Author's response to reviews


Authors:

David M Hondula (dmh4j@virginia.edu)
Robert E Davis (red3u@virginia.edu)
Matthew J Leisten (mjl2fj@virginia.edu)
Michael V Saha (mvs4f@virginia.edu)
Lindsay M Veazey (lmv3p@virginia.edu)
Carleigh R Wegner (crw9a@virginia.edu)

Version: 2 Date: 30 January 2012

Author's response to reviews: see over
Reply to Reviews
MS: 102582161155424
David M. Hondula et al.

January 30, 2012

Dear Dr. Forastiere:

Thank you very much for providing reviews for our submission to *Environmental Health*. We were very encouraged to read the positive and constructive commentary from you and the reviewers. Enclosed you will find a new version of the manuscript with substantial changes that we hope you and the reviewers will find have significantly improved our work. We have also provided a point-by-point reply to each critique in the reviews and a highlighted version of the manuscript identifying places where changes have been made.

We carefully considered and responded to each critique in the reviews. In nearly every case we have made changes to the text in accordance with the ideas provided by the reviewers, and in several cases adjustments have been made to the study methodology. The majority of the changes are in line with your concerns related to our assessment of the study scope and limitations. The discussion section now provides a more detailed perspective of the suite of variables that one might include in a study of heat-related mortality, including variables that we did not directly include in our approach.

As you evaluate our reply and subsequent perspective from the reviewers, we would like to reiterate that ours is one of the first studies to adopt a local-scale perspective to heat-related mortality, and amongst those, stands apart with the use of a multi-decade period of record of mortality observations and a large suite of potential explanatory variables. Advancing research conducted at this scale can directly lead to better public health through the enhancement of alert systems and longer-term strategies that promote efficient resource allocation and reduce thermal stress. There are many well-respected researchers in our field who are working to refine elegant statistical approaches for assessing the temperature-mortality relationship. A thorough understanding of the impact of all factors that modify the temperature-mortality linkage will certainly advance public health applications as well, but we maintain that temperature and humidity are the main drivers of the summertime heat-related mortality. We have chosen to focus our attention in this manuscript on extending this basic knowledge to a more useful scale for practical action, with the recognized tradeoff that other factors may also be important. We hope that you and the reviewers share our perspective that advancing research in this direction has considerable benefits and is of interest to the readership of *Environmental Health*.

We greatly appreciate your continued feedback on our manuscript and look forward to your review of our response.

Respectfully,

David M. Hondula et al.
REVIEWER #1 COMMENTS:

This manuscript represents an innovative approach to defining heat vulnerability. As the authors note, just about all fine-scale spatial models of heat-related mortality are based on case studies (and often limited to only ‘official’ heat-related deaths), which have inherent limitations. The research presented in this manuscript, by taking advantage of more than two decades’ worth of health data and by use of all-cause mortality, has the potential to convey far more salient information on human vulnerability, and more useful information on where mitigation efforts can be concentrated. I believe the manuscript is publishable with only minor revisions. My concerns are outlined below:

What is the direct citation for the Steadman model with the wind speed correction? It doesn’t seem to be in the manuscript (The O’Neill et al. citation doesn't directly use a wind correction). I’m not sure it would impact results much but I’m just wondering where this comes from.

We have updated the citation in the text to reference three separate manuscripts: O’Neill et al. (2003), Steadman (1979), and Kalkstein and Valimont (1986). Apparent temperature regression equations with units of dry bulb and dew point temperature appear in both the O’Neill et al. and Kalkstein and Valimont manuscripts; we used the version printed by O’Neill et al. because it corresponds more closely to the apparent temperature values listed Table 2 of Steadman (1979) at a range of values that we tested. The addition of a wind constant is suggested in all three manuscripts, but Kalkstein and Valimont most clearly articulate that the correction is a simple addition or subtraction from the regression equation. The values for the wind speed correction are interpolated from Table 2 in Steadman (1979).

Though the study spans 1983-2006, the zoning, and socioeconomic variables all seem to come from the latter end of the period of study (either the 2000 census or what I assume are the most recent zoning maps). As Philadelphia has gentrified and otherwise changed in the meanwhile, perhaps a comment should address that these variables themselves are not static, and that there may be some uncertainty in their use.

Our thanks to the reviewer for identifying this important concern. While we do hope to examine the within- and across-zip code relationship between various risk factors and heat-related mortality over time (as well as space), our limited sample sizes in some zip codes requires additional methodological considerations that we believe fall outside of the scope of this particular analysis. We have added the following text to the discussion section:
“The sociodemographic and zoning variables were derived from data available at a fixed point in time (e.g., the year 2000 census). However, the underlying demographics and zoning ordinances both change over time, a process we were unable to capture using this approach. This introduces some uncertainty into the results, and future research should explore local-scale mortality patterns over both space and time.”

