

Quantitative Assessment of the Contribution of Meteorological Variables to the Prediction of the Number of Heat Stroke Patients for Tokyo

Takuto Sato¹, Hiroyuki Kusaka², and Hideitsu Hino³

¹Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan

²Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan

³The Institute of Statistical Mathematics, Tokyo, Japan

Abstract

This study reveals the best combination of meteorological variables for the prediction of the number of emergency transport due to heat stroke over 64 years old in Tokyo metropolis based on a generalized linear model using 2008–2016 data. Temperature, relative humidity, wind speed, and solar radiation were used as candidates of the explanatory variables. The variable selection with Akaike's information criterion (AIC) showed that all the four meteorological elements were selected for the prediction model. Additional analysis showed that the combination of daily mean temperature, maximum relative humidity, maximum wind speed, and total solar radiation as explanatory variables gives the best prediction, with approximately 19% less error than the conventional single-variable model which only uses the daily mean temperature. Finally, we quantitatively estimated the relative contribution of each variable to the prediction of the daily number of heat stroke patients using standardized partial regression coefficients. The result reveals that temperature is the largest contributor. Solar radiation is second, with approximately 20% of the temperature effect. Relative humidity and wind speed make relatively small contributions, each contributing approximately 10% and 9% of the temperature, respectively. This result provides helpful information to propose more sophisticated thermal indices to predict heat stroke risk.

(Citation: Sato, T., H. Kusaka, and H. Hino, 2020: Quantitative assessment of the contribution of meteorological variables to the prediction of the number of heat stroke patients for Tokyo. *SOLA*, **16**, 104–108, doi:10.2151/sola.2020-018.)

1. Introduction

In recent years, the demand for ambulance transport due to heat strokes has been increasing in Japan. This has resulted in social problems such as the limitation of outdoor activities. Therefore, predicting the number of heat stroke patients is not only useful for human health risk assessment, but also for ambulance management and planning of operations. Studies have sought relationships between human comfort or the number of heat stroke patients and climatological conditions (e.g. Ishigaki et al. 2001; Honda et al. 2007). Piver et al. (1999) showed that heat stroke emergency transport demand could be estimated using the daily maximum temperature and NO₂ concentrations. Miyatake et al. (2012) also reported that ambulance transport due to heat strokes was positively correlated with air temperature. Ono (2013) showed a strong correlation between the daily number of ambulance calls for heat stroke patients and daily maximum temperature. Furthermore, Fujibe et al. (2018b) showed that the higher humidity caused the higher risk under the same daily maximum temperature condition. However, the relative contribution of each meteorological variable to the prediction of heat stroke occurrence is still unclear.

To focus on the prediction of heat stroke occurrence, Fuse

et al. (2014) developed a prediction model of the number of ambulance transports due to heat stroke, based on a single-variable least-squares method with the daily mean temperature as the explanatory variable. This model is useful for risk assessment, but there are some problems. First, Fuse et al. (2014) used temperature as the only explanatory variable. It is well known that there are four important meteorological elements (i.e. temperature, humidity, wind speed, and solar radiation) to determine the thermal condition. Indeed, researchers have reported a significant relationship between heat stroke occurrence or thermal stress and various thermal indices consisting of these multiple meteorological elements such as the wet bulb globe temperature (WBGT) and the universal thermal climate index (UTCI) (e.g., Ohashi et al. 2014; Suzuki-Parker and Kusaka 2016; Akatsuka et al. 2016; Di Napoli et al. 2018). The second problem is that Fuse et al. (2014) used only the correlation coefficients to select the predictive meteorological element. It is better to judge the appropriateness of elements by considering not only their correlation with the objective variable but also the information criterion score. The information criterion can evaluate the appropriateness of variables for prediction because it is calculated using likelihood and the number of variables.

The goal of this study is to propose the multi-variable prediction model with the best variable combination for predicting heat stroke patients. The unique aspect of this model is that both the Akaike's information criterion (AIC) and root mean squared error (RMSE) are used to judge the best prediction model. The other objective is to independently investigate the contribution of each meteorological element to the prediction.

