PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The role of temperature, influenza and other local characteristics in seasonality of mortality: a population-based time-series study in Japan
AUTHORS	Madaniyazi, Lina; Ng, Chris Fook Sheng; Seposo, Xerxes; Toizumi, Michiko; Yoshida, LM; Honda, Yasushi; Armstrong, Ben; Hashizume, Masahiro

VERSION 1 – REVIEW

REVIEWER	Rosano, Aldo Italian National Agency for Regional Healthcare Services
REVIEW RETURNED	24-Dec-2020

GENERAL COMMENTS	The main aim of the manuscript “Seasonality of mortality in Japan: the role of temperature, influenza and other local characteristics” is to investigate the contribution of temperature and influenza to seasonality of mortality in Japan. The manuscript is well structured, the objectives are clearly stated, but the methods and the discussion should be integrated with further details. In general, conclusions are consistent with the results, even if some of the reported associations need a deeper discussion. Here some comments and suggestions: Lines 6-10 page 4. In general the mortality is higher in cold seasons than in warm seasons, but more and more we observe peak of mortality in warm seasons (e.g. in France and Italy in 2003 etc…). This aspect, that is expected to occur even more frequently because of climate change, should be considered. Lines 51-58 page 6. To use the peak-to-trough ratio (PTR) as a measure of seasonality when applied on daily data may be influenced by “outliers”. It would be better to use weekly averages. Line 54 page 6. It would be useful to report exactly how the “predicted mortality” was calculated. Data analysis section: a) No diagnostic test was reported to evaluate the fit of the adopted model. A simple check of model adequacy may be based on diagnostic plots of residuals. Time series of incidence counts often show secular trends in addition to seasonal patterns. For this reason it would be preferable to adopt models taking in to account secular trends, as well as overdispersion and serial correlation between observations.

	b) The authors used as explanatory variable the count of daily deaths due to influenza as a measure of severe influenza circulating in the population. However, very few deaths are used to be reported with influenza as primary cause of death. Influenza is infrequently listed on death certificates of people who die from flu-related complications. On the other hand, a large number of deaths are an indirect cause of influenza infection, but not necessarily the primary cause of death. Seasonal influenza-related deaths (SIRD) are commonly used to analyse the mortality attributable to influenza. There are different approaches to calculate SIRD (see How CDC Estimates the Burden of Seasonal Influenza in the U.S. https://www.cdc.gov/flu/about/burden/how-cdc-estimates.htm) In alternative, the idea of using weekly ILI cases, instead of influenza mortality counts, for influenza adjustment looks more realistic and could be used not only as a sensitive analysis test. it must be said that the Goldstein index, which is the product of the percentage of patients seen with influenza-like illness (ILI) and percentage of influenza-positive specimens (if available), would be preferable to ILI (Goldstein E, et al.. Improving the estimation of influenza-related mortality over seasonal baseline. Epidemiology. 2012; 23:829-838). Line 8 page 9. The authors decided to exclude some variables from the analysis "for the sake of brevity", such as population density, proportion of individuals aged over 65 years old. Exclusions should be based on objective procedures, such as stepwise regression models or structural equation models. Lines 40-43 page 9. The authors concluded that "adjusting for temperature and influenza did not flatten the seasonal pattern or reduce the PTR to 1." This aspect has been emphasized in the discussion. In reality the adjusted PTR value is very close to 1. Measurement errors or scarce model fit are possible reasons for the significant departure from 1 of PTR after adjusting for temperature and influenza. Lines 23-25 page 10. The association between mean temperature and PTR, adjusted for temperature, looks like a circular relationship. The authors should better explain this aspect. Line 14-16 page 12. The authors concluded that "living in prefectures characterized by warm climate and low inequalities experienced larger seasonal variations of mortality". The rationale behind these findings is very hard to understand. The effect modification attributed to the Gini index looks like a statistical artefact, probably caused by using population data rather than individual data, specially if the relation between the individual risk of mortality and individual income is not linear. The authors should describe and justify and not simply defined as "counterintuitive" such an effect. The suggestion "preventive strategies targeting the impact of temperature may reduce the vulnerability of individuals living in prefectures characterized by warm climate and low inequality" is not based on solid evidence and should be removed or reformulated, excluding "low inequality".

	Lines 53-55 page 13 The reference of the study cited in lines 53-55 page 13, “A recent multi-country analysis found a positive association between Gini index and heat effect of temperature on mortality, whereas no evidence was observed for its association with cold effect. “ is missing.

REVIEWER	Schlüter, Benjamin-Samuel UCLouvain
REVIEW RETURNED	05-Feb-2021

GENERAL COMMENTS	I warmly thank the authors for this important and interesting piece of work. This manuscript investigates what is the amount of seasonality of mortality explained by temperature and influenza. Authors also investigated the modifications of a seasonality measure, the peak-to-trough ratio (PTR), by controlling for certain socio-economic variables at the prefecture level. The data has been clearly described. The main modeling part is well explained. The article greatly focus on both temperature and influenza, which is not frequently done. Results are well summarized in Fig1, 2 and 3.My opinion is that this is a well-written paper, however, some important points still need clarifications. Major comments: - Different ICD versions are used in the manuscript. Might these revisions impact the time series of death by creating some disruption at the time of revision (Pechholdová et al. (2017) Eur. J. Population)? If yes, could that impact the results? This might be added in the limitation or discussion. - There is no theory/cited literature behind the inclusion of the socio-economic variables included at the prefecture level to assess their association on PTR. In addition, some are dropped for "brevity" reason while they seem of importance for the research questions. More precisely, authors dropped proportion of population aged above 65 years old. As the share of population above 65 years old grew over time in Japan, this might have an important impact on the seasonality of mortality as older adults are more at risk of dying during winter. In my opinion this variable should be kept in the analysis. - In addition to FigS1, it would be of interest to have a figure of daily counts of death for all-causes, circulatory causes and respiratory causes over the studied period. For transparency, including a figure comparing the fit of the model to the daily death counts over time is also important. In the same vain of idea, having a plot of the residuals over time and their auto-correlation might improve the transparency of the paper. - One of the independant variable used in the model is daily mortality counts of influenza. When studying counts of death by respiratory disease, could that lead to an endogeneity bias? I imagine it depends on the burden of influenza in respiratory mortality in Japan? - When working at the prefecture level (Fig2), are the daily death counts due to respiratory disease not frequently equal to 0 (outside of the winter period)? If so, doesn't the Poisson model suffer from this high number of zero counts? It is known that Poisson models do not handle a high amount of 0 in the dependent variable well (hence the Zero-inflated model). Same remark for independent variable