The use of the different hourly AT values is quite interesting, and perhaps worth a paper in itself. Was daily mean AT considered as well, or max and min AT? To lend more credence to your use of the hourly AT values, it would be interesting to see if the mortality thresholds were better selected through your method than through the other, more standard metrics. (I’m not sure you need to add much on this in this paper beyond perhaps a comment, but I think it would be useful for future research.)

We appreciate the reviewer’s encouraging words regarding our approach with hourly AT values rather than the daily max, min, or mean, and do intend to continue this exploration in future work. The reviewer asks if the thresholds might be “better selected” through our method than the more traditional metrics. Our interpretation of this inquiry is that the reviewer is asking how the mortality rate associated with this threshold compares to the daily mean, max, and min. If the maximum or minimum apparent temperature were the “best” metrics of heat-related mortality, we should see a pattern in Figure 3b that is similar to (or directly opposite) of Figure 3a. Instead, the pattern seems much more complex. We need to examine how consistent the pattern in Figure 3b is within different time slices for Philadelphia as well as in other locations around the country before we can make a reasonable assertion on this topic. We have added the following text to the discussion:

“Examining effects by hour, rather than using more conventional metrics like daily maximum, minimum, or mean temperature allows for a more specific identification of hazardous periods. These patterns may arise in part from the threshold chosen for each hour: the threshold temperature might be expected to follow an even smoother pattern than that shown in Figure 3a.”

In several places, the association of heat mortality with recreational zoning is discussed. I’m not sure I see where this assertion is coming from. In Table 1, it is positive (.396) in PC5, but is negative (-.404; -.323) in PC1 and PC6. The other characteristics mentioned all make sense, but I am struggling to see your assertion in the results presented.

Upon reexamination of the PC loadings, in particular with respect to recreational zoning, we have removed the claims related to this zoning class from the text. Although recreational zoning is one of the variables most highly correlated with PC5, the correlation is still relatively low. Any interpretation of recreational zoning within PC5 is confounded by its loadings for other PCs with roughly the same magnitude. Our discussion of the lack of a “protective” role for recreational zoning is still relevant and remains in the text; we have made a minor change to the language to
reflect that we did not find a relationship (instead of reporting that more recreational zoning was linked with higher mortality):

“The lack of a strong relationship with recreational zoning is surprising because we expected places with more parks and green space to have lower surface temperatures…”

Line 304 – “Mortality rates were highest when thresholds were exceeded in both the morning and evening hours on the days immediately prior to death.”. Was this actually tested? It suggests you examined mortality when thresholds were exceeded both morning and evening, but I’m not sure this is what you did.

We apologize for the unclear language: we did NOT test for a “paired” exceedance when extreme temperatures were observed both in the morning and evening (although that is an interesting idea). Instead, we were attempting to report that we found higher mortality rates separating days based on morning OR evening temperatures compared with other times of day.
We have removed the word “both” and changed “and” to “or:”

“Mortality rates were highest when thresholds were exceeded in the morning or evening hours on the days immediately prior to death.”

Line 307 – I’m not sure you can be this specific in terms of cause-effect. It looks as though in the Lag 1 late morning window you have a relatively high threshold temperature, which may explain the higher response compared with surrounding hours. I also think this late morning increase may be due to the fact that very hot days will likely have very hot temperatures in the late morning – that is, you may have afternoon thunderstorms affecting the apparent temperature later on, but not during this time of day.

We have taken the reviewer’s advice to more clearly acknowledge that the mortality response is sensitive to the threshold chosen for each hour. Furthermore we have removed the text suggesting a specific cause for the peaks in the response curve. The revised text in the discussion section reads as follows:

“These patterns may arise in part from the specific threshold chosen for each hour: the threshold temperature might be expected to follow an even smoother pattern that shown in Figure 3a. In particular, the threshold temperature seems to increase rapidly on lag 1 between roughly 6:00 and 10:00 a.m., and thus the relatively high values here might be leading to the spike in the response at the same time in Figure 3b. The other peaks in the response curve (Figure 3b) seem less likely to be influenced by variations in the threshold curve. Future work might examine the mortality response above various percentiles of hourly temperature rather than a mortality-based threshold.”

Minor
Line 139, change ‘m/’ to ‘m/s’
Line 248, change ‘ATS’ to ‘ATs’
Line 287, ‘generally correct for nearly all’ is a rather vague term.
Line 299, ‘outstanding’ is an odd word here, both for its alternate definition and that there’s been quite a bit done in this area. Perhaps ‘unsettled’?
Table 3 should be presented more clearly – right now it looks pasted from SPSS.
I would also suggest that you change all the “.000” sigs to “<.001”.