2. Methods and data

2.1 Experiment overview

The present study develops several statistical models to predict the daily number of ambulance transports of heat stroke patients over 64 years old (hereafter referred to as the number of heat stroke patients), using meteorological variables. The reason why we focus on patients over 64 years old is that heat stroke occurrence in elderly people has become a social problem in Japan. Their susceptibility to thermal stress is greater than that of other age groups (Semanza et al. 1996; Nakai et al. 1999). The models are based on the generalized linear model (GLM, Dobson and Barnett 2008), shown in equation (1).

$$pat = \exp \left[\sum_{i=1}^N \beta_i x_i \right] \quad (1)$$

Here, pat , x_i , β_i , N denote the number of heat stroke patients, explanatory variables (meteorological variables), regression coefficients, and the number of meteorological variables, respectively. The maximum likelihood method is used for parameter estimation. The GLM can assume not only normal distribution but also other probability distributions as objective variables. In this study, the Poisson distribution is assumed as the objective variable (pat), and the parameters that maximize the log likelihood are estimated. In addition, the zero-inflated Poisson distribution (ZIP, Lambert 1992) was also used for fitting because the count of heat stroke

Corresponding author: Hiroyuki Kusaka, Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577. E-mail: kusaka@ccs.tsukuba.ac.jp.



patients is zero on many days. It is considered that the ZIP model is useful for data which seems to have excess zero like the data in this study. To clearly distinguish the contribution of the four meteorological elements to the prediction, the prediction models are developed in two steps as described below.

We first used AIC to determine the best combination of the explanatory variables of the statistical model from the three daily mean values (temperature, relative humidity, wind speed) and total solar radiation. It is important that daily maximum and minimum values are not used in this step. This is because the relative contribution cannot be discussed if the same elements are selected more than twice. Here, the result of this selection which has the smallest AIC is called the best combination of meteorological elements in this study. Note that, to focus on the individual effect of each explanatory variable for prediction, we do not consider the cross terms.

The next step is searching for the best prediction model by replacing the daily mean values with daily maximum and minimum values. At the end of this step, we can obtain the best prediction model (which has the smallest RMSE) in this study. Here, RMSE and MAE are calculated using observed and estimated number of heat stroke patients. Through this two-step model development, we can obtain an accurate model in which the relative contribution of each meteorological elements is easy to clarify.

To understand the method described in the above, the procedures are summarized as follows.

- (i) Investigation of the best combination of meteorological elements using AIC and daily mean of meteorological elements.
- (ii) Investigation of the best prediction models using the same combination in (i), but replacing the daily mean values with the daily maximum and minimum values.

After these two steps, the relative contribution of each meteorological elements will be investigated by calculating standardized partial regression coefficient of the best prediction model.

2.2 Data

As objective variable, we used the daily number of the heat stroke patients over 64 years old in Tokyo metropolis. The dataset is provided by the Fire and Disaster Management Agency web page (<https://www.fdma.go.jp/disaster/heatstroke/post3.html>). Here, the reason why we predict the number of heat stroke patients is that it is useful not only for risk assessment, but also for operation planning for ambulance management.

As candidate explanatory variables, hourly meteorological elements were used, which were obtained from the automated meteorological data acquisition system (AMeDAS) of the Japan Meteorological Agency. Note that temperature and wind speed were spatially averaged across all eight observation points in Tokyo metropolis without any weighting factors (shown in Fig. 1). On the other hand, relative humidity and solar radiation were not spatially averaged, but the values obtained at the Tokyo regional headquarters, JMA were used. The daily maximum, mean, and minimum values were calculated from these hourly data from 1:00 to 24:00 JST.

The training period was set to be July–September in the years 2008 to 2013 to estimate the model parameters. This choice entails the assumption that there were no yearly trends in the examined parameters and the model would not change from year to year. The experiments for model validation were conducted using July–September for the period 2014 to 2016.

3. Results

3.1 Preliminary analysis: correlation analysis

As a preliminary analysis, we investigated the relationships between the number of heat stroke patients and daily mean meteorological variables: temperature, relative humidity, wind speed, and total solar radiation.

