

1 **A machine learning model for predicting out-of-hospital cardiac arrest incidence**

2 **using meteorological, chronological, and geographical data from the United States**

3

4 Takahiro Nakashima, MD, PhD,^{a, b} Soshiro Ogata, PhD,^b Eri Kiyoshige, PhD,^b

5 Mohammad Z Al-Hamdan, PhD,^{c, d} Yifan Wang, MPH,^c Teruo Noguchi, MD, PhD,^f

6 Theresa A Shields, RN, BSN^a, Rabab Al-Araji, MPH,^g Bryan McNally, MD, MPH^h

7 Kunihiro Nishimura, MD, PhD^b and Robert W Neumar, MD, PhD^a

8

9 Drs. Nakashima and Ogata contributed equally to this work.

10

11 **Short title:** A ML model predicting daily OHCA incidence

12

13 a. Department of Emergency Medicine and the Max Harry Weil Institute for Critical

14 Care Research and Innovation, University of Michigan, Ann Arbor, United States

15 b. Department of Preventive Medicine and Epidemiologic Informatics, National Cere-

16 bral and Cardiovascular Centre, Suita, Japan

17 c. National Center for Computational Hydroscience and Engineering, School of Engi-

18 neering, University of Mississippi, Oxford, Mississippi, USA

19 d. Department of Civil Engineering, School of Engineering, University of Mississippi,

20 Oxford, Mississippi, USA

21 e. Gangarosa Department of Environmental Health, Rollins School of Public Health,

22 Emory University

23 f. Department of Cardiovascular Medicine, National Cerebral and Cardiovascular

24 Centre, Suita, Japan

25 g. Emory University, Woodruff Health Sciences Center Atlanta, Georgia

26 h. Emory University School of Medicine, Atlanta, Georgia

27

28 Corresponding author:

29 Takahiro Nakashima, MD, PhD

30 Department of Emergency Medicine and Michigan Center for Integrative Research in

31 Critical Care, University of Michigan

32 1500 E. Medical Center Drive, Ann Arbor, MI, 48109, United States

33 Tel: +1-734-936-0253

34 Email: takana@med.umich.edu

35

36 **Manuscript Total Word Count: xxxx** (excluding the abstract, references, tables, boxes,

37 and figures)

38 **Clinical Perspective**

39 **What is new?**

40 A machine learning predictive model developed with a high-resolution meteorological
41 dataset and chronological and geographical variables predicted the daily incidence of
42 out-of-hospital cardiac arrest (OHCA) in the U.S. population with high precision. The
43 predictive accuracy at the state level was greater in medium and high-temperature areas
44 than in the low-temperature area.

45

46 **What are the clinical implications?**

47 This predictive model revealed complex associations between meteorological,
48 chronological, and geographic variables in relation to predicting daily incidence of
49 OHCA. It might be useful for public health strategies in temperate regions, for example,
50 by providing a warning system for citizens and emergency medical services agencies on
51 high-risk days.

52 **Abstract**

53 **Background:** Despite advances in pre- and post-resuscitation care, percentage of
54 survival to hospital discharge after out-of-hospital cardiac arrest (OHCA) was extremely
55 low. Development of an accurate system to predict the daily incidence of OHCA might
56 provide a significant public health benefit. Here, we developed and validated a machine
57 learning (ML) predictive model for daily OHCA incidence using high-resolution
58 meteorological, chronological, and geographical data.

59 **Methods:** We analyzed a dataset from the United States that combined an OHCA
60 nationwide registry, high-resolution meteorological data, chronological data, and
61 geographical data. We developed a model to predict daily OHCA incidence with a
62 training dataset for 2013–2017 using the eXtreme Gradient Boosting algorithm. A
63 dataset for 2018–2019 was used to test the predictive model. The main outcome was the
64 predictive accuracy for the number of daily OHCA events, based on root mean squared
65 error (RMSE), mean absolute error (MAE), and mean absolute percentage error
66 (MAPE). In general, a model with MAPE less than 10% is considered highly accurate.

67 **Results:** Among the 446,830 OHCA of non-traumatic cause where resuscitative efforts
68 were initiated by a 911 responder, 264,916 in the training dataset and 181,914 in the
69 testing dataset were included in the analysis. The ML model with combined

70 meteorological, chronological, and geographical data had high predictive accuracy in
71 relation to nationwide incidence rate per 100,000 at the nationwide level) in the training
72 dataset (RMSE, 0.016; MAE, 0.013; and MAPE, 7.61%) and in the testing dataset
73 (RMSE, 0.018; MAE, 0.014; and MAPE, 6.52%).

74 **Conclusions:** A ML predictive model using comprehensive daily meteorological,
75 chronological, and geographical data allows for highly precise estimates of OHCA
76 incidence in the United States.