	"influenza death counts", it might have lots of zero. This leads to some doubt about extrem right plot of Fig2 that is indeed showing huge confidence intervals and estimates not in line with the two other plots on its left. - Fig2 shows impossible values. PTR uses as denominator, the minimum mortality prediction and hence, is by construction a measure always above 1. However, some of your confidence intervals in Fig2 go below 1 which is not possible. Please correct that and provide an explanation to the reader on how 95% confidence intervals for PTR were computed. Also, it has been shown that since PTR are by construction above 1, there can sometimes be positive PTR that are only due to random fluctuations in the data and not due to seasonality (Skajaa et al. (2018) Epidemiology). This has not been adressed by the authors, especially when facing low PTR (p.9 l.34-36). This should be at least discussed in the limitation section. - Again on Fig2, why does PTR have relatively much wider confidence intervals when controlling for influenza (difference between green/red and dark/blue)? It would be nice to add some explanations in the text. - FigS3 has some PTR with adjustment higher than without adjustment as transparently highlighted by the authors. However, there is no explanation for this observation in the text. Is it a sign of something wrong happening in the model? If yes, what is happening there? - It is supprising that authors did not incorporate an important confounder usually controlled for in the literature which is air pollution. This might have an important impact on the result and should be accounted for. - In order to control for seasonality, authors use cyclic cubic splines over the days of the year with 4 degrees of freedom (df). Does that mean that there are four df by year? Why the choice of four? It is usually advised to use 7df (Bhaskaran et al. (2013) Int. J. of Epidemiology).This needs to be specified in the text. This is of importance as the df directly impact the fit to the data, and by doing so, impact the estimated PTR. - The explanation provided by the authors on the negative relationship between seasonality and inequality at the prefecture level is not satisfactory. End of the 2nd page of discussion section: "A recent multi-country analysis found a positive association between Gini index and heat effect of temperature on mortality". Hence, a higher Gini index means more inequality, which is associated with a heat effect of temperature. However the authors conclude: "Therefore, prefectures characterized by low inequality may be more vulnerable to heat effect". This is the contrary to what has been said earlier. Minor comments: - The method section, despite the obvious willingness from the authors to make it clear, is hard to read through. The fact that the authors use several different models to adress different research questions is part of the explanation. It might be of interest to the

	reader to include some equations summarizing the models used. - In the discussion section, when mentioning that prefectures with high climate experienced larger seasonal variations it might be of interest to refer to what is called "the seasonality paradox" (McKee (1989) Euro. J. of Epidemiology). - After page 9 all pages are numbered 1. - Authors do not publicly provide any code to reproduce their results, it would be nice to do so from an Open science perspective.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Aldo Rosano, Italian National Agency for Regional Healthcare Services
Comments to the Author:
The main aim of the manuscript "Seasonality of mortality in Japan: the role of temperature, influenza and other local characteristics" is to investigate the contribution of temperature and influenza to seasonality of mortality in Japan.

The manuscript is well structured, the objectives are clearly stated, but the methods and the discussion should be integrated with further details. In general, conclusions are consistent with the results, even if some of the reported associations need a deeper discussion.

Response: We thank Dr. Aldo Rosano for the constructive comments and suggestions. Several major revisions have been made, including (1) using weekly influenza like illness (ILI) cases for influenza adjustment, (2) model checking and sensitivity analysis, and (3) discussion on the association between indicators on prefecture-specific characteristics and seasonality estimates.
Please see below our responses to Dr. Aldo Rosano's comments.

Here some comments and suggestions:

Lines 6-10 page 4. In general the mortality is higher in cold seasons than in warm seasons, but more and more we observe peak of mortality in warm seasons (e.g. in France and Italy in 2003 etc…). This aspect, that is expected to occur even more frequently because of climate change, should be considered.

Response: We thank Dr. Rosano for pointing out this issue. It is possible that seasonal pattern of mortality may change in the future under a changing climate with an increasing temperature and extreme weather events. However, no evidence is available on this hypothesis at the current moment, therefore we did not consider this topic in our current study.

Lines 51-58 page 6. To use the peak-to-trough ratio (PTR) as a measure of seasonality when applied on daily data may be influenced by "outliers". It would be better to use weekly averages.

Response: We agree that outliers may occur in daily counts of death, but it is unlikely to affect our results or assumptions.

We checked outliers which were more than two and a half interquartile ranges above the median number of daily mortality cases.[1] We detected one outlier, which was daily all-cause mortality on the day of the Great East Japan Earthquake (11 March 2011) and was excluded from our analysis. In addition, we assessed seaonality for each prefecture by using 17-year of data (i.e., between 1999 and

2015) and then for each year by using 47-prefecture of data. Given the large numbers of observations in our analysis, outliers are unlikely to affect our results or assumptions. Using weekly averages may be able to avoid the impact of outliers on seasonality estimates, but it may underestimate seasonality estimates.

Reference:

1. *Schwartz J. Air pollution and hospital admissions for heart disease in eight U.S. counties. Epidemiology. 1999;10(1):17-22. http://www.ncbi.nlm.nih.gov/pubmed/9888275. Accessed February 23, 2021.*

Line 54 page 6. It would be useful to report exactly how the "predicted mortality" was calculated.

Response: We are sorry for the confusion. We did not predict daily mortality in this study. "Predicted mortality" refers to the fitted value of the generalized liner models. We have revised the sentence accordingly in the text (Line 140).

"*The days-of-year with maximum and minimum mortality estimates from generalized liner models were identified as the peak and trough days, respectively.*"

Data analysis section:

a) No diagnostic test was reported to evaluate the fit of the adopted model. A simple check of model adequacy may be based on diagnostic plots of residuals. Time series of incidence counts often show secular trends in addition to seasonal patterns. For this reason, it would be preferable to adopt models taking in to account secular trends, as well as overdispersion and serial correlation between observations.

Response: Thank you for the comments and suggestions.

We agree with the suggestion to adopt models that take in account secular trends, overdispersion and serial correlation. The model we applied was a generalized linear model with quasi-likelihood estimation to assess seasonality of mortality in each prefecture. The quasi-Poisson family link was used to accommodate over-dispersion of the health outcomes (Line 134). In order to evaluate the variation in mortality that was explained by the seasonal components, we used two separate functions for seasonality and secular trends: a cyclic spline was used for seasonality, and indicators for year, day-of-week and their interaction were used to control for the long-term trend (i.e., secular trends) and the effect of day-of-week (Lines 146-147).

As suggested, we have performed additional diagnostic testing on the residuals. In most time series regression studies, which use a single high-dimensional spline to control for seasonality and long-term trend at the same time, residual autocorrelation will tend to be negligible.[1] Our model, however, showed a slow decay in partial autocorrelation function (PACF) plot (Figure S6). This is probably due to not including such a single high-dimensional time spline in our analysis, as it would have made the estimation of seasonality impossible. We conducted further sensitivity analyses using more flexible cyclic spline functions and the standard adjustment for autocorrelation using lagged model residuals.[2] We observed that the seasonality estimates (i.e., PTR) changed very little (< 1% , Table S4).

For simplicity, we have decided to retain the original model and to substantiate our results using the additional sensitivity analyses. Although some previous literatures have mentioned the use of more sophisticated modelling methods based on conditional or marginal models to reduce autocorrelation,[1-2] these were not considered here. This is because the additional complexity of these models may impact the interpretability of results, both in terms of the biological and social pathways.[3-4] Given our primary interest that lies in the estimates of seasonal variation in mortality using the day of the year as the exposure indicator, and their associations with mortality, we believe our current approach that incorporates sensitivity testing is adequate for the purpose.

References:
1. Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L. & Armstrong, B. Time series regression studies in environmental epidemiology. Int. J. Epidemiol. 42, 1187–1195 (2013).
2. Brumback, Babette A., et al. "Transitional regression models, with application to environmental time series." Journal of the American Statistical Association 95.449 (2000): 16-27.
3. Barnett, A. G., Stephen, D., Huang, C., & Wolkewitz, M. "Time series models of environmental exposures: Good predictions or good understanding." Environmental research 154 (2017): 222-225.
4. Peng, Roger D., and Francesca Dominici. "Statistical methods for environmental epidemiology with R." R: a case study in air pollution and health (2008).

