Author’s response to reviews

Title: Asthma incidence in children growing up close to traffic: a registry-based birth cohort.

Authors:

Anna Lindgren (anna.lindgren@med.lu.se)
Emilie Stroh (emilie.stroh@med.lu.se)
Jonas Björk (jonas.bjork@med.lu.se)
Kristina Jakobsson (kristina.jakobsson@med.lu.se)

Version: 2 Date: 7 October 2013

Author’s response to reviews:

We are greatful to the reviewers for their very valuable comments. We are not able to address all comments in the current study, but we will keep the comments in mind for further studies; especially the need to describe the potential socio-economic selection and the need of exposure comparability with other studies in the field.

Response to reviewer 1.
1. The cut-off levels for the categorization of the NOx estimates seem to be arbitrarily selected. The authors might justify these cut-off settings or alternatively use interquartile increments for the modeling of the health effects.

1. Author’s response: The number of cut-offs were chosen for power consideration, and to fulfill the proportional hazards assumptions, which is not fulfilled when using a large number of categories (probably because of numerical instability). Also, if we use more categories (as IQR) there will be some subgroups that does not have all exposure categories, which makes adjustment for confounding and subgroup stratification less meaningful. We wanted to have relatively large exposure contrasts between the categories, at the same time as we wanted to avoid numerical instability.

2. Asthma is a complex disease and not easy to diagnose during the first years of life. The authors might consider analyzing the potential adverse effects of exposure to ambient air pollutants in relation to the considered health outcomes beyond the age of 3 years. A sensitivity analysis restricted to those outcomes (greater than 3 years) would be informative for the reader, as corticosteroids can be prescribed for other reasons than those analyzed in this manuscript.

2. Author’s response: The incidence is lower in relation to traffic, in the children >3 years of age, as can be seen in the Kaplan-Meier estimates in Additional file 2. The proportional hazards-assumptions were fulfilled for all variables used, with
no crossing over, so it is implied that the effect is the same for the older subgroup.

However, this is a very special group of children with first time use of asthma medication in an older age. Since most cases of allergic asthma, starts with respiratory symptoms in early age that persists into older age, “persistent” asthma is the most interesting variable. We considered analyzing this but there is so much fewer cases in the older age groups, so there is not enough power certainly not for proper adjustments. Instead – we will follow up this group of children within a few years where analyzes can be properly made of the persistent cases.

3. In the tables, no data is provided on moving activities, the frequency of #2 agonists and corticosteroid prescriptions and doctor diagnoses. This information could be easily added to table 1 or in the text.

3. Author’s response: We have now added life table data to Additional file 1 (eTable 5), with information on the incidence of outcomes and first-time move, in different age-groups. Since we only have yearly registered addresses, none are considered to have moved during their first year of life.

4. How precisely was the information on medication prescriptions and doctor diagnoses recorded in the database? If this information is only available for an interval of a year, than the authors might have considered a discrete hazard model instead of Cox PH.

4. Author’s response: The raw data on dispensed medication were in continuous time (date of dispense). However, since we only had residential address on a yearly basis, we categorized outcome data to yearly intervals. We agree in hindsight that discrete COX time may have been most theoretically appropriate. In practice however, since data approximately fulfill the proportional hazard assumption, and also since we have time-invariant predictors, the main reason for using a discrete PH-model, as we understand it, would be to get more appropriate standard errors. However, this difference in standard error seems to be very marginal (We have tried on a few of the outcomes). Since we already give estimates rounded to very few digits, it makes no practical difference. We thus prefer to keep the Cox PH model.

5. The traffic intensity was given in numbers of cars/min. Most similar studies have provided DTV in numbers of cars/24h. The authors should consider this, and come to their own conclusion regarding which metric is best.

5. Author’s response: Traffic intensity is now stated in cars/24h throughout the manuscript, to facilitate for the reader.

