Author’s response to reviews

Title: Estimating policy-relevant health effects of ambient heat exposures using spatially contiguous reanalysis data

Authors:

TEMILAYO ADEYEYE (Temilayo.Adeyeye@health.ny.gov)
Tabassum Insaf (tabassum.insaf@health.ny.gov)
Mohammad Al-Hamdan (mohammad.alhamdan@nasa.gov)
Seema Nayak (Seema.Nayak@health.ny.gov)
Neil Stuart (neil.stuart@noaa.gov)
Stephen DiRienzo (stephen.dirienzo@noaa.gov)
William Crosson (bill.crosson@nasa.gov)

Version: 1 Date: 30 Nov 2018

Author’s response to reviews:

Reviewer Response Document

We would like to thank both reviewers for their extensive and helpful comments. We have made extensive revisions to address the concerns raised in their reviews. Please find below our systematic responses to the specific questions.

Reviewer #1: The conducted study aims to assess the necessity to use spatially resolved temperature exposure data in health impact studies to be able to accurately set temperature thresholds for health interventions. Spatially resolved temperature and air pollution data was linked to addresses to get exposure data on an individual level. The health outcomes investigated in the study was emergency department visits due to heat stress, dehydration, acute kidney failure and cardiovascular diseases in New York State.

I believe that the study is well conducted but there are needs for clarification, reformulation and maybe reductions and/or additions in the manuscript. My points are mainly on larger issues and further comments may be added at a later stage.

Main points
The aim is to prove the usefulness of remote sensing data when refining exposure-response functions for regions where observed data from air-monitors may be sparse. This is also claimed to have been proven in the conclusion part of the abstract. To do this however, I would think that a comparison with an analysis using a fixed weather station as exposure is needed to claim that any additional information is gained by this method. The possibility to use a more individual exposure is, intuitively, a better way to estimate exposure but I do not think that this study shows that. It is also a bit odd to state that the study demonstrated something in the aims section.

We apologize for the lack of clarity. The reviewer is correct that we do not provide comparisons with a study of regional air monitors to fully validate the utility of the NLDAS data. We have reworded the aims to clarify this distinction. However, most studies that use in situ air monitor data are restricted to urban areas. In New York State, there are large regions especially in rural areas where air monitors did not exist for much of the study period. In such areas lacking air monitors, reanalysis data is an attractive option and provides uniform spatial coverage for rural as well as urban areas. We have done extensive ground-truthing of the NLDAS dataset in NYS (Eleizer et al, under review at Remote Sensing) to demonstrate that the NLDAS dataset correlates well with in situ data in areas with air monitoring stations. We have reworded the aims to reflect that this study was “to demonstrate the usefulness of the incorporation of re-analysis data (12-km NLDAS products) in refining exposure-response functions for regions where observed data from air-monitors are unavailable” (Page 5, paragraph 2). We have also added a comment in the limitations section to suggest that monitoring data from dense networks are needed to provide evidence that the uniform spatial scale of the NLDAS dataset provides a better estimate of heat-health associations (Page 18, paragraph 2).

In the methods section it is unclear whether the temperatures at different lags where all included at the same time in the analysis or separately. This should be stated and depending on how temperatures were included in the model more questions might arise.

In the manuscript, we present results where the temperature lags were modeled separately. This has been further clarified in the manuscript (Page 11, paragraph 4). Recent studies of health impacts of high ambient temperatures have used either distributed [1, 2] or single lags [3] with case-crossover analysis. Modeling multiple lags simultaneously may produce high levels of multicollinearity and unstable results as temperature values are generally correlated [4]. However, there is some evidence that single lag models may be negatively biased if exposure values are not measured daily [5]. We do not expect this to be the case as the NLDAS data does not have any temporal gaps. We present below a comparison of effect estimates from separate models (Table A) and effect estimates for a distributed lag model with all lags included in the same model (Table B). The estimates for the model with the distributed lag structure are smaller in magnitude and there is some loss of precision, however, the conclusions remain virtually unchanged using the alternative specification. We have retained the single lag models in the manuscript but include the distributed-lag results for the reader’s reference in the appendix (A7).