We have included model description and diagnosis in the supplementary material and discussed the limitation in the main text (Lines 275-277). The changes are reproduced below for easy reference:

Description of models

We applied a generalized linear model with a quasi-Poisson family to assess seasonality of mortality in each prefecture.

Adjusting for temperature

Adjusting for ILI

Adjusting for both temperature and ILI

t: the day of the observation;
Y_t: the observed daily numbers of mortality on day t;
: the intercept;
doy: day of year, which was fitted using cyclic cubic spline with 4 degrees of freedom (df);
: the daily numbers of ILI on day t, which was controlled using natural cubic spline with 3 df;
$Strata_t$: strata defined by year, day of week, and their interaction to control for the long-term trend and the effect of day of week, and λ is the vector of coefficients;
$Temp_{t,l}$: a matrix obtained by using cross basis function to temperature; l is the lag days, and β is the vector of coefficients. (For the cross-basis function, a natural cubic B-spline basis with threinternal knots at the 25th, 50th, and 75th percentiles of temperature distribution was used for exposure-response association, and another natural cubic B-spline basis with 3 df with extended lag up to 21 days was used for the lag-response association.)

Model Checking and sensitivity analysis

We used scatter plot of deviance residuals vs time and partial autocorrelation function plot of the deviance residuals to check the models. In addition, sensitivity analysis was conducted to check the robustness of our estimates.
We used the largest prefecture (i.e., Tokyo) for model evaluation, as the statistical uncertainty for the estimates was small.

- Scatter plot of deviance residuals vs time

In general, the plot shows an even band of points over the time, although we observed a few spikes, for example, in 1999. This pattern did not change significantly when we use more flexible modelling for seasonality, temperature, and influenza.

Figure S5. Deviance residuals over time from the analysis in Tokyo (without adjustment for temperature and/or influenza)

- Partial autocorrelation function (PACF) plot of the deviance residuals

PACF shows a slow decay and a high degree of autocorrelation around a 1-week lag. This pattern remained when we included temperature and/or ILI in the model. In order to reduce the autocorrelation, we tried more flexible functions for seasonality by increasing the degree of freedom, and then we added lagged deviance residuals to the model in several different ways. For example, 1-day lagged deviance residuals, 1- to 6-day lagged deviance residual, and a moving average of 6 days lagged deviance residuals, respectively. The autocorrelation remained without much reduction after many attempts, but the coefficient and its standard error from cyclic spline functions for seasonality changed very little (Table S4).

Figure S6. Partial autocorrelation function plot of the deviance residuals from the analysis in Tokyo (without adjustment for temperature and/or influenza)

Table S4. Seasonality estimates for Tokyo without adjusting for temperature and/or influenza like illness

Models	Peak-to-Trough (95% confidence interval)
Main model	1.254 (1.249, 1.259)
Model 1	1.249 (1.237, 1.255)
Model 2	1.244 (1.237, 1.252)
Model 3	1.253 (1.249, 1.258)
Model 4	1.253 (1.248, 1.257)
Model 5	1.252 (1.248, 1.257)
Model 6	1.250 (1.247, 1.254)

Main model:
(Stratat: strata defined by year, day of week, and their interaction to control for long-term trend and effect of day of week)
Model 1:
Model 2:
Model 3:
Model 4:
Model 5:
Model 6:

Discussion in the main text (Lines 275-277)

"We observed some autocorrelation in the model residuals despite our attempts to model it (Figure S6). However, sensitivity testing showed that it had limited impacts on the estimate of seasonality (Table S4)."

b) The authors used as explanatory variable the count of daily deaths due to influenza as a measure of severe influenza circulating in the population. However, very few deaths are used to be reported with influenza as primary cause of death. Influenza is infrequently listed on death certificates of people who die from flu-related complications. On the other hand, a large number of deaths are an indirect cause of influenza infection, but not necessarily the primary cause of death. Seasonal influenza-related deaths (SIRD) are commonly used to analyse the mortality attributable to influenza. There are different approaches to calculate SIRD (see How CDC Estimates the Burden of Seasonal Influenza in the U.S. https://www.cdc.gov/flu/about/burden/how-cdc-estimates.htm)

In alternative, the idea of using weekly ILI cases, instead of influenza mortality counts, for influenza adjustment looks more realistic and could be used not only as a sensitive analysis test. it must be said that the Goldstein index, which is the product of the percentage of patients seen with influenza-like illness (ILI) and percentage of influenza-positive specimens (if available), would be preferable to ILI (Goldstein E, et al.. Improving the estimation of influenza-related mortality over seasonal baseline. Epidemiology. 2012; 23:829-838 PubMed).

Response: We thank Dr. Rosano for his insightful comments and suggestions.

We agree that daily deaths due to influenza may not be the best indicator for influenza adjustment. As suggested, our analysis of using weekly ILI cases for influenza adjustment has been moved to the main text (Table 2, Figure 1 to Figure 3), and the analysis of using influenza mortality counts has been removed from the current manuscript.

Although Goldstein index would be preferable to weekly ILI cases, the data on influenza-positive specimens are not available for the current study.

Line 8 page 9. The authors decided to exclude some variables from the analysis “for the sake of brevity”, such as population density, proportion of individuals aged over 65 years old. Exclusions should be based on objective procedures, such as stepwise regression models or structural equation models.

Response: We thank Dr. Rosano for pointing out this issue.

We included all the indicators in the analysis and updated results in the revision (Figure S4). The updated results did not show strong evidence for any associations. Although previous studies suggested a larger seasonal variation in mortality in warmer and less developed locations,[1-3] we did not find any evidence for the modifying effect of prefecture-specific indicators on seasonality of mortality. This could be partially explained by the limited range of variations in the indicators and possible confounding effect between them. Furthermore, our data on the indictors are population-level, and future investigations with individual-level data is recommended to examine these issues.

Notably, the updated results are different from the findings in our first submission, where we observed associations of PTR with averaged annual mean temperature and Gini index. This is probably related with the study period: we used data for 17 years (1999 - 2015) in our revision while 44-yr of data (1972 – 2015) was used in our original submission.

References:

1. *Healy, J. D. Excess winter mortality in Europe: a cross country analysis identifying key risk factors. J. Epidemiol. Community Health **57**, 784–9 (2003).*
2. *Stewart, S., Keates, A. K., Redfern, A. & McMurray, J. J. V. Seasonal variations in cardiovascular disease. Nat. Rev. Cardiol. **14**, 654–664 (2017).*
3. *Gemmell, I., McLoone, P., Boddy, F., Dickinson, G. J. & Watt, G. Seasonal variation in mortality in Scotland. Int. J. Epidemiol. 29, 274–279 (2000).*

Figure S4 is reproduced below for easy reference:

Figure S4. Associations between each indicator and PTR before and after adjusting for influenza like illness (ILI) and temperature

Coefficient and 95% confidence intervals were obtained from liner mixed effect models adjusting for latitude and longitude, except for when we investigated averaged annual mean temperature as the indicator, due to their high correlation. Results are expressed aslog (PTR) change for standard deviation increase in each indicator.

Lines 40-43 page 9. The authors concluded that “adjusting for temperature and influenza did not flatten the seasonal pattern or reduce the PTR to 1.” This aspect has been emphasized in the discussion. In reality the adjusted PTR value is very close to 1. Measurement errors or scarce model fit are possible reasons for the significant departure from 1 of PTR after adjusting for temperature and influenza.