Figure 2 shows the scatter diagram used to understand the relationship between daily mean meteorological elements and the number of heat stroke patients. A positive correlation exists

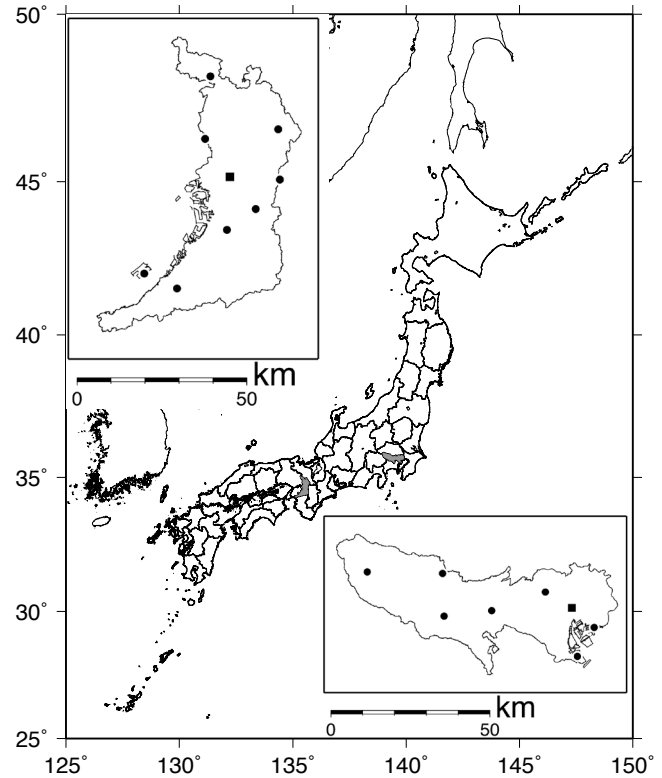


Fig. 1. Target areas and observation sites used in this study. Circle: AMeDAS (temperature and wind speed), Squares: Tokyo and Osaka headquarters, JMA (all four meteorological elements).

between the temperature and the number of heat stroke patients (correlation coefficient, $R = 0.61$). A similar result was obtained for the total solar radiation ($R = 0.48$). On the other hand, there is no clear correlation between relative humidity and the number of heat stroke patients ($R = -0.19$). Wind speed has also a small correlation with heat stroke patients ($R = 0.10$). Here, no collinearity was confirmed (scatter diagrams are shown in Supplement 1).

Relative humidity and wind speed are not necessary to use as explanatory variables in terms of correlation with the objective variable. However, for better prediction, the appropriate variable combination should be confirmed using AIC. Therefore, the variable selection using daily mean meteorological elements (temperature, relative humidity, wind speed) and the total solar radiation was carried out.

3.2 Variable selection using daily mean values and AIC

Results of AIC indicate that all four meteorological elements are able to contribute to the prediction of heat stroke patients. Thus, all four elements should be used as explanatory variables. The model with the best combination is derived as equation (2).

$$\text{pat} = \exp[-14.22 + 0.5260T_{\text{ave}} - 0.2017WS_{\text{ave}} + 0.0292RH_{\text{ave}} + 0.0555S_0] \quad (2)$$

T , WS , RH , and S_0 indicate temperature ($^{\circ}\text{C}$), wind speed (m/s), relative humidity (%), and total solar radiation (MJ/m^2), respectively. The RMSE and mean absolute error (MAE) of this prediction model were 12.6 and 6.9, respectively.

Here, the same experiment (searching for the best combination) was executed based on mixed model which assumed ZIP. The RMSE and MAE of this model were 12.8 and 7.0, respectively. These values are almost the same as for the GLM based model, thus we use the GLM based model for the following analysis because of its simplicity and ease of understanding.

