Assessing seasonality and the role of its potential drivers in environmental epidemiology: a tutorial

Lina Madaniyazi ![ORCID] ,[1,2]* Aurelio Tobias,[1,3] Yoonhee Kim,[4]
Yeonseung Chung ![ORCID] ,[5] Ben Armstrong[6] and Masahiro Hashizume[2,7]

[1]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan, [2]Department of Pediatric Infectious Diseases, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan, [3]Institute of Environmental Assessment and Water Research (IDAEA), Spanish Council for Scientific Research (CSIC), Barcelona, Spain, [4]Department of Global Environmental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, [5]Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, [6]Centre for Statistical Methodology, London School of Hygiene & Tropical Medicine, London, UK and [7]Department of Global Health Policy, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

*Corresponding author. School of Tropical Medicine and Global Health, Nagasaki University, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan. E-mail: lina.madaniyazi@nagasaki-u.ac.jp

Abstract

Several methods have been used to assess the seasonality of health outcomes in epidemiological studies. However, little information is available on the methods to study the changes in seasonality before and after adjusting for environmental or other known seasonally varying factors. Such investigations will help us understand the role of these factors in seasonal variation in health outcomes and further identify currently unknown or unmeasured risk factors. This tutorial illustrates a statistical procedure for examining the seasonality of health outcomes and their changes, after adjusting for potential environmental drivers by assessing and comparing shape, timings and size. We recommend a three-step procedure, each carried out and compared before and after adjustment: (i) inspecting the fitted seasonal curve to determine the broad shape of seasonality; (ii) identifying the peak and trough of seasonality to determine the timings of seasonality; and (iii) estimating the peak-to-trough ratio and attributable fraction to measure the size of seasonality. Reporting changes in these features on adjusting for potential drivers allows readers to understand their role in seasonality and the nature of any residual seasonal pattern. Furthermore, the proposed approach can be extended to other health outcomes and environmental drivers.

Key words: Seasonality, time-series, peak to trough, attributable fraction, mortality, temperature

---

Key Messages

- Examining the seasonality of a health outcome before and after adjusting for the potential environmental drivers may provide an insight into the aetiology of the disease.
- We propose a set of statistical procedures to summarize and compare the seasonality of a health outcome in three features: shape, timings and size, before and after the adjustment.
- We recommend two indicators to measure the size of seasonality: peak-to-trough ratio and attributable fraction to estimate the amplitude and impact of seasonality, respectively.
- Considering the extent and the peak of residual seasonality can also help identify further environmental or behavioural risk factors.

Introduction

Seasonal variation in health outcomes (hereafter, seasonality) has been largely recognized.[1–8] Seasonality is widely caused by various environmental factors, such as weather and air pollution, and others such as holidays. Several methods have been used to describe and assess the seasonality in epidemiological studies, including statistical tests and graphical methods.[9]

In recent environmental epidemiology studies, seasonality is usually considered as one of the main confounders when examining the health impact of environmental factors.[10] However few studies have focused on seasonality, and little attention has been paid to studying seasonality changes in the health outcome after adjusting for environmental or other known seasonally varying factors. Such investigation is of interest because it provides an insight into the aetiology of the disease. The changes in seasonality after adjustment will help us understand these factors' contributions to seasonality. In addition, the periodic and generally regular patterns in the residual time series may provide clues as to the presence and importance of other currently unknown or unmeasured causes, such as human behaviour.

We aim to introduce a set of statistical procedures to assess the seasonal variation of a health outcome and its changes after accounting for the potential environmental drivers. Throughout, concepts and methods will be illustrated through an example dataset for investigating the change in seasonality of all-cause mortality before and after adjusting for the short-term and direct effect of temperature in London. We used temperature as an example for the environmental driver because of its well-documented associations with mortality[11] and the easy accessibility of the data.

Example dataset

We collected daily counts of all-cause mortality and daily mean temperature in London between 1993 and 2006. Here, the number of all-cause deaths shows a repeating seasonal pattern that appears approximately sinusoidal, with an increase in cold seasons followed by a decrease in warm seasons (Figure 1 and Table 1). This dataset has been previously used as an example elsewhere.[10,12]

All the statistical analyses were conducted using R, version 3.6.3. The example dataset, along with the R code to reproduce the analyses, are available as Supplementary Material at *IJE* online.