Response: The PTR for all-cause mortality reduced from 1.29 to 1.07 on adjusting for temperature and influenza. Though it is true that a rate ratio of 1.07 would be considered small in many contexts, in environmental epidemiology RRs are frequently lower (see especially much air pollution epidemiology), and the narrow CI (1.06,1.07) further indicates that this PTR of 1.07 is highly statistically significantly different from 1.

Though we are not sure, we take the reference to measurement error being a reason for the departure of the adjusted PTR from 1 to refer to the possibility that confounding by temperature and/or influenza may have been incompletely controlled because of error in measuring these variables. We acknowledge this is a reasonable concern, and have added text to the discussion to address it, in which we note the reasons why any bias in estimated PTR from this cause is likely to be minimal.

Addition to the Limitation paragraph in discussion (Lines 278-281):
"It is possible that the PTR on adjusting for influenza and temperature may be overestimated due to residual confounding as a result of error in measuring these variables.[1] However, any such overestimation would be believed to be slight, as the main error here would be of Berkson type, which does not cause bias and hence not compromise confounder control.[2]"

Reference:
1. Armstrong, Ben G. "Effect of measurement error on epidemiological studies of environmental and occupational exposures." Occupational and environmental medicine 55.10 (1998): 651-656.
2. Dominici, Francesca, Scott L. Zeger, and Jonathan M. Samet. "A measurement error model for time-series studies of air pollution and mortality." Biostatistics 1.2 (2000): 157-175.

Lines 23-25 page 10. The association between mean temperature and PTR, adjusted for temperature, looks like a circular relationship. The authors should better explain this aspect.

Response: We are sorry for the confusing explanation.

For each indicator, we computed the averaged value across the years 1999-2015 for each prefecture to assess its association with seasonality estimates. Hence, the association between mean temperature and PTR, adjusted for temperature, refers to the association between averaged annual mean temperature (from 1999 to 2015) and temperature adjusted PTR. We have clarified this issue in our revision (Line 121, and Lines 125-126).

Line 14-16 page 12. The authors concluded that "living in prefectures characterized by warm climate and low inequalities experienced larger seasonal variations of mortality". The rationale behind these findings is very hard to understand. The effect modification attributed to the Gini index looks like a statistical artefact, probably caused by using population data rather than individual data, specially if the relation between the individual risk of mortality and individual income is not linear. The authors should describe and justify and not simply defined as "counterintuitive" such an effect. The suggestion "preventive strategies targeting the impact of temperature may reduce the vulnerability of individuals living in prefectures characterized by warm climate and low inequality" is not based on solid evidence and should be removed or reformulated, excluding "low inequality".

Response: We thank Dr. Aldo Rosano for pointing out this important concern.

In the revision, we assessed the association between prefecture-specific indicators and seasonality estimates (log(PTR)) before and after adjusting for temperature and/or ILI by using data between 1999 and 2015 (Figure S4). The updated results did not show strong evidence for any associations. Although previous studies suggested a larger seasonal variation in mortality in warmer and less developed locations,[1-3] we did not find any evidence for the modifying effect of prefecture-specific indicators on seasonality of mortality. This could be partially explained by the limited range of variations in the indicators and possible confounding effect between them.

Furthermore, our data on the indictors are population-level, and future investigations with individual-level data is recommended to examine these issues.

Notably, the updated results are different from the findings in our first submission, where we observed associations of PTR with averaged annual mean temperature and Gini index. This is probably related with the study period: we used data for 16 years (1999 - 2015) in our revision while 44-yr of data (1972 – 2015) was used in our original submission.

References:

1. *Healy, J. D. Excess winter mortality in Europe: a cross country analysis identifying key risk factors. J. Epidemiol. Community Health* ***57****, 784–9 (2003).*
2. *Stewart, S., Keates, A. K., Redfern, A. & McMurray, J. J. V. Seasonal variations in cardiovascular disease. Nat. Rev. Cardiol.* ***14****, 654–664 (2017).*
3. *Gemmell, I., McLoone, P., Boddy, F., Dickinson, G. J. & Watt, G. Seasonal variation in mortality in Scotland. Int. J. Epidemiol. 29, 274–279 (2000).*

Lines 53-55 page 13 The reference of the study cited in lines 53-55 page 13, "A recent multi-country analysis found a positive association between Gini index and heat effect of temperature on mortality, whereas no evidence was observed for its association with cold effect. " is missing.

Response: We have included the reference in the revision (Line 266; ref# 23).

Reviewer: 2
Dr. Benjamin-Samuel Schlüter, UCLouvain
Comments to the Author:
I warmly thank the authors for this important and interesting piece of work. This manuscript investigates what is the amount of seasonality of mortality explained by temperature and influenza. Authors also investigated the modifications of a seasonality measure, the peak-to-trough ratio (PTR), by controlling for certain socio-economic variables at the prefecture level. The data has been clearly described. The main modeling part is well explained. The article greatly focus on both temperature and influenza, which is not frequently done. Results are well summarized in Fig1, 2 and 3.My opinion is that this is a well-written paper, however, some important points still need clarifications.

Response: We greatly appreciate the constructive comments and suggestions form Dr. Benjamin-Samuel Schlüter. We carefully revised our manuscript according to all these comments and suggestions. Please see below our responses to the Dr. Schlüter.

Major comments:
- Different ICD versions are used in the manuscript. Might these revisions impact the time series of death by creating some disruption at the time of revision (Pechholdová et al. (2017) Eur. J. Population)? If yes, could that impact the results? This might be added in the limitation or discussion.
Response: We agree that changes in the classification of causes of death may disrupt the time series of death. In our first submission, we conducted sensitivity analysis by using data between 1999 and 2015 where only ICD-10 was used for the classification of causes of death, and our seasonality estimates changed little.

In the revision, we restricted our main analysis to the period between 1999 and 2015 (ICD-10 was used), whereas 44-yr of data between 1972 and 2015 was used in our first submission. The reason is that we followed the suggestion from Reviewer #1 and used influenza like illness (ILI) for influenza adjustment, and that ILI data was not available until 1999.

We reproduced the comment from Reviewer #1 below.

"b) The authors used as explanatory variable the count of daily deaths due to influenza as a measure of severe influenza circulating in the population. However, very few deaths are used to be reported with influenza as primary cause of death. Influenza is infrequently listed on death certificates of people who die from flu-related complications. On the other hand, a large number of deaths are an indirect cause of influenza infection, but not necessarily the primary cause of death. Seasonal influenza-related deaths (SIRD) are commonly used to analyse the mortality attributable to influenza. There are different approaches to calculate SIRD (see How CDC Estimates the Burden of Seasonal Influenza in the U.S. https://www.cdc.gov/flu/about/burden/how-cdc-estimates.htm)

In alternative, the idea of using weekly ILI cases, instead of influenza mortality counts, for influenza adjustment looks more realistic and could be used not only as a sensitive analysis test. it must be said that the Goldstein index, which is the product of the percentage of patients seen with influenza-like illness (ILI) and percentage of influenza-positive specimens (if available), would be preferable to ILI (Goldstein E, et al.. Improving the estimation of influenza-related mortality over seasonal baseline. Epidemiology. 2012; 23:829-838 PubMed)."

- There is no theory/cited literature behind the inclusion of the socio-economic variables included at the prefecture level to assess their association on PTR. In addition, some are dropped for "brevity" reason while they seem of importance for the research questions. More precisely, authors dropped proportion of population aged above 65 years old. As the share of population above 65 years old grew over time in Japan, this might have an important impact on the seasonality of mortality as older adults are more at risk of dying during winter. In my opinion this variable should be kept in the analysis.