77 **271 / 300 words**

78

79 **Keywords**

80 Prediction, out-of-hospital cardiac arrest, machine learning, meteorology

81

82 The estimated global incidence of out-of-hospital cardiac arrest (OHCA) treated by
83 emergency medical service (EMS) is 62.0 per 100,000 person-years, with a specific
84 incidence of 53.1 per 100,000 person-years in North America.¹ Despite advances in pre-
85 and post-resuscitation care, a recent systematic review showed that survival to hospital
86 discharge after OHCA was extremely low, at 8.8% (95% confidence interval [CI], 8.2–
87 9.4%).² Prevention and early reaction in the prehospital setting are important for
88 improving the prognosis of patients with OHCA. Development of an accurate system to
89 predict the daily incidence of OHCA might provide a significant public health benefit.

90 Several studies have shown associations between ambient temperature and
91 cardiovascular events³⁻⁶ and between day of week or season and cardiovascular events.⁷⁻
92 ¹² However, many of those studies used conventional linear regression, which might not
93 be suitable for assessing the influence of ambient temperature on OHCA incidence and
94 handling large amounts of high-resolution meteorological data. To date, meteorological
95 and chronological data have not been applied to real-world practice. Machine learning
96 (ML) can use advanced analytics to integrate multiple quantitative variables and
97 identify associations not identified with conventional one-dimensional statistical
98 approaches, which might allow us to develop a predictive model. Recently, our team
99 developed a ML predictive model for daily OHCA incidence based on combined

100 meteorological and chronological data with high accuracy for the Japanese population.¹³

101 In this study, we develop and validated the ML predictive model for robust
102 estimation of daily OHCA incidence of cardiac origin for the U.S. population; the
103 United States is close in latitude to Japan. This model was evaluated using the OHCA
104 dataset from the Cardiac Arrest Registry to Enhance Survival (CARES) from the United
105 States,¹⁴ as well as comprehensive meteorological data from National Aeronautics and
106 Space Administration (NASA), chronological data, and geographic data.

107

108 **Methods**

109 The study was approved by the University of Michigan Hospital's institutional review
110 board (HUM00189913). The requirement for written informed consent was waived
111 because the researchers only analyzed deidentified (anonymized) data.

112

113 **Data source for out-of-hospital cardiac arrest data**

114 OHCA data was provided by CARES, which is a prospective multicenter registry of
115 patients with OHCA from 30 state-based registries, the District of Columbia, and more
116 than 45 community sites in 16 additional U.S. states. It has a catchment area of
117 approximately 175 million residents in 2021 (**Figure 1**). The design of the registry,

118 which was established by the U.S. Centers for Disease Control and Prevention and
119 Emory University, has been previously described.¹⁴ Patient-level data were collected by
120 EMS agencies using standardized international Utstein definitions for clinical variables
121 and outcomes to ensure uniformity.

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

133

134 **Meteorological data**

135 We analyzed meteorological data from NASA's North American Land Data

136 Assimilation System, which provides hourly gridded data with 12-km spatial
137 resolution.^{15, 16} Meteorological variables included average daily values of ambient
138 temperature (°C), precipitation (mm), relative humidity (%), and wind speed (m/s)
139 during the study period. Those values were averaged by EMS agency areas.

140

141 **Chronological and geographical data**

142 Chronological variables included year (2013 was considered year 0), month (February
143 to December as categorical variables, with January as the referent), day of the week
144 (with Sunday as the referent), holidays, start and end of daylight saving time, and
145 holiday season from December 24 to January 1 (categorical variable with a value of 0 or
146 1).

147 Geographical data were collected at the census tract level included median age
148 (four categories: <35, 35–39, 40–44, and ≥45 year, near its quartiles), proportion of men
149 (two categories: <50% or ≥50%, which approximate the median), race (proportion of
150 Blacks [three categories: <1%, 1–6%, and ≥6%, near its tertiles], proportion of Asians
151 [three categories: <1%, 1–2%, and ≥3%, near its tertiles]), proportion of individuals
152 with a high school diploma or higher (three categories: <55%, 55–60%, 60–64%, and
153 ≥65%, near its quartiles), unemployment rate (four categories: <4%, 4–4.9%, 5–6%, and

154 $\geq 6\%$, near its quartiles), percentage of individuals living below the poverty level (four
155 categories: $< 7\%$, 7–10%, 11–14%, and $\geq 15\%$, near its quartiles), average household size
156 (four categories: < 2.4 , 2.4–2.55, 2.56–2.69, and ≥ 2.7 , near its quartiles), and population
157 per square mile (four categories: < 500 , 500–1,699, 1700–3,200, and 3,200 and over,
158 near its quartiles). The geographical data at the census tract level were merged into EMS
159 agency areas. In cases where an EMS agency area was covered by multiple census tract
160 areas, we used the geographical from the tract with the highest number of OHCA cases
161 among the several tract areas.

