Bayesian analysis of the short-term association of NO₂ exposure with local burden of asthmatic symptoms in children

Vera van Zoest, Gerard Hoek, Frank Osei, Alfred Stein

Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, 7500 AE Enschede, the Netherlands
Institute for Risk Assessment Sciences (IRAS), Utrecht University, PO Box 80178, 3508 TD Utrecht, the Netherlands

HIGHLIGHTS

• Positive associations between NO₂ and lower respiratory symptoms were observed.
• Credible intervals narrowed when including prior information in a Bayesian analysis.
• Prior distributions of the effect estimates were obtained from a meta-analysis.
• No significant relation was observed between NO₂ and dry cough or phlegm.
• Burden of disease maps showed a strong spatial variability in asthmatic symptoms.

GRAPHICAL ABSTRACT

ABSTRACT

Short-term exposure to air pollution has been associated with exacerbation of respiratory diseases such as asthma. Substantial heterogeneity in effect estimates has been observed between previous studies. This study aims to quantify the local burden of daily asthma symptoms in asthmatic children in a medium-sized city. Air pollution exposure was estimated using the nearest sensor in a fine resolution urban air quality sensor network in the city of Eindhoven, the Netherlands. Bayesian estimates of the exposure response function were obtained by updating a priori information from a meta-analysis with data from a panel study using a daily diary. Five children participated in the panel study, resulting in a total of 400 daily diary records. Positive associations between NO₂ and lower respiratory symptoms and medication use were observed. The odds ratio for any lower respiratory symptoms was 1.07 (95% CI 0.92, 1.28) expressed per 10 μg m⁻³ for current day NO₂ concentration, using data from the panel study only (uninformative prior). Odds ratios for dry cough and phlegm were close to unity. The pattern of associations agreed well with the updated meta-analysis. The meta-analytic random effects summary estimate was 1.05 (1.02, 1.07) for LRS. Credible intervals substantially narrowed when adding prior information from the meta-analysis. The odds ratio for lower respiratory symptoms with an informative prior was 1.06 (0.99, 1.14). Burden of disease maps showed a strong spatial variability in the number of asthmatic symptoms associated with ambient NO₂ derived from a regression kriging model. In total, 70 cases of asthmatic symptoms can daily be associated with NO₂ exposure in the city of Eindhoven. We conclude that Bayesian estimates are useful in estimation of specific local air pollution effect estimates and subsequent local burden of disease calculations. With the fine resolution air quality network, neighborhood-specific burden of asthmatic symptoms was assessed.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords:
Asthma
Panel study
Nitrogen dioxide
Respiratory symptoms
Meta-analysis
Air pollution

Article history:
Received 7 January 2020
Received in revised form 20 February 2020
Accepted 24 February 2020
Available online 24 February 2020

Editor: Lidia Morawska

* Corresponding author.
E-mail address: v.m.vanzoest@utwente.nl (V. van Zoest).
1. Introduction

Air pollution has major effects on human health including respiratory and cardiovascular diseases (Brook et al., 2010; Brunekeef and Holgate, 2002; Goldizen et al., 2016; Guarnieri and Balmes, 2014). In the recent Health Risks of Air Pollution In Europe (HRAPIE) report, the World Health Organization (WHO) presents the key health endpoints (WHO, 2013). Short-term exposure to air pollutants and incidence of daily asthma symptoms in children require further quantification for a full health impact assessment of outdoor air pollution (WHO, 2013). Asthma symptoms cause a significant societal and financial burden (Martinez and Vercelli, 2013). A number of panel studies have been conducted with the aim to quantify the short-term effects of air pollutants on asthmatic symptoms, particularly in Europe and North-America (Weinmayr et al., 2010). There is, however, a substantial heterogeneity between the effect estimates of previous studies, due to differences in study design and study area.

To assess the burden of disease related to asthmatic symptoms in a specific city, generic concentration-response functions can be used, based on previous studies conducted in other areas. Alternatively, data from typically a single epidemiological study in the specific city can be used. Both approaches have pros and cons, including more robust evidence from multiple studies versus differences in the magnitude of effect between different study areas. An interesting option is to combine both approaches in a Bayesian analysis. Previous studies have suggested that accurate estimates and narrow credible intervals can be acquired with Bayesian estimates that include prior information in addition to local effect estimates on the exposure-effect relationship (Post et al., 2001; Le Tetre et al., 2005; Liu et al., 2009; Beach et al., 2012). Potentially, Bayesian estimates may also be useful in estimation of health effects in a panel study, strengthening local information from a small sample with estimates from a combination of previous studies.

Daily exposure is estimated in various ways in panel studies, including measurements at a central monitoring site, personal monitoring and satellite imagery (Chambers et al., 2018). Personal monitoring of air quality is expensive and time-consuming (Brandt et al., 2015), so its rare use is mostly limited to short measurement campaigns (e.g. Linn et al., 1996; Spira-Cohen et al., 2011). In the majority of panel studies, one or two central monitors in a city are used to estimate exposure (Dales et al., 2009; Roemer et al., 1993; van der Zee et al., 1999; van der Zee et al., 2000). However, some air pollutants like NO$_2$ typically show a strong short-distance variability within urban areas, because of the variety of road types and land uses within the city (Hoek et al., 2008). A more accurate estimate of exposure can be acquired by estimating exposure at or near the individual residence and work or school address.