Response: We thank Dr. Schlüter for pointing out this important issue.

As mentioned in the introduction (Lines 92-97) and discussion (Line 263), previous studies[1-6] suggested that the spatial variation in seasonality of mortality and the health effect of temperature may be related with the socio-economic variables. Therefore, we investigated the association between socio-economic variables and PTR.

In the revision, we included all the variables in the analysis and updated the results (Figure S4, attached below). Although previous studies suggested a larger seasonal variation in mortality in warmer and less developed locations, we did not find any evidence for the modifying effect of prefecture-specific indicators on seasonality of mortality. This could be partially explained by the limited range of variations in the indicators and possible confounding effect between them. Furthermore, our data on the indictors are population-level, and future investigations with individual-level data is recommended to examine these issues.

Notably, the updated results are different from the findings in our first submission, where we observed associations of PTR with averaged annual mean temperature and Gini index. This is probably related with the study period: we used data for 16 years (1999 - 2015) in our revision while 44-yr of data (1972 – 2015) was used in our original submission. Figure S4 is reproduced below for easy reference:

References:
1. Healy, J. D. Excess winter mortality in Europe: a cross country analysis identifying key risk factors. J. Epidemiol. Community Health ***57****, 784–9 (2003).*
2. Gemmell, I., McLoone, P., Boddy, F., Dickinson, G. J. & Watt, G. Seasonal variation in mortality in Scotland. Int. J. Epidemiol. ***29****, 274–279 (2000).*
3. Hajat, S. et al. Public health vulnerability to wintertime weather: time-series regression and episode analyses of national mortality and morbidity databases to inform the Cold Weather Plan for England. Public Health ***137****, 26–34 (2016).*

4. Medina-Ramón, M. & Schwartz, J. Temperature, temperature extremes, and mortality: a study of acclimatisation and effect modification in 50 US cities. Occup. Environ. Med. ***64****, 827–33 (2007).*
5. Braga, A. L. F., Zanobetti, A. & Schwartz, J. The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities. Environ. Health Perspect. ***110****, 859–63 (2002).*
6. McKee, C. M. Deaths in Winter: Can Britain Learn from Europe? Eur. J. Epidemiol. ***5****, 178–82 (1989).*

Figure S4. Associations between each indicator and PTR before and after adjusting for influenza like illness (ILI) and temperature

Coefficient and 95% confidence intervals were obtained from liner mixed effect models adjusting for latitude and longitude, except for when we investigated averaged annual mean temperature as the indicator, due to their high correlation. Results are expressed aslog (PTR) change for standard deviation increase in each indicator.

- In addition to FigS1, it would be of interest to have a figure of daily counts of death for all-causes, circulatory causes and respiratory causes over the studied period. For transparency, including a figure comparing the fit of the model to the daily death counts over time is also important. In the same vain of idea, having a plot of the residuals over time and their auto-correlation might improve the transparency of the paper.

Response: We thank Dr. Schlüter for the suggestions. We have included a figure of daily counts of deaths over the studied period and added model checking in the supplementary material. The changes are reproduced below for easy reference:

Figure S1**. Time series of national wide daily mortality cases from all-cause, circulatory, respiratory disease and influenza between April 1999 and 2015**

Model Checking and sensitivity analysis

We used scatter plot of deviance residuals vs time and partial autocorrelation function plot of the deviance residuals to check the models. In addition, sensitivity analysis was conducted to check the robustness of our estimates.

We used the largest prefecture (i.e., Tokyo) for model evaluation, as the statistical uncertainty for the estimates was small.

- Scatter plot of deviance residuals vs time

In general, the plot shows an even band of points over the time, although we observed a few spikes, for example, in 1999. This pattern did not change significantly when we use more flexible modelling for seasonality, temperature, and influenza.

Figure S5. Deviance residuals over time from the analysis in Tokyo (without adjustment for temperature and/or influenza)

- Partial autocorrelation function (PACF) plot of the deviance residuals

PACF shows a slow decay and a high degree of autocorrelation around a 1-week lag. This pattern remained when we included temperature and/or ILI in the model. In order to reduce the autocorrelation, we tried more flexible functions for seasonality by increasing the degree of freedom, and then we added lagged deviance residuals to the model in several different ways. For example, 1-day lagged deviance residuals, 1- to 6-day lagged deviance residual, and a moving average of 6 days lagged deviance residuals, respectively. The autocorrelation remained without much reduction after many attempts, but the coefficient and its standard error from cyclic spline functions for seasonality changed very little (Table S4).

Figure S6. Partial autocorrelation function plot of the deviance residuals from the analysis in Tokyo (without adjustment for temperature and/or influenza)

Table S4. Seasonality estimates for Tokyo without adjusting for temperature and/or influenza like illness

Models	Peak-to-Trough (95% confidence interval)
Main model	1.254 (1.249, 1.259)
Model 1	1.249 (1.237, 1.255)
Model 2	1.244 (1.237, 1.252)
Model 3	1.253 (1.249, 1.258)
Model 4	1.253 (1.248, 1.257)
Model 5	1.252 (1.248, 1.257)
Model 6	1.250 (1.247, 1.254)

Main model:
($Strata_t$: strata defined by year, day of week, and their interaction to control for long-term trend and effect of day of week)
Model 1:
Model 2:
Model 3:
Model 4:
Model 5:
Model 6:

In most time series regression studies, which use a single high-dimensional spline to control for seasonality and long-term trend at the same time, residual autocorrelation will tend to be negligible.[1] Our model, however, showed a slow decay in partial autocorrelation function (PACF) plot (Figure S6). This is probably due to not including such a single high-dimensional time spline in our analysis, as it would have made the estimation of seasonality impossible. We conducted further sensitivity analyses using more flexible cyclic spline functions and the standard adjustment for autocorrelation using lagged model residuals.[2] We observed that the seasonality estimates (i.e., PTR) changed very little (< 1% , Table S4).

For simplicity, we have decided to retain the original model and to substantiate our results using the additional sensitivity analyses. Although some previous literatures have mentioned the use of more sophisticated modelling methods based on conditional or marginal models to reduce autocorrelation,[1-2] these were not considered here. This is because the additional complexity of these models may impact the interpretability of results, both in terms of the biological and social pathways.[3-4] Given our primary interest that lies in the estimates of seasonal variation in mortality using the day of the year as the exposure indicator, and their associations with mortality, we believe our current approach that incorporates sensitivity testing is adequate for the purpose.

References:
1. Bhaskaran, K., Gasparrini, A., Hajat, S., Smeeth, L. & Armstrong, B. Time series regression studies in environmental epidemiology. Int. J. Epidemiol. 42, 1187–1195 (2013).
2. Brumback, Babette A., et al. "Transitional regression models, with application to environmental time series." Journal of the American Statistical Association 95.449 (2000): 16-27.
3. Barnett, A. G., Stephen, D., Huang, C., & Wolkewitz, M. "Time series models of environmental exposures: Good predictions or good understanding." Environmental research 154 (2017): 222-225.
4. Peng, Roger D., and Francesca Dominici. "Statistical methods for environmental epidemiology with R." R: a case study in air pollution and health (2008).

We included this in our limitations as below (Lines 275-277)

"We observed some autocorrelation in the model residuals despite our attempts to model it (Figure S6). However, sensitivity testing showed that it had limited impacts on the estimate of seasonality (Table S4)."