Table A. Effect estimates for lags 0-7 for multiple health outcomes using individual lag models.
### Table B. Effect estimates for lags 0-7 for multiple health outcomes with all lags in the same model.

<table>
<thead>
<tr>
<th>Maximum Temperature Lag</th>
<th>Heat Stress</th>
<th>Dehydration</th>
<th>Acute Kidney Failure</th>
<th>CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.366 (1.347, 1.386)</td>
<td>1.024 (1.021, 1.028)</td>
<td>1.012 (1.008, 1.016)</td>
<td>0.997 (0.996, 0.998)</td>
</tr>
<tr>
<td>1</td>
<td>1.216 (1.204, 1.229)</td>
<td>1.018 (1.015, 1.021)</td>
<td>1.017 (1.014, 1.021)</td>
<td>0.997 (0.997, 0.998)</td>
</tr>
<tr>
<td>2</td>
<td>1.143 (1.133, 1.153)</td>
<td>1.013 (1.010, 1.015)</td>
<td>1.013 (1.010, 1.016)</td>
<td>0.999 (0.998, 1.000)</td>
</tr>
<tr>
<td>3</td>
<td>1.107 (1.097, 1.117)</td>
<td>1.011 (1.008, 1.014)</td>
<td>1.010 (1.007, 1.013)</td>
<td>1.000 (0.999, 1.001)</td>
</tr>
<tr>
<td>4</td>
<td>1.073 (1.064, 1.082)</td>
<td>1.009 (1.006, 1.012)</td>
<td>1.006 (1.003, 1.009)</td>
<td>1.001 (1.000, 1.002)</td>
</tr>
<tr>
<td>5</td>
<td>1.046 (1.037, 1.054)</td>
<td>1.009 (1.006, 1.011)</td>
<td>1.003 (1.001, 1.006)</td>
<td>1.000 (1.000, 1.001)</td>
</tr>
<tr>
<td>6</td>
<td>1.015 (1.006, 1.023)</td>
<td>1.007 (1.004, 1.009)</td>
<td>1.005 (1.002, 1.008)</td>
<td>1.001 (1.001, 1.002)</td>
</tr>
<tr>
<td>7</td>
<td>0.976 (0.968, 0.984)</td>
<td>1.002 (0.999, 1.005)</td>
<td>1.005 (1.002, 1.008)</td>
<td>1.002 (1.001, 1.002)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maximum Heat Index Lag</th>
<th>Heat Stress</th>
<th>Dehydration</th>
<th>Acute Kidney Failure</th>
<th>CVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.208 (1.198, 1.219)</td>
<td>1.023 (1.020, 1.025)</td>
<td>1.010 (1.007, 1.013)</td>
<td>0.997 (0.996, 0.997)</td>
</tr>
<tr>
<td>1</td>
<td>1.135 (1.127, 1.142)</td>
<td>1.017 (1.014, 1.019)</td>
<td>1.015 (1.012, 1.017)</td>
<td>0.998 (0.997, 0.999)</td>
</tr>
<tr>
<td>2</td>
<td>1.094 (1.088, 1.101)</td>
<td>1.011 (1.009, 1.013)</td>
<td>1.011 (1.009, 1.013)</td>
<td>0.999 (0.999, 1.000)</td>
</tr>
<tr>
<td>3</td>
<td>1.075 (1.069, 1.081)</td>
<td>1.010 (1.007, 1.012)</td>
<td>1.009 (1.007, 1.011)</td>
<td>1.000 (0.999, 1.001)</td>
</tr>
<tr>
<td>4</td>
<td>1.050 (1.044, 1.056)</td>
<td>1.007 (1.005, 1.009)</td>
<td>1.006 (1.003, 1.008)</td>
<td>1.001 (1.000, 1.001)</td>
</tr>
<tr>
<td>5</td>
<td>1.026 (1.020, 1.032)</td>
<td>1.006 (1.004, 1.008)</td>
<td>1.003 (1.000, 1.005)</td>
<td>1.000 (1.000, 1.001)</td>
</tr>
<tr>
<td>6</td>
<td>1.003 (0.998, 1.009)</td>
<td>1.004 (1.002, 1.006)</td>
<td>1.004 (1.001, 1.006)</td>
<td>1.001 (1.000, 1.001)</td>
</tr>
<tr>
<td>7</td>
<td>0.974 (0.969, 0.980)</td>
<td>1.000 (0.998, 1.002)</td>
<td>1.003 (1.000, 1.005)</td>
<td>1.001 (1.000, 1.001)</td>
</tr>
<tr>
<td>Lag</td>
<td>Maximum Heat Index</td>
<td>Heat Stress</td>
<td>Dehydration</td>
<td>Acute Kidney Failure</td>
</tr>
<tr>
<td>-----</td>
<td>--------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>0</td>
<td>1.283 (1.262, 1.305)</td>
<td>1.017 (1.012, 1.021)</td>
<td>1.001 (0.996, 1.006)</td>
<td>0.998 (0.997, 0.999)</td>
</tr>
<tr>
<td>1</td>
<td>1.064 (1.048, 1.081)</td>
<td>1.007 (1.003, 1.012)</td>
<td>1.015 (1.010, 1.020)</td>
<td>0.998 (0.997, 0.999)</td>
</tr>
<tr>
<td>2</td>
<td>1.040 (1.024, 1.056)</td>
<td>1.001 (0.997, 1.005)</td>
<td>1.001 (0.996, 1.005)</td>
<td>1.000 (0.999, 1.002)</td>
</tr>
<tr>
<td>3</td>
<td>1.000 (0.984, 1.016)</td>
<td>1.003 (0.999, 1.007)</td>
<td>1.006 (1.001, 1.010)</td>
<td>1.000 (0.998, 1.001)</td>
</tr>
<tr>
<td>4</td>
<td>1.012 (0.997, 1.028)</td>
<td>0.999 (0.995, 1.004)</td>
<td>0.999 (0.994, 1.003)</td>
<td>1.002 (1.001, 1.003)</td>
</tr>
<tr>
<td>5</td>
<td>0.997 (0.982, 1.012)</td>
<td>1.003 (0.999, 1.008)</td>
<td>0.997 (0.993, 1.002)</td>
<td>0.999 (0.998, 1.000)</td>
</tr>
<tr>
<td>6</td>
<td>1.000 (0.985, 1.014)</td>
<td>1.002 (0.998, 1.006)</td>
<td>1.002 (0.998, 1.007)</td>
<td>1.001 (1.000, 1.002)</td>
</tr>
<tr>
<td>7</td>
<td>0.962 (0.951, 0.973)</td>
<td>1.000 (0.996, 1.003)</td>
<td>1.003 (1.000, 1.007)</td>
<td>1.001 (1.000, 1.002)</td>
</tr>
</tbody>
</table>