The objective of this study is to estimate the spatially explicit burden of air pollution exposure on daily symptom prevalence in asthmatic children in a medium-sized urban area. We used a Bayesian analysis for effect estimation, with the aim to combine prior information from a meta-analysis with data from a panel study conducted in the city of Eindhoven, the Netherlands. In this city, a fine resolution urban air quality sensor network is located which was used for exposure estimation.

2. Methods

Our study included the following components to assess the local burden of disease (Fig. 1):

1. Panel study among asthmatic children in Eindhoven, combining a daily symptom diary and a fine resolution sensor network to obtain local effect estimates
2. Update of the meta-analysis by Weinmayr et al. (2010) to obtain prior information on the association between air pollution and daily asthma symptoms in children
3. Bayesian analysis with and without informative prior to obtain the concentration-response function to be used in the burden of asthma symptom calculation
4. Spatial modelling of outdoor air pollution across the city of Eindhoven, averaged to annual mean concentrations per neighborhood
5. Collection of data on number of children per neighborhood, proportion of children with asthma, and daily asthma symptom prevalence rate
6. Calculation of neighborhood-specific and overall city burden of daily asthma symptoms, combining prevalence of daily asthma symptoms with the attributable fraction related to the average neighborhood air pollution

2.1. Panel study

The design of the panel study is based upon previous panel studies in the Netherlands (Roemer et al., 1993; van der Zee et al., 1999; van der Zee et al., 2000) and elsewhere (Weinmayr et al., 2010). An important lesson from previous studies is that short observation periods (~2 months) are prone to confounding by infection episodes which are difficult to characterize (Weinmayr et al., 2010). Therefore, in our panel study, each participant was asked to participate for 4 months. To increase variability in air pollutant concentrations, participants started at different moments in time. The total study period was 13 months (29/03/2017–22/04/2018).

Participants were recruited in different ways. Articles advertising about the panel study were published in local newspapers, on social media and on websites. Flyers were spread door-to-door to ~5000 households, and flyers were displayed in schools, at general practitioners and at pulmonologists in hospitals in Eindhoven. In order to be eligible to participate in the panel study, a subject had to meet all of the following criteria: (1) the child has current asthma, (2) the child is 7 to 11 years old at the start of his or her participation, and (3) the child lives and attends school in the municipality of Eindhoven. A child was considered to have current asthma when at least two of the following three criteria were met: (1) asthma diagnosis by a physician ever, (2) wheeze in the past 12 months, and (3) use of asthma medication in the past 12 months. Children living in a house in which people smoke indoors were excluded from participation, to avoid noise related to this strong risk factor. Children aged 7–11 years were selected because a minimum age of 7 is required for accurate lung function measurements. Young children with asthma are furthermore very sensitive to the effects of air pollution (Guarnieri and Balmes, 2014). Increased sensitivity compared to adults is due to a combination of increased time spent exercising outside, high ventilation rates per body weight, developing lungs and immature metabolic pathways (Guarnieri and Balmes, 2014). Finally, school age children have more predictable time activity patterns, allowing more precise exposure assessment based on outdoor monitors.

The parents of the children were asked to fill out an electronic diary with the child every evening during the 4 months of participation. Electronic diaries are considered more accurate than paper diaries, since the electronic diaries could only be filled in retrospectively for a limited amount of days up to a week (Velicka et al., 2015). The diary questions were based on previous panel studies in children (Roemer et al., 1993; van der Zee et al., 1999; van der Zee et al., 2000). The symptoms included wheeze, shortness of breath in rest, shortness of breath after exercise, waking up during the night with breathing problems, dry cough, phlegm, and nose complaints. In the diary, the participants reported on-demand medication use and the presence of asthma symptoms as absent, mild, or moderate/severe. The participants were visited at home three times: at the start of the 4-month study period, after 2 months, and at the end of the study period. During the first visit, the study was explained to the
child and parent(s), there was an opportunity to ask questions, informed consent forms were signed, and instructions for filling out the daily diary were given. Furthermore, the parent(s) filled out a baseline questionnaire to characterize the child’s health status, medication use, daily activities and indoor sources of air pollutants. After 2 months the research assistant visited the participants to keep motivation up, answer potential questions and monitor general progress. All parents signed an informed consent form. The study was approved by the Medical Ethical Committee of University Medical Centre Utrecht.