REPLY: The suggested changes have been incorporated into the manuscript. The sentence including the “rather vague term” in Line 287 has been removed. The critical values from Table 3 have been incorporated into the body of the text and the original table has been removed.

Level of interest: An article of outstanding merit and interest in its field
Quality of written English: Acceptable
Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

REVIEWER #2 COMMENTS

Article is of interest as it poses to explore the geographical differences in excess mortality within urban areas due to local socio-demographical and environmental characteristics. However there are some things which are not clear in the analysis which should definitely be clarified and explained more clearly in the text.

Introduction
Introduction is clear but should focus more on features in study to be more effective and to the point. It is worth mentioning that in urban areas especially in the US, high density areas as often those with high UHI intensity and low socio-economic level so when studying these together, especially as a factor analysis is carried out it is worth mentioning the fact they are correlated, especially Often one can be a proxy of the other so what is the benefit/limit?
Please address this!

We have removed the following portions of text from the introduction to meet the reviewer’s suggestion that the focus be improved:

“For most heat-related mortality research to date, the greater metropolitan area has served as the smallest unit of analysis; however…”

“In metropolitan areas, higher temperatures arise from a suite of processes including reduced radiant heat loss, lower wind velocities, increased exposure to radiation, anthropogenic heat sources, atmospheric contaminants, and reduced evaporation [1].”
Our thanks to the reviewer for raising an issue we are very concerned with and interested in—separating the effects of income, education, and exposure when all three are very highly correlated. In a city like Philadelphia where these fields are so closely related, it is nearly impossible using any statistical technique to pinpoint the quantitative impact of each variable on heat-related risk. We hope that our analysis of other cities in the future might reveal clues as to whether this relationship is generalizable. For the purposes of this manuscript, however, we agree with the reviewer that this issue needs to be addressed to the readers. We believe that the text is very explicit that we are not attributing a causal relationship to any of the individual variables, but instead highlighting characteristics of places that are associated with higher mortality during heat events. We have added text to the discussion section addressing the limitations imposed by the high degree of correlation between several of the variables:

“...We observed a high correlation between the surface temperature field and several socioeconomic variables, as evident by the high loadings on the first principal component. Although principal components allows for the examination of potential effects of a large suite of variables believed to be associated with risk, one tradeoff can be difficulty in interpreting the results. We can definitively say that places with higher surface temperatures are associated with higher mortality risk, but those places are also have a high percentage of residents living in poverty and a high percentage of residents without a high school diploma. This pattern has been observed for other cities in the United States [5] and makes it difficult to pinpoint a causal relationship between the individual predictor variables and the health response. Even if it is difficult to separate the effects of individual variables, identifying characteristics of places associated with higher heat-related mortality can lead to improvements in the allocation of medical resources during dangerous conditions. Our future analysis in other cities in the United States where socioeconomic status and surface temperatures may not be as highly correlated may shed light on the relative impact of exposure, education, and income on heat-related risk.”

We have also added to the end of the introductory paragraph in the hopes of both making the study focus more clear and emphasizing the major contribution:

“Past work has primarily focused on a larger scale examining the response of an entire metropolitan area, but the allocation of resources intended to protect citizens from the dangerous effects of heat and humidity could be improved with a more specific knowledge of where the risk is highest within urban areas. The current state of reporting and data availability makes it possible to assess this risk with a multidecadal record of geographically-specific observations. In this manuscript we utilize such a record to evaluate intra-city mortality risk within a major United States metropolitan area..”

Data
Satellite data: why use only 2 days if 47 clear-sky images? An “average” map would have given a more robust idea of UHI phenomenon in Philadelphia, not only during two extreme days”, suggesting the phenomenon is always present but aggravates during extreme conditions. Satellite images are daily while UHI effect is a nighttime phenomenon, hence not convinced this aspect is reported correctly here.
We appreciate the reviewer’s concern and interest this topic as well. We share the reviewer’s sentiment that improvements are possible over the recent “status quo” in the literature of using only one or two images to represent the urban heat island, as originally noted in the text in the discussion section:

“We believe that the use of remote imagery in our study, and others, could be greatly enhanced if more surface temperature images were used. In just the two used in our study, there is variability in the surface temperature pattern that may be linked to seasonal differences, synoptic-scale conditions, or other environmental controls.”