The results section needs quite a substantial revision. The use of heat-impacted illness, heat-related illness (mixed up with heat stress on line 190?), heat-stress related, heat-related hospital visits etc makes the results section very hard to follow. This needs to be better structured.

We apologize for the lack of clarity. The use of the heat-impacted and heat-related illness relative to heat stress has been reviewed and restructured to clearly identify all heat-health outcomes throughout the manuscript.
In the results section the authors present a comparison between the exposure-response functions between maximum temperatures, the heat index and heat stress (Figure 3). This analysis is only commented briefly in the results section and not in the discussion section. This should either be removed or be revisited at some point.

All analyses were carried out on both maximum temperature and heat index. Similar trends were observed for both temperature indicators. The NWS uses heat index to issue advisories and therefore the final recommendations were based on the heat index results. We have removed the comparison from the results section.

The added analysis is based on heat stress, the health outcome that is by far the smallest investigated group with 8000 out of almost 1 million cases. It might be of more interest to present another health outcome that might have larger health impacts.

We have added the threshold analysis results for other health outcomes for the reader’s information (Figure 3).

The joint effects are only presented for heat stress (table 2) and dehydration (table s1) while cardiovascular disease is not mentioned at all. However, if the associations were estimated they should at least be mentioned in the results section especially as cardiovascular ED visits/hospitalisations where by far the largest group investigated.

The joint effect results for other outcomes such AKF and CVD have been added to the supplementary document. We are limited by the journal requirements on total number of tables and figures. (Appendices A3 – A5, Tables S1 – S3)

I have a hard time trying to find the actual temperature knots or thresholds, from which the mortality increase is estimated for the different health outcomes. I can't find them, so they are either missing or they are too hard to find.

