
Working Paper 1: Systematic review and meta-analysis of cohort studies of NO₂ and all-cause mortality

Richard W Atkinson & Barbara K Butland
AIM
To recommend a coefficient for long-term exposure to NO₂ and all-cause mortality for use in health impact assessment.

To achieve this aim the following objectives were set:

- to undertake a systematic literature search to identify cohort studies reporting hazard ratios (HRs) for long-term exposure to NO₂ and all-cause mortality
- to select studies for meta-analysis and derive a summary HR
- to identify studies reporting HRs for NO₂ and all-cause mortality adjusting for PM₂.₅
- to estimate the reduction in the HR for NO₂ after adjustment for PM₂.₅

METHODS
To identify publications reporting results for cohort studies of NO₂ and mortality we conducted a broad search of the online medical databases EMBASE and MEDLINE, supplemented with citation searches of recently published literature reviews and COMEAP papers.

Search strategy
Three search strings were applied to Ovid Medline (R) without Revisions for the period 1996 to October Week 1 2015 and to Embase for the period 1996 to 2015 Week 41. The search strings were: a) "cohort" & "no2" & "mortality"; b) "cohort" & "air pollution" & "mortality"; and c) "long-term" & "no2" & "mortality". These searches were supplemented by citation searches in 6 review articles.¹⁻⁶

Our search strategy excluded conference abstracts, conference papers, notes, editorials and letters. Cohort studies were selected if they included a ‘long-term’ exposure metric for NO₂—studies using daily or monthly exposures were excluded. Cohort studies were also required to have individual-level covariate information. Cross sectional, case control and nested case-control studies were excluded from the review. The outcome studied had to be all-cause or cause-specific mortality (not disease incidence).
Inclusion/exclusion criteria
Studies providing HRs for NOX were excluded from the review. Studies providing quantitative HRs (i.e. not graphically) together with either standard errors or 95% confidence intervals were selected. Adequate information had to be provided to allow presentation of such estimates as per 10 µg/m³ increase in pollutant. Where results for various follow-up periods were provided we chose the follow-up period that was most up to date.

If results for the same outcomes were available for the full cohort or a subset we used results from the full cohort unless these results were considered to be out of date (e.g. statistical analysis, exposure assessment, date of last follow-up). Two studies from the same cohort were only included if they provided results for different outcomes.

Data extraction
Cohort and estimate level information were extracted from each paper/online supplement. These data included cohort name, country, cohort description, date of enrolment of cohort members, age at enrolment, number of subjects, follow-up period, exposure period and exposure assessment method (measured/modelled).

All HRs were standardised to 10µg/m³ increase in NO₂. Where the units used in the original study were ppb, a conversion factor of 1.88µg/m³ per 1ppb was used (assuming 25ºC and 1013mb atmospheric pressure).

Quality criteria
Covariate adjustment should include individual-level age, sex, smoking and BMI. Also adjustment for some marker of socioeconomic status (e.g. education level, income etc.) at either the individual or ecological level.

Meta-analysis
All analyses were conducted in STATA Version 12. All studies reported HRs together with 95% confidence intervals. Therefore, estimates of the standard error were derived using each limit value in turn and the two estimates averaged. Forest plots were used to display study information and HRs graphically. Meta-analytic summary estimates were calculated using fixed/random effects models using the program ‘metan’ in STATA. Heterogeneity was assessed using the I² statistic. Small study bias was assessed using Begg7 and Egger8 tests and the Trim and Fill procedure9.
RESULTS

Literature search

996 records were identified from the database searches and other sources. After removal of duplicates and application of the inclusion/exclusion criteria, 66 articles were identified for full-text review. After exclusion of studies that reported results for NO\textsubscript{x} (n=4), replicated results reported elsewhere (n=6), or did not report HRs quantitatively (n=11), 45 articles remained, of which 20 did not adjust fully for all confounders.

Causes of ‘all’ deaths were variously described as ‘All cause’, ‘Natural causes’ and ‘Non Accidental’ and were re-coded as ‘All Cause’. Studies only of cause-specific deaths were not considered further. 28 articles analysing 21 cohorts (including the ESCAPE consortium of individual cohorts) reported results for all-cause mortality.

Cohorts comprising selected subgroups defined by pre-existing disease were excluded as being unrepresentative of the general population and therefore unsuitable for the purpose of a health impact assessment exercise. Consequently, 6 publications from 5 cohorts (stroke survivors\textsuperscript{10}, CHD survivors\textsuperscript{11}; attendees at respiratory clinic\textsuperscript{12}; ACS survivors\textsuperscript{13}; and hypertensive US veterans\textsuperscript{14,15}) were excluded.

