Author’s response to reviews

Title: Changes in triggering of ST-elevation myocardial infarction by particulate air pollution in Monroe County, New York over time: a case-crossover study

Authors:

Meng Wang (meng_wang@urmc.rochester.edu; meng.wang@heart.rochester.edu)

Philip Hopke (Philip_Hopke@URMC.Rochester.edu)

Mauro Masiol (mauro.masiol@gmail.com)

Sally Thurston (sally_thurston@urmc.rochester.edu)

Scott Cameron (Scott_Cameron@URMC.Rochester.edu)

Frederick Ling (Fred_Ling@URMC.Rochester.edu)

Edwin van Wijngaarden (Edwin_van_Wijngaarden@URMC.Rochester.edu)

Daniel Croft (Daniel_Croft@URMC.Rochester.edu)

Stefania Squizzato (stefania.squizzato81@gmail.com)

Kelly Thevenet-Morrison (Kelly_thevenet-morrison@urmc.rochester.edu)

David Chalupa (David_Chalupa@URMC.Rochester.edu)

David Rich (David_Rich@urmc.rochester.edu)

Version: 1 Date: 12 Apr 2019

Author’s response to reviews:

Note: all page and line numbers listed below apply to the clean version of the manuscript.

Review #1

Major revisions:

1. The air quality policies and economic changes cited in the title are not well defined, and simple time windows are used as a proxy of the interventions. This is a clear limitation of this study, which should be discussed by the authors and accounted for in interpreting the results of the analyses. I think that also the title should be more adherent to the actual analysis performed
in the paper: a comparison between time-periods characterized by different levels of air pollution/particles composition.

As suggested, we have added text to describe this limitation in the discussion section (clean version, page 20 line 417 – 430), with a figure included in the supplemental material showing SO2 emissions from coal-fired power plants over time as an example of emission changes. The title was also changed to “Changes in triggering of ST-elevation myocardial infarction by particulate air pollution in Monroe County, New York over time: a case-crossover study”

2. Due to fact that the authors perform a large number of analyses, selecting the "significant" results on the basis of the p-value can be misleading (Figure 1). In fact, one can argue that the "significant" results could be false positives. This point should be discussed in the paper and the results interpreted accounting for this problem.

For our main research question of effect modification by period, we draw our conclusions primarily based on patterns of associations across lag times, rather than just significance testing. Further, the pattern of a larger increase in the rate of STEMI associated with each IQR increase in pollutant concentrations in the AFTER period compared to the BEFORE and DURING period was consistently observed for multiple air pollutants (UFP, UFP 11-50nm, BC, and SO2). However, as suggested, we have discussed the limitation of inflated type I error in the assessment of the main effect of air pollution in the discussion section (page 21, line 451 – 454).

Figure 1 is to illustrate the pattern of effect modification by period for different pollutants. For Figure 1, we only choose pollutants with an observed main effect, and lag hours for these pollutants with the largest main effect that was statistically significant. We could include all pollutants and all lag hours in the figure, but then it would be redundant with Table 4.

3. In the analysis of epidemiological time series, influenza epidemics are usually considered a confounder of the air pollution effect. Did the authors include this term in the model? Was information about influenza available at the individual level?

Similarly, the potential confounding effect of holidays should be accounted for.
First, the study design was a time-stratified case-crossover study rather than time series. A case-crossover design is analogous to a matched case-control study. However, instead of contrasting air pollutant exposures between someone with disease (case) and someone without disease (control), it contrasts pollutant concentration immediately before the acute STEMI (case period) to other time periods when the same patient did not have an acute STEMI (control period). In this study design, control periods were matched to case period by calendar year, month, weekday, and hour of the day, and thus confounding by temporal factors (e.g. hour of the day, weekday, season, and long term time trend) and any non-time varying personal characteristics (e.g. race, health history, genetic traits, etc.) is controlled by design without any need for further statistical adjustment.

We do not think it is likely that Influenza is a confounder in this study. If influenza is associated with ambient air pollutant concentrations, a mechanism may be that high levels of air pollution increase the risk of respiratory infection. If so, influenza is a result of air pollution rather than a cause. Adjusting for a descendant of exposure is not appropriate since it will lead to a bias towards the null if influenza lies on the causal pathway from air pollution to STEMI. Therefore, a patient’s influenza status at the case and control periods should not be controlled for in the model assessing the main effect of air pollution on STEMI.