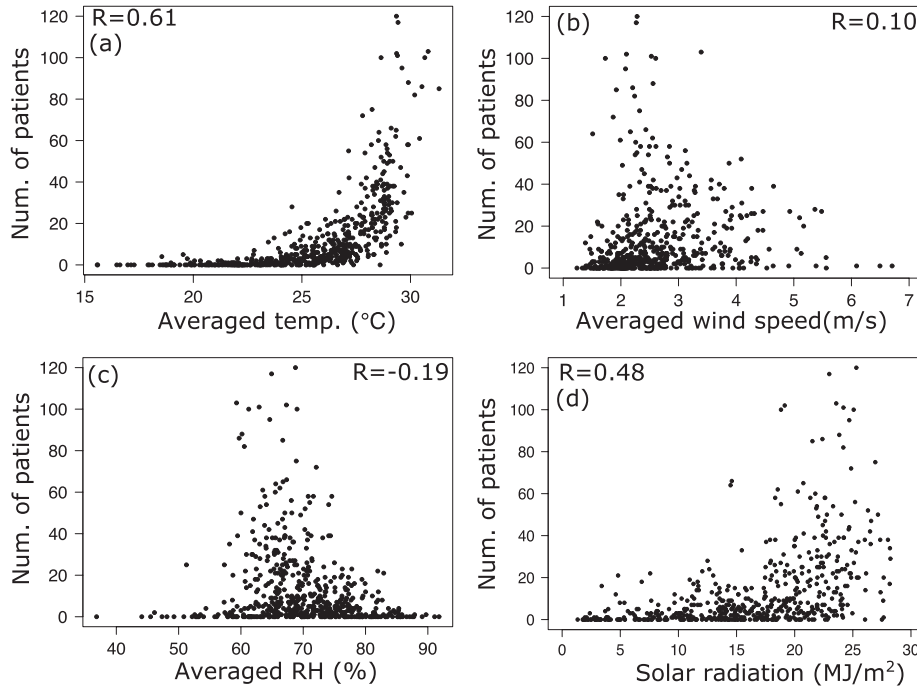


Fig. 2. Relationships between thermal elements and heat stroke patients in the training period (2008–2013). (a) daily mean temperature, (b) daily mean wind speed, (c) daily mean relative humidity, and (d) total daily solar radiation

3.3 The model with the smallest error

In Section 3.2, the model with four meteorological elements was selected. In this section, we seek the best prediction model by replacing the daily mean values with daily maximum and minimum values. Regarding S_0 , the daily total value and daily maximum value are considered. Here, the best prediction model in this study means that the model shows the smallest RMSE and MAE in three target years. The errors of all models considered in this research are summarized in Fig. 3.

The GLM with daily mean temperature, daily maximum relative humidity, daily maximum wind speed, and total solar radiation produces the smallest error. We therefore call equation (3) the best prediction model.

$$\begin{aligned} pat = \exp[-13.91 + 0.5234T_{ave} - 0.1100WS_{max} \\ + 0.0228RH_{max} + 0.0479S_0] \end{aligned} \quad (3)$$

The RMSE and MAE of this model are 11.8 and 6.7, respectively.

Hereafter, the results of the best prediction model and conventional single-variable model with daily mean temperature are compared. Figure 4 compares the time series of predicted heat stroke patients using the two models. The best prediction model in this study can reproduce spike shaped fluctuations more accurately compared with the single-variable model, in addition to reducing total error (RMSE and MAE). However, both models tend to underestimate the number of patients especially in July and early August. This is probably because these two models don't include the acclimatization effect. Here, acclimatization means that people are less adapted to heat in early summer (rapid temperature rise) compared with mid or end of summer (mentioned by Nairn and Fawcett 2015; Fujibe et al. 2018a). The error of heat stroke patients also varies in years. Possible reasons are the difference of characteristics of the summer (i.e., hot and cold summer) and/or psychological effect.

