

<https://doi.org/10.1038/s41746-025-02235-4>

Development and evaluation of a machine learning model predicting out-of-hospital cardiac arrest using environmental factors

Check for updates

Takahiro Nakashima^{1,2,3,10} ✉, Soshiro Ogata^{3,10}, Eri Kiyoshige³, Mohammad Z. Al-Hamdan^{4,5}, Yifan Wang⁶, Teruo Noguchi⁷, Theresa A. Shields¹, Rabab Al-Araji⁸, Bryan McNally⁹, Kunihiro Nishimura³ & Robert W. Neumar¹

Development of an accurate system to predict the daily incidence of out-of-hospital cardiac arrest (OHCA) might provide a significant public health benefit. We developed a machine learning model to predict daily OHCA incidence at the regional level using high-resolution meteorological data, chronological data, and sociodemographic data. Among OHCA of non-traumatic cause, 196,735 in the training dataset (2013-2017), 119,455 in the testing dataset (2018-2019) from internal areas, and 37,160 in the testing dataset from external areas were included in the analysis. Application of invariant causal prediction (ICP) successfully reduced the number of variables to 17 while maintaining high predictive performance for nationwide incidence per 100,000 per day in both the training and testing datasets, whether from internal or external regions, comparable to the non-ICP model using the initial 34 variables. Furthermore, the prediction model retained a satisfactory performance up to seven days in advance.

The estimated global incidence of out-of-hospital cardiac arrest (OHCA) treated by emergency medical service (EMS) is 62.0 per 100,000 person-years, with a specific incidence of 53.1 per 100,000 person-years in North America¹. Despite advances in pre- and post-resuscitation care, a recent systematic review showed that survival to hospital discharge after OHCA was extremely low, at 8.8%². Development of an accurate system to predict the daily incidence of OHCA might provide a more opportunity to prevent and react in the prehospital setting.

Several studies have shown associations between ambient temperature and cardiovascular events³⁻⁶ and between day of week or season and cardiovascular events⁷⁻¹². However, many of those studies used conventional linear regression, which may not be suitable for handling large amounts of high-resolution meteorological data and not be applied to real-world practice. Machine learning (ML) can use advanced analytics to integrate

multiple quantitative variables and identify associations not identified with conventional one-dimensional statistical approaches. Recently, our team developed a ML prediction model for daily OHCA incidence based on combined meteorological and chronological data with high accuracy for the Japanese population¹³. Moreover, given that several studies have demonstrated the association between social determinants and cardiovascular risk, we considered it valuable to incorporate sociodemographic variables into the prediction model^{14,15}.

In this study, we developed and evaluated the ML prediction model for robust estimation of daily OHCA incidence of cardiac origin for the U.S. population based on comprehensive meteorological data, chronological data, and sociodemographic data. In addition, applying the invariant causal prediction (ICP), we identified variables that consistently contributed to the predicted OHCA incidence across all sub-regions stratified by the

¹Department of Emergency Medicine and the Max Harry Weil Institute for Critical Care Research and Innovation, University of Michigan, Ann Arbor, MI, USA.

²Department of Cardiovascular Medicine, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan. ³Department of Preventive Medicine and Epidemiology, National Cerebral and Cardiovascular Centre, Suita, Japan. ⁴National Center for Computational Hydroscience and Engineering, School of Engineering, University of Mississippi, Oxford, MS, USA. ⁵Department of Civil Engineering, School of Engineering, University of Mississippi, Oxford, MS, USA.

⁶Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. ⁷Department of Cardiovascular Medicine, National Cerebral and Cardiovascular Centre, Suita, Japan. ⁸Emory University, Woodruff Health Sciences Center Atlanta, Atlanta, GA, USA. ⁹Emory University School of Medicine, Atlanta, GA, USA.

¹⁰These authors contributed equally: Takahiro Nakashima, Soshiro Ogata. ✉e-mail: takana@med.umich.edu

sociodemographic status of the population size, the proportion of people living in poverty, and the proportion of people who have achieved an education, and reconstructed the prediction model for greater generalizability.

Results

Characteristics of the training and testing datasets

From the CARES registry, 421,531 EMS-treated OHCA of non-traumatic cause between 2013 and 2019 were matched with meteorological data. Of these, 196,735 cases from 2013 to 2017 were included in the training dataset, and 156,615 cases from 2018 to 2019 were assigned to the testing dataset. Within the testing dataset, 119,455 cases were from internal areas (i.e., these areas were the same areas as the training data), and 37,160 cases were from external areas (i.e., these areas were not included in the training dataset). The characteristics of the datasets are summarized in Table 1. The median age of OHCA onset increased slightly from 64 (IQR, 52–76) to 65 years (52–76), and the proportion of males increased from 61 to 62%; however, these differences were modest. The proportion of individuals living below the poverty level was 11% of the population throughout the study period.

The median of the mean ambient temperature within a day decreased in the low, intermediate, or high-temperature regions. In the low-temperature region, differences between maximum and minimum ambient temperatures within a day (diurnal temperature range) increased from 9.4 °C (6.8–12.1) to 9.6 °C (7.0–12.3) in the internal area and 8.9 °C (6.2–11.6) to 9.6 °C (7.0–12.3) in the external area between 2013 and 2019. The diurnal temperature range was greater in higher temperature regions. Relative humidity also increased throughout the study period in all regions.

Between 2013–2017 and 2018–2019, the incidence of OHCA per 100,000 person-years increased from 63.7 to 76.7 in internal areas and from 65.3 to 73.7 in external areas. The incidence of OHCA by meteorological condition is shown in Supplementary Fig. 1.

Model diversity

To select the optimal analytical algorithms for model development, we developed ML prediction models including all 34 variables, using the generalized additive model (GAM) modeling time series, eXtreme Gradient Boosting (XGBoost), CatBoost, and random forest. Predicted and observed incidence of OHCA with a cardiac origin for each model are shown in Fig. 1. All ML prediction models effectively identified days with significant increases in OHCA incidence at the nationwide level, demonstrating strong concordance between predicted and observed values. In both the training dataset and the testing dataset from internal areas, predictive performance of all models was similar. In contrast, in the testing dataset from external areas, the XGBoost gradient boosting algorithm had the highest predictive performance at the nationwide level (root mean squared error (RMSE), 0.032, [95% confidence interval {CI}: 0.030–0.033]; mean absolute error (MAE), 0.025 [0.024–0.026]; mean absolute percentage error (MAPE), 13.17% [12.13–14.22]), the state level (RMSE, 0.212 [0.125–0.272]; MAE, 0.144 [0.101–0.187]), and the agency level (RMSE, 1.324 [0.967–1.603]; MAE, 0.345 [0.308–0.383]) among all models (Table 2).

Predictive performance of the model after ICP

Through the application of ICP to the XGBoost model, we identified 17 variables (**ICP model**) from the initial 34 (**non-ICP model**) that consistently contributed to the prediction for OHCA incidence across all sub-areas, as determined by deciles of agency population sizes, tertiles of proportion living below the poverty level, and tertiles of proportion with a high school diploma or higher (Supplementary Table 1). These 17 variables included mean ambient temperature, diurnal ambient temperature, mean wind speed, difference in wind speed, mean relative humidity, difference in relative humidity, mean precipitation, difference in precipitation, year, January, February, median age, proportion of men, proportion of Blacks, proportion of Asians, proportion with a high school diploma or higher, and proportion living below the poverty level.

Figure 2 and Table 3 shows that the ICP model maintained high predictive accuracy at the nationwide level in the training dataset (RMSE, 0.022 [0.021–0.023]; MAE, 0.018 [0.017–0.019]; and MAPE, 11.42% [10.76–12.09]), testing dataset with the internal area (RMSE, 0.021 [0.019–0.024]; MAE, 0.017 [0.015–0.018]; and MAPE, 7.80% [7.22–8.38]), and testing dataset with the external area (RMSE, 0.033 [0.030–0.035]; MAE, 0.026 [0.024–0.028]; and MAPE, 13.92% [12.70–15.14]), as did the non-ICP model with the initial 34 variables. At the state and agency level, both non-ICP and ICP models yielded similar predictive accuracy. To further clarify the predictive performance of this model at varying time intervals, we also evaluated 3-day-ahead and 7-day-ahead predictive performance. The model retained a satisfactory level of performance up to seven days in advance.

Contribution of each predictor to the predicted value of OHCA incidence

The predictive importance of variables in the ICP model is shown in Fig. 3. With regards to meteorological variables, mean ambient temperature within a day was the variable most strongly contributing to the predicted OHCA incidence, followed by mean relative humidity, diurnal temperature range, difference and mean in wind speed, and difference and mean precipitation. Among sociodemographic variables, proportion living below the poverty level was the variable most strongly contributing to the predicted OHCA incidence, followed by proportion with a high school diploma or higher, percentage of Black persons and Asian, and median age contributed to the predicted OHCA incidence.

