of particulate matter on long-term survival after stroke in South London. International Journal of Stroke. 2015;5):61-2.

Di, Q., Zanobetti, A., Wang, Y., Koutrakis, P., Choirat, C., Dominici, F., et al., 2017. Air pollution and mortality in the medicare population. N. Engl. J. Med. 376 (26), 2513–2522.

Dimakopoulou, K., Samoli, E., Beelen, R., Stafoggia, M., Andersen, Z.J., Hoffmann, B., et al., 2014. Air pollution and nonmalignant respiratory mortality in 16 cohorts

within the ESCAPE project. Am. J. Respir. Crit. Care Med. 189 (6), 684–696.
Dominici, F., Peng, R.D., Barr, C.D., Bell, M.L., 2010. Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. Epidemiology 21 (2), 187–194.
Egger, M., Davey Smith, G., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. BMJ 315 (7109), 629–634.
Egger, M., Schneider, M., Davey, Smith G., 1998. Spurious precision? Meta-analysis of observational studies. BMJ. 316 (7125), 140–144.
EPA. US. Integrated Science Assessment for Oxides of Nitrogen – Health Criteria (2016 Final Report). 2016 Contract No.: EPA/600/R-15/068.
Faustini, A., Rapp, R., Forastiere, F., 2014. Nitrogen dioxide and mortality: review and meta-analysis of long-term studies. Eur. Respir. J. 44 (3), 744–753.
Filleul, L., Rondeau, V., Vandentorren, S., Le Moual, N., Cantagrel, A., Annesi-Maesano, I., et al., 2005. Twenty five year mortality and air pollution: results from the French PAARC survey. Occup. Environ. Med. 62 (7), 453–460.
Fischer, P.H., Marra, M., Ameling, C.B., Hoek, G., Beelen, R., De Hoogh, K., et al., 2015. Air pollution and mortality in seven million adults: The dutch environmental longitudinal study (DUELS). Environ. Health Perspect. 123 (7), 697–704.
Gan, W.Q., FitzGerald, J.M., Carlsten, C., Sadatsafavi, M., Brauer, M., 2013. Associations of ambient air pollution with chronic obstructive pulmonary disease hospitalization and mortality. Am. J. Respir. Crit. Care Med. 187 (7), 721–727.
Gehring, U., Heinrich, J., Kramer, U., Grote, V., Hochadel, M., Sugiri, D., et al., 2006. Long-term exposure to ambient air pollution and cardiopulmonary mortality in women. Epidemiology. 17 (5), 545–551.
Greenbaum, D., Shaikh, R., 2010. First steps toward multipollutant science for air quality decisions. Epidemiology. 21 (2), 195–197.
Guyatt, G.H., Oxman, A.D., Sultan, S., Glasziou, P., Akl, E.A., Alonso-Coello, P., et al., 2011. GRADE guidelines: 9. Rating up the quality of evidence. J. Clin. Epidemiol. 64 (12), 1311–1316. https://doi.org/10.1016/j.jclinepi.2011.06.004.
Guyatt, G.H., Oxman, A.D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., et al., 2011. GRADE guidelines: 8. Rating the quality of evidence — indirectness. J. Clin. Epidemiol. 64 (12), 1303–1310. https://doi.org/10.1016/j.jclinepi.2011.04.014.
Haneuse, S., VanderWeele, T.J., Arterburn, D., 2019. Using the E-value to assess the potential effect of unmeasured confounding in observational studies. JAMA 321 (6), 602–603. https://doi.org/10.1001/jama.2018.21554.
Hanigan, I.C., Rolfe, M.I., Knibbs, L.D., Salimi, F., Cowie, C.T., Heyworth, J., et al., 2019. All-cause mortality and long-term exposure to low level air pollution in the '45 and up study' cohort, Sydney, Australia, 2006–2015. Environ. Int. 126, 762–770.
Hart, J.E., Garshick, E., Dockery, D.W., Smith, T.J., Ryan, L., Laden, F., 2011. Long-term ambient multipollutant exposures and mortality. Am. J. Respir. Crit. Care Med. 183 (1), 73–78.
Hart, J.E., Rimm, E.B., Rexrode, K.M., Laden, F., 2013. Changes in traffic exposure and the risk of incident myocardial infarction and all-cause mortality. Epidemiology. 24 (5), 734–742.