The knots were placed in 5-degree regularly placed increments. We have revised the threshold analysis to identify trigger points as suggested by reviewer 2 (Page 10, paragraph 3).

On line 263 the authors claim that the effects persist up to 4 days after exposure. It might extend further but as the study have limited the lag times to 4 days the study can only show that the effects persist at least 4 days. The choice to use 4 days as the maximum exposure lag time is not explained or strengthened by any references.

Thank you for the suggestion. We have extended analysis to a 7-day lag period to highlight the extended effect of temperature on heat-related illnesses. For heat stress, dehydration, and AKF, the lag effect extended up to 6-days and has been highlighted in the manuscript text (Figure 2).

The public health implications of using the spatial temperatures are not as substantiated as the authors claim. The difference between the proposed threshold and the old one is mainly that new data and evidence is available rather than the difference between spatial and fixed temperature data.
Air monitor data was not available for all parts of NYS for the period of the study. Data from regional air monitors located sparsely across the state may not be reflective of the actual exposure of New Yorkers in rural and sub-urban areas where there are no monitors. We have revised the public health implications section based on the reviewer’s comments (Page16, paragraph 2).

Reviewer #2: The manuscript "Estimating policy-relevant health effects of ambient heat exposures using spatially contiguous remote sensing reanalysis data" presents an effort to advance our understanding of variability in heat-health risks by incorporating spatially-varying temperature data into a case-crossover analysis. The authors leverage multiyear health records available from New York State to examine how risks vary across a number of outcomes and independent variables. This work is framed around a goal of improving National Weather Service heat advisory criteria in the study region, which the authors note was accomplished as a byproduct of their research and collaboration with the NWS. Based on the health data they analyzed, the authors also report that rural and urban areas had similar risk for heat related illness, which is an important finding given a prevailing emphasis on urban areas in much of the environmental health literature related to extreme temperatures.

The manuscript addresses a topic that is relevant for this journal and the premise of the manuscript should be of interest to the journal's readership. The manuscript does not offer a transformative scientific advance, but does improve knowledge along several different topics of importance to the temperature-health research community.

Regrettably, I found a few shortcomings with the manuscript that constrain my ability to recommend it for publication in its current stage. At times, parts of the manuscript feel misaligned with the overall objective or incompletely address the objective. Not all of the study findings are well supported by or detailed in the methods and results sections. I also suggest that the authors missed some opportunities to contribute new knowledge about temperature-health effects in New York State. Most importantly, I think that there are two important threads to this work that get muddied into a bit of a confusing story. It seems as though the authors are trying to both inform weather service criteria with analysis of health records, and separately, to determine if spatially varying temperature data change and improve their ability to inform those thresholds. I recommend that they separate these goals (and the appropriate methods, results, etc.) in a future iteration of this manuscript. I hope that the authors will consider addressing these comments and wish them well in their ongoing research. Detailed comments follow below.

TITLE

The authors refer to "spatially contiguous remote sensing reanalysis data," but nowhere in the manuscript is it explained how and why it is appropriate to consider NLDAS "remote sensing" data. I think it would be fine to simply refer to the data as a "reanalysis" product. (This same comment applies to all uses of "remote sensing" throughout the manuscript, unless there is a different data set included in the paper that is not well introduced).
The authors acknowledge that the NLDAS is a reanalysis product that was derived from remote sensing and in situ data along with other inputs. We have changed all statements to refer to NLDAS as reanalysis data throughout the manuscript.

ABSTRACT

The first sentence notes that weather service criteria were (historically) based on the frequency of heat events estimated by sparse monitoring data. While this may be true, is it also the case that the criteria were NOT based on analysis of health records? If so, it seems like the manuscript could be oriented around two objectives. First, to inform the weather service criteria with health-based thresholds, and second, to determine if there is value added from spatially resolved temperature data (and how that might influence the thresholds).

We have rephrased the objectives of the manuscript to better reflect the research work and results (Page 2, Paragraph 1).