2.2. Statistical analysis

Multiple logistic regression models are often used to model the relation between air pollutant concentrations and occurrence of asthma-related symptoms (Forsberg et al., 1998; Migliore et al., 2009; Ranzi et al., 2015; Roemer et al., 1993; Schinasi et al., 2011; van der Zee et al., 1999; van der Zee et al., 2000). The binary outcome \( y_{ij} \) denotes the presence or absence of an outcome reported by participant \( i \) on day \( j \). Both diary options “mild” and “moderate/severe” were combined to symptom presence, \( y_{ij} = 1 \). Separate models were built for each of the symptoms and for any lower respiratory symptoms (LRS), including wheeze, shortness of breath in rest, shortness of breath after exercise, and waking up during the night with breathing problems. Daily exposure to \( \text{NO}_2 \), \( \text{PM}_{10} \), \( \text{PM}_{2.5} \) and \( \text{PM}_1 \) was estimated by an urban air quality sensor network consisting of 35 airboxes (Close, 2016; van Zoest et al., 2018; van Zoest et al., 2019b), using the airbox closest to the house and school of the child. All airboxes measure particulate matter, whereas 25 airboxes measure \( \text{NO}_2 \). Only six airboxes measure ultrafine particles (UFP), so the distance from the house of the child to the nearest UFP sensor could be > 2.5 km in some cases, while UFP varies strongly over shorter distances. Therefore, daily exposure to UFP was estimated by the mean of the six airboxes. After careful data quality evaluation and outlier detection (van Zoest et al., 2019b; van Zoest et al., 2018), the data obtained from the urban air quality sensor network reflected the spatio-temporal variability in air pollutant concentrations in the city well (van Zoest et al., 2019a). We evaluated the associations of air pollution concentrations of the same day (lag 0), day before (lag 1), and average of lag 0–2 to allow for delayed associations on the various symptoms. The effect estimates were calculated per 10 \( \mu \text{g} \text{m}^{-3} \) increase in air pollutant concentrations for \( \text{NO}_2 \), \( \text{PM}_{10} \), \( \text{PM}_{2.5} \) and \( \text{PM}_1 \), and per 10,000 particles increase for UFP. Separate models were built for each lag. The presence of a symptom \( y_{ij} \) follows a Bernoulli distribution with probability of occurrence \( p_{ij} \): \[
y_{ij} \sim \text{bernoulli}(p_{ij})
\]
where

$$\logit(p_i) = \beta_{0,s} + \beta_{0,c} \gamma_i^{\text{fl}} + \beta_{0,d} \gamma_i^{\text{day}} + \beta_{0,h} \gamma_i^{\text{RH}} + \beta_1 x_i^{\text{T}} + \beta_{0,\text{wday}} x_i^{\text{wday}}. \tag{2}$$

Here, $\beta_k$ are the coefficients and $\gamma$ are the covariate values for the different covariates $c$, consisting of the air pollutant $ap$ in $\{\text{NO}_2, \text{PM}_1, \text{PM}_{2.5}, \text{PM}_{10}, \text{UPF}\}$ and confounders. Confounders were included as a priori based upon previous studies (Forsberg et al., 1998; Ranzi et al., 2015; Roemer et al., 1993; Van der Zee et al., 1999; van der Zee et al., 2000): day of follow-up ($s$), daily mean relative humidity ($RH$), daily mean temperature ($T$), daily reported flu ($\text{flu}$), and day of the week categorized as weekdays/weekend days ($\text{wday}$). Daily temperature and relative humidity were obtained from the Royal Netherlands Meteorological Institute weather station in Eindhoven (KNMI, 2019). Characteristics of the child that are expected to remain constant over the full study period, such as gender, socio-economic status, and presence of pets in the household, were not included in the analysis because the form of the model considers each subject as its own control. To adjust for differences in baseline symptom reporting, individual-specific intercepts ($\beta_{0,s} = \alpha + \gamma_i$) were specified for each individual, consisting of an overall intercept $\alpha$ and a random intercept $\gamma_i$ for each participant $i$. Our interest is in the odds ratio (OR): $\text{OR} = \exp(\beta_{0,c})$.

We first estimated the parameters using restricted maximum likelihood (REML). Based on this preliminary analysis (Tables S1–S5 in Supplementary Materials), we found no informative results (i.e. wide confidence intervals) for PM$_{10}$, PM$_{2.5}$, PM$_1$ and UPF. We therefore continued our Bayesian analysis and burden of disease study on NO$_2$ only.

Bayesian estimation of the parameters was performed using JAGS (Plummer, 2003), through the ‘rjags’ package in R (Su and Yajima, 2015). We used the Gelman-Rubin diagnostic $R$ to evaluate convergence (Gelman and Rubin, 1992), by comparing variances of different Markov Chain Monte Carlo (MCMC) simulation chains. Values close to 1 indicate convergence. Two chains of MCMC simulations were run until convergence was achieved for all parameters in the model ($R$<1.1).

### 2.3. Meta-analysis and prior selection

First, we used an uninformative prior for all parameters in the model (Table 1). The individual-specific intercepts were considered as exchangeable random effects $\gamma_i \sim N(\mu = 0, \sigma = \sigma_{\gamma})$ with an uninformative prior on $\sigma_{\gamma}$.