Hartiala, J., Breton, C.V., Tang, W.H.W., Lurmann, F., Hazen, S.L., Gilliland, F.D., et al., 2016. Ambient air pollution is associated with the severity of coronary atherosclerosis and incident myocardial infarction in patients undergoing elective cardiac evaluation. J. Am. Heart Assoc. 5 (8).
Health Effects Institute. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project. Health Effects Institute, 2000.
Heinrich, J., Thiering, E., Rzehak, P., Kramer, U., Hochadel, M., Rauchfuss, K.M., et al., 2013. Long-term exposure to NO2 and PM10 and all-cause and cause-specific mortality in a prospective cohort of women. Occup. Environ. Med. 70 (3), 179–186.
Higgins, J.P.T.G.S., 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration.
Hoek, G., Brunekreef, B., Goldbohm, S., Fischer, P., Van Den Brandt, P.A., 2002. Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study. Lancet 360 (9341), 1203–1209.
Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., et al., 2013. Long-term air pollution exposure and cardio- respiratory mortality: a review. Environ. Health. 12 (1), 43.
Hultcrantz, M., Rind, D., Akl, E.A., Treweek, S., Mustafa, R.A., Iorio, A., et al., 2017. The GRADE Working Group clarifies the construct of certainty of evidence. J. Clin. Epidemiol. 87, 4–13. https://doi.org/10.1016/j.jclinepi.2017.05.006.
Hvidtfeldt, U.A., Sorensen, M., Geels, C., Ketzel, M., Khan, J., Tjonneland, A., et al., 2019. Long-term residential exposure to PM2.5, PM10, black carbon, NO2, and ozone and mortality in a Danish cohort. Environ. Int. 123, 265–272.
IntHout, J., Ioannidis, J.P., Rovers, M.M., Goeman, J.J., 2016. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open. 6 (7), e010247.
Ioannidis, J.P.A., Tan, Y.J., Blum, M.R., 2019. Limitations and misinterpretations of E-values for sensitivity analyses of observational studies. Ann Intern Med. 170 (2), 108–111.
Jerrett, M., Burnett, R.T., Pope, C.A., Ito, K., Thurston, G., Krewski, D., et al., 2009. Long-Term Ozone Exposure and Mortality. N. Engl. J. Med. 360 (11), 1085–1095.
Jerrett, M., Finkelstein, M.M., Brook, J.R., Arain, M.A., Kanaroglou, P., Stieb, D.M., et al., 2009. A cohort study of traffic-related air pollution and mortality in Toronto, Ontario, Canada. Environ. Health Perspect. 117 (5), 772–777.
Jerrett, M., Burnett, R.T., Beckerman, B.S., Turner, M.C., Krewski, D., Thurston, G., et al., 2013. Spatial analysis of air pollution and mortality in California. Am. J. Respir. Crit. Care Med. 188 (5), 593–599.
Katanoda, K., Sobue, T., Satoh, H., Tajima, K., Suzuki, T., Nakatsuka, H., et al., 2011. An association between long-term exposure to ambient air pollution and mortality from lung cancer and respiratory diseases in Japan. J. Epidemiol. 21 (2), 132–143.
Kazemiparkouhi, F., Eum, K.D., Wang, B., Manjourides, J., Suh, H.H., 2019. Long-term ozone exposures and cause-specific mortality in a US Medicare cohort. J. Eposure Sci. Environ. Epidemiol.
Klompmaker, J.O., Hoek, G., Bloemsma, L.D., Marra, M., Wijga, A.H., van den Brink, C., et al., 2020. Surrounding green, air pollution, traffic noise exposure and non-accidental and cause-specific mortality. Environ. Int. 134, 105341.
Krewski, D., Jerrett, M., Burnett, R.T., Ma, R., Hughes, E., Shi, Y., et al., 2009(140):5–114;. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. Research report (Health Effects Institute) discussion 5–36.
Lim, C.C., Hayes, R.B., Ahn, J., Shao, Y., Silverman, D.T., Jones, R.R., et al., 2019. Long-term exposure to ozone and cause-specific mortality risk in the United States. Am. J. Respir. Crit. Care Med. 200 (8), 1022–1031.
Lipfert, F.W., Wyzga, R.E., Baty, J.D., Miller, J.P., 2006. Traffic density as a surrogate measure of environmental exposures in studies of air pollution health effects: Long-term mortality in a cohort of US veterans. Atmos. Environ. 40 (1), 154–169.
Lipfert, F.W., Baty, J.D., Miller, J.P., Wyzga, R.E., 2006. PM2.5 constituents and related air quality variables as predictors of survival in a cohort of U. S. military veterans. Inhalation Toxicol. 18 (9), 645–657.