- **The fit of the model to the daily death counts over time**

Figure S7. Daily mean number of observed all-cause, circulatory, and respiratory mortality in Japan averaged from 47 prefectures over the study period and estimated number of daily mortality from time series regression models (Main model without adjusting for temperature and/or influenza)
Grey dot: daily mean number of observed mortality cases averaged from 47 prefectures over the study period;
Red: pooled estimates with 95% confidence intervals obtained from prefecture-specific estimates from models without temperature adjustment

Figure S7 suggests that our models before adjusting for temperature and/or influenza fitted seasonality of circulatory mortality better and may underestimate the seasonal variation in all-cause and respiratory mortality. The discrepancy between observed and fitted values may be explained by the risk of temperature, infectious disease, and other factors (e.g., human behaviour).

- One of the independant variable used in the model is daily mortality counts of influenza. When studying counts of death by respiratory disease, could that lead to an endogeneity bias? I imagine it depends on the burden of influenza in respiratory mortality in Japan?

Response: Daily death due to influenza is part of all-cause and respiratory mortality. In Japan, the influenza-related respiratory mortality rate was estimated at 0.2, 3.5, and 27.5 per 100 000 individuals aged <65 years, 65-74 years, and ≥75 years, respectively, which is lower than many other countries.[1] In our original submission, we log-transformed the daily mortality counts of influenza in our models to minimize endogeneity bias, and checked our findings by replacing daily mortality counts of influenza with influenza-like-illness from surveillance data.

Our first reviewer, Dr. Aldo Rosano, has suggested that daily deaths due to influenza may not be the best indicator for influenza adjustment. As suggested, our analysis of using weekly ILI cases for influenza adjustment has been moved to the main text (Table 2, Figure 1 to Figure 3), and the analysis of using influenza mortality counts for influenza adjustment has been removed from the current manuscript.

Reference:

1. Danielle Iuliano A, Roguski KM, Chang HH, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. www.thelancet.com. 2018;391:1285. PubMed doi:10.1016/S0140-6736(17)33293-2

- When working at the prefecture level (Fig2), are the daily death counts due to respiratory disease not frequently equal to 0 (outside of the winter period)? If so, doesn't the Poisson model suffer from this high number of zero counts? It is known that Poisson models do not handle a high amount of 0 in the dependent variable well (hence the Zero-inflated model). Same remark for independent variable "influenza death counts", it might have lots of zero. This leads to some doubt about extrem right plot of Fig2 that is indeed showing huge confidence intervals and estimates not in line with the two other plots on its left.

Response: We applied a generalized linear model with a quasi-Poisson family to assess seasonality of mortality in each prefecture. The quasi-Poisson family was used to accommodate over-dispersion of the observations. As the reviewer pointed out, there are zeros in the daily death counts due to respiratory diseases at prefectural level. However, we did not consider Zero-inflated model here. The numbers of zero count in our study is small (Table below), which may have little or no impact on our estimates. Indeed, in an event whereby the data generation process (DGP) for zeros is dependent on an earlier condition, then zero-inflated Poisson (ZIP) models would be suitable. In brief, ZIP accounts for these two DGPs, 1) generation of zeros, and 2) usual Poisson distribution generating counts (some may have zeros). This, however, is not the case in our study, since the occurrence of zero counts in respirtory mortality is from the natural process, governed by Poisson distribution, and is not due to design, survey, or observer errors.[1]

Table. Numbers of zero counts in daily respiratory mortality for each prefecture

Prefecture	Numbers of zero counts in daily respiratory mortality (n)	Proportion of zero counts (n/N, where N= 6208 days from 1999 to 2015)
Aichi	0	0.0%
Akita	55	0.9%
Aomori	38	0.6%
Chiba	0	0.0%
Ehime	18	0.3%
Fukui	217	3.5%
Fukuoka	0	0.0%
Fukushima	4	0.1%
Gifu	6	0.1%
Gunma	4	0.1%
Hiroshima	1	0.0%
Hokkaido	0	0.0%
Hyogo	0	0.0%
Ibaraki	1	0.0%
Ishikawa	104	1.7%
Iwate	26	0.4%

Kagawa	65	1.0%
Kagoshima	1	0.0%
Kanagawa	0	0.0%
Kochi	139	2.2%
Kumamoto	7	0.1%
Kyoto	1	0.0%
Mie	14	0.2%
Miyagi	9	0.1%
Miyazaki	62	1.0%
Nagano	4	0.1%
Nagasaki	12	0.2%
Nara	61	1.0%
Niigata	2	0.0%
Oita	30	0.5%
Okayama	2	0.0%
Okinawa	147	2.4%
Osaka	0	0.0%
Saga	156	2.5%
Saitama	0	0.0%
Shiga	108	1.7%
Shimane	217	3.5%
Shizuoka	0	0.0%
Tochigi	16	0.3%
Tokushima	135	2.2%
Tokyo	0	0.0%
Tottori	564	9.1%
Toyama	69	1.1%
Wakayama	89	1.4%
Yamagata	47	0.8%

As for independent variable "influenza death counts", it could have lots of zero. Therefore, in our original submission, we added one to daily mortality counts for influenza before the log-transformation (i.e., log(flu_death+1)). In the revision, the results by using daily mortality from influenza was removed from the manuscript (please refer to our responses to your previous comment).

The large confidence intervals for respiratory mortality were somewhat larger, which may be due to the far fewer number of respiratory deaths than other causes (Table S1)

Reference:

1. *Zuur A, Ieno E N, Walker N, et al. Mixed effects models and extensions in ecology with R[M]. Springer Science & Business Media, 2009.*

- Fig2 shows impossible values. PTR uses as denominator, the minimum mortality prediction and hence, is by construction a measure always above 1. However, some of your confidence intervals in Fig2 go below 1 which is not possible. Please correct that and provide an explanation to the reader on how 95% confidence intervals for PTR were computed. Also, it has been shown that since PTR are by construction above 1, there can sometimes be positive PTR that are only due to random fluctuations in the data and not due to seasonality (Skajaa et al. (2018) Epidemiology). This has not been adressed by the authors, especially when facing low PTR (p.9 l.34-36). This should be at least discussed in the limitation section.

Response: We appreciate this constructive comment from Dr. Schlüter.

Previous studies[1,2] which used PTR as a measure of seasonality have indeed enforced the boundary constraint by truncating the lower confidence limit at one. However, Skajja et al. was concerned that this may introduce a positive bias into the PTR, as even if the null hypothesis (PTR=1) is correct, any statistical variability in the risk will lead to a non-null estimate (PTR>1).[3] However, we did not truncate the lower confidence limit at one in our analysis, in order to show the statistical variation of our PTR estimates. Therefore, positive PTR that are only due to random fluctuations in the data and not due to seasonal pattern is unlikely to happen in our analysis.

The concept of PTR in our study is similar to relative risk (RR) which has been used widely to quantify the effect of temperature/air pollutant on mortality. For example, when assessing thessociation between temperature and mortality, the temperature at which the mortality estimate is the lowest is identified as the reference/threshold, and mortality estimates at any temperature is naturally expressed relative to that risk at the reference/threshold (i.e., 1). Notably, the shape of exposure-response curve does not depend on the choice of the centering point, but confidence intervals do change. Previous studies have explained this topic in detail.[4]

We have clarified in our method that we did not truncate the lower confidence limit at one (Lines 142-146):

"When constructing confidence intervals for PTR, previous studies enforced the boundary constraint by truncating the lower confidence limit at one for PTR. However, doing that may introduce a positive bias into the PTR. In order to show the statistical variability in PTR, therefore, we did not truncate the lower confidence limit at one for PTR."