Predictive performance based on annual average of daily mean ambient temperature

We compared the predictive performance of the XGBoost models stratified by annual average of ambient temperature (Supplementary Table 2) at aggregate level for all regions falling within that temperature category. Population was lower in the low-temperature region, followed by the intermediate-temperature region and the high-temperature region. The predictive accuracy of both non-ICP and ICP models was higher in the intermediate- and high-temperature regions than in the low-temperature region across both the training and testing datasets, consistent with the population. However, there was not much difference in the predictive accuracy between summer and winter in the regions stratified by average ambient temperature (Supplementary Table 3).

Discussion

In this study, using an ML prediction model developed with the combination of meteorological, chronological, and sociodemographic variables, we successfully predicted the daily incidence of non-traumatic OHCA in the United States with high precision at the nationwide, state, and agency level, respectively. In addition, ICP identified meteorological and socio-demographic variables as consistently important predictors of daily OHCA incidence, ensuring nationwide generalizability, regardless of population sizes of the EMS agencies.

An association between ambient temperature and incidence of cardiovascular events has been previously reported^{3–10}. However, since these studies focused on ambient temperature or season alone, diversity in comprehensive meteorological, chronological, and sociodemographic variables were not considered. Recently, our team reported that an ML prediction model for OHCA incidence in Japan based on a comprehensive meteorological dataset and chronological variables had high predictive accuracy¹³. In the present study, when sociodemographic data were added to the prediction model, we achieved a high predictive accuracy in the U.S. population. Model evaluation demonstrated a decline in predictive performance when applied across both external periods and areas, suggesting areas participating in the OHCA registry could receive higher merit for predictive accuracy when our ML models are implemented compared to areas not participating in the registry. Nevertheless, the model retained a satisfactory performance under these external validation conditions.

Table 1 | Characteristics of daily data in the training dataset (2013–2017) and testing dataset (2018–2019)

Year	Excluded dataset	Training dataset	Testing dataset		SMD
	2013–2017		2018–2019		
Area	External		Internal	External	
Demographic variables					
N (incident cases)	68,181	196,735	119,455	37,160	
Number of participating EMS agencies	183	728	728	183	
OHCA incidence, per 100,000 person-years	65.3	63.7	76.7	73.7	
Patient characteristics					
Age, years, median (IQR)	64 (52, 76)	64 (52, 76)	65 (52, 76)	64 (52, 76)	0.004
Man, n (%)	41,567 (61)	120,878 (61)	74,110 (62)	23,078 (62)	0.015
Race, n (%)					0.280
White	28,905 (42)	96,487 (49)	61,839 (52)	16,591 (45)	
Black	19,142 (28)	38,625 (20)	28,192 (23)	10,021 (27)	
Asian	1361 (2)	3514 (2)	2387 (2)	920 (3)	
Hispanic/Latino	7583 (11)	7333 (4)	5447 (5)	4177 (11)	
Other	11,190 (16)	50,776 (26)	21,590 (18)	5451 (15)	
First documented rhythm, n (%)					0.025
Shockable	13,232 (19)	38,800 (20)	22,384 (19)	6774 (18)	
Non-shockable	54,945 (81)	157,916 (80)	97,046 (81)	30,385 (82)	
Sociodemographic variables ^a					
Median age, years	41 (38, 45)	41 (37, 44)	41 (38, 44)	42 (38, 46)	0.181
Man, %	49 (48, 50)	49 (49, 50)	49 (49, 50)	49 (49, 50)	0.041
Black, n	1092 (267, 7554)	1647 (282, 10,329)	1185 (186, 7242)	912 (137, 5200)	0.091
Asian per, n	1032 (253, 3776)	1110 (227, 4469)	689 (134, 3372)	747 (132, 3424)	0.070
Individuals living below the poverty level	11 (6, 15)	11 (7, 15)	10 (7, 15)	11 (6, 15)	0.035
Individuals with a high school diploma or higher, n	33,100 (14,706, 67,723)	31,412 (12,987, 83,010)	24,191 (9354, 60,461)	23,924 (9376, 59,776)	0.100
Meteorological variables ^a					
Ambient temperature, °C					
Mean value within a day					0.439
Low-temperature area	9.0 (– 0.1, 17.8)	9.3 (– 0.1, 18.1)	7.9 (–7.8, 18.3)	7.9 (– 0.7, 18.5)	
Intermediate temperature area	11.9 (4.9, 19.2)	11.8 (4.3, 19.4)	11.4 (4.2, 20.1)	11.4 (4.6, 19.3)	
High-temperature area	17.7 (11.6, 23.9)	17.1 (10.4, 23.4)	16.9 (10.1, 24.1)	17.0 (10.3, 23.9)	
Differences between maximum and minimum values within a day					0.168
Low-temperature area	8.9 (6.2, 11.6)	9.4 (6.8, 12.1)	9.6 (7.0, 12.3)	9.6 (7.0, 12.3)	
Intermediate temperature area	9.0 (6.3, 11.7)	9.2 (6.5, 11.8)	9.9 (7.4, 12.7)	9.8 (7.2, 12.6)	
High-temperature area	9.8 (7.2, 12.2)	10.3 (7.8, 12.7)	10.3 (7.7, 13.0)	10.3 (7.7, 12.9)	
Precipitation during the previous hour, mm					
Low-temperature area	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.038
Intermediate temperature area	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	
High-temperature area	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	
Wind speed, m/s					
Low-temperature area	3.4 (2.3, 4.7)	3.4 (2.2, 4.7)	3.3 (2.2, 4.7)	3.4 (2.3, 4.7)	0.071
Intermediate temperature area	3.2 (2.1, 4.5)	3.2 (2.1, 4.5)	3.2 (2.2, 4.5)	3.2 (2.1, 4.4)	
High-temperature area	3.2 (2.2, 4.2)	3.1 (2.2, 4.2)	3.1 (2.2, 4.3)	3.2 (2.3, 4.3)	
Relative humidity, %					
Low-temperature area	75.2 (63.6, 83.5)	75.5 (64.5, 83.6)	78.8 (70.4, 85.7)	79.0 (70.9, 85.8)	0.218
Intermediate temperature area	75.6 (66.4, 83.7)	74.8 (65.2, 83.2)	77.8 (69.8, 84.6)	78.0 (70.2, 84.8)	
High-temperature area	71.3 (59.4, 80.4)	71.4 (60.3, 80.2)	75.0 (65.6, 82.7)	74.0 (63.0, 82.2)	

Continuous values are presented as medians (IQR).

^aValues are presented at the EMS agency level.

IQR interquartile range, OHCA out-of-hospital cardiac arrest.