Selecting the proper number of images to include in this analysis was a particular challenge, because, as the reviewer correctly notes, we had a reasonably large sample to choose from even after limitations associated with data quality and cloudiness. However, our analysis focuses on extreme days: 160 days in the record met our apparent temperature threshold criterion out of a total sample of 9490, roughly 3000–4000 of which can be said to fall in the “warm season.” Thus the 160 “high AT days” represent approximately the top 4–5% of all days in the record. We attempted to mirror this percentage with the image selection (settling on 4.2%). Including more images would result in the inclusion of more relatively cool days in the mapping of the urban heat island. Ongoing analysis by the authors is investigating the consistency of the SUHI effect for Philadelphia and other cities at varying air temperatures, and we hypothesize that the effect is not consistent. As the air temperature increases, we believe that the surface temperature in some areas increases at a faster rate than others—one could make this argument based upon Bowen ratio variability alone. Accordingly, including more images might lead to unrepresentative picture of the SUHI effect on the days we are investigating. Additionally, of the 47 images originally downloaded, at least 50% still contained cloud features that obscured one or more zip codes in our study region, reducing their usefulness.

The reviewer makes the point that the canopy layer urban heat island (air temperature) is an effect most pronounced at night. We of course agree with this assertion and wish that we had a high-density thermometer network across the city for continuous spatial and temporal ambient air temperature. There is some support for the use of surface temperature measurements as a proxy for air temperature during the daytime hours, including Nichol’s (2005) assessment:

“The image temperatures recorded appear to represent Oke’s (1976) urban canopy layer heat island, of which it is stated: ‘It would appear that the canopy layer thermal climate is dominated by the immediate site character (especially building geometry and materials) and not by thermally modified air from upwind.’”

Furthermore Imhoff (2010) reports that the surface urban heat island effect is greater during the daylight hours than at night, consistent with previous studies. We have clarified the section in our text specifying that our measurements are of the surface heat island and added the Nichol reference with respect to the possible correlation between the surface and canopy layer fields:

“The surface temperature field may be much different from the air temperature field over the same area, and we are not claiming that the two are identical, although some research indicates a
high degree of spatial correlation between the two fields during daylight hours [24]. There are many aspects of the urban heat island worthy of consideration in the context of urban health, including day/night variability and the contrast between the surface heat island and that of the canopy layer. We are investigating if residents in places with higher morning surface temperatures on hot days are at greater risk.”

At the very least, we believe that it is worthwhile to include a daytime assessment of surface temperatures (that do contribute to the human radiation balance) as a possible explanatory variable for heat stress even if this variable is not a good proxy for ambient air temperature (although it may be).

**Modeling**

A summer only analysis would have been better and standardization procedure more specific to within-summer mortality fluctuations and age-standardization. This part needs to be reworded, to explain what has been done and not necessarily all the single steps of the analysis.

We have adjusted the study methods so that all results and statistics are based on a summer-only analysis. The editor made this suggestion as well, and our reply is:

In all cases in the manuscript where the mortality response is reported as a percentage, that number is was based on the mean warm-season mortality of 0.26 deaths per 10,000 per day. The warm season was defined as days 150–275, and none of the days above the AT threshold fall outside of this window. In this respect, the analysis did focus on the summer months. We have added the following text to the methods section clarifying this point:

“In all cases the mortality response is expressed as a percent difference relative to the mean warm-season mortality in Philadelphia County (or within each ZCTA) of 0.26 deaths per 10,000 residents per day.”

The original significance testing using the randomization procedure was completed using data from the entire year. Random samples could include values from both inside and outside the warm season. Removing the seasonality, we believed, would cause there to be no difference in the random sample “baseline” for comparison based on which particular days were used in the sampling. Based on the reviewer’s and editor’s suggestion, however, we investigated the differences in the random samples drawn using year-round versus warm-season-only data. We found the distributions to be different such that the values drawn for the warm-season only tend to be lower than those for year-round throughout the distribution (Figure R1, below). Thus the test used in the original manuscript was in fact more conservative than needed (the “critical value” for identifying a significant response was higher than it should have been had we used only warm-season values).

We accordingly proceeded to re-run all of the test statistics for the manuscript using a random sampling procedure based on the warm season only. It is important to note that the actual mortality response is not impacted by this change (such as citywide mortality being 9.3% above normal when the AT threshold is met), but the critical values that identify significant differences from the background level could change...
We found that the results presented in the text were not impacted by this change: the overall citywide mortality response remains significant, and the same 10 zip codes that demonstrated a significant response continued to pass the randomization test. No new zip codes were added to the “high-mortality” group with this change. We added the following text to the methods section to reflect this change in the randomization test:

“Samples were drawn exclusively from days falling within the warm season, defined here as between calendar days 150–275 (approximately June–September).”