Second, as suggested, we performed additional analyses adjusting for holiday in the model assessing the main effect of air pollutants on the rate of STEMI. The results were essentially the same as those obtained from the models without holiday. We had added text in the Methods (page 10, line 190-191) and Results (page 13, line 255-256) section describing this sensitivity analysis. A table of the results with holiday adjusted was included in supplemental material (Table S2).

4. The authors write that the characteristics of the patients are not well balanced among time periods. Thus, the estimated interaction between time period and air pollution could be indicative of a different association between exposure and outcome in patients with different clinical/individual characteristics (frailer/aged people could be more susceptible to high exposures than less frail/younger ones), rather than of a difference among periods. In the supplemental material, the authors show that there are not "significant" interactions between patients' characteristics and exposure. I think that this analysis is not sufficient. Did they try to include in the main models interaction terms between patients' characteristics and exposure? Sensitivity analyses should be conducted to evaluate if the results are robust to the introduction of these terms in the model.
As suggested, we performed additional analyses adding interaction terms between air pollutants and patients' characteristics in models assessing effect modification by periods. Specifically, for each air pollutant and lag hour, we added an interaction term between the pollutant concentration and age (<65 years and ≥65 years) to the model. We then repeated this analysis for all patient characteristics with unbalanced distributions across the three periods (sex, smoking, diabetes, dyslipidemia, heart failure). The pattern of larger increased rate of STEMI associated with increased UFP, UFP-11-50nm, BC, and SO2 concentrations in the AFTER period compared to the BEFORE and DURING periods, when including these additional interaction terms little changed from our main analysis. We have now included these new results for the rate of STEMI associated with interquartile range increase in UFP concentration at lag hour 0 by period and by patient characteristics in the supplementary materials (Table S4). We also added text in the Methods (page 10 - 11, line 206-211), Results (page 13, line 267-269), and Discussion section (page 20-21, line 435-443) describing this sensitivity analysis.

5. I suppose that a problem in including terms involving personal characteristics in the main model is related to the presence of missing values. Did the authors consider the opportunity to perform some kind of imputation of the missing values?

We do not have sufficient information needed to build models to predict patient characteristics and thus cannot impute missing values.

6. Is a linear term sufficient to describe the effect of temperature? I would expect an excess of events during both very cold and very hot days. I suggest to include the results of the preliminary analysis on the temperature-outcome relationship in the supplemental material. Moreover, the sensitivity of the results when changing the shape of the temperature-outcome relationship should be evaluated as well as the presence of possible interactions between temperature and patient's characteristics.

As we described in the Statistical Analysis portion of the Methods section (page 9-10, line 183 – 186), we examined whether temperature and relative humidity should be modeled as linear terms or with natural splines with 2, 3, or 4 degrees of freedom. Based on Akaike’s information criterion, the best functional form (i.e. the model with the smallest AIC) for both temperature and relative humidity across all lag periods was one degree of freedom (linear). A table of model AIC for temperature and relative humidity by degrees of freedom and lag periods is shown below (Table 1). However, a thorough examination of the temperature, relative humidity, and air
pollution interactions and their independent and combined effects on the rate of STEMI in these patients is a separate and large analysis. This is beyond the scope of this study, and any such set of analyses will be described in a future publication.

Table 1. Model AIC by degrees of freedom and lag hours

<table>
<thead>
<tr>
<th>Lag</th>
<th>df</th>
<th>Temp</th>
<th>RH</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2642.411</td>
<td>2643.286</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2643.668</td>
<td>2644.751</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2645.188</td>
<td>2645.65</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2647.179</td>
<td>2645.478</td>
</tr>
<tr>
<td>0-2</td>
<td>1</td>
<td>2627.761</td>
<td>2629.107</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2629.273</td>
<td>2630.892</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2630.424</td>
<td>2630.786</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2632.451</td>
<td>2630.359</td>
</tr>
<tr>
<td>0-11</td>
<td>1</td>
<td>2639.429</td>
<td>2640.417</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2641.397</td>
<td>2642.416</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2641.651</td>
<td>2643.096</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2641.556</td>
<td>2642.504</td>
</tr>
<tr>
<td>0-23</td>
<td>1</td>
<td>2640.964</td>
<td>2641.958</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2642.938</td>
<td>2643.894</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2643.451</td>
<td>2645.567</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2641.926</td>
<td>2647.417</td>
</tr>
<tr>
<td>0-47</td>
<td>1</td>
<td>2626.889</td>
<td>2623.438</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2628.866</td>
<td>2625.349</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2628.738</td>
<td>2623.772</td>
</tr>
</tbody>
</table>
7. In Table 4, the interaction tests are not reported. Is there a reason to report such tests for the confounders but not for the exposures?