Assessing seasonality

Time series regression with a cyclic spline

A wide variety of methods can be applied to assess seasonality in a health outcome, such as regression models with indicator variables for the month, cosinor models (a single sine+cosine pair with annual periodicity) and extending this by adding harmonics, often called Fourier functions.[9] Recently, several studies have applied a cyclic spline function to model seasonality on a daily basis.[1,5] A cyclic spline function is a smoothing method to estimate periodic variation such as daily or annual pattern of time-series observations. Basically, it is a periodic piece-wise cubic function with continuity up to the second derivative, so the function of day-of-year changes continuously at the end of the year.[13]

Here, we illustrate the modelling approach using a time-series regression with a cyclic spline function to assess seasonality by following our previous work[5]:

$$Y_t \sim quasi - Poisson(E(Y_t))$$

without temperature adjustment:

$$Log(E(Y_t)) = \alpha + cs(Doy_t) + \lambda Strata_t$$

with temperature adjustment:

$$Log(E(Y_t)) = \alpha + cs(Doy_t) + \lambda Strata_t + cb(Temp_{t,l})$$

where: Y_t is mortality on day t assumed to follow a Poisson distribution with overdispersion (i.e. quasi-Poisson). Doy is the day of year on day t ranging from 1 to 366 to model seasonality. We used a cyclic spline (cs) with 4 degrees of freedom (df) for the day of year. *Strata* is the strata defined
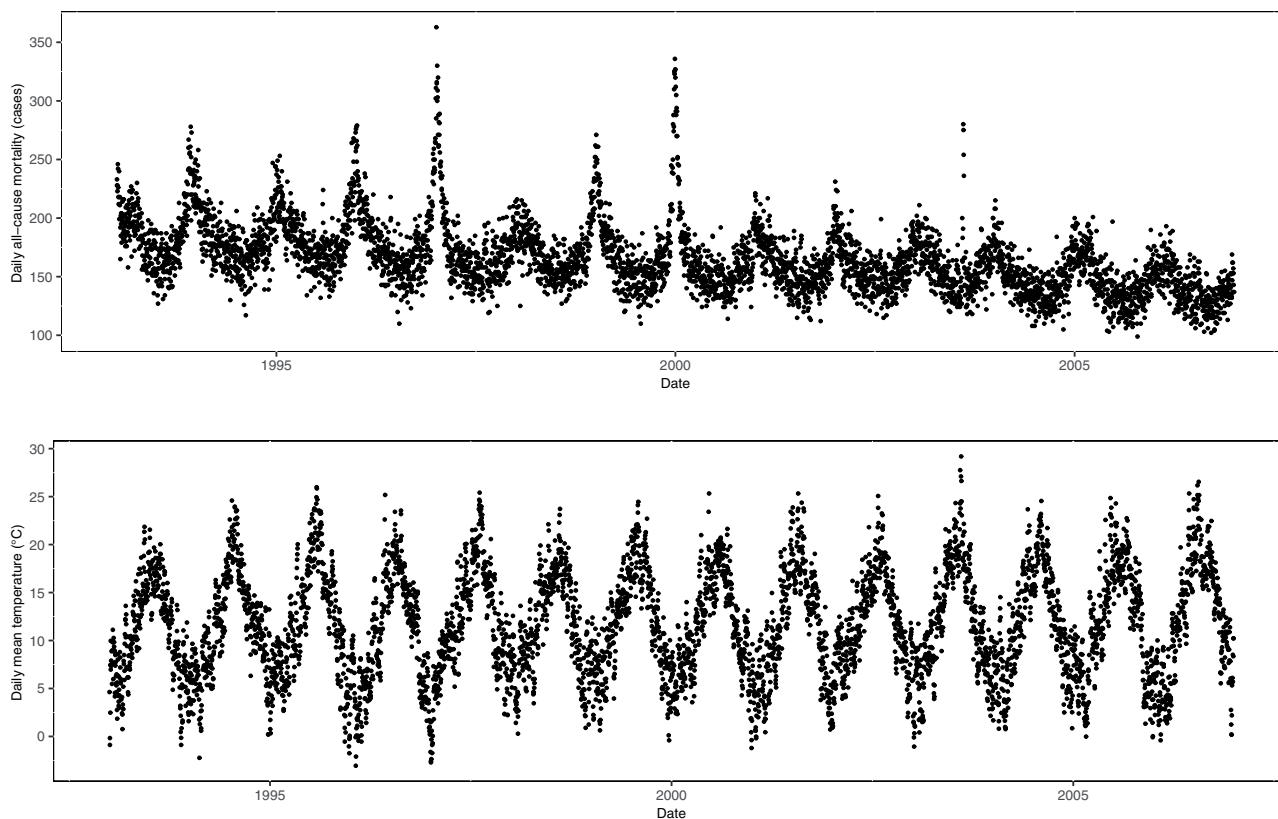
Figure 1 Daily time-series of all-cause mortality and mean temperature in London from 1993 to 2006

Table 1 Descriptive summary of all-cause mortality and ambient temperature by season, London 1993–2006 [mean (Standard Deviation, SD)]