To the best of our knowledge age-standardization for the summer months only would not change the results in any meaningful way. Population figures for the U.S. are reported in the national census once every ten years, and we are interpolating these values to estimate the age structure on an annual basis. Given that the summer months fall at the “midpoint” of each year in the record, the estimated age structure for the entire year (using our current method) would nearly exactly match what we would find if we attempted to interpolate by month. The errors associated with these estimates certainly are greater than any change we might find by shifting to a monthly standardization.

The reviewer suggests that the methods section be “reworded,” and not “list every single step in the analysis.” In several places the text in this section has been changed in light of methodological changes addressed elsewhere in this reply. With respect to the level of detail included, although we would welcome suggestions for specific areas that merit consolidation, we thought it particularly important to be precise in our descriptions because this is one of the first studies to rigorously examine heat-related mortality at this scale using a long time series.

**Why were 3°C AT bins chosen? Doesn’t this limit sample size? Especially for extreme temperatures? Maybe something could be said about other combinations tested, especially larger bins. A sensitivity analysis would be appropriate.**

We did examine several different combinations of bin size before ultimately settling on this particular AT interval. Bins that are too small, as the reviewer recognizes, leave too few samples to examine average mortality is like within a particular temperature range. Larger bins, however, decrease the precision associated with estimating a threshold temperature. If we were to examine the mortality response in the AT bin from 34–39°C, for example, we would find the mortality rate to be much higher than that reported in the manuscript. However it would be unclear whether the effect was felt at the lower end of the range or the higher end. Identifying the threshold apparent temperature associated with mortality rates significantly above a baseline required the bins to be as small as possible while still preserving sample size. We added the following text to the methods section:

“This particular bin size was chosen based on sensitivity analyses in an effort to balance threshold AT specificity and sample size within each bin.”
And later in the methods section, the following text was added to more clearly state what the final dependent variable is:

“We then calculated the overall mortality response when AT values in excess of the threshold were observed (instead of within each AT bin).”

Please note that the final dependent variable (mortality above a certain AT value) is not sensitive to the choice in bin size. The selection of the AT threshold is, of course, sensitive to the binning parameters. The binning interval we chose is relatively consistent with previously published studies, including [2], [11], and several of those cited in [7]. In Figure R2 we demonstrate that our choice of a bin width of 1.5°C is certainly a “middle ground” in terms of the response. As the bin size increases, more samples are added below the center value of the bin than above. This adds more days to each sample with a lower mortality, until eventually the pattern levels off around a bin width of 5°C where the entire set of hot days is included with no precision sensitive to the actual AT values included in the bin.

![Bin Width - Sensitivity Analysis](image)

**Figure R2.** A sample sensitivity analysis for choosing an apparent temperature threshold based on bin width. Here, the within-bin mortality response is shown for bins centered on 35.5°C ranging in size from 1–10°C. In the manuscript we chose to use a 3°C bin width.

Not too convinced about using different time intervals and how the potential lag effect of exposure to high temperatures is treated here. Is it repeated differently for each ZCTA? If yes, is this plausible and feasible to have differentiated lags by zone? This aspect would be complex to apply to HHWWs, which at this point should be differentiated by zone and hour?
Our apologies to the reviewer for this confusion. We did not examine different time intervals and lags within each zip code, although that could be an interesting topic for future analysis. We certainly agree that a varying time/lag model at the zip-code level would be quite complex in terms of HHWWS application. We only looked at different times of day and lag to find a suitable time for examining local-scale effects; we wanted to identify a specific time where we knew the citywide response was significant enough that there should be at least some zip codes that demonstrate elevated response. We removed the following sentence in the methods section that may have been the source of the miscommunication:

“The procedure was then repeated for each of the 48 ZCTAs in Philadelphia County comparing mortality on all days when the AT at the chosen hour occurred above the threshold to all days when lower ATs were observed”

And the following sentence was added:

“The randomization test was again used to identify significantly elevated mortality at both the County-wide and ZCTA scale, except in this instance the test statistic was based on the cumulative response above the threshold AT instead of within a particular AT bin.”

Results
Tables need to be improved and made more legible, SPSS output not appropriate.

Table 1 has been removed and the regression equation incorporated into the text. We have changed the font and increased the font size in new Table 1 (former Table 2) to improve readability.

Would have been better to include PC2 and PC3 as they do explain 16 and 10% of variance respectively. While 5 and 6 are much smaller and not very clear what they represent in terms of heat related mortality. Which Pc explains the elderly as PC1, and Pc have negative coefficients and Pc5 is very small?