We then evaluated the effect of choosing an informative prior for $\beta_{\text{NO}_2}$. To obtain these informative priors we performed a meta-analysis of the literature, based on the work by Weinmayr et al. (2010). They performed a robust systematic review and meta-analysis of literature published between 1990 and July 2008 on short-term health effects of NO$_2$ on respiratory health among children with asthma or asthma-like symptoms, providing estimates for asthmatic symptoms and cough. To get a more up-to-date estimate, we updated the meta-analysis to include publications from August 2008 to March 2019. We followed the same procedure applied by Weinmayr et al. (2010). The MEDLINE database was searched through the PubMed search engine, using the search string ‘([“asthma” OR “wheeze” OR “cough” OR “bronchitis” OR “lung function”] AND (“air” AND pollut)) AND (“NO2” OR “NO2Z” OR “nitrogen dioxide”]) and limits were set to retrieve only publications about children (0–18 years). Based on the abstracts, we excluded indoor air pollution studies, laboratory studies and studies on infants. The meta-analysis included only panel studies on asthmatic or symptomatic children which reported a quantitative effect estimate and which controlled for temperature and day of the week or temporal autocorrelation.

The effect estimates of all panel studies (1990 to March 2019) were combined in a random effects meta-analysis model (DerSimonian and Laird, 1986) using the ‘metafor’ package in R (Viechtbauer, 2010). The effect estimates of all studies were standardized to $\beta_{\text{NO}_2}$ coefficients per 10 $\mu g$ m$^{-3}$. Where needed, concentrations in ppb were converted to $\mu g$ m$^{-3}$ using the standard conversion at 20 °C: 1 ppb = 1.91 $\mu g$ m$^{-3}$. ORs were converted to $\beta_{\text{NO}_2}$ coefficients before standardization, using the natural logarithm $\beta_{\text{NO}_2} = \log(\text{OR})$. The combined effects estimate of $\beta_{\text{NO}_2}$ was used as a normally distributed prior in the Bayesian estimation.

#### 2.4. Burden of disease mapping

To obtain the burden of disease we calculated the potential health risk reductions when NO$_2$ exposure levels would be reduced from its actual levels per neighborhood to zero. The analysis was conducted per neighborhood because this was the smallest unit with data on number of children available. We determined the number of attributable cases ($AC_k$) for each neighborhood $k$:

$$AC_k = AF_k \times PR \times P_k \tag{3}$$

where the attributable fraction $AF_k = \frac{(RR_k - 1)}{RR_k}$ (Shaddick et al., 2018) is based on the neighborhood-specific relative risk $RR_k$, and $PR$ is the prevalence rate. We use the OR as a proxy for RR, as it typically represents RR well (Liu et al., 2009). Since we obtain the OR per 10 $\mu g$ m$^{-3}$, multiplication by 0.1 is required for standardization. $RR_k$ is then obtained as:

$$RR_k = \exp(0.1 \times \log(\text{OR}) \times \Delta x_{\text{NO}_2}^{\text{act}}) \tag{4}$$

where $\Delta x_{\text{NO}_2}^{\text{act}} = x_{\text{NO}_2}^{\text{act}} - x_{\text{NO}_2}^{\text{baseline}}$ is the annual average NO$_2$ concentration in neighborhood $k$ minus the baseline NO$_2$ concentration. We set $x_{\text{NO}_2}^{\text{baseline}} = 0$. The NO$_2$ concentrations were estimated on a 25 x 25 m grid using a regression kriging model for each hour of the day and weekdays/weekends separately. Details are described elsewhere (van Zoest et al., 2019a). The sensor network data was not available for the entire year 2016 due to a few months of maintenance. The lack of complete daily data for 2016 also precluded linking the measurements of the central monitoring stations in Eindhoven (van Zoest et al., 2019a). The sensor network data was not available for the entire year 2016 due to a few months of maintenance. The lack of complete daily data for 2016 also precluded linking the exposure response function with daily spatial exposure data. Therefore, an average of the modelled values for June and November was used. This is a reasonable proxy for the annual average when consulting the measurements of the central monitoring stations in Eindhoven (Table S6 in Supplementary Materials), which are part of the national ambient air quality monitoring network (RIVM, 2019).

### Table 1

Uninformative priors for Bayesian estimation of the parameters in the model (Eq. (2)).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Prior distribution$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_{\text{NO}_2}$, for each symptom s</td>
<td>$\beta_{\text{NO}<em>2} \sim N(\mu = 0, \sigma = \sigma</em>{\beta})$</td>
</tr>
<tr>
<td>$\beta_c$, for each confounder</td>
<td>$\beta_c \sim N(\mu = 0, \sigma = \sigma_{\beta})$</td>
</tr>
<tr>
<td>$\alpha$ (fixed intercept)</td>
<td>$\alpha \sim N(\mu = 0, \sigma = \sigma_{\alpha})$</td>
</tr>
<tr>
<td>$\gamma_i$ (random intercept)</td>
<td>$\gamma_i \sim N(\mu = 0, \sigma = \sigma_{\gamma})$</td>
</tr>
</tbody>
</table>

$^a$ $N_{\gamma}$ denotes a half-normal distribution truncated at zero, such that the standard deviation can only take positive values.
raster cells within neighborhood $k$ were used to obtain the average NO$_2$ concentration in neighborhood $k$.