In the Background section, the authors note that they use "exposure data with a higher temporal and spatial resolution." I understand how they are using exposure data with a higher spatial resolution, but cannot determine where higher temporally resolved data are involved. The authors report most of their results in terms of daily maximum, etc., which is consistent with most contemporary literature.

We agree with the reviewer's assessment, we are using daily data, we have reworded that statement (Page 2, Paragraph 1).

In the Methods section, the authors say, "maximum surface temperature," but I assume they mean "maximum near-surface air temperature given the data sets they are using? I believe they are working with estimates of air, rather than surface, temperature.

We have revised this to “maximum near-surface air temperature” (Page 2, Paragraph 2).

In the Results section, the authors report some of their results as risk ratios with confidence intervals, and other times only report percentage increases. Why are the results reported inconsistently?

All results have been reviewed and we now report effect estimates as risk ratios (Page 2, Paragraph 3).

In the Conclusion section, the authors note that thresholds were changed in the upstate region of New York State, but nothing in the sections leading up to the Conclusions referenced any type of spatially stratified analysis.

We have revised the abstract to include a statement about rural/urban heat-health associations and why a change was needed to the heat advisory threshold for the region (Page 3, Paragraph 2).
INTRODUCTION

Line 78: As noted previously, I am not sure if the remote sensing angle is appropriate to include in this manuscript. It also seems like a bit of a reach to claim that the application of remote sensing data has "overcome" these issues because of the very different nature of the physical variables measured by remote sensing tools (surface temperature) versus reanalysis products and station data (air temperature). The references provided to support this point seem marginally relevant at best.

The introduction section has been reviewed and we have made revisions to decrease the emphasis on remote sensing (Page 4, paragraph 1 & page 5, paragraph 1). The NLDAS data allowed us to estimate health effects in areas without extensive monitoring networks. We have removed superfluous references.

Line 87: Determining heat-health thresholds to inform weather service criteria is not dependent on the availability of fine-scale temperature data; others have done so with the monitoring data that are more commonly available. This is where I believe the authors could separate their manuscript into two different objectives, first to establish thresholds, and second to see if and how the incorporation of the reanalysis data change their conclusions about what the important thresholds might be (and for different regions, etc.). A few references related to threshold identification are below.


Thank you for the suggestion and the citations. We have made extensive changes to the methods based on the suggestion. We have also separated the two different objectives in the introduction (Page 5, paragraph 2) and revised the methods, results and discussion section to clarify the separate sections for each objective.

Line 102: The scale of the meteorological data set is mentioned here, but the scale of the health data is not. Could the authors mention that they used residential address-based health records?
The scale of the health data has been added as requested (Page 5, paragraph 2).

METHODS

Line 140: Although it is a subtle point, the authors might want to note that their maximum temperature (based on hourly data) slightly differs from the maximum temperature data that would be reported from a typical monitoring station with min-max capabilities. This subtle difference can translate into differences in estimated temperature-health effects (see Davis et al. 2015, as an example).


Thank you for the reference. We agree with the reviewer that we are unable to evaluate whether differences in the way maximum temperature data are estimated will influence results. We include Davis et al. in the discussion section to acknowledge this limitation (Page 18, paragraph 2).

Line 141: Could the authors explain how they calculated daily maximum heat index? I assume that they calculated the heat index for each hour of the day and then chose the maximum, but this is not articulated. In future work, perhaps the authors could look at the maximum, minimum, and mean of both dry bulb temperature and heat index?

We conducted preliminary analysis with all temperature indices including minimum and mean temperatures but decided to focus on maximum temperatures and maximum heat index for the final analysis as these measures were most relevant to policy changes under consideration. We have added a statement regarding additional future work needed in this area to the manuscript.

The daily maximum heat index (HI) was calculated from the down-scaled daily maximum temperature, along with the relative humidity at the time of the maximum temperature (one of the available NLDAS 12-km products), assuming that the maximum HI occurs simultaneously with maximum temperature, a reasonable but not perfect assumption[6].