162

163 **Data management and development of predictive models**

164 We matched the CARES data, meteorological data, chronological data, and geographic
165 data between January 1, 2013, and December 31, 2019 at the hourly level based on the
166 time of the emergency call. We classified data from January 1, 2013 to December 31,
167 2017 in this merged dataset as the training dataset for developing the predictive model.
168 Of EMS agencies in the training dataset, 30% were used as the validation dataset for
169 selecting hyperparameters. Data from January 1, 2018 to December 31, 2019 were used
170 as the testing dataset for assessing whether the predictive model can work in other
171 years.

172 To develop predictive model for the daily incidence of OHCA, we used the
173 eXtreme Gradient Boosting (XGBoost) gradient boosting algorithm.^{17, 18} We selected
174 XGBoost hyperparameters that maximized prediction performance in the validation
175 dataset. In other words, we minimized root mean squared error (RMSE) by developing a
176 model in the dataset between 2013 and 2017 in 70% of the participating agency areas
177 and checking its prediction performances according to its hyperparameters in the
178 validation dataset (i.e., data between 2013 and 2017 in the remaining 30% of agency
179 areas). Population size for each agency area was included in the XGBoost algorithm as
180 an offset term. We assessed prediction performance of the model developed with the
181 testing dataset (i.e., data between 2018 and 2019 from all participating agency areas).

182

183 **Primary and secondary outcomes**

184 The primary outcome was predictive accuracy of OHCA incidence rate per 100,000 at
185 nationwide level of the predictive model based on RMSE, mean absolute error (MAE),
186 and mean absolute percentage error (MAPE), which are generally used as measures of
187 predictive accuracy for a forecasting method. The secondary outcome was predictive
188 accuracy of OHCA incidence rate per 100,000 at the state level, which was limited to 24
189 state-based registries.

190

191 **Statistical analysis**

192 We performed a third-step analysis. First, we examined the concordance between the
193 predicted incidence of OHCA based on the ML model and the observed incidence of
194 OHCA in the testing dataset. Next, we investigated important predictors for predicting
195 the OHCA incidence in the developed prediction model. Finally, we assessed the
196 predictive accuracy of the ML model stratified into low-, intermediate-, and high-
197 temperature areas, further divided into summer (June–August) and winter (December–
198 February). Low-, intermediate-, and high-temperature areas were defined as regions
199 with mean ambient temperature in the 25th percentile or lower, in the 25–75th
200 percentiles, and 75th percentile or higher, respectively.

201 The characteristics of the present dataset were summarized with medians and
202 interquartile ranges (IQRs) for continuous variables, and numbers and percentages for
203 categorical variables by area and day in the training and testing datasets. We evaluated
204 the predictive accuracy of the predictive models based on RMSE, MAE, and MAPE
205 between predicted values calculated with the predictive models and observed daily
206 OHCA incidence at the EMS agency level. RMSE and MAE reflect the average
207 magnitude of differences between predicted values and observed values. RMSE and

208 MAE can range from zero to infinity. Lower RMSE and MAE values indicate higher
209 predictive performance. MAPE is an average of the absolute values of errors divided by
210 observed values. MAPE ranges from zero to infinity. Lower MAPE values indicate
211 higher model predictive performance. In general, MAPE less than 10% is considered
212 highly accurate predicting.¹⁹ Formulas are as follows;

213

$$214 \quad \text{RMSE} = \sqrt{\sum_{t=1}^n (|\text{observed value} - \text{predicted value}|^2 / N)}$$

$$215 \quad \text{MAE} = \frac{1}{n} \sum_{t=1}^n |\text{observed value} - \text{predicted value}|$$

$$216 \quad \text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|\text{observed value} - \text{predicted value}|}{\text{observed value}} * 100$$

217

218 To show important predictors of the OHCA incidence in the developed
219 prediction model, we used the Shapley Additive Explanations (SHAP) values
220 summarizing contribution of each predictor to the predicted value of an instance.²⁰ For a
221 given set of feature values, a SHAP value reflects how much a single variable, in the
222 context of its interaction with other variables, contributes to the difference between the
223 actual prediction and the mean prediction.

224 All statistical analyses were performed with R statistical software, version 4.1.2
225 (<https://www.R-project.org/>) and the xgboost package for R, version 1.5.0.1

226 (<https://CRAN.R-project.org/package=xgboost>). Missing values for continuous and
227 categorical variables were, respectively, imputed by a median value for each continuous
228 variable and treated as a missing category. These missing procedures work well when
229 using XGBoost due to the nature of decision tree algorithms.