Variable (daily)	Whole year	Season			
		Winter (Dec–Feb)	Spring (Mar–May)	Summer (Jun–Aug)	Autumn (Sep–Nov)
All-cause mortality (cases)	165.3 (29.2)	190.9 (34.3)	163.3 (20.0)	149.8 (19.5)	157.7 (22.6)
Mean temperature (°C)	11.7 (5.5)	6.0 (3.1)	10.5 (3.7)	18.0 (3.0)	12.2 (4.0)

by year, day of week and their interaction to control long-term trend and the effect of the day of week. $Temp_{t,l}$ is a vector obtained using a cross-basis function of daily mean temperature; l is the lag days. 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 is used for the exposure-response association, and another natural cubic B-spline basis with 3 *df* with extended lag up to 21 days is used for the lag-response association. We assessed the seasonality before and after temperature adjustment separately.
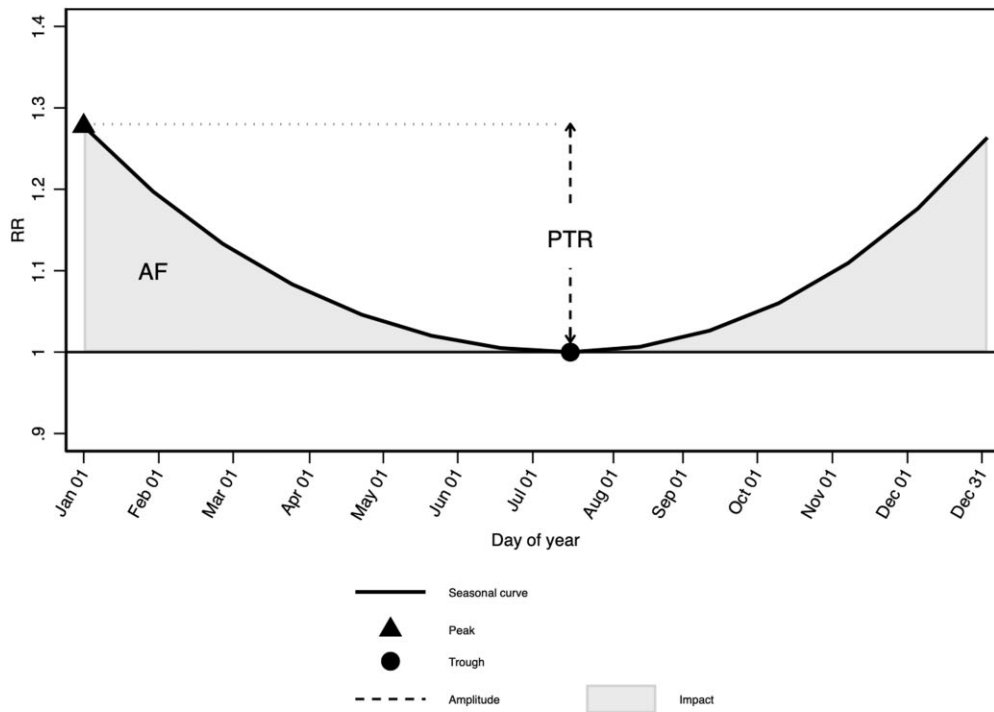
Summarizing and comparing seasonality

The key features of seasonality include its shape, timings (peak and trough) and size (amplitude and impact)

(Figure 2, panel a). We can summarize and compare the seasonality by each key feature before and after the adjustment (Figure 2, panel b).

The empirical confidence intervals (eCIs) for timings and impact can be obtained through Monte Carlo simulations.[16,17] In brief, random samples are taken from the original parameters of the cyclic spline function, which are assumed to follow a multivariate normal distribution with their point estimates and variance matrix derived from the regression model. The timings and impact are computed from these samples, which empirically reconstruct their distributions. The related 2.5th and 97.5th percentiles of these distributions are interpreted as 95% eCIs. The results from our example are presented in Table 2 and Figure 3 and discussed in detail below.