References:

1. Christensen, A. L., Lundbye-Christensen, S. & Dethlefsen, C. Poisson regression models outperform the geometrical model in estimating the peak-to-trough ratio of seasonal variation: A simulation study. Comput. Methods Programs Biomed. ***104****, 333–340 (2011).*

2. Brookhart, M. A. & Rothman, K. J. Simple estimators of the intensity of seasonal occurrence. BMC Med. Res. Methodol. ***8****, 67 (2008).*

3. N, S. et al. Forty-year Seasonality Trends in Occurrence of Myocardial Infarction, Ischemic Stroke, and Hemorrhagic Stroke. Epidemiology ***29****, (2018).*

4. Armstrong B. Models for the relationship between ambient temperature and daily mortality. Epidemiology. 2006 Nov;17(6):624-31. doi: 10.1097/01.ede.0000239732.50999.8f. PMID: 17028505.

- Again on Fig2, why does PTR have relatively much wider confidence intervals when controlling for influenza (difference between green/red and dark/blue)? It would be nice to add some explanations in the text.

Response: We believe that the reviewer is referring to adjustment for temperature rather than influenza. Certainly Figure 2 showed wider confidence intervals when controlling for temperature (green/red), and this pattern seems to be more evident for small prefectures (e.g., Tokyo vs Nagasaki). This is probably due to the stronger confounding of PTR by temperature than by influenza.

- FigS3 has some PTR with adjustment higher than without adjustment as transparently highlighted by the authors. However, there is no explanation for this observation in the text. Is it a sign of something wrong happening in the model? If yes, what is happening there?

Response: We thank Dr. Schlüter for pointing out this important finding.

In our updated results, the yearly analysis in 2000 showed a higher PTR for all-cause and respiratory mortality after including temperature in the adjustment (Figure S3). We further checked its sensitivity to temperature adjustment. Changing the lag period of 21 days in cross-basis function to 14 days reduced temperature-adjusted PTR, although it remained slightly higher than unadjusted

PTR with a largely overlapped confidence intervals (table below). The results for the other years did not change much.
We acknowledge that seasonality estimates in 2000 seems to be sensitive to temperature adjustment. Further investigations will be conducted to explore the potential reasons.

	Temperature adjustment	PTR
Unadjusted PTR for all-cause mortality	NA	1.35 (1.35, 1.36)
Temperature adjusted PTR for all-cause mortality	Cross-basis function with 21 lag days	1.44 (1.38, 1.50)
	Cross-basis function with 14 lag days	1.36 (1.32, 1.41)

This finding was reported and discussed in the revision (Lines 223-227 and 274-275):

"We further checked the sensitivity of our estimates to temperature adjustment. Changing the lag period of 21 days in cross-basis function to 14 days reduced temperature-adjusted PTR, although it remained slightly higher than unadjusted PTR with a largely overlapped confidence intervals. The results for the other years did not change much (results not shown)."

"Our findings for 2000 were sensitive to temperature adjustment."

- It is supprising that authors did not incorporate an important confounder usually controlled for in the literature which is air pollution. This might have an important impact on the result and should be accounted for.

Response: Air pollution is usually considered as a confounder when investigating the association between temperature and mortality. However, the aim of the current analysis is to estimate seasonal variation in mortality and the contribution of temperature and influenza to seasonality of mortality. Therefore, we did not consider air pollution in current analysis. We will investigate the contribution of air pollution to seasonal variation in mortality in our future studies.

- In order to control for seasonality, authors use cyclic cubic splines over the days of the year with 4 degrees of freedom (df). Does that mean that there are four df by year? Why the choice of four? It is usually advised to use 7df (Bhaskaran et al. (2013) Int. J. of Epidemiology).This needs to be specified in the text. This is of importance as the df directly impact the fit to the data, and by doing so, impact the estimated PTR.

Response: We thank Dr. Schlüter for the comments.
Bhaskaran et al. (2013) Int. J. of Epidemiology suggested to use a spline function with 7df to account for both seasonality and long-term trend at the same time. In order to estimate the seasonality in our analysis, we used separate functions for seasonality and long-term trend. We used a cyclic spline with 4df for seasonality(lines 139-140) and separate indicators for year, day-of-week and their interaction for long-term trend and the effect of day-of-week (lines 147-148). Furthermore, we checked the sensitivity of our seasonality estimates by varying df for the cyclic spline from 4 to 6, and the results changed little (Table S4).

Table S4. Seasonality estimates for Tokyo without adjusting for temperature and/or influenza like illness

Models	Peak-to-Trough (95% confidence interval)
Main model	1.254 (1.249, 1.259)

Model 1	1.249 (1.237, 1.255)
Model 2	1.244 (1.237, 1.252)
Model 3	1.253 (1.249, 1.258)
Model 4	1.253 (1.248, 1.257)
Model 5	1.252 (1.248, 1.257)
Model 6	1.250 (1.247, 1.254)

Main model:
($Strata_t$: strata defined by year, day of week, and their interaction to control for long-term trend and effect of day of week)
Model 1:
Model 2:
Model 3:
Model 4:
Model 5:
Model 6:

- The explanation provided by the authors on the negative relationship between seasonality and inequality at the prefecture level is not satisfactory. End of the 2nd page of discussion section: "A recent multi-country analysis found a positive association between Gini index and heat effect of temperature on mortality". Hence, a higher Gini index means more inequality, which is associated with a heat effect of temperature. However the authors conclude: "Therefore, prefectures characterized by low inequality may be more vulnerable to heat effect". This is the contrary to what has been said earlier.

Response: We thank Dr. Schlüter for pointing out this important concern. The same concern was also raised by Reviewer #1. We reproduced our response to Reviewer #1 below.

In the revision, we assessed the association between prefecture-specific indicators and seasonality estimates (log(PTR)) before and after adjusting for temperature and/or ILI by using data between 1999 and 2015 (Figure S4). The updated results did not show strong evidence for any associations. Although previous studies suggested a larger seasonal variation in mortality in warmer and less developed locations,[1-3] we did not find any evidence for the modifying effect of prefecture-specific indicators on seasonality of mortality. This could be partially explained by the limited range of variations in the indicators and possible confounding effect between them. Furthermore, our data on the indictors are population-level, and future investigations with individual-level data is recommended to examine these issues.

Notably, the updated results are different from the findings in our first submission. This is probably related with the study period: we used data for 16 years (1999 - 2015) in our revision while 44-yr of data (1972 – 2015) was used in our original submission.

References:
1. *Healy, J. D. Excess winter mortality in Europe: a cross country analysis identifying key risk factors. J. Epidemiol. Community Health* ***57****, 784–9 (2003).*
2. *Stewart, S., Keates, A. K., Redfern, A. & McMurray, J. J. V. Seasonal variations in cardiovascular disease. Nat. Rev. Cardiol.* ***14****, 654–664 (2017).*
3. *Gemmell, I., McLoone, P., Boddy, F., Dickinson, G. J. & Watt, G. Seasonal variation in mortality in Scotland. Int. J. Epidemiol. 29, 274–279 (2000).*

Minor comments:

- The method section, despite the obvious willingness from the authors to make it clear, is hard to read through. The fact that the authors use several different models to adress different research questions is part of the explanation. It might be of interest to the reader to include some equations summarizing the models used.