Lipsett, M.J., Ostro, B.D., Reynolds, P., Goldberg, D., Hertz, A., Jerrett, M., et al., 2011. Long-term exposure to air pollution and cardiorespiratory disease in the California teachers study cohort. Am. J. Respir. Crit. Care Med. 184 (7), 828–835.
Maheswaran, R., Pearson, T., Smeeton, N.C., Beevers, S.D., Campbell, M.J., Wolfe, C.D., 2010. Impact of outdoor air pollution on survival after stroke: population-based cohort study. Stroke 41 (5), 869–877.
Malig, B.J., Pearson, D.L., Chang, Y.B., Broadwin, R., Basu, R., Green, R.S., et al., 2016. A time-stratified case-crossover study of ambient ozone exposure and emergency department visits for specific respiratory diagnoses in California (2005–2008). Environ. Health Perspect. 124 (6), 745–753.
Moher D, Liberati A, Tetzlaff J, Altman DG, Group P, 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6 (7), e1000097.
Morgan, R.L., Thayer, K.A., Bero, L., Bruce, N., Falck-Ytter, Y., Ghersi, D., et al., 2016. GRADE: Assessing the quality of evidence in environmental and occupational health. Environ Int. 92–93, 611–616. https://doi.org/10.1016/j.envint.2016.01.004.
Morgan, R.L., Thayer, K.A., Santesso, N., Holloway, A.C., Blain, R., Eftim, S.E., et al., 2019. A risk of bias instrument for non-randomized studies of exposures: a users' guide to its application in the context of GRADE. Environ. Int. 122, 168–184.
Naess, O., Nafstad, P., Aamodt, G., Claussen, B., Rosland, P., 2007. Relation between concentration of air pollution and cause-specific mortality: four-year exposures to nitrogen dioxide and particulate matter pollutants in 470 neighborhoods in Oslo Norway. Am. J. Epidemiol. 165 (4), 435–443.
Nuvolone, D., Petri, D., Voller, F., 2018. The effects of ozone on human health. Environ. Sci. Pollut. Res. Int. 25 (9), 8074–8088.
Ostro, B., Lipsett, M., Reynolds, P., Goldberg, D., Hertz, A., Garcia, C., et al., 2010. Long-term exposure to constituents of fine particulate air pollution and mortality: results from the California Teachers Study. Environ. Health Perspect. 118 (3), 363–369. https://doi.org/10.1289/ehp.0901181.
Peters, J.L., Sutton, A.J., Jones, D.R., Abrams, K.R., Rushton, L., Moreno, S.G., 2010. Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. J. Roy. Statist. Soc.: Series A (Statistics in Society). 173 (3), 575–591.
Raaschou-Nielsen O, Andersen ZJ, Jensen SS, Ketzel M, Sorensen M, Hansen J, et al. Traffic air pollution and mortality from cardiovascular disease and all causes: A Danish cohort study. Environmental Health: A Global Access Science Source. 2012;11 (1) (no pagination)(60).
Rosenlund, M., Picciotto, S., Forastiere, F., Stafoggia, M., Perucci, C.A., 2008. Traffic-related air pollution in relation to incidence and prognosis of coronary heart disease. Epidemiology. 19 (1), 121–128.
Rothman, K.J., Greenland, S., 2018. Planning study size based on precision rather than power. Epidemiology. 29 (5), 599–603. https://doi.org/10.1097/EDE.0000000000000876.
Rush, B., McDermid, R.C., Celi, L.A., Walley, K.R., Russell, J.A., Boyd, J.H., 2017. Association between chronic exposure to air pollution and mortality in the acute respiratory distress syndrome. Environ. Pollut. 224, 352–356.
Schunemann H, Brożek J, Guyatt G, Oxman AD, editors. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. Hamilton, ON: The GRADE Working Group; 2013 (https://gdt.gradepro.org/app/handbook/handbook.html, accessed 21 February 2020).
Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ : Brit. Med. J. 349, g7647.
Smith, K.R., Jerrett, M., Anderson, H.R., Burnett, R.T., Stone, V., Derwent, R., et al., 2009. Public health benefits of strategies to reduce greenhouse-gas emissions: health implications of short-lived greenhouse pollutants. The Lancet. 374 (9707), 2091–2103.
StataCorp. Stata Statistical Software: Release 15 College Station, TX: StataCorp LLC; 2017.