6. Do the authors have the possibility to study potential health effects associated with moving from a polluted residential area to a clean area, and vice versa?
6. Author’s response: No we don’t have enough power to study the incidence of disease associated with moving from a polluted to a clean area and vice versa.

7. The authors speculated that there is an unmeasured risk factor behind the observed ‘protective effects’ on P17. To which unmeasured risk factors are the authors referring to?

7. Author’s response: Unfortunately, we don’t think of any particular risk factor that could explain this. However, any difference between groups has to depend either on different health-seeking behavior, or due to different disease incidence depending on some underlying factor.

Response to Reviewer 2.
1) Selection of population. 18230 subjects were excluded because of no CHC-questionnaire data. It’s about 65% of your initial cohort, and 70% of cohort with exposure data. I think this big reduction of the “registry based birth cohort” has to be accompanied by an analysis of possible changes in the structure of the population, in particular for potential confounders. No information on this is reported.

1. Authors response: The only confounder information we have on the overall level is sex and city. We can thus not present much information on the change in population structure. However we think it is strengthening our results that the unadjusted results for the larger cohort are the same as for the more selected population answering the CHC-questionnaires, despite possible differences in population structure.

2) Related to previous point, the description of study population (Table 1) reports 96% of population with no problems to pay bills and 73% of population with >12 years of education of any parents (2 of 3 indexes used to define individual SES). A previous paper, cited by authors, indicates that 82% of Scania population and 81% of Malmoe population is at low education. I understand the use of diverse metrics in the 2 studies (individual vs aggregate by census), but the difference is very relevant. Also the analysis restricted to children with high socio-economic status regarded 3464 cases (44%). There is a suspect of selection bias due to filling of questionnaire, and must be considered by authors. This shift towards high SES could explain also lower exposure levels to NOx, as reported by authors.

2. Author’s response: High level of education was by Stroh et al classified as post-secondary education for three years or longer (#15 years of school) (9.6%). 18.6 % of the population in Stroh et al had any post secondary education (>12
years of schooling). In, our article we classified high education as having “any parent” with a post secondary education (>12 years in any parent) (73%).

Also the age distribution in Stroh et al. is a working population of 25-64 year, with probably more equal distribution on the different age groups. In contrast, since our population consists of parents to young children, we are likely to have an age distribution among them that is younger. A younger generation is likely to have higher education due to the rapid expansion of the post-secondary education. However, this is only an assumption, since we don’t have access to parental age in the current data set.

The difference in socio-economic status is thus not as wide as it appears. However – it does not explain all of the difference, and we agree that there is likely to be a selection towards higher socio-economic strata in those who answered the CHC-questionnaire. We have now added a sentence in the beginning of the discussion section, acknowledging that there is likely socio-economic selection due to the CHC-questionnaires (page 16): “There are still some possibilities of selection bias due to questionnaire data in this study, since the confounder information was only available through CHC-questionnaires, which is likely to have lead to a selection towards high socio-economic status among those who answered the CHC-questionnaire..”

3) I understand that a birth cohort could be not representative of total population (although authors, in the introduction, correctly argued the need of registry based studies), but in this case this topic is important, because the influence of socioeconomic status on drug use (the main proxy of outcome) is reported in literature (see among others Laurent et al., 2009). Do the authors have more to say on the issue, based on their collection of data? [Laurent O. et al. (2009) Influence of socio-economic deprivation on the relation between air pollution and β-Agonist Sales for Asthma. CHEST, 135(3): 717-723]

3. Author’s response: Socio-economic status could be important for two reasons, either 1) because people with low SEI may be more susceptible to traffic exposure, or 2) because factors related to low SEI can confound the associations between traffic and disease.

We think that we based on our data, can try to exclude socio-economic confounding in our results, since we can adjust for many factors. However, we can not contribute to knowledge about potential effect modification – even if our data indicates that people with low SEI consume less B2-agonists, we can not say if this is due to less need, or due to different health-seeking behavior and low compliance.