Three studies\textsuperscript{16-18} were excluded as their results were included in ESCAPE meta-analysis\textsuperscript{19}. The ESCAPE study provided a meta-analytical result (of cohorts in the ESCAPE project) only and not individual cohort HRs.

A further 5 studies were excluded as the same cohorts were analysed in other publications included in the review.\textsuperscript{20-24} In one study\textsuperscript{25}, results for two cohorts were reported – the HR for the ACS CPS II cohort reported in this study was not used.

Meta-analysis

Following these exclusions, results from 14 separate cohorts (including the ESCAPE consortium of 22 individual cohorts) reported results for NO\textsubscript{2} and all-cause mortality.\textsuperscript{19,25-37} The majority of the 14 cohorts were in European populations (7 including the ESCAPE study); 6 cohorts were from North America and a single cohort from Japan. Key cohort characteristics and corresponding HRs are presented in Figure 1. There was substantial heterogeneity between effect estimates, I\textsuperscript{2}=96%.
Figure 1 HRs (95% CI) per 10μg/m$^3$ for cohort studies reporting associations between NO$_2$ and all-cause mortality

Analysis stratified by adults across a broad age range vs specific age groups is shown in Figure 2. Four studies focused on specific age groups. The summary HR for these studies was substantially larger than for studies with broader age ranges at recruitment. As the focus of our review was to derive a summary HR considered to be representative of the general population we excluded, from further analyses, the cohort studies (n=4) in adults restricted to narrow age ranges at cohort entry.¹

When the remaining cohorts (n=10) were stratified by level of covariate adjustment – i.e. those controlling for the required confounding factors and those that did not (Figure 3) there was a substantial difference in the NO$_2$ random-effects summary estimates; 1.008 (95% CI: 0.993, 1.024) vs 1.031 (95% CI: 1.025, 1.037) per 10μg/m$^3$ respectively.²

¹ Bentayeb et al 2015 does not report any age restriction on cohort members at recruitment. In the original meta-analysis conducted in 2015 this study was therefore coded as ‘adult’ rather than ‘restricted’. The meta-analytical summary estimate was then used in subsequent health impact calculations. On further investigation conducted in July 2017, a related paper was identified which indicated the age range of cohort participants in Bentayeb et al 2015 was restricted to ages 35-50. The coding for Bentayeb et al 2015 was changed therefore to reflect this new information, and the meta-analyses presented in this working paper use this coding.

² The recoding of Bentayeb et al 2015 did not alter materially this finding.
Figure 2 HRs (95% CI) for cohort studies reporting associations between NO₂ and all-cause mortality stratified by age groups

Figure 3 HRs (95% CI) per 10μg/m³ for cohort studies reporting associations between NO₂ and all-cause mortality stratified by covariate adjustment
Based upon the 10 studies presented in Figure 3, the fixed-effects summary HR for NO₂ was 1.010, (95% CI: 1.009, 1.012) per 10μg/m³. There was substantial heterogeneity between estimates, $I^2=97\%$. The corresponding random-effects summary HR was 1.021 (95% CI: 1.006, 1.036) per 10μg/m³.³

**Small study bias**

P-values for Begg’s and Egger’s tests were 0.37 and 0.32 respectively. Application of the trim and fill technique indicated the need to impute 3 additional study estimates (Figure 4) to adjust for small study bias assuming a fixed-random effects model, adjusted HR=1.010 (95% CI: 0.996, 1.023). Assuming a random-random effects model the Trim & Fill procedure did not indicate the need to impute additional estimates to achieve symmetry.

---

³ The incorrect coding of age range for Bentayeb et al. 2015 at the start of our review meant that the original meta-analyses was conducted on 11 studies. This analysis reported a fixed-effects summary HR of 1.010, (95% CI: 1.009, 1.012) per 10 μg/m³ with substantial heterogeneity between estimates, $I^2=97\%$. The corresponding random-effects summary HR was 1.023 (95% CI: 1.008, 1.037) per 10 μg/m³.
Independence from PM$_{2.5}$

Table 1 and Figure 5 show the HRs (95% CI) from single- and two-pollutant models for NO$_2$ and PM in the cohorts reporting results for all-cause mortality. Five studies adjusted for PM$_{2.5}$ and one for PM$_{10}$. Confidence intervals for unadjusted and adjusted HRs overlapped. None of the cohorts assessed the independence of NO$_2$ from traffic related particles (ultrafine, elemental carbon etc.).