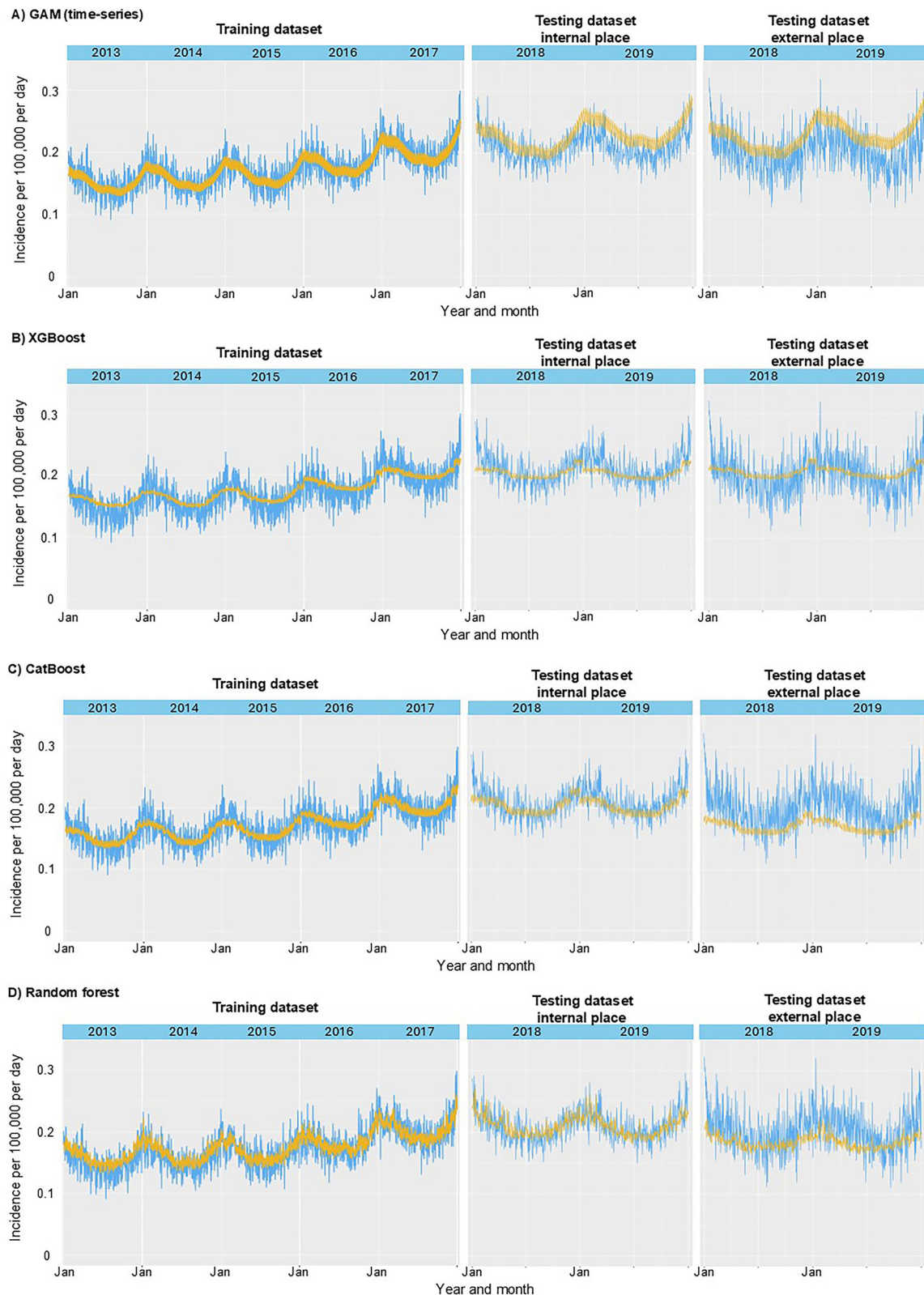


Fig. 1 | Observed versus predicted incidence of out-of-hospital cardiac arrest by various analysis methods. The results obtained using each method, including GAM (A), XGBoost (B), CatBoost (C), and Random Forest (D), are presented. The light blue lines indicate the observed daily incidence per 100,000 of out-of-hospital

cardiac arrests in the registry participating areas. The yellow lines indicate the predicted daily incidence per 100,000 based on combined meteorological, chronological, and sociodemographic variables. GAM generalized additive model, Jan means January, XGBoost eXtreme gradient boosting.

Table 2 | Performance of the prediction model by various analysis methods

		GAM (time series)	XGBoost	CatBoost	Random forest
Nationwide level					
Training dataset 2013–2017	RMSE	0.018 (0.017–0.018)	0.020 (0.019–0.020)	0.018 (0.017–0.019)	0.014 (0.014–0.015)
	MAE by day	0.014 (0.014–0.015)	0.016 (0.015–0.16)	0.014 (0.014–0.015)	0.011 (0.011–0.012)
	MAPE by day ^a	8.46 (8.12–8.81)	9.65 (9.17–10.14)	8.63 (8.26–9.00)	7.05 (6.66–7.44)
Testing dataset 2018–2019 Internal place	RMSE	0.026 (0.025–0.028)	0.021 (0.020–0.023)	0.020 (0.019–0.021)	0.020 (0.018–0.021)
	MAE by day	0.021 (0.020–0.023)	0.016 (0.015–0.017)	0.015 (0.014–0.016)	0.015 (0.014–0.016)
	MAPE by day ^a	10.65 (9.80–11.50)	7.47 (6.87–8.08)	7.09 (6.57–7.61)	7.31 (6.84–7.79)
Testing dataset 2018–2019 External place	RMSE	0.040 (0.038–0.043)	0.032 (0.030–0.033)	0.043 (0.040–0.047)	0.036 (0.033–0.038)
	MAE by day	0.032 (0.030–0.035)	0.025 (0.024–0.026)	0.035 (0.032–0.038)	0.028 (0.026–0.030)
	MAPE by day ^a	17.87 (16.33–19.42)	13.17 (12.13–14.22)	16.15 (15.27–17.04)	13.29 (12.56–14.02)
State level					
Training dataset 2013–2017	RMSE	0.146 (0.110–0.174)	0.145 (0.109–0.173)	0.141 (0.104–0.170)	0.104 (0.076–0.125)
	MAE by day	0.099 (0.080–0.119)	0.097 (0.078–0.116)	0.092 (0.073–0.112)	0.068 (0.054–0.083)
Testing dataset 2018–2019 Internal place	RMSE	0.140 (0.110–0.164)	0.136 (0.105–0.162)	0.133 (0.099–0.159)	0.144 (0.102–0.176)
	MAE by day	0.105 (0.086–0.124)	0.098 (0.078–0.117)	0.094 (0.092–0.095)	0.101 (0.077–0.125)
Testing dataset 2018–2019 External place	RMSE	0.213 (0.130–0.272)	0.212 (0.125–0.272)	0.211 (0.127–0.271)	0.230 (0.122–0.302)
	MAE by day	0.154 (0.107–0.201)	0.144 (0.101–0.187)	0.147 (0.103–0.192)	0.167 (0.104–0.229)
Agency level					
Training dataset 2013–2017	RMSE	1.004 (0.822–1.157)	1.003 (0.821–1.157)	1.002 (0.823–1.153)	0.767 (0.636–0.878)
	MAE by day	0.295 (0.285–0.305)	0.287 (0.277–0.297)	0.299 (0.285–0.313)	0.253 (0.237–0.269)
Testing dataset 2018–2019 Internal place	RMSE	1.279 (1.039–1.481)	1.278 (1.037–1.481)	1.279 (1.041–1.479)	1.311 (1.082–1.506)
	MAE by day	0.371 (0.354–0.388)	0.339 (0.322–0.355)	0.359 (0.339–0.378)	0.450 (0.421–0.480)
Testing dataset 2018–2019 External place	RMSE	1.324 (0.968–1.602)	1.324 (0.967–1.603)	1.320 (0.971–1.594)	1.377 (1.033–1.651)
	MAE by day	0.382 (0.345–0.420)	0.345 (0.308–0.383)	0.374 (0.325–0.422)	0.524 (0.444–0.605)

RMSEs and MAEs are shown as daily incidence per 100,000.

^aMAPEs are shown as %. In general, MAPE < 10% is considered highly accurate predicting; 10–20%, good predicting; 20–50%, reasonable predicting; and >50%, inaccurate predicting²². MAPE could not be calculated at the state or agency level because there were days with no cardiac arrests.

GAM generalized additive model, MAE mean absolute error, MAPE mean absolute percentage error, ML machine learning, RMSE root mean squared error, XGBoost eXtreme gradient boosting.

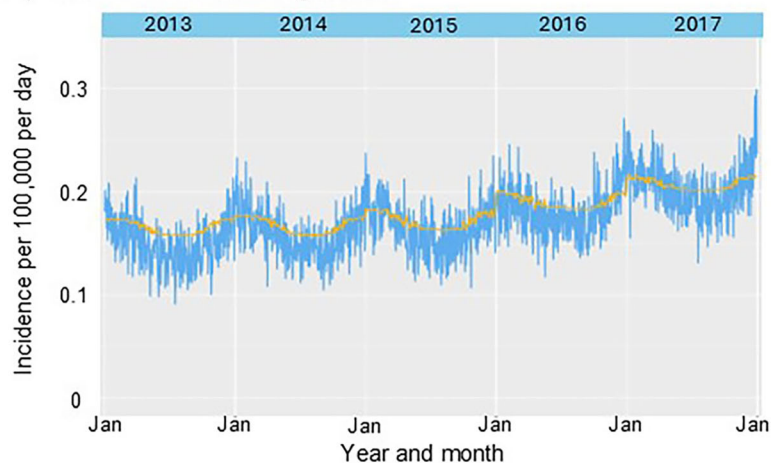
Conducting ICP, we identified 17 out of 34 variables that contributed to the predicted non-traumatic OHCA incidence with preserving a robust predictive accuracy regardless of population sizes of the EMS agencies, thereby enhancing the generalizability of the prediction model. As a result, all chronological variables, except for the year, January, and February, were excluded from the prediction model by ICP. SHAP analysis revealed that sociodemographic variables contributed more strongly to prediction for OHCA incidence than most meteorological factors. While the differences in sociodemographic characteristics between counties might shape vulnerability to weather conditions, they are not directly relevant to daily fluctuations in OHCA incidence because those variables do not change throughout the year. This finding is consistent with epidemiological evidence that socioeconomic status and racial disparities can be related to cardiovascular risks and might also amplify the impact of environmental exposures^{14,16–18}. In addition, meteorological predictors, except for mean ambient temperature and relative humidity, showed comparable contributions, suggesting that OHCA incidence is influenced by a complex interplay of weather factors, such as Heat Index and Wind Chill, rather than a single dominant variable. These results may emphasize the need for comprehensive inclusion of meteorological features in predictive models and also raise the hypothesis that social determinants of health might modulate the effects of climate on OHCA risk.