**(a)**   Features for summarizing seasonality.

(b)   Key features for comparing seasonality before and after adjusting for an environmental driver.
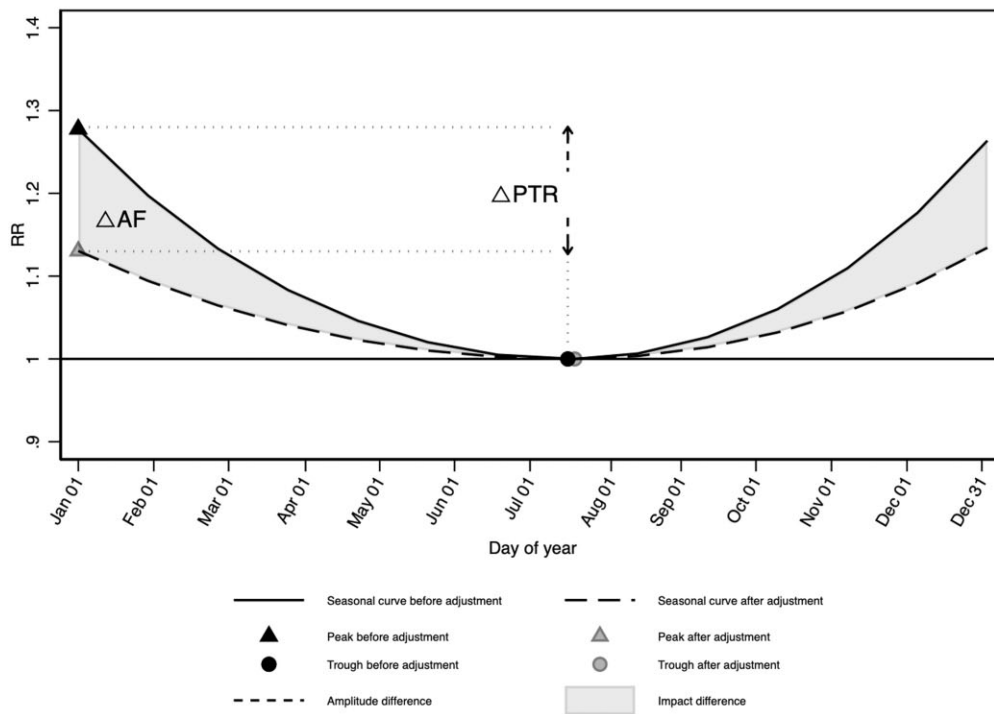
Figure 2 Key features for summarizing and comparing seasonality. PTR, peak-to-trough ratio; AF, attributable fraction; RR (relative risk) $= \frac{Mortality\ estimate\ on\ the\ day-of-year\ x}{Minimum\ mortality\ estimate\ at\ the\ trough}$; $\triangle$PTR (change in PTR after adjustment) $= \exp(\log(PTR_{after}) - \log(PTR_{before}))$; $\triangle$AF (change in AF after adjustment) $= AF_{after} - AF_{before}$

Table 2 Seasonality assessment of all-cause mortality before and after temperature adjustment

Temperature adjustment	Shape	Timings (day-of-year) (95% empirical confidence interval)		Size	
		Peak	Trough	Peak-to-trough ratio (95% confidence interval)	Attributable fraction (%) (95% empirical confidence interval)
Unadjusted	High mortality in cold seasons and low mortality in warm seasons	9 (8, 10)	250 (244, 255)	1.34 (1.32, 1.35)	10.6 (10.1, 11.1)
Adjusted	High mortality in cold seasons and low mortality in warm seasons; a smaller amplitude	1 (362, 3)	249 (101, 257)	1.14 (1.10, 1.17)	4.1 (3.8, 5.9)
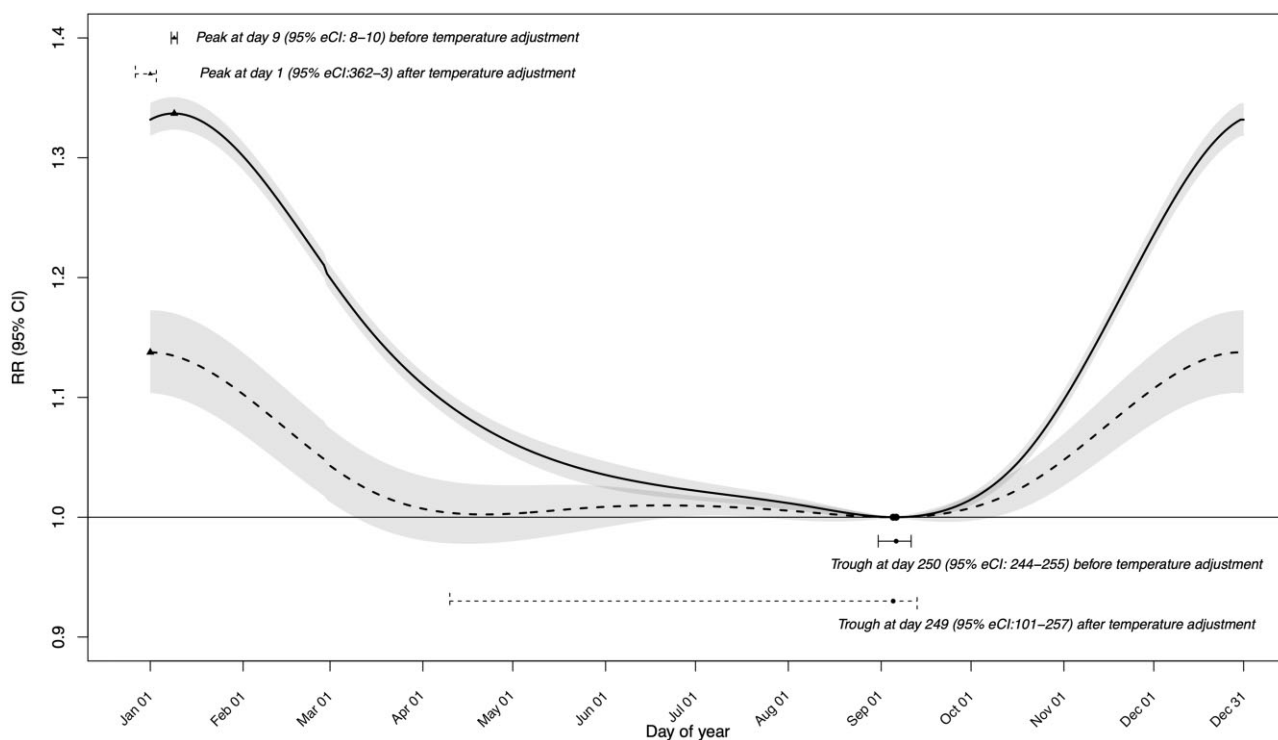
Figure 3 Seasonality of all-cause mortality, and its peak and trough days before (solid) and after (dashed) temperature adjustment. The seasonality is assessed using a time-series regression model with a cyclic spline function with 4 degrees of freedom. The relative risk (RR) is the ratio of mortality estimates on the day of year x to daily minimum mortality estimates at the trough day with 95% confidence intervals (95% CI):