The percentage reduction in the ln(HR) for NO$_2$ after adjustment for PM varied from 10% to 95% and in one study, a negative association between NO$_2$ and mortality reduced further upon adjustment for PM$_{2.5}$. For PM (PM$_{2.5}$ or PM$_{10}$) the reductions were between 0% and 82% and in one study the PM$_{2.5}$ coefficient increased by 42% upon adjustment for NO$_2$. Correlations between NO$_2$ and PM$_{2.5}$ (Table 1) were high in 2 studies (0.79 and 0.85), moderate in 2 studies (0.2-0.7 and 0.55) and weak (-0.08) in one study. In the single study using PM$_{10}$, rather than PM$_{2.5}$, the correlation with NO$_2$ was 0.58.
Table 1 Hazard ratios (HR) from single and two pollutant models for NO₂ and PM_{2.5} or PM_{10} (HRs are expressed per IQR)

<table>
<thead>
<tr>
<th>Study</th>
<th>Cohort</th>
<th>NO₂ IQR (μg/m³)</th>
<th>HR NO₂</th>
<th>NO₂ adj PM_{2.5}/PM_{10} IQR (μg/m³)</th>
<th>% ^5</th>
<th>HR PM_{2.5}/PM_{10} adj NO₂</th>
<th>HR PM_{2.5}/PM_{10} adj NO₂</th>
<th>Combined NO₂ adj/PM adj HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cesaroni et al</td>
<td>Rome 0.79</td>
<td>10.7</td>
<td>1.029</td>
<td>1.026 (1.015, 1.037)</td>
<td>10</td>
<td>1.023 (1.016, 1.031)</td>
<td>1.004 (0.994, 1.015)</td>
<td>82</td>
</tr>
<tr>
<td>Carey et al</td>
<td>CPRD 0.85</td>
<td>10.7</td>
<td>1.022</td>
<td>1.001 (0.959, 1.044)</td>
<td>95</td>
<td>1.023 (1.000, 1.046)</td>
<td>1.023 (0.989, 1.060)</td>
<td>0</td>
</tr>
<tr>
<td>Beelen et al</td>
<td>ESCAPE 0.2&lt;-0.7</td>
<td>10.0</td>
<td>1.015</td>
<td>1.007 (0.967, 1.049)</td>
<td>53</td>
<td>1.070 (1.016, 1.127)</td>
<td>1.060 (0.977, 1.150)</td>
<td>14</td>
</tr>
<tr>
<td>Fischer et al</td>
<td>DUELS 0.58</td>
<td>10.0</td>
<td>1.027</td>
<td>1.019 (1.015, 1.023)</td>
<td>29</td>
<td>1.019 (1.016, 1.022)</td>
<td>1.010 (1.007, 1.013)</td>
<td>46</td>
</tr>
<tr>
<td>HEI 2000</td>
<td>ACS CPS II -0.08</td>
<td>81.4</td>
<td>0.95</td>
<td>0.90 (0.84, 0.96)</td>
<td>105</td>
<td>1.15 (1.05, 1.25)</td>
<td>1.22 (1.11, 1.33)</td>
<td>-42</td>
</tr>
<tr>
<td>Jerret et al</td>
<td>ACS CPS II 0.55</td>
<td>7.7</td>
<td>1.031</td>
<td>1.025 (0.997, 1.054)</td>
<td>19</td>
<td>1.032 (1.002, 1.062)</td>
<td>1.015 (0.980, 1.050)</td>
<td>53</td>
</tr>
</tbody>
</table>

Notes:
1 PM_{2.5} results –personal communication.
2 Based on 14 cohorts in which correlation between NO₂ and PM_{2.5} was less than 0.7. HRs are presented per 10 μg/m³ NO₂ and 5 μg/m³ PM_{2.5}
3 PM_{10}
4 HR (95% CI) for min-max range of average concentrations in fine particulate cohort (41 cities).
5 % reduction in ln(HR)
(HR reported to 3 decimal places taken from publication or provided by personal communication)
Figure 5 Hazard ratios (95% CI) from single- and two-pollutant models for NO₂ (A) and PM₂.₅ or PM₁₀ (B) (HRs are expressed per IQR or selected increments in multi-centre studies).
DISCUSSION