In Eq. (3), $P_k$ is the exposed population at risk in neighborhood $k$. Assuming that all people are exposed to ambient NO$_2$ air pollutant concentrations, $P_k$ represents the population at risk $P_k = n_k \text{child} \times \text{Pasthma}$, where $n_k \text{child}$ is the number of children in neighborhood $k$, and Pasthma is the average proportion of asthmatic children, estimated at 0.126 based on the PIAMA birth cohort study in the Netherlands (Scholtens et al., 2009). In the absence of local data on the daily prevalence of asthmatic symptoms in asthmatic children, we estimated the prevalence rate $PR$ at 0.17 asthmatic symptoms per person-day in the exposed population at risk, based upon the HRAPIE report (WHO, 2013).

To obtain a measure of uncertainty propagated from the input data into the number of attributable cases, we ran MCMC simulations drawing samples from prior distributions of the input variables. For log(OR), this prior distribution equals the posterior distribution of $\beta_{\text{NO}_2}$, which was obtained through Bayesian estimation of the parameters in Eq. (2). An approximation of the number of children per neighborhood, $E(n_{\text{child}}) = p_{\text{child}} \times E(n_{\text{pop}})$, is retrieved from the Statistics Netherlands 2016 population data (CBS, 2018). Here, $p_{\text{child}}$ is the proportion of children between 0 and 14 years as part of the total population $n_{\text{pop}}$. The percentage of children is considered to be measured without uncertainty. However, to avoid privacy concerns, CBS rounds the population data to a multiple of five in a random manner. Therefore, CBS provides an approximation of the number of inhabitants $E(n_{\text{pop}})$ for each neighborhood, and we consider the uncertainty in the true number of inhabitants per neighborhood by posing $n_{\text{pop}} \sim \text{Unif}(E(n_{\text{pop}}) - 5, E(n_{\text{pop}}) + 5)$. We account for uncertainty in the proportion of asthmatic children by sampling from a truncated positive normal distribution $\text{Pasthma} \sim N(\mu = 0.126, \sigma = 0.05)$. For the prevalence rate we sample values from $PR \sim N(\mu = 0.17, \sigma = 0.05)$. The modelled NO$_2$ exposure at 25 m raster resolution is within the 30% uncertainty required for modelling annual average values (European Parliament and Council of the European Union, 2008). Averaging to neighborhood averages however leads to added uncertainty at measurement locations, since concentrations near main roads can be substantially higher than the background concentrations in the neighborhoods. As most people live in areas with background concentrations, we consider the average uncertainty at background locations only (van Zoest et al., 2018). We sample values from the distribution $X_{\text{NO}_2}^C \sim N(\mu = E(X_{\text{NO}_2}^C), \sigma = \sigma_{\text{NO}_2}^C)$, where $E(X_{\text{NO}_2}^C)$ is the average modelled NO$_2$ concentration of the raster cells in neighborhood $k$, and $\sigma_{\text{NO}_2}^C$ = 2.64 is the standard deviation of the absolute differences between modelled and observed annual average NO$_2$ concentration values at the background airbox locations.

We obtained the total number of LRS related to ambient NO$_2$ concentrations in the city, $A_{\text{city}}$, by summing the number of attributable cases from all neighborhoods:

$$A_{\text{city}} = \sum_k A_{\text{city}_k}$$ (5)

We then compared $A_{\text{city}}$ with the number of LRS which would have been obtained in a non-spatial analysis, in which only one central monitoring station is used to estimate exposure. Typically, when no fine resolution sensor network would be available, a central monitor of the national ambient air quality network would have been used for this purpose. However, the two central monitors in Eindhoven are both in traffic locations, which are not representative for the background concentrations in which most people live. We therefore use the airboxes from the sensor network in background locations to obtain a burden of disease estimate from one airbox at a time. We report the minimum and maximum values to show the variability in estimates depending on the location of the central monitoring site.

### 3. Results

#### 3.1. Panel study descriptives

Despite extensive recruitment efforts, only seven children could be recruited to participate in the panel study. Two participants stopped filling in the diary within the first three days of the study and were excluded from the analysis. The remaining five children were two boys and three girls with a mean age of 9.4 yr (range 7–11 yrs). The participants were living in different areas of the city with highly different NO$_2$ concentrations: the mean concentration during the study period varied between 19.7 and 49.3 µg m$^{-3}$ for the different participants, based on the closest airbox. Differences in mean NO$_2$ between children were accounted for using random intercepts in the model. Daily variability is therefore considered more important than variability between individuals and locations.

Days with missing NO$_2$ values were removed from the analysis. In total, $n = 394$ diary entries were included. Table 2 shows the frequency of symptoms reported in these diary entries. There were signs of clustering of symptoms within participants, but each individual symptom was reported by at least three of the five participants. One participant only reported a single symptom throughout the entire study period. This record was not used in the analysis due to missing NO$_2$ data from the nearest NO$_2$ sensor on that day. The remaining non-symptomatic days of this participant, however, were included in the analysis.