$$\text{Relative risk} = \frac{\text{mortality estimate on the day} - \text{of} - \text{year x}}{\text{minimum mortality estimate at the trough}}$$

The day of year with maximum and minimum mortality estimates is identified as the peak (triangle) and trough (circle) day, respectively, of the seasonality of mortality. Monte Carlo simulation was used to obtain empirical confidence intervals for peak and trough days

Shape: the fitted seasonal curve

The shape of seasonality can be estimated through the fitted seasonal curve from the regression model described above, and compared by visual inspection before and after adjusting for the environmental driver. Figure 3 shows the seasonal pattern for all-cause mortality in London with a unimodal shape: a sharp peak in winter and a more extended trough in warmer months. After adjusting for

temperature, the seasonal pattern remained similar with reduced amplitude.

Timings: peak and trough

The day-of-year maximum and minimum mortality estimates were identified from the fitted seasonal curve as the peak and trough, respectively.[15] The eCIs for peak and trough were obtained through Monte Carlo simulations.[16,17]

In our example (Table 2), the peak and trough estimates were observed at Days 9 (95% eCI = 8, 10) and 250 (95% eCI = 244, 255), respectively. After adjusting for temperature, the peak and trough days moved forward to Day 1 (95% eCI = 362, 3) and Day 249 (95% eCI = 101, 257). It should be noted that the uncertainty of the trough is higher after temperature adjustment.

Size: amplitude and impact

We propose to summarize the size of seasonality by measuring the amplitude and estimating its impact on the health outcome. The amplitude of seasonality can be measured as the ratio of the maximum mortality estimate at peak day to the minimum mortality estimate at trough day (i.e. peak-to-trough ratio, PTR),[5] and its 95% CI can be obtained from the variance matrix of the estimated coefficients from the cyclic spline function.[12] However, the PTR is not sensitive to seasonality shape and, more importantly, offers limited information on the impact of seasonality on mortality.[18] In this context, the attributable fraction (AF) may help understand the public health burden of seasonality. A general definition of AF_x for a given exposure x can be provided by $AF_x = 1 - \exp(-\beta_x)$, where β_x refers to the association of the outcome with a specific exposure intensity x compared with a reference value x_0. Here, we obtained the overall AF as[19]: $AF = \sum p_x[1 - \exp(-\beta_x)]$, where x represents the day of year, ranging from 1 to 366, β_x is the log relative risk of mortality on the day of year x compared with the minimum mortality estimated at the trough, and p_x is the percentage of cases on the day of year x. The empirical CIs for AF were estimated through Monte Carlo simulations.[17] Thus, the AF estimates the fraction by which mortality would be reduced in a counterfactual scenario where mortality risk never rose above its seasonal trough.

Table 2 reports an estimated PTR of 1.34 (95% CI = 1.32, 1.35), which is substantially reduced after adjusting for temperature to 1.14 (95% CI = 1.10, 1.17). The estimated AF indicates that 10.6% of deaths (95% eCI = 10.1, 11.1) are attributable to seasonality within the study period. After adjusting for temperature, the AF decreases to 4.1% (95% eCI = 3.8, 5.9).