230

231 **Results**

232 **Characteristics of the training and testing datasets**

233 From the CARES registry, 446,830 EMS-treated OHCA of non-traumatic cause
234 between 2013 and 2019 were matched with meteorological data; there were 264,916
235 cases in the training dataset. There were 181,914 cases in the testing dataset. The
236 characteristics of the datasets are summarized in **Table 1**. Between 2013 and 2019, the
237 median annual incidence of OHCA increased from 58.6 to 76.3 per 100,000 person-
238 years. The median age of OHCA onset increased from 64 (IQR, 52–77) to 65 years
239 (IQR, 53–76). The proportion of males increased from 61% to 62%.

240 The median of the mean ambient temperature within a day decreased from
241 8.9°C (IQR, –0.1 to 18.1) to 7.9°C (IQR, –0.7 to 18.3) in the low-temperature area. This
242 trend was not observed in the intermediate or high-temperature areas. Differences
243 between maximum and minimum ambient temperatures within a day (diurnal

244 temperature range) increased from 9.4°C (6.8–12.1) to 9.6°C (7.0–12.4) in the low-
245 temperature area, from 9.2°C (6.5–11.8) to 9.9°C (7.4–12.7) in the intermediate-
246 temperature area, and from 10.3°C (7.8–12.7) to 10.6°C (7.8–3.6) in the high-
247 temperature area. Relative humidity also increased throughout the study period in all
248 areas. The incidence of OHCA by meteorological condition is shown in **Figure 2**.

249

250 **Predictive performance of the models**

251 Predicted and observed incidence of OHCA with a cardiac origin for each model are
252 shown in **Figure 3**. The ML-predicted model, which incorporated meteorological,
253 chronological, and geographical variables, was able to accurately predict daily
254 fluctuations and a significant increase in OHCA incidence at the nationwide level,
255 demonstrating good concordance between predicted and observed values. The error rate
256 was less than 0.05 per 100,000 persons per day. The predictive accuracy of the models
257 is shown in **Table 2**. At the nationwide level, the ML predictive model had high
258 predictive accuracy in the training dataset (RMSE, 0.016; MAE, 0.013; and MAPE,
259 7.61%) and the testing dataset (RMSE, 0.018; MAE, 0.014; and MAPE, 6.52%). At the
260 state level, the model had RMSE of 0.216 and MAE of 0.128 with the training dataset
261 and RMSE of 0.209 and MAE of 0.131 with the testing dataset. At the agency level, the

262 model had RMSE of 1.018 and MAE of 0.287 with the training dataset and RMSE of
263 1.290 and MAE of 0.343 with the testing dataset.

264

265 **Contribution of each predictor to the predicted value of OHCA incidence**

266 Importance of meteorological, chronological, and geographical variables in the ML
267 predictive model is shown in **Figure 4**. In general, geographical and meteorological
268 variables contributed more to the ML model than chronological variables. With regards
269 to meteorological variables, lower mean ambient temperature within a day was the
270 variable most strongly contributing to the predicted OHCA incidence, followed by
271 mean relative humidity and larger diurnal temperature range. Among chronological
272 variables, year, day of week, and holiday contributed more strongly to the predicted
273 OHCA incidence than month, but their contributions were relatively small. Among
274 geographic variables, percentage of Black persons, percentage of individuals living
275 under the poverty level, unemployment rate, median age, and population per square
276 mile contributed strongly to the predicted OHCA incidence.

277

278 **Predictive performance based on annual average of daily mean ambient**
279 **temperature**

280 We compared the predictive accuracy of the ML model stratified by annual average of
281 ambient temperature (**Table 3**). Population and OHCA incidence per 100,000 person-
282 years were higher in the high-temperature area, followed by the intermediate-
283 temperature area and low-temperature area. Predictive accuracy was higher in the
284 intermediate- and high-temperature areas than in the low-temperature area in the
285 training and testing datasets, as was OHCA incidence. However, there was not much
286 difference in the predictive accuracy between summer and winter in any area.

287

288 **Discussion**

289 In this study, using an ML predictive model developed with the combination of
290 meteorological, chronological, and geographic variables, we successfully predicted the
291 daily incidence of OHCA due to a cardiac origin in the United States with high
292 precision at the nationwide, state, and agency level, respectively. Lower mean ambient
293 temperature within a day and larger difference between maximum and minimum
294 ambient temperatures within a day were strongly associated with daily OHCA
295 incidence.