Heat Index (HI) arose from Steadman [7] who referred to a quantity that combined temperature and humidity into an index of comfort both as ‘apparent temperature’ and a ‘temperature-humidity index’. Rothfusz [8] performed a multiple regression analysis using Steadman’s [7] tabular results to obtain an expression for what has become known as the ‘Heat Index’, but is sometimes still referred to as apparent temperature, given (in degrees F) by the following:

\[
HI = -42.379 + 2.049015T + 10.143331R - 0.2247554TR - 6.83783x10^{-3}T^2 - 5.481717x10^{-2}R^2 + 1.22874x10^{-3}T^2R + 8.5282x10^{-4}TR^2 - 1.99x10^{-6}T^2R^2
\] (1)

where \(T\) = air temperature (oF) and \(R\) = relative humidity. \(R\) is calculated from Iribarne and Godson [1981] as:
\[ R = q / \left[ (0.622 \cdot 10^{23.547 - \frac{2937.4}{TK} - 4.9283 \cdot \log_{10} TK} ) / p \right] \]  

(2)

where \( q \) = specific humidity (kg/kg), \( p \) = atmospheric pressure (mb), and \( TK \) = air temperature (Kelvins). NLDAS provides hourly near-surface (2 m above ground) specific humidity and atmospheric pressure data used in (2).

There are many different ways in which heat index has been calculated in environmental research studies. However, we do not expect much difference in heat index values in areas of low humidity and temperate climates such as in New York State[9]. We have included this information in the Appendix A1.

Line 175: Here the authors specify that they calculated risk ratios based on temperatures in degrees C, but earlier specified that their temperature data were in degrees F. In which step of the process did the unit conversion occur?

The temperature data were in degree Fahrenheit (°F) and all analyses were done in °F for evaluating NWS criteria and developing public health outreach materials. This was because NWS uses °F to define their heat advisories. However, for the purpose of this manuscript, the data was converted to °C and the analysis was rerun because of journal requirements that results be reported in S.I units. We have clarified this in the methods. (Page 9, paragraph 3).

In general, I think the methods section could use a few overview sentences that explain the different types of effect estimates the authors are going to produce. In line 180, they note that they assessed urban-rural differences, but exactly what this means is not entirely clear at this point in the manuscript. They also go on to report effect estimates across different demographic variables, and others, but this is never clearly established as part of the approach in the methods section. Please be sure that the methods section explains all of the different tests, models, and effect estimates that readers can expect to find in the results.

We apologize for the lack of clarity, we have made extensive revisions to the Methods and Results section to improve interpretation.

Are the authors combining the emergency department visit and hospitalization data into one time-series for each different diagnosis code? If so, what is the justification for doing so?

The authors combined both emergency department visits and hospitalization data into one time-series for each different health outcomes because of small sample size. For the reviewer’s information we separated the analysis by ED and non-ED cases. The effect estimates are similar in magnitude, so we retain the combined results in the manuscript.

Table C. Effect estimates for lags 0-7 for heat stress stratified by ED and non-ED cases

<table>
<thead>
<tr>
<th>Maximum Temperature Lag</th>
<th>Heat Stress (All)</th>
<th>Heat Stress (ED cases)</th>
<th>Heat Stress (non-ED cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.366 (1.347, 1.386)</td>
<td>1.366 (1.345, 1.388)</td>
<td>1.373 (1.324, 1.424)</td>
</tr>
<tr>
<td></td>
<td>1.216 (1.204, 1.229)</td>
<td>1.206 (1.192, 1.219)</td>
<td>1.275 (1.241, 1.310)</td>
</tr>
<tr>
<td>---</td>
<td>---------------------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>2</td>
<td>1.143 (1.133, 1.153)</td>
<td>1.132 (1.121, 1.143)</td>
<td>1.203 (1.175, 1.231)</td>
</tr>
<tr>
<td>3</td>
<td>1.107 (1.097, 1.117)</td>
<td>1.096 (1.085, 1.106)</td>
<td>1.176 (1.149, 1.205)</td>
</tr>
<tr>
<td>4</td>
<td>1.073 (1.064, 1.082)</td>
<td>1.067 (1.058, 1.077)</td>
<td>1.109 (1.084, 1.135)</td>
</tr>
<tr>
<td>5</td>
<td>1.046 (1.037, 1.054)</td>
<td>1.045 (1.036, 1.055)</td>
<td>1.049 (1.027, 1.071)</td>
</tr>
<tr>
<td>6</td>
<td>1.015 (1.006, 1.023)</td>
<td>1.014 (1.005, 1.023)</td>
<td>1.017 (0.997, 1.038)</td>
</tr>
<tr>
<td>7</td>
<td>0.976 (0.968, 0.984)</td>
<td>0.978 (0.970, 0.987)</td>
<td>0.965 (0.946, 0.985)</td>
</tr>
</tbody>
</table>

RESULTS

Line 190: Following from the previous comment, the different health events that were assessed should be outlined in the methods section. It may be helpful to repeat them here, but I strongly urge that the authors establish their suite of tests first in the methods section.