3.4 Contribution of meteorological variables to the prediction

After finding the best prediction model, standardized partial regression coefficients were calculated to investigate the relative contribution of each individual explanatory variable to the pre-

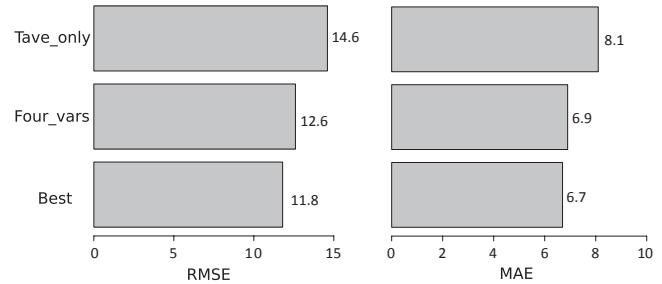


Fig. 3. Errors of all prediction models in this research. “Tave_only” denotes the single-variable model with daily mean temperature, “Four_vars” denotes the model with the best combination (Eq. 2), “Best” denotes the best prediction model (Eq. 3).

diction. Here, the coefficients were calculated by normalizing the explanatory variables in this research because we used a log-linear model. Here, normalization means that the average values are adjusted to 0, and the standard deviations are adjusted into 1. The coefficients from the best prediction model are shown in Table 1. In Tokyo, the contribution of the daily mean temperature was the largest, followed by that of the total solar radiation (20% of temperature's contribution). The contribution of the daily maximum relative humidity and daily maximum wind speed were approximately 10% and 9% of that of the daily mean temperature.

3.5 Comparison with other large city in Japan

As pointed out by previous research, good predictors or good models for heat stroke occurrence are defined by region (e.g. Bobb et al. 2011; Zhang et al. 2014). Therefore, the relative contribution of each meteorological element can be assumed by region. To investigate this, prediction model is developed in Osaka prefecture that contain large city in Japan, by the same method used for Tokyo metropolis.

The AIC results indicated that all four meteorological elements should be used as explanatory variables in Osaka. However, the contributions of the explanatory variables are different from

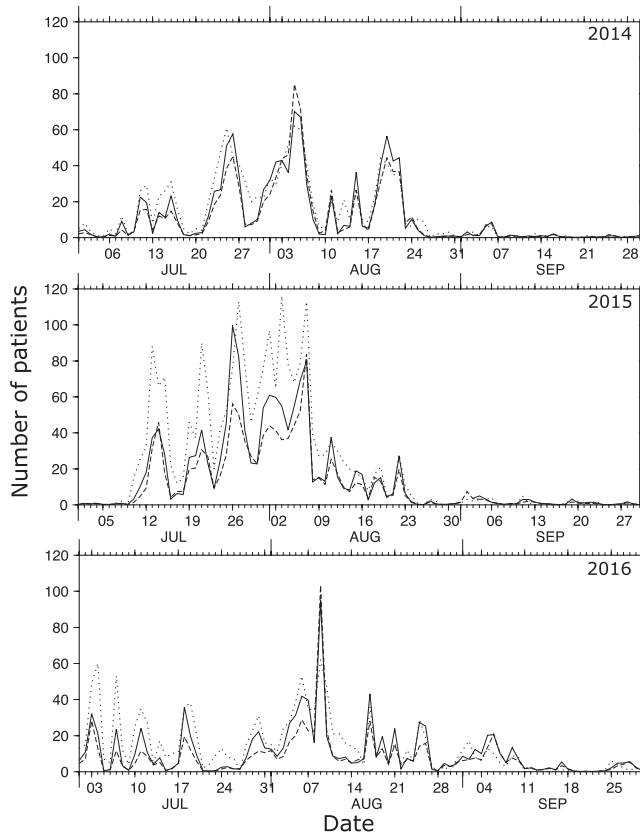


Fig. 4. Time series of heat stroke patients in the test period (2014–2016). Solid line: predicted number of patients using the “best model” (Eq. 3); Dashed line: predicted number of patients using daily mean temperature only; Dotted line: observed number of patients.

Tokyo. The standardized partial regression coefficients used in the best prediction models for the both prefectures are summarized in Table 1.

The result from Osaka prefecture shows different features compared with Tokyo metropolis, except that the temperature is still the largest contributor. The relative humidity is the second contributor, which is about 21% of the temperature. The third contributor is the total solar radiation, which is about 15% of the temperature. The last contributor is wind speed, with only 1% contribution. A possible reason of this difference is the larger fluctuation of relative humidity in Osaka. The standard deviation of temperature is almost same in both areas, but that of relative humidity is 1.7 times larger in Osaka compared with Tokyo.