296 An association between ambient temperature and incidence of cardiovascular
297 events has been previously reported.³⁻¹⁰ However, since these studies focused on

298 ambient temperature or season alone, diversity in climate (comprehensive
299 meteorological variables), chronological variables, and geography were not considered.
300 Recently, our team reported that an ML predictive model for OHCA incidence in Japan
301 based on a comprehensive meteorological dataset and chronological variables had high
302 predictive accuracy.¹³ In the present study, when geographical data were added to the
303 predictive model, we achieved a higher predictive accuracy in the U.S. population than
304 in the Japanese population (MAPE, 6.52% vs. 7.79%). Notably, patients with OHCA in
305 the United States were more likely to be younger (64 vs. 80 years) or male (62% vs.
306 57%) than patients with OHCA in Japan. When chronological and geographic data were
307 added to the model, the association between OHCA incidence and meteorological
308 conditions can be more widely generalizable for the general population, even in
309 countries with a different OHCA profile as long they are at a similar latitude range.

310 We found that geographic data as well as race, socioeconomic disparities, and
311 age are strongly associated with OHCA incidence in a SHAP analysis. However, while
312 geographic characteristics were related to differences between counties, they are not
313 relevant to daily fluctuations in OHCA incidence because those variables do not change
314 throughout the year. For meteorological variables, mean ambient temperature in a day
315 and a larger diurnal temperature range were associated with the incidence of OHCA

316 with a cardiac origin. Wolf et al., who analyzed 9,801 patients from a cardiovascular
317 disease registry in Germany, reported that a drop in ambient temperature of 10°C or
318 more within a day and ambient temperature itself are associated with myocardial
319 infarction.⁵ A similar trend was observed recently in our study of a ML predictive model
320 for OHCA in the Japanese population.¹³ The present results were also consistent with
321 the previously described relationship between ambient temperature and OHCA
322 incidence. Race and socioeconomic disparities might emphasize the relationship
323 between meteorological conditions and OHCA incidence.

324 The predictive accuracy of the model was generally lower at the state level than
325 at the nationwide level. Our ML predictive model had variations in predictive accuracy
326 across states. Analyses stratified by annual average ambient temperature showed that
327 predictive accuracy was the lowest in the low-temperature area, while the season
328 variable (summer or winter) did not substantially change predictive accuracy. These
329 results were partially explained by the population in the participating area. Collecting
330 more samples improves our ML model. In addition, populations residing in the low-
331 temperature area might be more habituated and better able to cope with climate change,
332 such as through building insulation and lifestyle habits. Curriero et al. reported a
333 latitude dependence of the temperature–mortality relationship in their analysis based on

334 11 eastern cities in the United States.²¹ More effective adaption to colder temperature
335 was observed in cities that are further north. In order to be more practical, it needs to be
336 further improved to predict OHCA incidence within a medical catchment area. Weather
337 forecasts can predict meteorological conditions approximately 10 days ahead. Thus, a
338 predictive model based on meteorological data might allow EMS agencies and hospitals
339 to prepare and reallocate medical resources, which leads to more rapid transport and
340 advanced post-arrest care. A future prospective study to evaluate the effectiveness of
341 this approach is needed.

342 This study has several inherent limitations. First, although the CARES registry is
343 the largest database of OHCA in the United States, it covered approximately 53% of
344 catchment areas as of 2022. CARES only includes EMS-treated OHCA. This
345 underestimates the overall incidence of OHCA, but is most relevant population on
346 which to focus public health interventions. Second, our data did not address the
347 potential variability in patients' preexisting medical conditions. Third, the predictability
348 of future OHCA events will depend on the accuracy of meteorological data. Finally,
349 external testing in other developed countries was not performed.

350

351 **Conclusion**

352 An ML predictive model using multiple meteorological, chronological, and
353 geographical variables could predict the incidence of OHCA with a cardiac origin with
354 high precision in the U.S. population. This predictive model might be useful for public
355 health prevention strategies in temperate regions, which can also be applied even in
356 countries with a different OHCA profile.

357

358 **Acknowledgments**

359 We thank all the EMS agencies for their cooperation in establishing and maintaining the
360 CARES registry.

361

362 **Data sharing statement**

363 The CARES registry and meteorological data were used with permission for this study.
364 Data sharing outside of the research team was not permitted. However, the analytic code
365 in R can be shared upon request.

366

367 **Contributors**

368 T. Nakashima contributed to formulation of the study design, interpretation of the
369 results, and preparation of the final report. T. Noguchi and R.W. Neumar contributed to

370 the final report. K. Nishimura, S. Ogata, and Kiyoshige contributed to data analysis. B.
371 McNally, R. Al-Araji, and T. Shields contributed to formulation of the concept of the
372 CARES registry, data collection, and data management. M. Al-Hamadan and W. Yifan
373 contributed to data collection and analysis of meteorological data. All authors approved
374 the final version.