4) Information on SES at census level for all with exposure at birth addresses (26128) could help in a more efficient sensitivity analysis adjusted for SES (eTable3), to understand the potential role of selection of study population.
4. Author’s response: Unfortunately, we have no information on SES on census level in the current study. However, we agree that this had been desirable, and will try to obtain this for follow-ups.

5) Case definition is based on dispensed medication. Is there relevant differences, in Swedish experience, between dispensed and prescribed data? A reason to prefer prescribed data (in addition to the one mentioned by the authors at the beginning of discussion) could be that you are closest to the moment of diagnosis: it is the doctor who prescribes because of asthma (or not). Some information on this issue could be useful.

5. Author’s response: We speculated in the article that dispense of medication is affected by SEI by affordability, which can of course be a large source of bias. However, a recent Swedish study did not find income to be a major determinant for dispense of drugs, however, they found indications for a prescription bias with higher prescription toward those with high socio-economic status. This shows that there may still be bias in prescription data, and of course there may also be different health-seeking behavior.

Page 16: We added a reference 32 (Nordin 2013, Socioeconomic inequalities in drug utilization for Sweden: Evidence from linked survey and register data.) to the discussion section, which is the recent Swedish study which found indications of a prescription bias toward those with higher socio-economic status.

6) The authors note a greater exposure high traffic for those born in 2006. This strange proportion should be discussed. Did authors find a similar pattern for exposure to NO2? The traffic data on which exposure definition is based are the same over the whole period, or different for each year? If so, traffic data should be verified for 2006.

6. Author’s response: Yes it is the same for NOx as for traffic intensity, approximately 40% of the children with high exposure were born in 2006. This is due to that a large absolute number of children in 2006 got a CHC-questionnaire at the CHC-centre, so there is also a large absolute number of children in Malmö that year (which is the most highly polluted municipality). There was also a slightly higher relative frequency of people in CHC-areas with higher traffic pollution compared to 2008 which also was high total number of participants. Peoples exposure are based on two different versions of the NOx-database updated in year 2001 (for 2005-2006) and year 2008 (for year 2007-2010). However, there is no large differences between those NOx-databases, and the results are fully explained by the large number of children in Malmö 2006. We adjust for year in the multivariable models. In addition, we conducted a sensitivity analysis where children from 2006 were excluded, which did not change the results. We now added a sentence in the manuscript that also NOx-exposure was high in 2006 (page 13, covariate description), and that
analysis excluding children born 2006, does not change the results (page 12 & page 15, sensitivity analysis).

7) Authors use an unusual metric to define high traffic roads (HTR). Following another study in the area, they choose a cut-off of 6 vehicles/min to distinguish between HTR and other roads. Most studies consider HTR as road with more than 10000 vehicles/day, that corresponds to 6.99 in authors’ metric. Could authors give a description of the distribution of this variable, to better understand if there is a big difference (or not, as reasonable) between these 2 definitions of HTR?

7. Author's response: The GIS-analyses were done so they measured distance to roads of the following categories (0-2, 2-6, 6-10, #10 cars per minutes (corresponding to 0-2880, 2880-8640, 8640-14400, #14400 cars/day) within 100m. We then identified which of these categories was the heaviest road within that distance. We can therefore not display the distribution of people with >10000 cars/day within 100 m with the current data set. We chose the categories of traffic counts to be consistent with our own previous studies, but in further studies we will consider to correspond better to what is the current most common practice. The prevalence of a car with >10 cars per minute within 100m of birth-address (14400 cars/day) was for the cohort with confounder information n=120, 1.5%. While the number of people with a traffic intensity 8640-14400 cars/day are n=1947, i.e. 24.7% in the the cohort with confounder information. These categories were thus merged for power reasons.

We corrected the exposure category limits (incorrectly displayed as 0-5 and >6, instead of 0-6 and #6 cars/min). To ease comparability with other studies for the reader, we now display this in cars/day (0-8640, # 8640) throughout the manuscript.