Strickland, M.J., Darrow, L.A., Klein, M., Flanders, W.D., Sarnat, J.A., Waller, L.A., et al., 2010. Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. Am. J. Respir. Crit. Care Med. 182 (3), 307–316.
Thurston, G.D., Kipen, H., Annesi-Maesano, I., Balmes, J., Brook, R.D., Cromar, K., et al., 2017. A joint ERS/ATS policy statement: what constitutes an adverse health effect of air pollution? An analytical framework. Eur. Respir. J. 49 (1).
Tonne, C., Wilkinson, P., 2013. Long-term exposure to air pollution is associated with survival following acute coronary syndrome. Eur. Heart J. 34 (17), 1306–1311.
Tonne, C., Halonen, J.I., Beevers, S.D., Dajnak, D., Gulliver, J., Kelly, F.J., et al., 2016. Long-term traffic air and noise pollution in relation to mortality and hospital readmission among myocardial infarction survivors. Int. J. Hyg. Environ. Health 219 (1),

72–78.
Turner, M.C., Jerrett, M., Pope, C.A., Krewski, D., Gapstur, S.M., Diver, W.R., et al., 2016. Long-term ozone exposure and mortality in a large prospective study. Am. J. Respir. Crit. Care Med. 193 (10), 1134–1142.
Urman, R., McConnell, R., Islam, T., Avol, E.L., Lurmann, F.W., Vora, H., et al., 2014. Associations of children's lung function with ambient air pollution: joint effects of regional and near-roadway pollutants. Thorax 69 (6), 540–547.
EPA US. Integrated Science Assessment (ISA) of Ozone and Related Photochemical Oxidants (Final Report, Feb 2013). 2013 EPA/600/R-10/076F.
VanderWeele, T.J., Ding, P., 2017. Sensitivity analysis in observational research: introducing the E-value. Ann. Intern. Med. 167 (4), 268–274.
VanderWeele, T.J., Mathur, M.B., Ding, P., 2019. Correcting misinterpretations of the E-value. Ann. Int. Med. 170 (2), 131–132. https://doi.org/10.7326/M18-3112.
Veroniki, A.A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., et al., 2016. Methods to estimate the between-study variance and its uncertainty in meta-analysis. Res. Synth. Methods 7 (1), 55–79.
Weichenthal, S., Pinault, L.L., Burnett, R.T., 2017. Impact of oxidant gases on the relationship between outdoor fine particulate air pollution and nonaccidental, cardiovascular, and respiratory mortality. Scientific Reports 7.
WHO. Air quality guidelines for Europe. 2000.
WHO. Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. 2005.
WHO. Review of evidence on health aspects of air pollution – REVIHAAP Project. 2013.
WHO Global Air Quality Guidelines Working Group on Risk of Bias Assessment. Risk of bias assessment instrument for systematic reviews informing WHO global air quality guidelines. Copenhagen: WHO Regional Office for Europe; 2020 (http://www.euro.who.int/_data/assets/pdf_file/0007/425518/Risk-of-bias-assessment-instrument-for-systematic-reviews-informing-who-global-air-quality-guidelines.pdf?ua=1, accessed 21 February 2020).
Woodruff, T.J., Sutton, P., 2011. Navigation Guide Work Group. An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. Health Aff (Millwood). 30 (5), 931–937. https://doi.org/10.1377/hlthaff.2010.1219.
World Health Organization. Burden of desease from ambient air pollution for 2012. Summary of results. 2014.
World Health Organization. WHO Handbook for Guideline Development, 2nd edition. Geneva: World Health Organization; 2014 (https://www.who.int/publications/guidelines/handbook_2nd_ed.pdf?ua=1; accessed 21 february 2020).
Yang, Y., Tang, R., Qiu, H., Lai, P.C., Wong, P., Thach, T.Q., et al., 2018. Long term exposure to air pollution and mortality in an elderly cohort in Hong Kong. Environ. Int. 117, 99–106.
Yorifuji, T., Kashima, S., Tsuda, T., Takao, S., Suzuki, E., Doi, H., et al., 2010. Long-term exposure to traffic-related air pollution and mortality in Shizuoka Japan. Occupat. Environment. Med. 67 (2), 111–117.
Yorifuji, T., Kashima, S., Tsuda, T., Ishikawa-Takata, K., Ohta, T., Tsuruta, K., et al., 2013. Long-term exposure to traffic-related air pollution and the risk of death from hemorrhagic stroke and lung cancer in Shizuoka, Japan. Sci. Total Environ. 443, 397–402.