The difference in the PTR and AF before and after the adjustment can be interpreted as the contribution of temperature
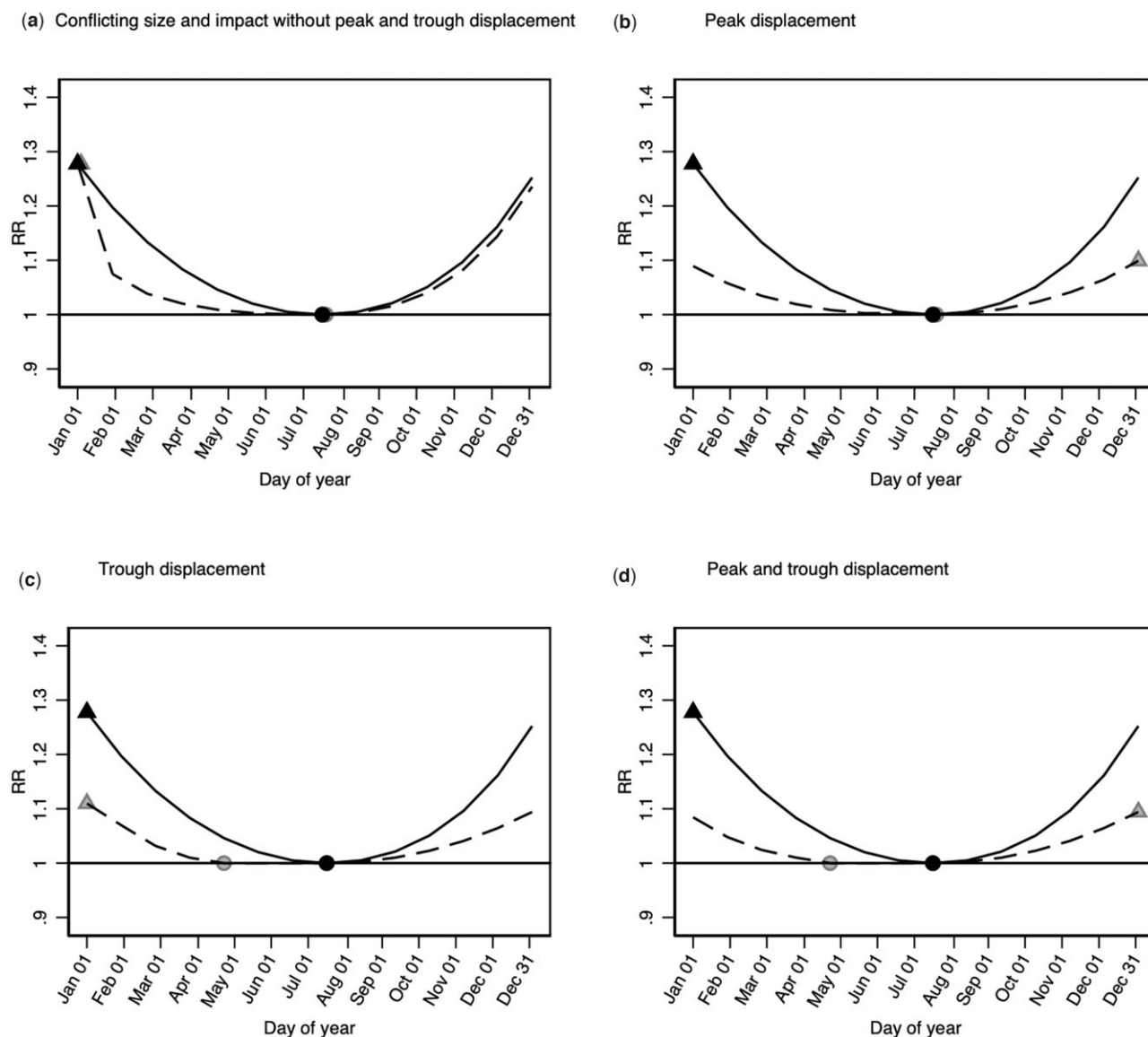
to the size of the seasonality of mortality and can be calculated as[20]: $d = \hat{E}_{after} - \hat{E}_{before}$ and $SE(d) = \sqrt{S\hat{E}_{after}^2 + S\hat{E}_{before}^2}$, where $\hat{E}$ are the estimates of the size before and after the adjustment, and $S\hat{E}$ are their respective standard errors. Here, we assume $\hat{E}_{before}$ and $\hat{E}_{after}$ are independent, which may overestimate $SE(d)$. Consequently, in our example, we observed an absolute reduction in the log(PTR) of 0.16 (95% CI = 0.14, 0.17), and an absolute reduction in the AF of 6.5% (95% eCI = 4.6, 8.5)

We recommend reporting the changes in PTR and AF jointly since we may find a particular set of situations in which, for example, we do not observe a change in the size of seasonality through the PTR, but with different impact through the AF, before and after the adjustment (Figure 4, panel a). Likewise, when the timings of seasonality are displaced notably before and after the adjustment (Figure 4, panels b, c and d), the estimation of PTR and AF after adjustment will be based on different shapes and timings of seasonality. Therefore, the changes would require additional descriptions and careful interpretations. The peak and/or trough displacement must be reported to make the readers understand how seasonality changed after adjustment.