The predictive accuracy of the model was acceptable at the state level, albeit it exhibited diminished performance compared to the nationwide level. Our ML prediction model had variations in predictive accuracy across states. Analyses stratified by average ambient temperature in the training and testing datasets showed that predictive performance was the lowest in

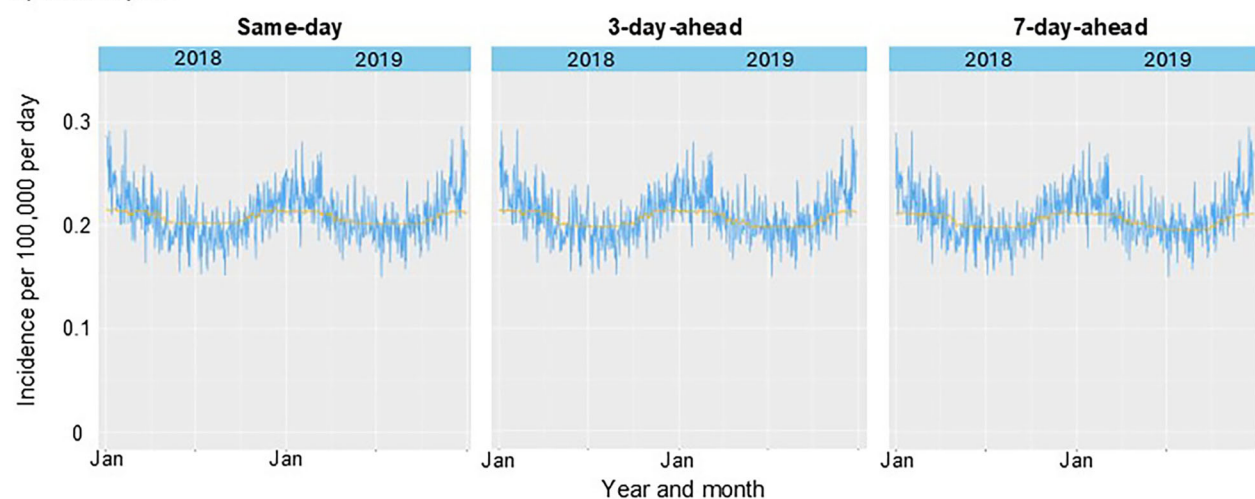
the low-temperature region, while the season variable (summer or winter) did not substantially change predictive accuracy. These results were partially explained by the population size in the participating area. Collecting more samples may improve our ML model. In addition, populations residing in the low-temperature area might be more habituated and better able to cope with climate change, such as through building insulation and lifestyle habits. Curriero et al. reported a latitude dependence of the temperature–mortality relationship in their analysis based on 11 eastern cities in the United States¹⁹. More effective adaption to colder temperature was observed in cities that are further north. In order to be more practical, it needs to be further improved to predict OHCA incidence within a medical catchment area.

Our prediction model retained a satisfactory level of performance up to seven days in advance, which holds potential value for proactive operational planning. Furthermore, the model’s predictive accuracy is inherently influenced by the quality of weather forecasts. Given that meteorological conditions can typically be predicted with reasonable accuracy up to seven days ahead, it is plausible that the model may sustain higher predictive performance over a similar temporal range. Beyond its predictive accuracy, our model has substantial potential for real-world application in EMS and hospital operations. In EMS, dynamic ambulance deployment informed by predictive analytics has been shown to reduce response times and improve patient outcomes^{20,21}. Integrating our weather-sensitive predictions into such EMS operational workflows could enable proactive ambulance positioning before anticipated high-risk periods, thereby facilitating more rapid transport and enabling advanced post-arrest care. In the hospital setting, a systematic review highlighted that various preparedness activities, such as resource reallocation, can improve hospitals’ surge capacity during demand

A) The ICP model in training dataset



B) Internal place



C) External place

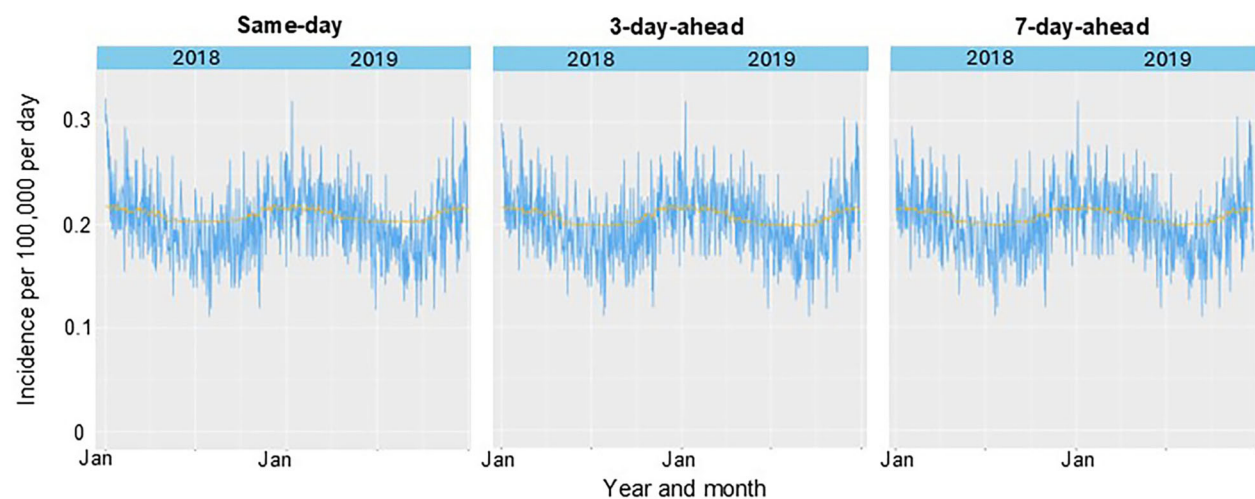


Fig. 2 | Observed versus predicted incidence of out-of-hospital cardiac arrest in the invariant causal prediction model at varying time intervals. A shows the results of the training dataset obtained using the ICP model. B and C present the testing results for same-day, 3-day-ahead, and 7-day-ahead predictions in internal and external settings, respectively. The light blue lines indicate the observed daily

incidence per 100,000 of out-of-hospital cardiac arrests in the registry-participating areas. The yellow lines indicate the predicted daily incidence per 100,000 by the XGBoost gradient boosting model using predictors selected by ICP. ICP denotes invariant causal prediction; Jan, January.

Table 3 | Performance of the invariant causal prediction model for out-of-hospital cardiac arrest at varying time intervals

Year	Training dataset		Testing dataset		
	2013–2017		2018–2019		
	Same-day		7-day-ahead		
Time intervals	Internal	External	Internal	External	
Primary outcomes					
Nationwide level					
RMSE	0.022 (0.021–0.023)	0.021 (0.019–0.024)	0.033 (0.030–0.035)	0.021 (0.020–0.022)	0.031 (0.030–0.033)
MAE by day	0.018 (0.017–0.019)	0.017 (0.015–0.018)	0.026 (0.024–0.028)	0.017 (0.016–0.018)	0.025 (0.023–0.027)
MAPE by day, % ^a	11.42 (10.76–12.09)	7.80 (7.22–8.38)	13.92 (12.70–15.14)	10.70 (10.13–11.27)	13.32 (12.22–14.43)
Secondary outcomes					
State level					
RMSE	0.145 (0.110–0.173)	0.136 (0.105–0.162)	0.211 (0.125–0.271)	0.137 (0.105–0.162)	0.211 (0.125–0.271)
MAE by day	0.098 (0.079–0.188)	0.098 (0.079–0.117)	0.145 (0.102–0.189)	0.098 (0.079–0.117)	0.144 (0.101–0.187)
Agency level					
RMSE	1.003 (0.821–1.157)	1.278 (1.037–1.481)	1.324 (0.967–1.603)	1.276 (1.036–1.478)	1.320 (0.964–1.599)
MAE by day	0.291 (0.281–0.302)	0.342 (0.326–0.359)	0.350 (0.312–0.387)	0.341 (0.324–0.358)	0.348 (0.310–0.385)

All models were developed by the XGBoost gradient boosting algorithm with predictors selected by ICP.

RMSEs and MAEs are shown as incidence per 100,000 per day.

^aIn general, MAPE < 10% is considered highly accurate predicting; 10–20%, good predicting; 20–50%, reasonable predicting; and >50%, inaccurate predicting⁴². MAPE could not be calculated at the state or agency level because there were days with no cardiac arrests. ICP invariant causal prediction, MAE mean absolute error, MAPE mean absolute percentage error, ML machine learning, RMSE root mean squared error, XGBoost eXtreme gradient boosting.