Response: Thank you for the suggestions. We included the equations in the supplementary material as below:

- Seasonality assessment without and with adjustments for temperature and/or influenza like illness (ILI)

We applied a generalized linear model with a quasi-Poisson family to assess seasonality of mortality in each prefecture.

Adjusting for temperature

Adjusting for ILI

Adjusting for both temperature and ILI

t: the day of the observation;
Y_t: the observed daily numbers of mortality on day t;
: the intercept;
doy: day of year, which was fitted using cyclic cubic spline with 4 degrees of freedom (df);
: the daily numbers of ILI on day t, which was controlled using natural cubic spline with 3 df;
$Strata_t$: strata defined by year, day of week, and their interaction to control for the long-term trend and the effect of day of week, and λ is the vector of coefficients;
$Temp_{t,l}$: a matrix obtained by using cross basis function to temperature; l is the lag days, and β is the vector of coefficients. (For the cross-basis function, a natural cubic B-spline basis with three internal knots at the 25th, 50th, and 75th percentiles of temperature distribution was used for exposure-response association, and another natural cubic B-spline basis with 3 df with extended lag up to 21 days was used for the lag-response association.)

- Modification of seasonal variation in mortality by prefecture-specific indicators

We applied linear mixed effects models (LMEMs) to investigate associations of PTR with each prefecture-specific indicator separately. We fitted LMEMs with random intercepts for prefectures and the inverse of squared SE as weight. The longitude and latitude for the capital city of each prefecture were included to reduce spatial correlation, except for when we investigated annual mean temperature as the indicator, due to their high correlation.

is the estimated coefficient for seasonality (i.e., log(PTR)) in prefecture
is the prefecture-specific indicator for prefecture (e.g., latitudes, longitudes, and averaged annual mean temperature)
and are estimated using least squares regression with inverse-variance weights.
is the variation within prefecture , with the variance as
represents the heterogeneity among prefectures with a variance of estimated using the restricted maximum likelihood approach.

- In the discussion section, when mentioning that prefectures with high climate experienced larger seasonal variations it might be of interest to refer to what is called "the seasonality paradox" (McKee (1989) Euro. J. of Epidemiology).
Response: Done.

- After page 9 all pages are numbered 1.

Response: Done.

- Authors do not publicly provide any code to reproduce their results, it would be nice to do so from an Open science perspective.

Response: We adopted the codes from https://github.com/gasparrini/2015_gasparrini_Lancet_Rcodedata and https://github.com/gasparrini/mixmeta , which were publicly available. Our codes are available from the first author upon request.

VERSION 2 – REVIEW

REVIEWER	Rosano, Aldo Italian National Agency for Regional Healthcare Services
REVIEW RETURNED	21-Apr-2021

GENERAL COMMENTS	The revisions made by the authors improved the manuscript, however there are still some concepts to be clarified. Here some details about the things to be clarified: Lines 92-94 "For example, a smaller seasonal amplitude was observed in areas with milder climates, suggesting that individuals living in warm areas might be more vulnerable to seasonal variations in mortality" This sentence seems contradictory, please check it or explain it. Line 96-97. Are the authors really interest only in "effect modifications"? In this case, you should test the hypothesis that the magnitude of the effect of the exposure variable (temperature and/or flu) on the outcome variable differs depending on a third variable through interactions terms (this is the epidemiological meaning of effect modification). Probably here the authors meant by the term "effect modification" in general sense, but it is better to avoid possible misunderstanding. Line 147. What is the rationale to control for the effect of the "day-of-week"? Do you suppose it is more probable to die in a specific day of the week? Why? Line 168-169. I suggest to better explain the hypothesis of spatial correlation mentioned in the lines 168-169. Line 241. Influenza epidemics may have an effect on mortality only in the winter seasons (and is highly correlated with cold temperatures), this should be considered in the discussion. Lines 278-279. This sentence is not clear.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Dr. Aldo Rosano, Italian National Agency for Regional Healthcare Services

Comments to the Author:

The revisions made by the authors improved the manuscript, however there are still some concepts to be clarified.

Here some details about the things to be clarified:
Lines 92-94 “For example, a smaller seasonal amplitude was observed in areas with milder climates, suggesting that individuals living in warm areas might be more vulnerable to seasonal variations in mortality” This sentence seems contradictory, please check it or explain it.

Response: We thank Dr. Rosano for pointing out this sentence. We have revised the sentence accordingly in the text (Lines 90-92):

“For example, a larger seasonal amplitude was observed in areas with milder climates, suggesting that individuals living in warm areas might be more vulnerable to seasonal variations in mortality”

Line 96-97. Are the authors really interest only in “effect modifications”? In this case, you should test the hypothesis that the magnitude of the effect of the exposure variable (temperature and/or flu) on the outcome variable differs depending on a third variable through interactions terms (this is the epidemiological meaning of effect modification). Probably here the authors meant by the term “effect modification” in general sense, but it is better to avoid possible misunderstanding.

Response: Thank you for pointing out this important issue. We have clarified this issue in our revision (Line 94 and Lines 101-102):

“However, only a few studies have evaluated their impact on seasonality of mortality.”

“In the current study, we collected daily mortality data between 1999 and 2015 from 47 prefectures in Japan to investigate the contribution of temperature and influenza to seasonality of mortality as well as to study the associations between prefecture-specific indicators and seasonality of mortality.”

Line 147. What is the rationale to control for the effect of the “day-of-week”? Do you suppose it is more probable to die in a specific day of the week? Why?

Response: The “day-of-week” is included to control for its potential confounding effect. Such confounding is possible if outcome and exposure are dependent on day of the week. Mortality has been shown to vary by day of the week,[1,2] and temperature is generally lower on weekends due to the low anthropogenic heat emissions.[3,4]

References:
1. Willich SN, Löwel H, Lewis M, Hörmann A, Arntz HR, Keil U. Weekly variation of acute myocardial infarction. Increased Monday risk in the working population. Circulation. 1994 Jul;90(1):87-93.
2. Chenet L, Britton A, Kalediene R, Petrauskiene J. Daily variations in deaths in Lithuania: the possible contribution of binge drinking. Int J Epidemiol. 2001 Aug;30(4):743-8.
3. Earl N, Simmonds I, Tapper N. Weekly cycles in peak time temperatures and urban heat island intensity[J]. Environmental Research Letters, 2016, 11(7): 074003.
4. Fujibe, F. Day-of-the-week variations of urban temperature and their long-term trends in Japan. Theor Appl Climatol **102,** 393–401 (2010).

Line 168-169. I suggest to better explain the hypothesis of spatial correlation mentioned in the lines 168-169.

Response: Locations close to each other exhibit more similar outcomes (here PTRs) than those farther apart. If this spatial correlation remains present in the residuals, one of the key assumptions of linearegression models, that residuals are independent, is violated. Therefore, we considered spatial correlation in our model.

Line 241. Influenza epidemics may have an effect on mortality only in the winter seasons (and is highly correlated with cold temperatures), this should be considered in the discussion.

Response: Thank you for the suggestion. We have included this in our discussion (Lines 247-248): *"The transmission of influenza virus is most efficient under cold and dry conditions, which may lead to considerable increase in mortality during winter."*

Lines 278-279. This sentence is not clear.

Response: We have modified the sentence as below (Lines 280-281):
"It is possible that temperature and influenza adjusted PTR may be overestimated due to the measurement error in temperature and influenza."