#### 3.2. Meta-analysis and prior selection

Weinmayr et al. (2010) included 20 studies on NO$_2$ and symptoms in their meta-analysis. Based on our extended literature search, we added 5 more publications with one study population each. All studies included in the meta-analysis reported effect estimates for asthmatic symptoms, with varying definitions. From the 39 study populations, effect estimates for cough were reported in 32 study populations.

Based on the combined meta-analysis, the combined ORs (95% confidence intervals) were 1.048 (1.023, 1.074) for asthmatic symptoms (Fig. 2) and 0.995 (0.973, 1.018) for cough (Fig. 3). We used prediction intervals instead of confidence intervals to obtain prior distributions. A prediction interval represents possible outcomes of single studies rather than the overall OR, and therefore allows for heterogeneity in individual studies. The prediction intervals of the meta-analysis resulted in a prior $\beta_{\text{NO}_2}(\mu = 0.047, \sigma = 0.040)$ for any lower respiratory symptoms and $\beta_{\text{NO}_2}(\mu = 0.005, \sigma = 0.039)$ for cough. We used the prior distribution for cough for the effect estimates of both dry cough and phlegm, as both were combined in the original meta-analysis (Weinmayr et al., 2010).

#### 3.3. Effect estimates

Table 3 shows the odds ratios (95% CI) for the different symptoms related to NO$_2$, obtained with uninformative priors. Despite the small number of children, we observed some associations between NO$_2$ and lower respiratory symptoms and medication use. ORs for cough were

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any lower respiratory symptoms (LRS)</td>
<td>111 (28.2%)</td>
</tr>
<tr>
<td>Wheezing</td>
<td>7 (1.8%)</td>
</tr>
<tr>
<td>Shortness of breath in rest</td>
<td>65 (16.5%)</td>
</tr>
<tr>
<td>Shortness of breath after exercise</td>
<td>38 (9.7%)</td>
</tr>
<tr>
<td>Waking up with breathing problems</td>
<td>17 (4.3%)</td>
</tr>
<tr>
<td>Dry cough</td>
<td>98 (24.9%)</td>
</tr>
<tr>
<td>Phlegm</td>
<td>44 (11.2%)</td>
</tr>
<tr>
<td>Nose complaints</td>
<td>138 (35.0%)</td>
</tr>
<tr>
<td>Medication use</td>
<td>37 (9.4%)</td>
</tr>
</tbody>
</table>
close to unity. This pattern of associations agrees well with the meta-analysis of previous studies. Table 4 shows the ORs for lag 0 obtained using informative priors based on the meta-analysis for LRS, dry cough and phlegm, in comparison to the ORs obtained using an uninformative prior. The ORs for lag 1 and mean lag 0\–2, using the same informative priors, are shown in Table S7 in the Supplementary Materials. Convergence was achieved for all parameters in the model ($^\text{R}_b$1: 1) and convergence was strong for the ORs of all symptoms ($^\text{R}_b$1: 0.2). The fastest and strongest convergence was found for the models using informative priors. The credible intervals were also much smaller when using informative priors, indicating less uncertainty about the estimates. ORs with informative prior for lag 1 and mean of lag 0\–2 were similar to ORs for lag 0 (Table S7), reflecting the large impact of the prior.

3.4. Burden of disease calculations

Fig. 4 shows maps of the number of children per neighborhood, the mean NO2 exposure per neighborhood, and the number of attributable cases per neighborhood per day, based on the association between NO2 and LRS at lag 0 using an informative prior. In most neighborhoods, between 0 and 1 LRS per day are associated with ambient NO2 concentrations. In some more populated neighborhoods, this can increase up to 3 LRS per day. Some neighborhoods show ‘No Data’, where the percentage of children is unavailable due to privacy issues. This only occurs if the number of inhabitants $b 5 0$. Since the number of children is also expected to be small here ($b 8$), the attributable number of cases is expected to be close to zero.

The number of attributable cases of LRS per neighborhood strongly reflects the number of children per neighborhood, as expected (Pearson’s correlation coefficient $r = 0.96$). NO2 exposure is modelled using population density as one of the covariates and is therefore also, though less strongly, related to the number of children ($r = 0.27$). NO2 exposure has a stronger relation with the number of LRS ($r = 0.45$).

Summing the $AC_k$ for each neighborhood, $AC_{city} = 70$ cases of LRS associated with ambient NO2 exposure on a daily basis. When only one background concentration monitor would have been consulted to estimate the burden of disease, the number of attributable cases would vary between 45 and 73, depending on the location of the background monitor. The spatial variability in exposure and the related burden is large, however, and $AC_{city}$ therefore represents a much better estimate of the true burden of disease. We chose a normal distribution to
represent the uncertainty in $PR$ and a uniform distribution to represent the uncertainty in $P_k$. Even though no parametric distribution is posed on the OR, the posterior density function of the OR is close to a normal distribution (Fig. 5a). This is also reflected in $AF_k$, which has a symmetrical distribution with long tails on both sides. Since Eq. (3) is a multiplicative function in which the distributions of $PR$, $P_k$ and $AF_k$ are multiplied, the uncertainty propagated in $AC_k$ is skewed (Fig. 5b) – a characteristic of multiplying normal distributions. The 95% credible interval of $AC_k$ is therefore highly unsymmetrical around the mean.