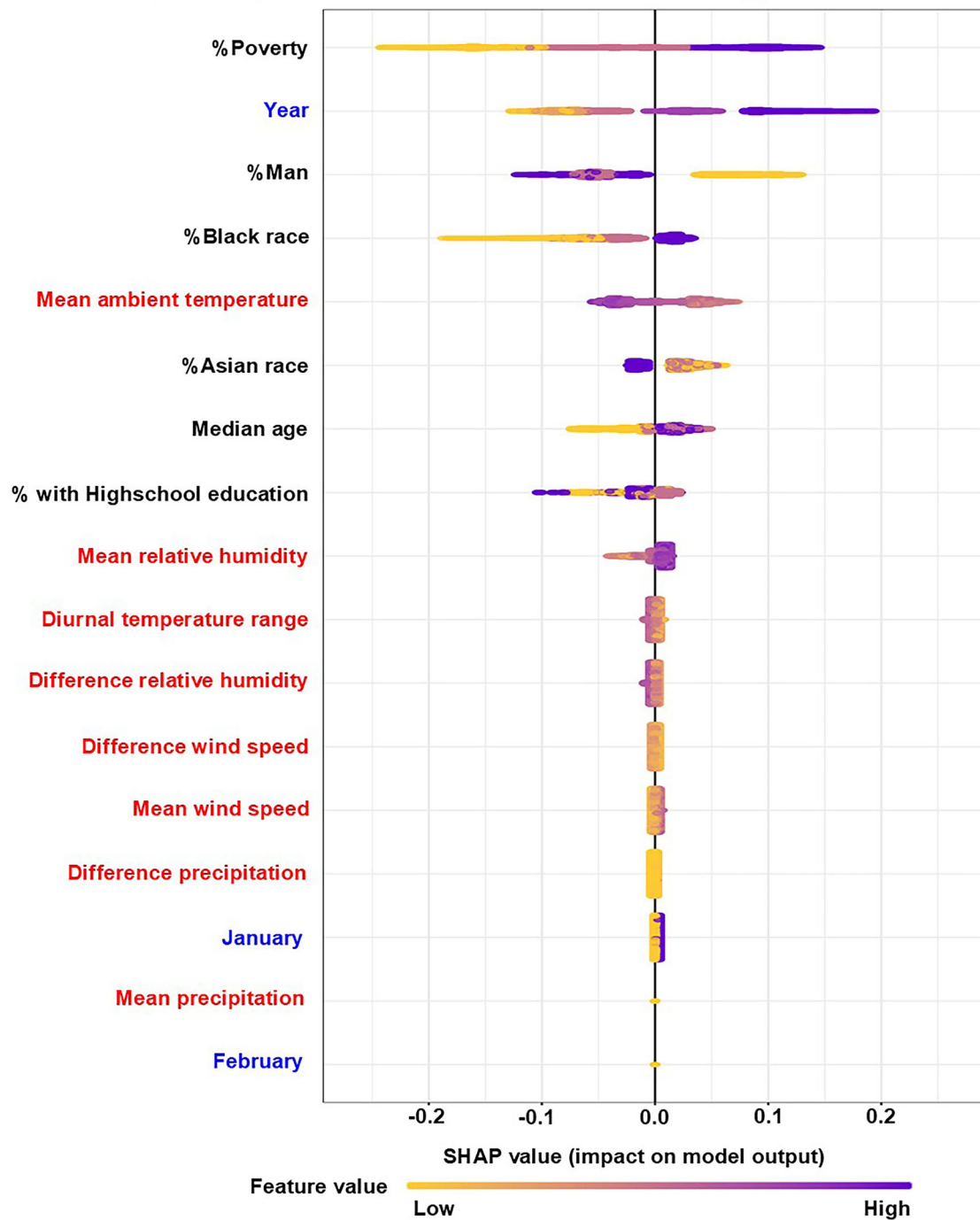


Fig. 3 | Importance of variables in a machine learning prediction model. This figure shows a variable importance plot for meteorological variables (red), chronological variables (blue), and sociodemographic variables (black) in a machine learning prediction model using XGBoost. The yellow to purple dots in each row represent low to high values for each predictor normally scaled. The x-axis shows the

Shapley value, indicating the variable’s impact on the model. Positive SHAP values tend to drive predictions toward more cases of OHCA and negative SHAP values tend to drive the prediction toward fewer cases of OHCA. * In the model, 2013 was considered year 0. OHCA denotes out-of-hospital cardiac arrest; SHAP, Shapley Additive Explanations; XGBoost, eXtreme Gradient Boosting.

spikes²². Furthermore, a UK teaching hospital reported that its ML-based prediction pipeline for patient admissions substantially outperformed the conventional six-week rolling average benchmark in predictive accuracy, even after the COVID-19 outbreak²³. Moreover, public health agencies could also leverage forecast outputs for targeted messaging campaigns, such as issuing warnings on predicting high-risk days, directed toward vulnerable populations. Therefore, these applications indicate that our model could be feasibly integrated into operational decision-making pipelines, strengthening both public health strategies and clinical preparedness in responses to

weather-related health risks. Such system-level improvements have the potential not only to prevent cardiac arrest events but also to increase rates of prehospital return of spontaneous circulation and overall survival. A future prospective study to evaluate the effectiveness of this approach is needed.

This study has several inherent limitations. First, although the CARES registry is the largest database of OHCA in the United States, it covered approximately 53% of catchment areas as of 2022. The CARES registry includes only EMS-treated OHCA cases, which may introduce selection bias by excluding untreated cases. Thus, sociodemographic variables such as race

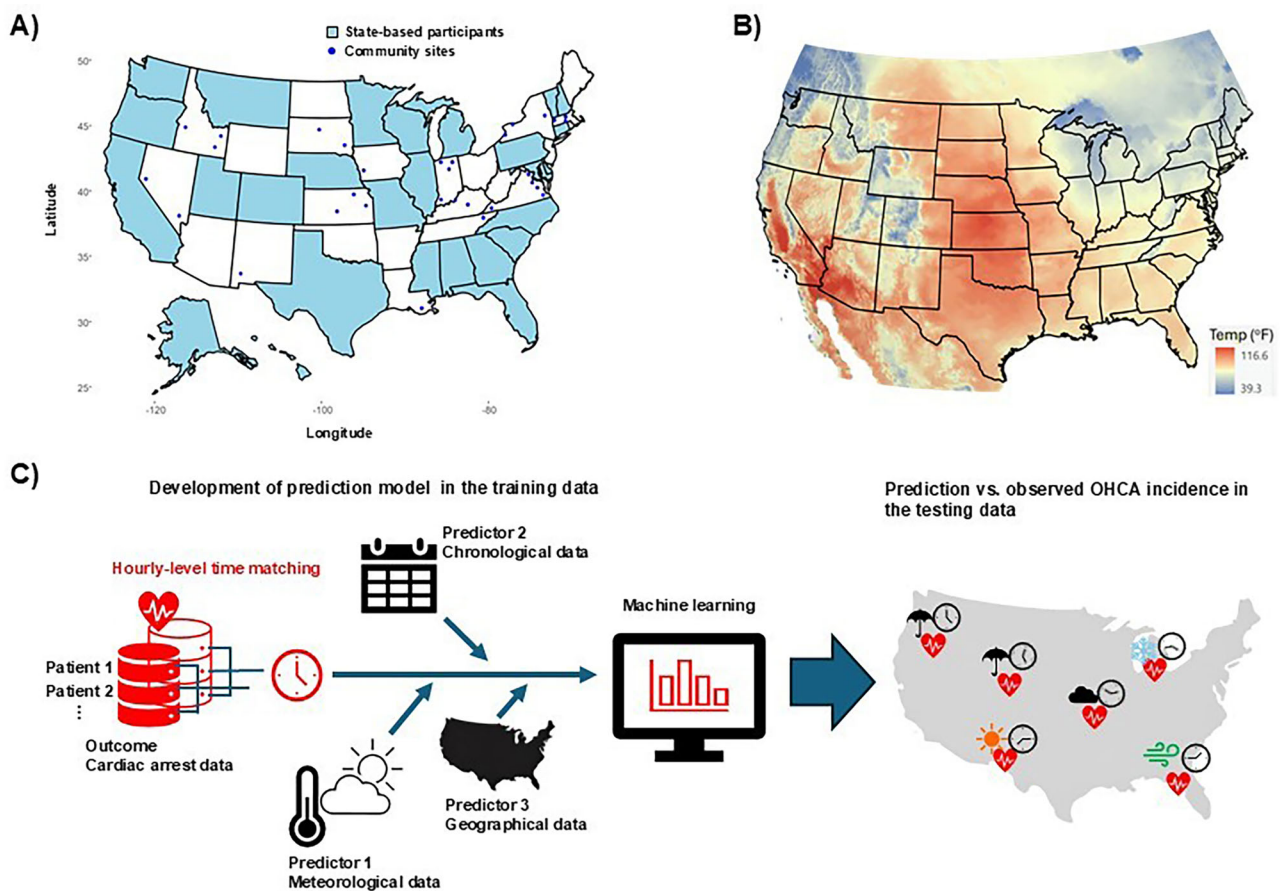


Fig. 4 | Data sources and overview of the machine learning model for predicting out-of-hospital cardiac arrest incidence. A illustrates the states and communities participating in the Cardiac Arrest Registry to Enhance Survival (CARES). B displays an example of daily maximum ambient temperature data on July 15, 2022,

obtained from the North American Land Data Assimilation System (NLDAS). C summarizes the development process of the machine learning model used to predict out-of-hospital cardiac arrest incidence.