We appreciate the reviewer’s interest in this aspect of the manuscript. While the second and third principal components do explain a higher percentage of the variance of the original variable pool, they were not significantly related to the spatial mortality pattern using multiple linear regression. When any of these variables are included in the regression model, although the overall model significance improves, the partial significance values for PC2, PC3, and PC4 values are all well above 0.05. It would be improper to include them in the final regression model. Additionally, even though PC6 only accounts for a small portion of the total variance in the independent variables, it still could be important in accounting for spatial variability in heat-related mortality. We need to eliminate colinearity between the independent variables but are most interested in finding joint correlation between the independent and dependent variables.

Our interpretation of the principal components is that PC6 is most closely related to the spatial pattern in the elderly population, with a loading of -0.620. The reviewer is correct that both PC5 and PC6 are very small, and clearly the elderly variable (% over age 65) is partially related to
PCs 1 and 2. PC6, however, appears to almost exclusively an age-related variable (with a reasonably-high loading for % mixed use zoning as well). Text in the results section identifies PC6 as that associated with spatial variability in age:

“PC6 is most closely related to age and mixed-use zoning (see Table 1).”

Discussion
Results presented are of interest however should be discussed more in detail. The scope of the study isn’t that clear sometimes and limits of study should be addressed more explicitly. The correlation between area indicators, especially UHI and low socio-economic conditions or high density in identifying areas at risk should be addressed as there doesn’t seem to be a great benefit in using them all as they are predominantly all included in PC1.

Please see our above reply to a similar suggestion made by this reviewer relevant to the introduction section. Text has been added to the discussion section highlighting the challenges associated with the high correlation between several of the variables and the associated limitations. We believe that this change, along with many of the others that have been made in light of the reviewers’ and editor’s suggestions, have helped clarify the scope of the study.

Implications of results especially for HHWW and heat public health are important, indications on how results could be used in terms of definition of at risk areas, at risk susceptible subgroups should be included.

Our thanks to the reviewer for identifying that there may be significant implications for HHWWS and public health in general. We have added the following text at the end of the discussion section:

“Intra-city variability in the response to high heat and humidity conditions indicates an opportunity for the improvement of heat-health watch-warning systems (HHWWS) that have been deployed in cities across the globe. When a dangerous event is forecast, for example, emergency managers might reprioritize allocation of medical resources to those geographic areas responsible for the largest portion of the heat-related deaths in the past. A more thorough effort to build and validate a predictive model of both the timing and placement of heat-related deaths is recommended prior to operational changes in any HHWWS. Longer-term strategies to reduce the heat stress and health burden in these localities might be considered as well, such as the implementation of building weatherization programs, adding green space to the city landscape, adoption of low-albedo and/or green building practices, and location of future healthcare facilities.”

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I declare a have no competing interests.

Sections of amended or added manuscript text are embedded in quotations.

**EDITOR COMMENTS**

We also would add and request that you address the comments below. The strengths of this project include an interesting and less studied question, use of a long-term data (26 years), examining timing of temperature, and assessing neighborhood demographics and land use. However, this paper suffers some major problems in study design and methodology including uncontrolled for some important confounders such as air pollution, time-varying variables, and some factors at individual level (medical care access and treatment etc.).

We appreciate that the editor considers many aspects of the project as strengths. In the subsequent reviews and revisions below we more specifically address several other concerns focused on study design and methodology, but here will we focus on the specific confounders suggested by the editor.

We agree that the extent to which air quality variations (both over time and space) might impact these results would be a worthwhile investigation, albeit one that falls outside the scope of the present study. There are two potential air quality impacts: (1) the apparent temperature thresholds and mortality response associated with an exceedence might be different if air quality (namely ozone) was accounted for, and (2) the local-scale mortality patterns may partially arise as a result of differing air quality within the County (in this case, particulate matter, as ozone is more regionally homogeneous).

In the case of identifying a threshold and response, clearly this is an area of ongoing debate in the literature. We believe that temperature, humidity, and air quality are responsible for negative health but that, in the warm season, extreme high temperature plays the largest role. We recognize that many studies using techniques like generalized additive models often include ozone effects and often find a linear relationship between mortality and ozone aside from temperature effects, but this effect is often small compared with the quasi-exponential relationship with temperature above some minimum mortality level. The recent study by Smargiassi et al. (2009), which is one of the projects we believe our work builds upon, states that “there was little confounding from the effects of ozone” in their assessment of temperature-related mortality for Montreal. Vaneckova et al. (2008) similarly report that the impact of ozone and particulates on heat-related mortality “was not clear.” For Philadelphia in particular, Smoyer et al. (2000) found that under the most extreme weather conditions “neither (particulates) nor ozone appear to contribute an ‘add on’ effect for weather-related mortality.’ Just to be clear, we acknowledge that there is some colinearity between high temperatures and ozone as well as the other weather conditions that can contribute to high ozone, and thus it is not a simple matter to separate out these effects. In fact, part of our ongoing work is attempting to address this question in detail.