We have removed the P value for interactions between pollutants and patient characteristics in Table S3 as this was not used for inferences. As described above in our response to comment #2, we made conclusions on effect modification by period based on similar patterns of effect modification across the three periods (i.e. a larger increased rate of STEMI associated with increased air pollutant concentrations in the AFTER period compared to the BEFORE and DURING periods) observed for multiple air pollutants (UFP, UFP-11-50nm, BC, and SO2), rather than significance testing.

Minor comments

1. pag 7. "If a patient experienced multiple STEMIs during the study period, we only included STEMI that occurred at least 72 hours after a previous STEMI, resulting in 921 STEMI events in 912 patients available for analyses". The correlation arising from considering multiple episodes for the same patient is likely negligible (the number of multiple events is very low), but this point should be discussed in the paper. In particular, did the authors account for within subject correlation or not?

Do sensitivity analysis accounting for the within-subject correlation?
As suggested, we have clarified in the Methods section (page 10, line 196-199) that since the correlation among data resulting from multiple STEMI events contributed by the same patient is likely negligible, we did not account for within-subject correlation in the analyses.

2. pag 8. "Case periods of this study were defined as the 1, 3, 12, 24, 48, and 72-hour periods prior to the time of STEMI symptom onset, with control periods (3-4 per case, depending on the number of days in calendar month) matched by calendar year, month, weekday, and hour of the day". This description is not completely clear. It seems to me that it confounds the definition of the exposure with the definition of the control periods.

In a case-crossover design, individuals within a cohort who experience the outcome events (STEMI in our study) are identified, and information about each subject’s exposure (air pollution concentrations in our study) during a case period prior to the event is compared with that individual’s exposure distribution at other times (control period). For example, if a STEMI occurred at 12pm on Monday May 10th, we would compare the mean air pollutant concentrations from 12pm on Monday May 6th to 12pm on Monday May 10th (96 hour lag) to the mean air pollutant concentrations in the 96 hours before 12pm Monday May 3rd, 17th, 24th, and 31st. Since the case and control periods are separated by 168 (24*7=168) hours, the 96 hour mean air pollutant concentration of the case period will not overlap with the 96 hours of any of the control periods. Since all the other lags assessed are for shorter than 96 hours, there is no overlap between periods. We have included text on page 9 (lines 165-167) to make this clearer.

3. pag. 10. Defining the average concentrations of the pollutant in the 1 hours prior to each symptom "lag hour 0" seems to me misleading.

If the time of symptom onset of a STEMI event was estimated to occur in the first 29 minutes of the hour (e.g. 12:28 or 12:02), then lag hour 0 was defined as the previous hour (i.e. 11:00-11:59). If the STEMI symptom onset time was estimated in the 30th minute or after (e.g. 12:39 or 12:58), then lag hour 0 was defined as that same hour (i.e. 12:00-12:59). We have added text to the manuscript to clarify this description (page 9 lines 176-179).

4. pag. 10. "The IQRs used to scale effect estimates were the IQRs of the pollutant specific lag times, from the control periods during the entire study years (2005-2016)". This statement is not
clear. Did the authors use only the control periods to calculate the IQRs? Is there a reason for this?

For example, the IQR used to scale the relative rate estimate for UFP at lag hour 0 is 3702 particles/cm³. This is the interquartile range of the UFP concentration at lag hour 0 in all control periods from 2005 to 2016. Control period IQR was used, because in a case-crossover design, control periods are meant to represent the exposure distribution in the person-time that give rise to cases. Thus, these are valid estimates of pollutant distributions.

In addition to the changes in response to the reviewer’s comments, we made a few other minor changes to update references and aid in completeness and interpretation in results.