375

376 **Funding**

377 This research was partially supported by a Grant-in-Aid for Young Scientists (A)
378 (20K17914) from the Japan Society for the Promotion of Science. All authors had full
379 access to all datasets. The corresponding author had the ultimate responsibility for the
380 decision to submit for publication.

381

382 **Competing Interests**

383 All authors declare no competing interests.

384

385 **Figure legends**

386 **Figure 1. States and communities participating in CARES**

387 CARES denotes Cardiac Arrest Registry to Enhance Survival.

388

389 **Figure 2. Incidence of out-of-hospital cardiac arrest by meteorological condition**

390 The plots indicate the mean daily OHCA incidence per 100,000 persons.

391 OHCA denotes out-of-hospital cardiac arrest.

392

393 **Figure 3. Observed versus predicted incidence of out-of-hospital cardiac arrest**

394 The light blue lines indicate the observed total number of out-of-hospital cardiac arrests

395 per day in the United States. The yellow lines indicate the predicted number based on

396 combined meteorological, chronological, and geographical variables.

397 Apr denotes April; Jan, January; Jul, July; Oct, October.

398

399 **Figure 4. Importance of meteorological, chronological, and geographical variables**

400 **in a machine learning predictive model**

401 This figure shows a variable importance plot for meteorological variables (red),

402 chronological variables (blue), and geographical variables (black) in a machine learning

403 predictive model using XGBoost. The yellow to purple dots in each row represent low
404 to high values for each predictor normally scaled. The x -axis shows the Shapley value,
405 indicating the variable's impact on the model. Positive SHAP values tend to drive
406 predictions toward an OHCA event and negative SHAP values tend to drive the
407 prediction toward no OHCA event.

408 * In the model, 2013 was considered year 0.

409 OHCA denotes out-of-hospital cardiac arrest; SHAP, Shapley Additive Explanations;

410 XGBoost, eXtreme Gradient Boosting.

411 **Table 1. Characteristics of daily data in the training dataset (2013–2017) and testing**
 412 **dataset (2018–2019)**

Variable	Training dataset (n=264,916)		Testing dataset (n=181,914)
	2013–2015	2016–2017	2018–2019
Demographic variables			
Number of participating EMS agencies	575	845	1222
OHCA incidence, per 100,000 person-years	58.6	70.4	76.3
Patient characteristics			
Age, years, median (IQR)	64 (52, 77)	64 (52, 76)	65 (53, 76)
Male sex, n (%)	81,903 (61)	86,761 (62)	115,674 (62)
Race, n (%)			
White	61,961 (46)	67,537 (48)	94,518 (51)
Black	27,408 (20)	31,402 (22)	43,049 (23)
Asian	2,531 (2)	2,903 (2)	4,046 (2)
Hispanic/Latino	7,952 (6)	7,506 (5)	12,199 (7)
Other	34,529 (26)	31,074 (23)	32,212 (17)
First documented rhythm, n (%)			
Shockable	27,547 (20)	26,599 (19)	34,718 (19)
Non-shockable	106,823 (79)	113,810 (81)	151,279 (81)
Meteorological variables			
Ambient temperature, °C			
Mean value within a day			
Low-temperature area	8.9 (−0.1, 18.1)	9.6 (0.2, 18.0)	7.9 (−0.7, 18.3)
Intermediate temperature area	11.9 (4.5, 19.5)	11.9 (4.4, 19.4)	11.6 (3.99, 20.0)
High-temperature area	17.4 (10.8, 24.0)	17.1 (10.7, 23.3)	17.6 (10.7, 24.5)
Differences between maximum and minimum values within a day			
Low-temperature area	9.4 (6.8, 12.1)	8.8 (6.1, 11.2)	9.6 (7.0, 12.4)
Intermediate temperature area	9.2 (6.5, 11.8)	9.0 (6.2, 11.6)	9.9 (7.4, 12.7)
High-temperature area	10.3 (7.8, 12.7)	9.6 (7.1, 12.0)	10.6 (7.8, 13.6)
Precipitation during the previous hour, mm			
Low-temperature area	0.00 (0.00, 0.06)	0.00 (0.00, 0.09)	0.00 (0.00, 0.09)
Intermediate temperature area	0.00 (0.00, 0.07)	0.00 (0.00, 0.09)	0.00 (0.00, 0.10)
High-temperature area	0.00 (0.0, 0.10)	0.00 (0.00, 0.05)	0.00 (0.00, 0.05)
Wind speed, m/s			
Low-temperature area	3.2 (2.1, 4.5)	3.5 (2.3, 4.8)	3.4 (2.2, 4.8)
Intermediate temperature area	3.0 (2.0, 4.3)	3.4 (2.3, 4.7)	3.2 (2.2, 4.5)
High-temperature area	3.1 (2.2, 4.2)	3.2 (2.2, 4.3)	3.1 (2.1, 4.3)
Relative humidity, %			
Low-temperature area	70.8 (55.5, 81.6)	77.6 (70.2, 84.7)	78.7 (69.9, 85.5)
Intermediate temperature area	74.0 (62.2, 82.8)	75.8 (67.7, 83.7)	77.9 (69.8, 84.5)
High-temperature area	69.3 (55.7, 79.1)	72.3 (63.3, 81.0)	74.1 (61.7, 82.1)

413 Continuous values are presented as medians (IQR).

414 *See **eFigure 1 in the Supplement** for more information.