The second potential confounder is the interesting notion that fine-scale variations in air quality produce a differential heat-mortality response. We have obtained air quality data from several monitors proximate to Philadelphia County, however the total number of monitors is far too few to properly quantify spatial variability at the scale of this study. Computer models of emissions and air quality might be useful for this sort of investigation, or the implementation of proxy variables such as road network density (Jerrett et al. 2005). The question is not whether individuals who regularly experience poorer air quality are at greater
risk, as that is already fairly well established. The question is whether individuals in poorer air quality regions especially are at higher risk when temperatures are extremely high because of the air quality in particular, apart from other factors. A study that investigated this topic would be of great interest to the authors and within the field in general, but is an undertaking that clearly cannot fit in what is already a lengthy manuscript by many standards and is beyond the scope of our specific hypotheses.

Although in our initial submission we acknowledged that air quality was not accounted for, we have added the following text to the discussion section to make this concern more straightforward:

“We did not incorporate air quality data into this study but encourage future study of the interactive effects of heat and air quality on summertime mortality as well as the potential for differential mortality over space as a result of local-scale air quality variability. Both are topics of active ongoing investigation by the authors and many others.”

If the editor is referring to the census variables as those that may be “time-varying,” we agree that this is a concern, and this is a point one reviewer made as well. Detailed annual data are limited or completely unavailable for many of these variables. We addressed the reviewer’s concern in the following manner:

Our thanks to the reviewer for identifying this important concern. While we do hope to examine the within- and across-zip code relationship between various risk factors and heat-related mortality over time (as well as space), our limited sample sizes in some zip codes requires additional methodological considerations that we believe fall outside of the scope of this particular analysis. We have added the following text to the discussion section:

“The sociodemographic and zoning variables were derived from data available at a fixed point in time (e.g., the year 2000 census). However, the underlying demographics and zoning ordinances both change over time, a process we were unable to capture using this approach. This introduces some uncertainty into the results, and future research should explore local-scale mortality patterns over both space and time.”

We have not been able to locate or develop a variable to represent medical care access as the editor suggests that would be suitable for the scale of this study. We have added this acknowledgement to the discussion section:

“Access and willingness to use medical care is a potential confounder at the individual level that we were not able to represent at the scale of this study, although it may be highly correlated with the socioeconomic variables included.”

Furthermore, it is unclear if the temperature assessment was through an only monitoring site in Philadelphia or through the Landsat imagery from this site to cover 48 areas. It is difficult to judge if a truly fine-scale temperature was used in this study.

Thanks to the editor for identifying an area for clarification in the text. We used both an airport monitoring site and Landsat imagery in this study. The first-order weather station data from Philadelphia
airport were used to identify periods when the regional air temperature and humidity indicated stressful conditions, i.e., the identification of “hot days.” This is not possible from the satellite imagery because images are only available every 16 days, and many of those images are obscured by clouds or otherwise unavailable. Thus the continuous airport record is the best and only available data source for identifying days associated with dangerous weather conditions on which we wanted to evaluate the mortality response.

Once these hot days are identified, however, we then wished to evaluate whether the mortality rate within the city is sensitive to fine-scale variability in thermal stress. This evaluation cannot be completed using the first-order weather station data because the station represents one fixed point. Satellite imagery provides the opportunity for fine-scale assessment of thermal stress. Although we believe that there is some variability in the surface temperature pattern across the city on days associated with high temperature and humidity conditions, it is not necessary to have an image for every day we identified as “hot” from the airport data because the underlying physical landscape that ultimately determines the surface temperature pattern does not significantly change over short time scales. This approach is comparable to that over Johnson et al. (2009) and Smargiassi et al. (2009).

We added the following text to the Data section:

“The airport AT measurements serve as the basis for the identification of days associated with exceptionally high heat and humidity conditions across Philadelphia County.”

A sentence in the methods section has been revised as well:

“The de-seasoned, age-standardized daily mortality departures were sorted into AT groups based on the hourly airport data to examine the heat-mortality relationship.”

Furthermore, it seems a year-round data (rather than summer months) was used, which was not appropriate for such study.