The different health outcomes assessed have been outlined in both the methods (page 5, paragraph 3) and results section (Page 11, paragraph 2). Results for all health outcomes are now included in Figures 2 and 3. We have also elaborated on all statistical analyses including effect modification and threshold analysis.

Line 206: As was the case in the abstract, here the results are reported as percentage increases but elsewhere results are reported at relative risks. Would it be an improvement to report the results more consistently?

All results are now reported as risk ratios in the Results section.

Line 217: Some of the text in this paragraph feels more appropriate for the methods section - a new methodological technique is introduced related to spline terms, etc.

We have added a section on the threshold analysis to the methods section and edited the results section as suggested.

Line 222: The text here feels fairly subjective amidst what is otherwise quite an objective analysis. The authors report "only shows small increases below" and "much steeper slopes at higher temperatures" and "slope levels off." Would it be possible to use a more objective technique to quantify where these important thresholds are and how the effect estimate changes at different temperature ranges? The Petitti reference above provides one possible approach, but also contains references for other techniques.
The Petitti et al. reference has been reviewed and a more objective method of reporting the threshold analysis results adopted (Page 10, paragraph 3). We provide minimum and excess risk temperatures to better delineate thresholds of health effects. We also provide a comparison of effect estimates at the old and revised heat advisory thresholds (Page 13, Paragraph 1).

Line 226: I'm not sure that it is so interesting that the slope for the heat index is shifted to the right given that, during the summer months in New York State, the heat index will be higher than the dry bulb temperature. I think it would be more interesting to discuss how the model fit compares between different exposure variables. Of the four exposure variables the authors considered, which best fits the data?

The model fit for both maximum temperature and heat index were reviewed. Heat index had a smaller AIC and better model fit. The NWS also issue their heat advisories and warnings using the heat index. Therefore, our recommendations on heat advisories were based on maximum heat index.

Line 236: The language here is confusing - what do the authors mean by "elevated risk albeit smaller in magnitude." To what does "elevated" refer?

The sentence has been revised to: “Although the highest risk of heat stress occurred in the summer months of June, July and August, the cooler shoulder months of May and September also show elevated risk of heat stress.” (Page 12, paragraph 2)

DISCUSSION

Line 248: It is not clear that the references provided here strongly support the claim that inter-individual variation is a "major" source of uncontrolled confounding in heat exposure studies, and one reference is not even topically relevant as it focuses on air pollution. My sense of our contemporary understanding is that inter-individual variation in exposure (at least assigning exposures from spatially resolved temperature data, ignoring true personal exposure), could result in small to moderate shifts in effect estimates. A more appropriate reference might be;

The use of a case-crossover design not the NLDAS data minimized bias due to inter-individual variation. The text has been modified to reflect this. We have added references that discuss the advantages of using spatially resolved data for exposure measurement (Page 13, paragraph 2).

Line 252: The discussion here seems a little too constrained in terms of the factors that should be considered in setting thresholds for NWS heat products and the message content. I agree with the authors that heat morbidity data are useful to include as part of the decision-making guidance, but what about other concerns? Heat products are not only issued to minimize morbidity, but also mortality. Would the results be different had the authors examined mortality data? How does variability in the thresholds found for different morbidity outcomes get accounted for? Beyond those concerns, what about effects outside of the health sector? And what evidence could the authors provide that more warnings would result in more behavior change,
given the statistics reported about behavior change from heat warnings in other research, as well
as concerns about warning fatigue?