4. Discussion

In this study we estimated the burden of NO$_2$ exposure on daily symptom prevalence in asthmatic children in the city of Eindhoven.

Table 3

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Lag 0</th>
<th>Lag 1</th>
<th>Mean lag 0–2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any lower respiratory symptoms (LRS)</td>
<td>1.07 (0.92, 1.28)</td>
<td>1.16 (0.96, 1.40)</td>
<td>1.20 (0.95, 1.52)</td>
</tr>
<tr>
<td>Wheezing$^a$</td>
<td>1.02 (0.80, 1.32)</td>
<td>1.00 (0.66, 1.43)</td>
<td>1.00 (0.63, 1.50)</td>
</tr>
<tr>
<td>Shortness of breath in rest</td>
<td>1.12 (0.95, 1.44)</td>
<td>1.23 (0.97, 1.58)</td>
<td>1.24 (0.93, 1.66)</td>
</tr>
<tr>
<td>Shortness of breath after exercise</td>
<td>1.15 (0.96, 1.55)</td>
<td>1.30 (0.99, 1.69)</td>
<td>1.50 (1.06, 2.06)</td>
</tr>
<tr>
<td>Waking up with breathing problems$^a$</td>
<td>0.98 (0.73, 1.23)</td>
<td>0.79 (0.46, 1.10)</td>
<td>0.81 (0.46, 1.18)</td>
</tr>
<tr>
<td>Dry cough</td>
<td>1.03 (0.87, 1.23)</td>
<td>1.02 (0.83, 1.22)</td>
<td>1.01 (0.80, 1.27)</td>
</tr>
<tr>
<td>Phlegm</td>
<td>1.07 (0.87, 1.40)</td>
<td>1.17 (0.88, 1.59)</td>
<td>1.16 (0.80, 1.65)</td>
</tr>
<tr>
<td>Nose complaints</td>
<td>1.04 (0.88, 1.26)</td>
<td>1.02 (0.81, 1.25)</td>
<td>1.02 (0.78, 1.31)</td>
</tr>
<tr>
<td>Medication use</td>
<td>1.03 (0.84, 1.29)</td>
<td>1.35 (1.00, 1.95)</td>
<td>1.35 (0.97, 1.94)</td>
</tr>
</tbody>
</table>

All OR are adjusted for daily temperature, relative humidity, day of follow-up, daily flu, weekday/weekend day and participant ID.

$^a$ Symptom prevalence <5% (Table 1); results should be interpreted with care.
the Netherlands. We used a Bayesian method for estimation of the effect estimates, combining prior information from an updated meta-analysis with data acquired in a local panel study. Odds ratios for lower respiratory symptoms for an increment of 10 μg m⁻³ NO₂ were 1.07 (0.92, 1.28) and 1.06 (0.99, 1.14) using an uninformative and informative prior respectively. The OR based on the meta-analysis was 1.05 (1.02, 1.07).

Although the number of participants in the panel study was low, we saw that the data affected the posterior distributions for all symptoms, by modestly pushing up the estimates of the prior distributions. The modest difference between posterior and prior distributions was due to the small size of the panel study but also because of the similarity in effect estimates between the local panel study and the meta-analytic combined estimate. The use of prior information narrowed the credible interval compared to the use of the panel study alone, in line with previous studies (Le Tertre et al., 2005; Liu et al., 2009; Post et al., 2001). Where feasible, this combination of local and generic exposure response data seems preferable to using only one of the two sources. The amount of work to recruit subjects for a panel study may however be problematic, unless strong cooperation with medical specialists is obtained. In the current setting, medical doctors were only willing to passively inform their patients by allowing leaflets in their facilities.

The priors used in this study were obtained by updating the meta-analysis by Weinmayr et al. (2010) with more recent panel studies on associations of NO₂ with daily asthmatic symptoms. The inclusion of 5 more recent studies modestly increased the summary OR: from OR = 1.031 (95% C.I. 1.001, 1.062) in the published meta-analysis to 1.048 (1.023, 1.074) in the current review.

There were signs of heterogeneity between studies, for example due to differences in study design, exposure estimation and study area. The effect estimates from the meta-analysis were based on the most significant lag in each study, which varied between different studies. To obtain a general overall estimate to be used as prior information, we combined all studies despite their heterogeneity and different lags, and used the same combined prior for all lags. For the priors we used the prediction interval rather than the confidence interval of the random effects estimates from the meta-analysis. The prediction interval is wider, representing the uncertainty of one single study rather than the uncertainty of the mean of all studies. The first is more appropriate, as we use the prior to predict the effect estimates of a single study.