Modelling choice

Alternative functions for seasonality

In our example, we have used a cyclic spline function to assess seasonality in daily mortality. Alternative models with different specifications can also be used: for example, a stationary cosinor, non-stationary cosinor, loess smoothing, Fourier function and conditional autoregression with the month as a random effect.[9] Barnett and Dobson offer a thorough overview of these methods.[9] The readers should select the appropriate model based on their data, research question and model fit. For instance, a cosinor is a single cosine/sine couple and can be considered as a special case of the Fourier function. A stationary cosinor is preferred for data covering a short period. A non-stationary cosinor is more appropriate for irregularly spaced data and will enable us to investigate the temporal changes of seasonality over a long period. This method can also be updated further to estimate the temporal changes.[9] However, these two methods assume a sinusoidal seasonal pattern. Whereas loess smoothing is useful for fitting a non-sinusoidal seasonal pattern, it only estimates the mean but not confidence intervals. On the other hand, spline functions provide a flexible and efficient way to fit irregular and/or complex seasonal patterns, even for those irregularly spaced data with few parameters. Fourier functions

(a) Conflicting size and impact without peak and trough displacement

(b) Peak displacement

(c) Trough displacement

(d) Peak and trough displacement

Figure 4 Example s of particular issues when assessing changes in the seasonal pattern before (solid) and after (dashed) adjusting for an environmental driver

have similar properties but may need more degrees of freedom (basis variables) than a cyclic spline to capture typical seasonal patterns.

Model choice may be based on model fit criteria such as deviance, Akaike's information criterion (AIC) or tests for white noise in residuals.[12] In addition to the function for seasonality, it is also desirable to consider model choices for the adjustment of environmental drivers, long-term trend and other specific factors.

Model checking

It is important to examine the models' residuals to check key assumptions of the regression models, including a scatter plot of residuals against the independent variable, the independence of residuals over time (e.g. autocorrelation) and the distribution of residuals.[10,14]

Sensitivity analysis

Since the modelling process involves many decisions, multiple sensitivity analyses are recommended to check the robustness of the main conclusions.[1,10,12] A series of sensitivity analyses on *df* of the cyclic spline and temperature adjustment has been conducted in our previous study[5] where the example dataset was included. In this tutorial, we compared the cyclic spline and a cosinor function (Supplementary Figure S1, available as Supplementary data at *IJE* online). Our sensitivity analysis showed a lower quasi-AIC for the model with a cyclic spline

Figure 5 Summary of key steps in quantifying seasonality changes. PTR, peak-to-trough ratio; AF, attributable fraction

(Supplementary Table S1, available as Supplementary data at *IJE* online), indicating that the cyclic spline fits the data better than a cosinor function, each with the same degrees of freedom.

Discussion

This article outlines a set of statistical procedures for examining the seasonality of health outcomes and their changes after adjusting potential environmental drivers (Figure 5). We illustrate the procedure by modelling the seasonal curve using a time-series regression with cyclic splines and summarizing the seasonality in different aspects from the fitted curve. In particular, we recommend summarizing and comparing seasonality by shape, timings of peaks and troughs, and size (PTR and AF) before and after the adjustment.

We believe that these procedures are applicable to a wide range of contexts, though some work would be needed for such generalizations, and there are some limitations. In particular, our example is tailored to a unimodal seasonal pattern of all-cause mortality and temperature. However, it can be extended to multimodal seasonality and other outcomes and environmental drivers, sometimes

applying alternative modelling choices. For example, a recent study[8] applied a quasi-Poisson regression model with a cyclic spline function of 3 *df* to investigate the impact of temperature on the bimodal seasonality of sports injuries in Madrid, Spain. Also, a previous study[18] illustrated the contribution of climatic factors to the bimodal seasonal pattern of cholera incidence in Dhaka, Bangladesh, using a Poisson regression model with Fourier functions. As this cholera study[18] suggests, the procedure described can more generally be applied to infectious diseases with some model variations.[21,22] In addition, the proposed approach here can also be used to analyse data from multiple locations, and the differences in location-specific seasonality estimates can be explored further through a meta-analytical technique.[1,5]

An important issue that should be addressed carefully is the optimal adjustment of the environmental drivers. This can be particularly critical for infectious diseases, as they usually exhibit a complex association with environmental factors.[14,21] In addition, although the cyclic spline function in our example is flexible for seasonality assessment, it is more mathematically complex and difficult to interpret than some alternatives, especially when the fitted curve is very wiggly. Therefore, the readers should critically assess the potential modelling alternatives and adapt the statistical procedure to their investigations.