and poverty rate may reflect biases in EMS activation or record keeping driven by socioeconomic factors; however, it represents the population most relevant for targeting public health interventions. Second, our data did not address the potential variability in patients' preexisting medical conditions. Third, the 12 km resolution has limitations, particularly in areas with complex geography. In mountainous regions, for example, it tends to smooth elevation differences, potentially misrepresenting temperature extremes in valleys or at high altitudes^{24–27}. Likewise, weather variability near coastlines or large water bodies, where conditions can change rapidly, is often averaged out, masking important local phenomena such as lake-effect snow or coastal heat stress. Nevertheless, a key advantage of 12 km data is its broad availability and consistency for public health applications. Fourth, the predictability of future OHCA events will depend on the accuracy of meteorological data. Finally, external testing in other developed countries was not performed.

In conclusion, an ML prediction model incorporating multiple meteorological, chronological, and sociodemographic variables could predict the incidence of non-traumatic OHCA with high precision in the U.S. population. Through ICP, the variables were refined to focus on meteorological and sociodemographic factors, rather than chronological ones, while maintaining high predictive accuracy. This prediction model might be useful for public health prevention strategies in temperate regions.

Methods

The study was approved by the University of Michigan Hospital's institutional review board (HUM00189913). The requirement for written

informed consent was waived because the researchers only analyzed de-identified (anonymized) data.

Data source for OHCA data

OHCA data was provided by the Cardiac Arrest Registry to Enhance Survival (CARES)²⁸, which is a prospective multicenter registry of patients with OHCA from 30 state-based registries, the District of Columbia, and more than 45 community sites in 16 additional U.S. states. It has a catchment area of approximately 175 million residents in 2021 (Fig. 4A). The design of the registry, which was established by the U.S. Centers for Disease Control and Prevention and Emory University, has been previously described²⁸. Patient-level data were collected by EMS agencies using standardized international Utstein definitions for clinical variables and outcomes to ensure uniformity.

CARES includes non-traumatic OHCA where resuscitative efforts were initiated by a 911 responder. The following patient information was collected and analyzed in this study: age, sex, etiology of arrest (i.e., presumed cardiac etiology, respiratory/asphyxia, drowning/submersion, electrocution, exsanguination/hemorrhage, drug overdose, or others), and location of cardiac arrest. Each patient in the CARES registry was geocoded to a U.S. county based on the ZIP code for the location of the OHCA through a crosswalk file from the U.S. Department of Housing and Urban Development. Data were submitted in two ways: via a data entry form on the CARES website (<https://mycares.net/>) or daily uploads from an EMS agency's electronic patient care record system. The CARES analyst (R.A.-A.) reviewed records for completeness and accuracy. Due to the nature of the

data resources, there were no missing values for developing and evaluating our prediction models.

Meteorological data

We analyzed meteorological data from National Aeronautics and Space Administration (NASA)'s North American Land Data Assimilation System (NLDAS), which provides hourly gridded data with 12-km spatial resolution (Fig. 4B)^{29,30}. Agencies such as NASA and NOAA maintain long-term archives at this resolution, allowing researchers to examine climate–health trends across time and large geographic areas^{24–27}. For public health applications, the 12 km grid aligns reasonably well with many health datasets, which are typically reported at the county, ZIP code, or hospital level. This compatibility supports regional exposure assessment and facilitates studies linking environmental exposures, such as ambient temperature or humidity, to health outcomes including cardiovascular and respiratory conditions.

Meteorological variables included eight factors: mean daily values and differences between maximum and minimum values within a day of ambient temperature (°C), precipitation (mm), relative humidity (%), and wind speed (m/s) during the study period. Those values were averaged by EMS agency areas. Due to the nature of the data resources, there were no missing values of those meteorological predictors.

Chronological and sociodemographic data

Chronological variables included 20 factors: year (2013 was considered year 0), month (January to December as categorical variables), and day of the week. Sociodemographic data were collected for six factors at the census tract level, including median age (three categories: < 38.8, 38.8–42.9, and ≥42.9 year, near its tertiles), proportion of men (three categories: <48.8%, 48.8–50.1, and ≥50.1%, near its tertiles), race (proportion of Blacks [three categories: <1%, 1–5%, and ≥5%, near its tertiles], proportion of Asians [three categories: <0.9%, 0.9–3%, and ≥3%, near its tertiles]), proportion of individuals with a high school diploma or higher (three categories: <60.5%, 60.5–65.3%, and ≥65.3%, near its tertiles), and proportion of individuals living below the poverty level (three categories: <7.9%, 7.9–13.1%, and ≥13.1%, near its tertiles). The sociodemographic data at the census tract level were merged into EMS agency areas. In cases where an EMS agency area was covered by multiple census tract areas, we used the sociodemographic from the tract with the highest number of OHCA cases among the several tract areas. Due to the nature of the data resources, there were no missing values of those sociodemographic predictors.

Data management and development of prediction models

In this study, we focused on EMS agencies that had data available in both periods: (A) at least part of January 1, 2013 to December 31, 2017, and (B) at least part of January 1, 2018 to December 31, 2019. By restricting the analysis to EMS agencies with data spanning both periods, we ensured comparability over time and reduced potential biases from agencies that appeared in only one timeframe, thereby strengthening the robustness of the evaluation. We then matched the CARES data, meteorological, chronological, and sociodemographic data between January 1, 2013, and December 31, 2019. From this merged dataset, 80% of the data from January 1, 2013 to December 31, 2017 were used to construct the training dataset for developing the prediction model, while data from January 1, 2018 to December 31, 2019 were reserved as the testing dataset to assess the temporal generalizability of the model (Fig. 4C). In addition, the testing dataset was stratified into internal (i.e., these areas were the same areas as the training data) and external areas (i.e., these areas were not included in the training dataset).

To develop prediction model for the daily incidence of OHCA, we used GAM modeling time-series effect, eXtreme Gradient Boosting (XGBoost) gradient boosting algorithm, CatBoost, and random forest^{31,32}. Note that, as a reference model, we employed the GAM modeling time-series effect with a negative binomial distribution to analyze the temporal dynamics of event counts. The reference model incorporated long-term temporal trends, seasonal variation with a yearly cycle, and day-of-week effects, with

population size included as an offset to model incidence rates. Those algorithms can model non-linear associations between the predictors and the outcome variable. GAM can model non-linear associations of the predictors with the outcome by using spline functions. Tree-based models, XGBoost, CatBoost, and random forest, can also model non-linear associations between predictors and the outcome, and are inherently more robust to multicollinearity than GAM. We selected each model's hyperparameters with minimum RMSE while developing models in the training dataset between 2013 and 2017 with the 5-fold cross validation of EMS agencies. Using the hyperparameters with minimum RMSE at the EMS agency level, we developed prediction models for the OHCA incidence per day per EMS agency. The candidate hyperparameters as well as the best-performing hyperparameters are provided in Supplementary Table 4.

All of the aforementioned meteorological, chronological, and socio-demographic variables were used as initial predictors. The predictor variables were standardized based on the mean and standard deviation values of each predictor in the training dataset. Log-transformed population size for each agency area was included as predictors for CatBoost and random forest. This was included as an offset term for the GAM modeling time-series effect and the XGBoost algorithm. We assessed predictive performance of the model using the testing dataset (i.e., data between 2018 and 2019 from all participating agency areas) stratified into internal and external areas. We used mgcv package for R, version 1.8-42 (<https://cran.r-project.org/web/packages/mgcv/index.html>) for the GAM modeling time-series effect, ranger package for R, version 0.15.1 (<https://cran.r-project.org/web/packages/ranger/index.html>) for the random forest model, bonsai package for R, version 0.4.0 (<https://cran.r-project.org/web/packages/bonsai/index.html>) of the CatBoost model, and xgboost package for R, version 1.7.7.1 (<https://cran.r-project.org/web/packages/xgboost/index.html>) for the XGBoost models.