415 †Snowfall values were calculated for the winter (December–February).

416 IQR denotes interquartile range; hPa, hectopascal; OHCA, out-of-hospital cardiac arrest.

417 **Table 2. Performance of the predictive model for out-of-hospital cardiac arrest**

418 **based on combined meteorological and chronological data**

Measure of predictive model performance	Training dataset	Testing dataset
Primary outcomes		
Nationwide level		
RMSE, incidence rate per 100,000 per day	0.016	0.018
MAE by day, incidence rate per 100,000 per day	0.013	0.014
MAPE by day, % ^a	7.61	6.52
Secondary outcomes		
State level ^b		
RMSE, incidence rate per 100,000 per day	0.216	0.209
MAE by day, incidence rate per 100,000 per day	0.128	0.131
MAPE by day, % ^a	NA	NA
Agency level		
RMSE, incidence rate per 100,000 per day	1.018	1.290
MAE by day, incidence rate per 100,000 per day	0.287	0.343
MAPE by day, % ^a	NA	NA

- 419 a. In general, MAPE <10% is considered highly accurate predicting; 10–20%, good predicting;
 420 20–50%, reasonable predicting; and >50%, inaccurate predicting.²² MAPE could not be
 421 calculated at the state or agency level because there were days with no cardiac arrests.
- 422 b. State-level calculations were limited to 24 states that participated in state-based data collection
 423 through 2017 (training dataset).
- 424 MAE denotes mean absolute error; MAPE, mean absolute percentage error; ML machine learning;
 425 NA, not applicable; RMSE, root mean squared error.

426

427 **Table 3. Accuracy of the predictive model for out-of-hospital cardiac arrest stratified by temperature**

Measure of predictive model performance	Low-temperature area (25th percentile)		Intermediate-temperature area (25–75th percentile)		High-temperature area (75th percentile)	
	Training dataset	Testing dataset	Training dataset	Testing dataset	Training dataset	Testing dataset
Number of participating EMS agencies	228	417	455	496	228	309
Population	11,956,637	21,604,072	43,947,825	45,568,174	48,807,506	63,166,710
OHCA incidence, per 100,000 person-years	52.5	69.2	64.0	78.1	66.8	77.4
All seasons						
RMSE ^a	0.043	0.034	0.023	0.027	0.024	0.024
MAE by day ^a	0.035	0.027	0.019	0.021	0.019	0.019
MAPE by day, % ^b	31.82	14.64	11.81	9.36	10.78	8.92
Summer						
RMSE ^a	0.041	0.031	0.023	0.024	0.022	0.023
MAE by day ^a	0.033	0.025	0.018	0.019	0.017	0.019
MAPE by day, % ^b	31.97	15.14	12.81	9.89	11.12	9.12
Winter						
RMSE ^a	0.045	0.037	0.025	0.034	0.026	0.026
MAE by day ^a	0.036	0.029	0.019	0.026	0.021	0.021
MAPE by day, % ^b	29.15	13.70	10.91	10.53	10.24	8.74

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

429 b. In general, MAPE <less than 10% is considered highly accurate predicting; 10–20%, good predicting; 20–50%, reasonable predicting; and >50%,
 430 inaccurate predicting.²²

431 MAE denotes mean absolute error; MAPE, mean absolute percentage error; ML, machine learning; RMSE, root mean squared error.