In all cases in the manuscript where the mortality response is reported as a percentage, that number was based on the mean warm-season mortality of 0.26 deaths per 10,000 per day. The warm season was defined as days 150–275, and none of the days above the AT threshold fall outside of this window. In this respect, the analysis did focus on the summer months. We have added the following text to the methods section clarifying this point:

“In all cases the mortality response is expressed as a percent difference relative to the mean warm-season mortality in Philadelphia County (or within each ZCTA) of 0.26 deaths per 10,000 residents per day.”

The original significance testing using the randomization procedure was completed using data from the entire year. Random samples could include values from both inside and outside the warm season. Removing the seasonality, we believed, would cause there to be no difference in the random sample “baseline” for comparison based on which particular days were used in the sampling. Based on the editor’s suggestion, however, we investigated the differences in the random samples drawn using year-
round versus warm-season-only data. We found the distributions to be different such that the values drawn for the warm-season only tend to be lower than those for year-round throughout the distribution (Figure R1, below). Thus the test used in the original manuscript was in fact more conservative than needed (the “critical value” for identifying a significant response was higher than it should have been had we used only warm-season values).

We accordingly proceeded to re-run all of the test statistics for the manuscript using a random sampling procedure based on the warm season only. It is important to note that the actual mortality response is not impacted by this change (such as citywide mortality being 9.3% above normal when the AT threshold is met), but the critical values that identify significant differences from the background level could change (and likely go down). We found that the results presented in the text were not impacted by this change: the overall citywide mortality response remains significant, and the same 10 zip codes that demonstrated a significant response continued to pass the randomization test. No new zip codes were added to the “high-mortality” group with this change. We added the following text to the methods section to reflect this change in the randomization test:

“Samples were drawn exclusively from days falling within the warm season, defined here as between calendar days 150–275 (approximately June–September).”

![Figure R1. A comparison of the random-sampling distributions for identifying a significant mortality response using year-round versus warm-season only data.](image_url)

In addition, the descriptions of methods are unclear in multiple locations. As the major findings could be explained by the biases due to various confounding factors and other major methodology problems described above, I don’t think this paper has adequate scientific merit and don’t recommend it to be published.
The reviewers and editor provided several specific suggestions for improvements and clarifications in the methods used and associated text that we have addressed in significant detail. The scientific merit has been strengthened based on these changes. Furthermore the discussion section has been lengthened to address the editor’s and reviewers’ valid and important concerns about confounding factors. We hope that these revisions have, on the whole, alleviated the editor’s concerns both about clarity in the writing and potential impacts from confounding factors.

We respectfully suggest that it is very unlikely that the major findings in this manuscript can be explained by biases associated with certain confounding factors including air quality, access to health care, and time trends in certain predictor variables. These are important areas for investigation, however, and we would welcome such discussion in the literature. In light of the fact that there are so few studies that have adopted this fine-scale approach to evaluating heat-related mortality (and results at this scale could eventually improve efficiency in the delivery of medical services), we believe that the major finding of our study, that mortality following high apparent temperature days measurably and significantly varies within a large U.S. metropolitan area is an important one to communicate to the readership of *Environmental Health* who might then build upon our methodological approach in terms of assessing mortality and relating it to various predictor variables.

We would be grateful if you could address the comments in a revised manuscript and provide a cover letter giving a point-by-point response to the concerns. Please also highlight (with 'tracked changes'/coloured/underlines/highlighted text) all changes made when revising the manuscript to make it easier for the Editors to give you a prompt decision on your manuscript.

As noted in the letter to the editor, we have provided both a clean version of our resubmitted manuscript as well as one with highlighted text where changes have been made.

As for the formatting, on the title page you change the title to read Fine-scale spatial variability of a heat-related mortality in a Philadelphia County, USA from 1983?2008: A case-series analysis; The title has been changed per the editor’s suggestion.

The title has been changed per the editor’s suggestion.

each affiliation should be written in full, in the following format: department, institute, city, and country; a superscript number should be placed before the institutional address(es) and the corresponding number placed after the respective author(s) name; and that an * be placed before the heading Corresponding author, all other information removed and the * placed after the institutional superscript number. The final heading on the title page should read Email addresses and the colon removed.

The title page has been updated accordingly.

In the Abstract please remove the colon after the headings and move the text beneath the headings. See published articles for examples.

These changes have been made.
Please remove the line numbering throughout the text and the headers in the upper right corner. The final heading in the text should be plural i.e. Conclusions. The heading after the Conclusions should be revised to read List of abbreviations. Again, the colons after the headings should be removed.

The manuscript has been revised to the correct format.

The numbers in the References section should be number, period space without the square brackets. Please also ensure otherwise that your revised manuscript conforms to the journal style (http://www.ehjournal.net/info/instructions/). It is important that your files are correctly formatted.

The references section has been changed in accordance with the journal style.