The present study used the ICP to identify key predictors of the number of cardiac arrest cases under varying environmental conditions as a combination of deciles of agency population sizes, tertiles of proportion living below the poverty level, and tertiles of proportion with a high school diploma or higher. The ICP identifies variables that consistently predict outcomes across different conditions by assuming that the underlying causal structure remains constant, even as environmental factors may vary. As the environmental factors, we used a combination of deciles of population sizes, tertiles of proportion living below the poverty level, and tertiles of proportion with a high school diploma or higher at EMS agency area level based on the previous evidence^{33,34}. Among socioeconomic determinants associated with cardiovascular events, the evidence regarding these two factors has been inconsistent¹⁴. Including too many variables as environmental factors would create an excessive number of strata, resulting in small sample sizes within each stratum and potentially overlooking useful predictors. Therefore, we limited the environmental factors to key sociodemographic variables only. The predictor selection by the ICP analysis is achieved by statistical tests to evaluate candidate models. We applied the ICP approach three times: 1) to weather variables, 2) to sociodemographic variables, and 3) to calendar variables, under varying the combination categories of the environmental factors, to identify important predictors for each type of variable. We used the InvariantCausalPrediction package (version: 0.8) for R (<https://CRAN.R-project.org/package=InvariantCausalPrediction>). Then, to clarify the robustness of this prediction model at varying time intervals, we also tested 3-day-ahead and 7-day-ahead predictive performance.

Primary and secondary outcomes

The primary outcome was predictive accuracy of OHCA incidence per 100,000 per day at nationwide level of the prediction model based on RMSE, MAE, and MAPE, which are generally used as measures of predictive accuracy for a forecasting method. The secondary outcome was predictive accuracy of OHCA incidence per 100,000 per day at the state level and EMS agency level, which was limited to 24 state-based registries, and at the agency level.

The prediction model developed in this study is based on meteorological factors, which influence entire regions, and is ultimately intended for implementation at a medical catchment area. Therefore, we evaluated model performance using continuous-outcome metrics that are appropriate for aggregated data, along with correlation-based statistics to assess concordance between observed and predicted regional incidence.

Sample size calculation

We evaluated the sample size required for the testing dataset based on the precision of the RMSE, targeting a 95% CI relative half-width of ±10%. Because the data have a region × day structure, we accounted for within-region and within-day correlations using a design effect (DEFF) approximation^{35,36}. This two-way DEFF approach aligns with the multi-way clustering framework but omits cross-cluster terms, providing a conservative estimate³⁷. Using the intraclass correlation coefficient (ICC) and DEFF estimated from the training data, we determined that the required sample size for the unobserved testing data is approximately 91,120 observations. We estimated ICCs from the training dataset (842,534 observations, 728 regions, 1826 days), yielding a substantial intra-region ICC (0.34) and negligible intra-day ICC (0.0002). The resulting DEFF was approximately 399. To achieve the target precision (±10%), about 196 effective degrees of freedom are required; adding 33 model parameters yields an effective sample size of ~229. Accounting for the DEFF, this corresponds to approximately 91,120 observations needed for the testing dataset. Our testing datasets had 511,000 observations for the internal areas and 129,940 observations for the external areas. Thus, we confirmed that our testing datasets had enough sample size to evaluate the developed prediction models. The details are shown in **Supplementary Methods**.

Sample Size Calculation

The prediction evaluation metric was the root mean squared error (RMSE). We defined the sample size required for the testing dataset as the number of observations necessary for the 95% confidence interval (CI) of RMSE to have a relative half-width (half-width / true value) within ±10%. Since the RMSE estimator is based on the error variance, its distribution follows a chi-squared law.

Because our dataset had a day × region panel structure, correlations are expected within regions and within days. To avoid overestimating precision, we applied a design effect (DEFF) correction^{35,36}. With two clustering dimensions, DEFF can be approximated as:

$$DEFF \approx 1 + (mr - 1)\rho_r + (md - 1)\rho_d, \tag{1}$$

where *mr* is the average number of days per region, *md* the average number of regions per day, and ρ_r, ρ_d the intraclass correlation coefficients (ICCs). ICCs were estimated by fitting a linear mixed-effects model with region and day as random effects. The DEFF was then computed, and the effective sample size was defined as:

$$neff = N/DEFF \tag{2}$$

with effective degrees of freedom $veff = neff - p$ (*p*: number of fixed-effect parameters).

In this study, to account for clustering in two dimensions (region × day), we adopted a sum-approximation DEFF^{35,36}. This approach is consistent with the framework of multi-way clustering inference proposed by Cameron et al., but it ignores cross-cluster interaction terms and thus represents a conservative approximation³⁷.

First, we evaluated the correlation structure of the training data (842,534 observations, 728 regions, 1826 days). On average, each region contained 1157 days, and each day included 461 regions. Variance component estimates from a linear mixed model indicated that the intra-region ICC was relatively high (0.34), whereas the intra-day ICC was negligible (0.0002). Based on these results, the design effect (DEFF) was calculated to be approximately 399. To ensure that the relative half-width of the 95%

confidence interval (CI) for the RMSE remains within ±10%, a theoretical effective degrees of freedom of $v_{target} \approx 196$ is required. Given the number of estimated model parameters ($p = 33$), the required effective sample size was $v_{target} + p \approx 229$. Taking the DEFF into account, the corresponding number of observations is

$$N_{required} \approx (v_{target} + p) \times DEFF \approx 229 \times 398.6 \approx 91,120. \tag{3}$$

Therefore, by using the ICC and DEFF estimated from the training data, we determined that the required sample size for the unobserved testing data is approximately 91,120 observations. Our testing datasets had 511,000 observations for the internal areas and 129,940 observations for the external areas. Thus, we confirmed that our testing datasets had enough sample size to evaluate the developed prediction models.

Statistical analysis

The characteristics of the present dataset were summarized with medians and interquartile ranges (IQRs) for continuous variables, and numbers and percentages for categorical variables by area and day in the training and testing datasets.

We examined the concordance between the predicted incidence of OHCA based on the ML model and the observed incidence of OHCA in the training and testing dataset. The predictive accuracy of the prediction models was evaluated based on RMSE, MAE, and MAPE between predicted values calculated with the prediction models and observed daily OHCA incidence at levels of the EMS agencies, the states, and the nation. RMSE and MAE reflect the average magnitude of differences between predicted values and observed values. RMSE and MAE can range from zero to infinity. Lower RMSE and MAE values indicate higher predictive performance. MAPE is an average of the absolute values of errors divided by observed values. MAPE ranges from zero to infinity. Lower MAPE values indicate higher model predictive performance. In general, MAPE less than 10% is considered highly accurate predicting³⁸. Formulas are as follows;

$$RMSE = \sqrt{\sum_{i=1}^n (observed\ value - predicted\ value)^2 / N} \tag{4}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |observed\ value - predicted\ value| \tag{5}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|observed\ value - predicted\ value|}{observed\ value} * 100 \tag{6}$$

Additionally, we estimated point estimates and 95% CIs for prediction accuracy metrics (RMSE, MAE, MAPE). For each metric, we constructed a constant-only regression model using the corresponding error series as the dependent variable and obtained robust standard errors with two-way clustering (by region and time). Estimation was conducted in R using the *fixest* package, version 0.11.1 (<https://cran.r-project.org/web/packages/fixest/index.html>). This approach accounts for spatiotemporal dependence in the data when constructing CIs.

To show important predictors of the OHCA incidence in the developed model with ICP, we used the Shapley Additive Explanations (SHAP) values summarizing contribution of each predictor to the predicted value of an instance^{39,40}. For a given set of feature values, a SHAP value reflects how much a single variable, in the context of its interaction with other variables, contributes to the difference between the actual prediction and the mean prediction. As an additional study, we assessed the predictive accuracy of the ML model stratified into low-, intermediate-, and high-temperature areas, further divided into summer (June–August) and winter (December–February). Low-, intermediate-, and high-temperature regions were defined as regions with mean ambient temperature in the 25th percentile or lower, in the 25–75th percentiles, and 75th percentile or higher, respectively.

All statistical analyses were performed with R statistical software, version 4.2.3 (<https://www.R-project.org/>). Missing values for continuous and categorical variables were, respectively, imputed by a median value for each continuous variable and treated as a missing category. These missing procedures work well when using XGBoost due to the nature of decision tree algorithms.

Data availability

The datasets analyzed during the current study include third-party data obtained from the CARES and NASA's NLDAS. These data are not publicly available due to data use agreements, but may be made available from the corresponding author upon reasonable request, subject to approval from the data providers and in accordance with ethical and legal requirements.

Code availability

The code supporting this prediction model is publicly available through the GitHub repository (<https://github.com/ssrogata/Prediction-Models-for-Cardiac-Arrest-in-the-United-States-of-America>). A detailed description of our data preprocessing parameter settings and equations is shown in the Supplementary Table 4.