Symptom reporting in a daily diary is inherently subjective. It is unlikely that children or parents were aware of the daily variation of air pollution in our setting of moderate air pollution without a clear local point source. Systematic bias related to subjective reporting in our study of the association between daily variation of air pollution and daily reporting of symptoms is therefore unlikely. Systematic differences in symptom reporting between children were adjusted for in our analysis.

We accounted for uncertainty in the number of attributable cases of LRS per neighborhood by sampling from prior distributions representing the uncertainty in the population at risk, attributable fraction and prevalence rate. The resulting posterior distribution was highly skewed. However, there may still be remaining sources of uncertainty, and the burden of disease may likely be an underestimation of the actual number of attributable cases in children with asthmatic

### Table 4

<table>
<thead>
<tr>
<th>Symptom</th>
<th>OR based on uninformative prior</th>
<th>OR based on informative prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any lower respiratory symptoms (LRS)</td>
<td>1.07 (0.92, 1.28)</td>
<td>1.06 (0.99, 1.14)</td>
</tr>
<tr>
<td>Dry cough</td>
<td>1.03 (0.87, 1.23)</td>
<td>1.00 (0.93, 1.08)</td>
</tr>
<tr>
<td>Phlegm</td>
<td>1.07 (0.87, 1.40)</td>
<td>1.00 (0.93, 1.08)</td>
</tr>
</tbody>
</table>

All OR are adjusted for daily temperature, relative humidity, day of follow-up, daily flu, weekday/weekend day and participant ID.
symptoms. For example, we here only used the number of children between 0 and 14 years old, based on available CBS data. Also, the NO2 exposure estimates used here are an average over each neighborhood, while concentrations near traffic roads may be substantially higher than background concentrations. Compared to the use of one central monitor to represent exposure, however, we believe that our method shows a more realistic picture of the burden of disease per neighborhood, dependent on the neighborhood-specific population at risk and relative risk.

This study has a number of strengths and limitations. Strengths include the Bayesian analysis combining prior and local information of exposure response functions and the use of a spatially refined low-cost sensor network to assess exposure more refined than in typical epidemiological panel studies (Weinmayr et al., 2010). Limitations include the small size of the local panel study, the inability to derive daily spatial maps that could have been linked with exposure response functions, and the lack of local data on number of children with asthma and the frequency of daily asthma symptoms in children with asthma. Although linking the annual average concentration with the exposure response function does not result in identical AF values compared to averaging daily AF values in a full year, the difference is likely small as \( \beta_{\text{NO2}} \) is small and thus \( \exp(\beta_{\text{NO2}}) \) is close to \( 1 + \beta_{\text{NO2}} \).

With 70 cases of LRS per day associated with ambient NO2 exposure only in the city of Eindhoven, the air pollution problem is one that should not be neglected even in cities where the air pollutant concentrations are generally below European limit values. We cannot evaluate whether the associations we found with NO2, are due to a direct causal effect of NO2 or to correlated other traffic-related air pollutants such as black carbon. The absolute number of cases is affected by the choice to calculate the burden compared to a zero NO2 concentration. We do not know whether associations with NO2 extend to zero, but so far there is little evidence of a threshold in the relationship between outdoor air pollution and respiratory symptoms.

5. Conclusions

We conclude that a Bayesian analysis is useful to estimate location-specific air pollution effects and subsequent local burden of disease. Despite the small number of participants in the panel study, we were able to derive narrow credible intervals around the effect estimates, by incorporating prior information from updating an existing meta-analysis using a Bayesian framework. With the help of a fine resolution urban air quality sensor network we were able to obtain air pollution exposure estimates close to the houses of the participants, rather than at a central location in the city. This allowed for spatial variability in the exposure of different participants. We created burden of disease maps, showing the spatial variability in the number of LRS associated with ambient NO2 exposure. The uncertainty propagation analysis showed that the uncertainty in the number of LRS is positively skewed. Imposing a normal distribution would have led to a biased mean and an unrealistically symmetrical 95% credible interval. By means of the Bayesian analysis we obtained more realistic estimates of the number of LRS associated with ambient NO2 exposure.

Declarations

This study was approved by the Medical Ethical Committee of UMC Utrecht. The parents of all participants signed an informed consent form. The authors declare that they have no competing interests. This work was supported by the Netherlands Organization for Scientific Research (NWO). NWO was not involved in study design, data collection, data analysis, data interpretation or writing of the manuscript.

CRediT authorship contribution statement

Vera van Zoest: Conceptualization, Methodology, Formal analysis, Writing - original draft. Gerard Hoek: Conceptualization, Methodology, Writing - review & editing. Frank Osei: Conceptualization, Methodology, Writing - review & editing. Alfred Stein: Conceptualization, Writing - review & editing.

Declaration of competing interest

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

This work was supported by the Netherlands Organisation for Scientific Research (NWO). The authors kindly thank Marielle Oldenwening for her support in the panel study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.137544.

References