The selection of maximum, mean, minimum values is hard to explain. Thus, further analysis is needed to reveal the reason of this heterogeneity.

Considering the relative contribution of four meteorological elements in Tokyo and Osaka, this study quantitatively supports the following two findings. The first is that the temperature contributes to the prediction of the number of heat stroke patients the most out of the four variables examined. The second is that the relative contribution of the meteorological elements to the model

differs. Especially, the contributions of relative humidity and wind speed are remarkably different by region. The relative contribution estimated in this study is particularly useful to determine such regional variations for better prediction. With regards to the second finding, which sees relative humidity varying greatly according to location, the study also suggests that this variable has a small contribution compared to the temperature in both regions. Generally, humidity is considered an important element to determine the heat stroke risk. However, in this research, it is shown that the relative humidity is not so determinant for prediction models.

4. Conclusions

In this study, we developed new multi-variable statistical models to predict the number of daily heat stroke patients in Tokyo metropolis. A unique aspect is that we used the AIC and not the correlation coefficient, to determine the effective meteorological elements for prediction. The best prediction model developed in this study includes daily mean temperature, daily maximum relative humidity, daily maximum wind speed, and total solar radiation as explanatory variables. This model can reduce the RMSE by 19% compared with the single-variable model.

Using the best prediction model, the contributions of each meteorological elements to the prediction were investigated quantitatively in Tokyo. The analysis confirmed that the temperature is a key element for predicting heat stroke occurrence. The total solar radiation contributes to the prediction by 20% of the contribution of temperature. On the other hand, the contribution of relative humidity and wind speed are relatively small compared with temperature (each of their contributions is about 10% and 9% to that of temperature). Especially, it should be noted that the relative humidity has a smaller contribution than the previous studies expected.

With the same method, the best prediction model for Osaka is also built to investigate the regional variability. A common feature in two prefectures is that the temperature has the largest contribution to the prediction. On the other hand, the relative contribution of the other variables in the other regions differs from that encountered in Tokyo. For instance, in Osaka, the relative humidity is the second contributor. However, even in Osaka, relative humidity contributes to the prediction by only 21% of temperature’s contribution. Thus, the contribution of the relative humidity is smaller than expected by previous studies.

5. Remarks

This research has some limitations. The effect of acclimatization is not included. This is one of the possible reasons leading to underestimation of the number of heat stroke patients in July and early August. Another limitation is that the performance of the prediction model compared to other thermal indices such as WBGT is not discussed. It may be beneficial to consider such thermal indices as possible explanatory variables to develop a more practical and applicable prediction model. This research reveals that relative humidity has smaller contribution than other meteorological elements, which is in contrast with the general perspective. Considering that, we should discuss new weight values of each meteorological variable to derive a more appropriate index to predict the heat stroke risk in the future study.

Table 1. The standardized partial regression coefficients of the thermal elements in each prefecture (“mean” daily mean, “max” daily maximum, “min” daily minimum).

Site	Element			Standardized partial regression coefficient			
	Temperature	RH	Wind speed	Temperature	RH	Wind speed	Solar radiation
Tokyo	mean	max	max	1.57	0.15	−0.14	0.31
Osaka	max	min	min	1.34	0.28	−0.02	0.20

Acknowledgements

This work was supported by the Social Implementation Program on Climate Change Adaptation Technology (SI-CAT) Grant Number JPMXD0715667165 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan and JSPS KAKENHI Grant number 17H01748. A part of figures are drawn by General Mapping Tools (GMT).

Edited by: T. Nakaegawa

Supplements

Supplement 1: The scatter diagram of explanatory variables.