Received: 23 July 2025; Accepted: 30 November 2025;

Published online: 22 December 2025

References

- Berdowski, J., Berg, R. A., Tijssen, J. G. & Koster, R. W. Global incidences of out-of-hospital cardiac arrest and survival rates: systematic review of 67 prospective studies. *Resuscitation* **81**, 1479–1487 (2010).
- Yan, S. et al. The global survival rate among adult out-of-hospital cardiac arrest patients who received cardiopulmonary resuscitation: a systematic review and meta-analysis. *Crit. Care* **24**, 61 (2020).
- Yamaji, K. et al. Relation of ST-segment elevation myocardial infarction to daily ambient temperature and air pollutant levels in a Japanese nationwide percutaneous coronary intervention registry. *Am. J. Cardiol.* **119**, 872–880 (2017).
- Cold exposure and winter mortality from ischaemic heart disease, cerebrovascular disease, respiratory disease, and all causes in warm and cold regions of Europe. The Eurowinter Group. *Lancet* **349**, 1341–1346 (1997).
- Wolf, K. et al. Air temperature and the occurrence of myocardial infarction in Augsburg, Germany. *Circulation* **120**, 735–742 (2009).
- Danet, S. et al. Unhealthy effects of atmospheric temperature and pressure on the occurrence of myocardial infarction and coronary deaths. A 10-year survey: the Lille-World Health Organization MONICA project (Monitoring trends and determinants in cardiovascular disease). *Circulation* **100**, E1–E7 (1999).
- Herlitz, J., Eek, M., Holmberg, M. & Holmberg, S. Diurnal, weekly and seasonal rhythm of out of hospital cardiac arrest in Sweden. *Resuscitation* **54**, 133–138 (2002).
- Arntz, H. R. et al. Diurnal, weekly and seasonal variation of sudden death. Population-based analysis of 24,061 consecutive cases. *Eur. Heart J.* **21**, 315–320 (2000).
- Kloner, R. A., Poole, W. K. & Perritt, R. L. When throughout the year is coronary death most likely to occur? A 12-year population-based analysis of more than 220000 cases. *Circulation* **100**, 1630–1634 (1999).
- Marchant, B., Ranjadayan, K., Stevenson, R., Wilkinson, P. & Timmis, A. D. Circadian and seasonal factors in the pathogenesis of acute myocardial infarction: the influence of environmental temperature. *Br. Heart J.* **69**, 385–387 (1993).
- Sandhu, A., Seth, M. & Gurm, H. S. Daylight savings time and myocardial infarction. *Open Heart* **1**, e000019 (2014).
- Manfredini, R. et al. Daylight saving time and myocardial infarction: should we be worried? A review of the evidence. *Eur. Rev. Med. Pharmacol. Sci.* **22**, 750–755 (2018).
- Nakashima, T. et al. Machine learning model for predicting out-of-hospital cardiac arrests using meteorological and chronological data. *Heart* **107**, 1084–1091 (2021).
- Schultz, W. M. et al. Socioeconomic status and cardiovascular outcomes: challenges and interventions. *Circulation* **137**, 2166–2178 (2018).
- Morris, A. A. et al. 2024 ACC/AHA key data elements and definitions for social determinants of health in cardiology: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical data standards. *Circ. Cardiovasc. Qual. Outcomes* **17**, e000133 (2024).
- Carnethon, M. R. et al. Cardiovascular health in African Americans: a scientific statement from the American Heart Association. *Circulation* **136**, e393–e423 (2017).
- Huang, Z., Chan, E. Y. Y., Wong, C. S. & Zee, B. C. Y. Spatiotemporal relationship between temperature and non-accidental mortality: Assessing effect modification by socioeconomic status. *Sci. Total Environ.* **836**, 155497 (2022).
- Chan, E. Y. Y., Goggins, W. B., Kim, J. J. & Griffiths, S. M. A study of intracity variation of temperature-related mortality and socioeconomic status among the Chinese population in Hong Kong. *J. Epidemiol. Community Health* **66**, 322–327 (2012).
- Curriero, F. C. et al. Temperature and mortality in 11 cities of the eastern United States. *Am. J. Epidemiol.* **155**, 80–87 (2002).
- Vecina, M. Á, Villa, F., Vallada, E. & Karpova, Y. Enhancing management efficiency in emergency vehicles logistic using realistic data from Geographic Information System. *Oper. Res. Data Anal. Logist.* **45**, 200478 (2025).
- Selvan, C., Anwar, B. H., Naveen, S. & Bhanu, S. T. Ambulance route optimization in a mobile ambulance dispatch system using deep neural network (DNN). *Sci. Rep.* **15**, 14232 (2025).
- Hasan, M. K., Nasrullah, S. M., Quattrocchi, A., Arcos González, P. & Castro-Delgado, R. Hospital surge capacity preparedness in disasters and emergencies: a systematic review. *Public Health* **225**, 12–21 (2023).
- King, Z. et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. *npj Digit. Med.* **5**, 104 (2022).
- Crosson, W. L., Al-Hamdan, M. Z. & Insaf, T. Z. Downscaling NLDAS-2 daily maximum air temperatures using MODIS land surface temperatures. *PLoS One* **15**, e0227480 (2020).
- Eliezer, H., Johnson, S., Crosson, W. L., Al-Hamdan, M. Z. & Insaf, T. Z. Ground-truth of a 1-km downscaled NLDAS air temperature product using the New York City Community Air Survey. *J. Appl. Remote Sens.* **13**, 024516–024516 (2019).
- Jung, J. et al. Evaluation of NLDAS-2 and downscaled air temperature data in Florida. *Phys. Geogr.* **43**, 562–588 (2022).
- Estes, M. G. Jr, Insaf, T., Al-Hamdan, M. Z., Adeyeye, T. & Crosson, W. Validation of North American land data assimilation system Phase 2 (NLDAS-2) air temperature forcing and downscaled data with New York State station observations. *Remote Sens. Appl. Soc. Environ.* **25**, 100670 (2022).
- McNally, B., Stokes, A., Crouch, A. & Kellermann, A. L. CARES: cardiac arrest registry to enhance survival. *Ann. Emerg. Med.* **54**, 674–683.e672 (2009).
- Al-Hamdan, M. Z. et al. Environmental public health applications using remotely sensed data. *Geocarto Int* **29**, 85–98 (2014).
- Cosgrove, B. A. et al. Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *J. Geophys. Res. Atmos.* **108**, 3118 (2003).
- Chen, T. Q. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Kdd'16: Proc. 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794 (ACM, 2016).
- Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **7**, 21 (2013).
- Akwo, E. A. et al. Neighborhood deprivation predicts heart failure risk in a low-income population of blacks and whites in the Southeastern

- United States. *Circ. Cardiovasc. Qual. Outcomes* **11**, e004052 (2018).
34. Dong, W. et al. Risk factors and geographic disparities in premature cardiovascular mortality in US counties: a machine learning approach. *Sci. Rep.* **13**, 2978 (2023).
 35. Kish, L. Sampling organizations and groups of unequal sizes. *Am. Sociol. Rev.* **30**, 564–572 (1965).
 36. Heeringa, S. G., West, B. T., Heeringa, S. G. & Berglund, P. A. *Applied Survey Data Analysis* (Chapman and Hall/CRC, 2017).
 37. Cameron, A. C., Gelbach, J. B. & Miller, D. L. Robust inference with multiway clustering. *J. Bus. Econ. Stat.* **29**, 238–249 (2011).
 38. Zhang, T., Wang, K. & Zhang, X. Modeling and analyzing the transmission dynamics of HBV epidemic in Xinjiang, China. *PLoS One* **10**, e0138765 (2015).
 39. Lundberg, S. M. & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neur.* **30**, 4768–4777 (2017).
 40. Chen, H., Covert, I. C., Lundberg, S. M. & Lee, S.-I. Algorithms to estimate Shapley value feature attributions. *Nat. Mach. Intell.* **5**, 590–601 (2023).

Acknowledgements

We thank all the EMS agencies for their cooperation in establishing and maintaining the CARES registry. This research was partially supported by a Grant-in-Aid for Young Scientists (A) (20K17914) from the Japan Society for the Promotion of Science and Takeda Science Foundation. The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. All authors had full access to all datasets. The corresponding author had the ultimate responsibility for the decision to submit for publication.

Author contributions

T.N. contributed to formulation of the study design, interpretation of the results, and preparation of the final report. S.O. and K. contributed to data analysis and preparation of the final report. K.N., T.N., and R.W. Neumar contributed to the final report. B.M., R.A., and T.S. contributed to formulation of the concept of the CARES registry, data collection, and data management. M.A. and W.Y. contributed to data collection and analysis of meteorological data. All authors approved the final version.

Competing interests

The authors declare no competing interests.

Patient and public involvement

No patients or members of the public were directly involved in the design, or conduct, or reporting, or dissemination plans of this research.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s41746-025-02235-4>.

Correspondence and requests for materials should be addressed to Takahiro Nakashima.

Reprints and permissions information is available at

<http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025