Conclusion

In conclusion, the proposed framework covers key steps and important issues involved in seasonality assessment, and provides an opportunity to advance through general methodological steps for a further examination of the underlying drivers of seasonality.

Ethics approval

Not applicable.

Data availability

Example dataset and the R code for the analysis are available as Supplementary data at *IJE* online.

Supplementary data

Supplementary data are available at *IJE* online.

Author contributions

L.M. analysed the data and drafted the manuscript. A.T. designed the study and directed the study's implementation. Y.K. helped to design the analytical strategy and to interpret the findings. Y.C. helped with data analysis and the revision of the manuscript. B.A.

helped to analyse and interpret the data and revise the article critically. M.H. made a substantial contribution to the concept of the study and interpretation of data.

Conflict of interest

None declared.

References

1. Yu J, Yang D, Kim Y *et al.* Seasonality of suicide: a multi-country multi-community observational study. *Epidemiol Psychiatr Sci* 2020;**29**:e163.
2. Stewart S, Keates AK, Redfern A, McMurray JJ. Seasonal variations in cardiovascular disease. *Nat Rev Cardiol* 2017;**14**:654–64.
3. Paireau J, Chen A, Broutin H, Grenfell B, Basta NE. Seasonal dynamics of bacterial meningitis: a time-series analysis. *Lancet Glob Health* 2016;**4**:e370–77.
4. Althouse BM, Flasche S, Thiem VD *et al.* Seasonality of respiratory viruses causing hospitalizations for acute respiratory infections in children in Nha Trang, Vietnam. *Int J Infect Dis* 2018;**75**:18–25.
5. Madaniyazi L, Armstrong B, Chung Y *et al.*; Multi-Country Multi-City (MCC) Collaborative Research Network. Seasonal variation in mortality and the role of temperature: a multi-country multi-city study. *Int J Epidemiol* 2022;**51**:122–33.
6. Momiyma M. Biometeorological study of the seasonal variation of mortality in Japan and other countries. *Int J Biometeor* 1968; **12**:377–93.
7. Momiyama M, Katayama K. Deseasonalization of mortality in the world. *Int J Biometeorol* 1972;**16**:329–42.
8. Tobías A, Casals M, Saez M, Kamada M, Kim Y. Impacts of ambient temperature and seasonal changes on sports injuries in Madrid, Spain: a time-series regression analysis. *BMJ Open Sport Exerc Med* 2021;**7**:e001205.
9. Barnett AG, Dobson AJ. *Analysing Seasonal Health Data*. Berlin: Springer, 2010.
10. Bhaskaran K, Gasparrini A, Hajat S, Smeeth L, Armstrong B. Time series regression studies in environmental epidemiology. *Int J Epidemiol* 2013;**42**:1187–95.
11. Gasparrini A, Guo Y, Hashizume M *et al.* Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *Lancet* 2015;**386**:369–75.
12. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology* 2006;**17**:624–31.
13. Wood SN. *Generalized Additive Models: An Introduction with R*. 2nd edn. New York, NY: Chapman and Hall/CRC, 2017.

14. Imai C, Hashizume M. A systematic review of methodology: time series regression analysis for environmental factors and infectious diseases. *Trop Med Health* 2015;**43**:1–9.

15. Christiansen CF, Pedersen L, Sørensen HT, Rothman KJ. Methods to assess seasonal effects in epidemiological studies of infectious diseases—exemplified by application to the occurrence of meningococcal disease. *Clin Microbiol Infect* 2012;**18**: 963–69.

16. Tobías A, Armstrong B, Gasparrini A. Brief Report: investigating uncertainty in the minimum mortality temperature. *Epidemiology* 2017;**28**:72–76.

17. Gasparrini A, Leone M. Attributable risk from distributed lag models. *BMC Med Res Methodol* 2014;**14**:55–58.

18. Hashizume M, Faruque AS, Wagatsuma Y, Hayashi T, Armstrong B. Cholera in Bangladesh: climatic components of seasonal variation. *Epidemiology* 2010;**21**:706–10.

19. Steenland K, Armstrong B. An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology* 2006;512–19.

20. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ* 2003;**326**:219.

21. Imai C, Armstrong B, Chalabi Z, Mangtani P, Hashizume M. Time series regression model for infectious disease and weather. *Environ Res* 2015;**142**:319–27.

22. Fisman DN. Seasonality of infectious diseases. *Annu Rev Public Health* 2007;**28**:127–43.