Thank you for the comment. This manuscript is part of an ongoing collaboration between the
New York State Department of Health and regional NWS offices to improve heat advisory
definitions and messaging. We intend to continue evaluating advisory criteria and explore effects
on mortality as well as effects outside the health sector. We have planned collaborations with
NWS to assess whether a change in threshold for heat advisories leads to lowering of morbidity.
We have included statements in the limitations and conclusions section acknowledging these
concerns (page 18, paragraph 2 & Page 19, paragraph 1).

Line 282: The claim that gender-related effects vary by study could be supported by a few references.
The references have been updated (Page 15, paragraph 2).

Line 291: This is the first time, as best I can tell, that information specific to New York City appears in the manuscript - the last sentence of the results section only mentions "rural and urban" areas more generally. Can the authors add to the methods and results section to more explicitly describe any separate analysis they ran for New York City? In the same section, why does the coastal location contribute to higher risk? And what basis do the authors have for claiming that the urban heat island effect results from evapotranspiration, which is a cooling process?
The manuscript has been edited to reflect rural/urban (upstate NYS/NYC) differences or similarities. The methods and results sections now clearly state the separate analyses for NYC and the rest of NYS. We have edited the manuscript to remove statements regarding contribution of evapotranspiration or coastal location to Urban Heat Island effects.

Line 338: By what objective basis did the authors come to conclude that a threshold of 35
C was appropriate?

At the pre-existing NWS threshold of 37.8°C (100°F) the risk ratio for heat stress was 3.73 while
the risk ratio for other health outcomes ranged from 1.73 for dehydration, 1.53 for AKF and 1.41
for CVD. In contrast at a reduced heat advisory criterion of 35°C (95°F), the risk ratio for heat
stress is 1.93 and ranges from 1.73 for dehydration, 1.33 for AKF and 1.29 for CVD. The table D
below provides the observed days above various temperatures during the study period.

The number of days above 100°F ranged from 0 to 4 per year for each county. In contrast the
number of days above 95°F ranged from 0 to 9 per year and days above 90°F ranged from 1 to 21
days per year. With the old criteria of 100°F heat warnings were only issued infrequently in the
region. It was decided that a threshold of 95°F would reduce morbidity due to heat but would not
be too frequent to induce warning fatigue. Lower criteria are already in place or being worked on
for high-risk groups such as outdoor workers, student athletes and school children [10-12]. We
have included this information in the manuscript (Page 16-17)
Table D: Average number of days above various temperature thresholds during the study period

<table>
<thead>
<tr>
<th>Year</th>
<th>Days Above 90°F</th>
<th>Days Above 95°F</th>
<th>Days Above 100°F</th>
<th>Days Above 105°F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>2008</td>
<td>5.60</td>
<td>3.67</td>
<td>1.73</td>
<td>1.68</td>
</tr>
<tr>
<td>2009</td>
<td>3.89</td>
<td>2.65</td>
<td>0.66</td>
<td>0.90</td>
</tr>
<tr>
<td>2010</td>
<td>14.63</td>
<td>6.85</td>
<td>4.53</td>
<td>2.41</td>
</tr>
<tr>
<td>2011</td>
<td>10.16</td>
<td>4.50</td>
<td>4.68</td>
<td>2.75</td>
</tr>
<tr>
<td>2012</td>
<td>11.48</td>
<td>7.45</td>
<td>4.23</td>
<td>3.32</td>
</tr>
</tbody>
</table>

CONCLUSIONS

Line 361: I do not believe that this study is the first to use spatially contiguous exposure data to inform heat-health efforts. See, for example, Shi et al. 2016.


Thank you for the reference. We have included this and other references to acknowledge the growing body of literature in this field (Page 13, paragraph2).

Line 365: The claim about transdisciplinary research seems irrelevant to the rest of the manuscript; this conclusion is not supported by the methods and results.

We have removed that statement.

FIGURES AND TABLES

Figure 2, it may be helpful to separate the different outcomes into different figures so the variability in the effect estimates for the outcomes other than heat stress are more easily discernible.

We agree with the reviewer that separating the figures for multiple outcomes improves clarity. Please see the revised Figure 2 for comparison of results.

References


11. Using the Heat Index to Protect Workers [https://www.osha.gov/SLTC/heatillness/heat_index/acclimatizing_workers.html]