References

- Akatsuka, S., T. Uno, and M. Horiuchi, 2016: The relationship between the heat disorder incidence rate and heat stress indices at Yamanashi prefecture in Japan. *Adv. Meteorol.*, **2016**, 1–11.
- Bobb, J. F., F. Dominici, and R. D. Peng, 2011: A bayesian model averaging approach for estimating the relative risk of mortality associated with heat waves in 105 U.S. cities. *Biometrics*, **67**, 1605–1616.
- Di Napoli, C., F. Pappenberger, and H. L. Cloke, 2018: Assessing heat-related health risk in Europe via the Universal Thermal Climate Index (UTCI). *Int. J. Biometeorol.*, **62**, 1155–1165.
- Dobson, A. J., and A. G. Barnett, 2008: *An Introduction to Generalized Linear Models. Third edition*, CRC press, pp. 307
- Fujibe, F., J. Matsumoto, and H. Suzuki, 2018a: Spatial and temporal features of heat stroke mortality in Japan and their relation to temperature variations, 1999–2014. *Geographical Review of Japan Series B*, **91**, 17–27.
- Fujibe, F., J. Matsumoto, and H. Suzuki, 2018b: Regional features of the relationship between daily heat-stroke mortality and temperature in different climate zones in Japan. *SOLA*, **14**, 144–147.
- Fuse, A., S. Saka, R. Fuse, T. Araki, S. Kin, M. Miyauchi, and H. Yokota, 2014: Weather data can predict the number of heat stroke patient. *Journal of Japanese Association for Acute Medicine*, **25**, 757–765 (in Japanese with English abstract).
- Honda, Y., M. Kabuto, M. Ono, and I. Uchiyama, 2007: Determination of optimum daily maximum temperature using climate data. *Environ. Health Prev. Med.*, **12**, 209–216.
- Ishigaki, H., T. Matsubara, S. Gonda, and T. Horikoshi, 2001: Experimental trial on the seasonal differences of the combined effect of air temperature and humidity on the human physiological and psychological responses. *J. Archit. Plann. Environ. Eng.*, **543**, 49–56 (in Japanese with English abstract).
- Lambert, D., 1992: Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1–14.
- Miyatake, N., N. Sakano, and S. Murakami, 2012: The relation between ambulance transports stratified by heat stroke and air temperature in all 47 prefectures of Japan in August, 2009: Ecological study. *Environ. Health Prev. Med.*, **17**, 77–80.
- Nakai, S., T. Itoh, and T. Morimoto, 1999: Deaths from heat-stroke in Japan: 1968–1994. *Int. J. Biometeorol.*, **43**, 124–127.
- Nairn, J. R., and R. J. B. Fawcett, 2015: The excess heat factor: A metric for heatwave intensity and its use in classifying heatwave severity. *Int. J. Environ. Res. Public Health*, **12**, 227–253.
- Ohashi, Y., Y. Kikegawa, T. Ihara, and N. Sugiyama, 2014: Numerical simulations of outdoor heat stress index and heat disorder risk in the 23 wards of Tokyo. *J. Appl. Meteor. Climatol.*, **53**, 583–597
- Ono, M., 2013: Heat stroke and the thermal environment. *Japan Medical Assoc. J.*, **56**, 199–205.
- Piver, W. T., M. Ando, F. Ye, and C. J. Portier, 1999: Temperature and air pollution as risk factors for heat stroke in Tokyo, July and August 1980–1995. *Environ. Health Perspect.*, **107**, 911–916.
- Semanza, J. C., C. H. Rubin, K. H. Falter, J. D. Selanikio, W. D. Flanders, H. L. Howe, and J. L. Wilhelm, 1996: Heat-related deaths during the July 1995 heat wave in Chicago. *The New England J. Medicine*, **335**, 84–90.
- Suzuki-Parker, A., and H. Kusaka, 2016: Future projections of labor hours based on WBGT for Tokyo and Osaka, Japan, using multi-period ensemble dynamical downscale simulations. *Int. J. Biometeorol.*, **60**, 307–310.
- Zhang, K., Y. Li, J. D. Schwartz, and M. S. O'Neill, 2014: What weather variables are important in predicting heat-related mortality? A new application of statistical learning methods. *Environ. Res.*, **132**, 350–359.

Manuscript received 9 March 2020, accepted 2 May 2020
SOLA: <https://www.jstage.jst.go.jp/browse/sola/>