1. Adjustment for individual confounders
Having excluded four cohorts with a restricted age range, four of the remaining 10 studies selected for meta-analysis did not control for individual measures of smoking and BMI. When stratified by level of covariate adjustment – i.e. those controlling for the required confounding factors and those that did not (Figure 3) the HRs differed substantially: 1.008 vs 1.031 per 10 μg/m³ respectively. Both Cesaroni et al 2013 and Fischer et al 2015 note this limitation of their studies. Cesaroni et al presented results from a small subset (7845) for which individual smoking measures were available and noted that adjustment for smoking did not alter associations between NO₂ and mortality. Cesaroni et al also adjusted for smoking related comorbidities. Fischer et al conducted a sensitivity analysis adjusted for regional age-standardised smoking-attributable mortality and noted an attenuation of the association from 1.03 (95% CI: 1.02, 1.03) to 1.02 (95% CI: 1.02, 1.03). A sensitivity analysis using the English cohort found that adjustment for individual level smoking status and BMI after adjustment for a small area marker of socio economic status attenuated the HRs by a further 15% (personal communication). The possibility remains, therefore, that studies unable to control for individual confounders may be overstating the size of the association between long-term NO₂ and all-cause mortality.

2. Heterogeneity
For the 10 single pollutant HRs selected for meta-analysis, the fixed- and random-effects summary estimates differed substantially: 1.010 (95% CI: 1.009, 1.012) and 1.021 (95% CI: 1.006, 1.036) per 10μg/m³ respectively. Under the fixed-effects model, all studies are assumed to estimate a common HR. In the meta-analysis therefore, only study precision determines study weight. In a random-effects model however, it is assumed that the study populations (and the study methodology) can differ in ways that can impact on the estimated HRs and a distribution of HRs is assumed. Weights in a random-effects meta-analysis are determined not just by study size, but also between-study variance. Hence, greater weight can be given to smaller studies and less weight to larger studies. In this meta-analysis, 97% of the variation in the HRs was attributable to between-study variance. The two models have different conceptual frameworks and when interpreting their results, it is important to understand possible reasons for the between-study variability. These may relate to population characteristics, baseline population risk, exposure assessment concentrations and sources of co-pollutants and variability in model specification including potential confounders. Careful interpretation of the model results are warranted therefore.
3. Small study bias

Small study bias encompasses publication bias – the publication of adverse, imprecise study results. Publication bias can arise from a number of stages in the process of publication of research findings. These include analyst decisions in model selection and the reporting of null results, decisions by study investigators to submit results for peer review and decisions by journal editors to publish study findings. Small study bias can also be due to heterogeneity between studies and differences in study methodology. It can be identified using a number of graphical and statistical tests.7-9 The presence of small study bias in air pollution epidemiology has been noted previously.38

The results from the Begg and Egger tests reported non-significant results whilst the Trim and Fill procedure required the imputation of additional results to achieve symmetry suggesting the presence of small study bias. The degree of adjustment to the summary estimate varied under the different model specifications available within the Trim and Fill technique. The performance of the Trim and Fill procedure, especially in the presence of heterogeneity between study estimates, has been assessed in simulation studies39,40 Given the substantial heterogeneity between our study estimates the interpretation of the results from the procedure require further investigation. We therefore do not recommend adjustment for small study bias until further assessment of the causes of heterogeneity have been identified but note the possibility that the unadjusted HR may be subject to some bias as a result.

4. Multi-pollutant models

The difficulty in interpreting regression coefficients for correlated variables in multivariate regression is well documented in the statistical literature. More recently, the difficulties in interpreting coefficients in multi-pollutant models has received attention.41,42 These difficulties include: 1) correlation between pollutants (arising due to common sources and meteorological conditions) can lead to unstable parameter estimation; 2) differential measurement error between pollutants can lead to the ‘transfer’ of an association from the less well measured (but true) pollutant to the better measured (but incorrect) pollutant; and 3) statistical models do not generally assess interactions between pollutants and these assessments are required to interpret correctly model main effects.

Given the characteristics of the 6 studies reporting multi-pollutant model results and the problems in interpreting coefficients from multi-pollutant models, the validity of the adjusted coefficients is questionable. Table 1 also shows the combined HRs for the two adjusted pollutant coefficients (i.e. NO₂ adjusted for PM and PM adjusted for NO₂). In four studies the
combined HRs were similar to the NO$_2$ or PM single pollutant HRs. In one study the combined NO$_2$ and PM HRs was larger than both single pollutant model HRs and in another the combined HR lay between the two single pollutant HRs. The combined HRs provide more stable and reliable estimates of the associations between exposure to the two correlated pollutants and all-cause mortality (notwithstanding the lack of interaction terms). These combined estimates could be used in formulating a multi-pollutant approach to regulatory policy as advocated by Greenbaum and colleagues.$^{43}$
REFERENCES