432

433

434

References

435

- 436 1. Berdowski J, Berg RA, Tijssen JG and Koster RW. Global incidences of out-of-
437 hospital cardiac arrest and survival rates: Systematic review of 67 prospective
438 studies. *Resuscitation*. 2010;81:1479–87.
- 439 2. Yan S, Gan Y, Jiang N, Wang R, Chen Y, Luo Z, Zong Q, Chen S and Lv C. The
440 global survival rate among adult out-of-hospital cardiac arrest patients who received
441 cardiopulmonary resuscitation: a systematic review and meta-analysis. *Crit Care*.
442 2020;24:61.
- 443 3. Yamaji K, Kohsaka S, Morimoto T, Fujii K, Amano T, Uemura S, Akasaka T,
444 Kadota K, Nakamura M, Kimura T and Investigators JPR. Relation of ST-Segment
445 Elevation Myocardial Infarction to Daily Ambient Temperature and Air Pollutant Levels
446 in a Japanese Nationwide Percutaneous Coronary Intervention Registry. *Am J Cardiol*.
447 2017;119:872–880.
- 448 4. Cold exposure and winter mortality from ischaemic heart disease,
449 cerebrovascular disease, respiratory disease, and all causes in warm and cold regions
450 of Europe. The Eurowinter Group. *Lancet*. 1997;349:1341–6.
- 451 5. Wolf K, Schneider A, Breitner S, von Klot S, Meisinger C, Cyrys J, Hymer H,
452 Wichmann HE and Peters A. Air temperature and the occurrence of myocardial infarction
453 in Augsburg, Germany. *Circulation*. 2009;120:735–42.
- 454 6. Danet S, Richard F, Montaye M, Beauchant S, Lemaire B, Graux C, Cottel D,
455 Marécaux N and Amouyel P. Unhealthy effects of atmospheric temperature and pressure
456 on the occurrence of myocardial infarction and coronary deaths. A 10-year survey: the
457 Lille–World Health Organization MONICA project (Monitoring trends and determinants in
458 cardiovascular disease). *Circulation*. 1999;100:E1–7.
- 459 7. Herlitz J, Eek M, Holmberg M and Holmberg S. Diurnal, weekly and seasonal
460 rhythm of out of hospital cardiac arrest in Sweden. *Resuscitation*. 2002;54:133–8.
- 461 8. Arntz HR, Willich SN, Schreiber C, Brüggemann T, Stern R and Schultheiss HP.
462 Diurnal, weekly and seasonal variation of sudden death. Population-based analysis of
463 24,061 consecutive cases. *Eur Heart J*. 2000;21:315–20.
- 464 9. Kloner RA, Poole WK and Perritt RL. When throughout the year is coronary
465 death most likely to occur? A 12-year population-based analysis of more than 220 000
466 cases. *Circulation*. 1999;100:1630–4.
- 467 10. Marchant B, Ranjadayalan K, Stevenson R, Wilkinson P and Timmis AD.
468 Circadian and seasonal factors in the pathogenesis of acute myocardial infarction:
469 the influence of environmental temperature. *Br Heart J*. 1993;69:385–7.

- 470 11. Sandhu A, Seth M and Gurm HS. Daylight savings time and myocardial
471 infarction. *Open Heart*. 2014;1:e000019.
- 472 12. Manfredini R, Fabbian F, De Giorgi A, Zucchi B, Cappadona R, Signani F,
473 Katsiki N and Mikhailidis DP. Daylight saving time and myocardial infarction: should
474 we be worried? A review of the evidence. *Eur Rev Med Pharmacol Sci*. 2018;22:750–755.
- 475 13. Nakashima T, Ogata S, Noguchi T, Tahara Y, Onozuka D, Kato S, Yamagata Y,
476 Kojima S, Iwami T, Sakamoto T, Nagao K, Nonogi H, Yasuda S, Iihara K, Neumar R and
477 Nishimura K. Machine learning model for predicting out-of-hospital cardiac arrests
478 using meteorological and chronological data. *Heart*. 2021.
- 479 14. McNally B, Stokes A, Crouch A and Kellermann AL. CARES: Cardiac Arrest
480 Registry to Enhance Survival. *Ann Emerg Med*. 2009;54:674–683. e2.
- 481 15. Al-Hamdan MZ, Crosson WL, Economou SA, Estes MG, Jr., Estes SM, Hemmings SN,
482 Kent ST, Puckett M, Quattrochi DA, Rickman DL, Wade GM and McClure LA. Environmental
483 Public Health Applications Using Remotely Sensed Data. *Geocarto Int*. 2014;29:85–98.
- 484 16. Cosgrove BA, Lohmann D, Mitchell KE, Houser PR, Wood EF, Schaake JC, Robock
485 A, Marshall C, Sheffield J and Duan Q. Real - time and retrospective forcing in the
486 North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical*
487 *Research: Atmospheres*. 2003;108.
- 488 17. Chen TQ and Guestrin C. XGBoost: A Scalable Tree Boosting System. *Kdd' 16:*
489 *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery*
490 *and Data Mining*. 2016:785–794.
- 491 18. Natekin A and Knoll A. Gradient boosting machines, a tutorial. *Front*
492 *Neurobot*. 2013;7.
- 493 19. Zhang T, Wang K and Zhang X. Modeling and Analyzing the Transmission
494 Dynamics of HBV Epidemic in Xinjiang, China. *PLoS One*. 2015;10:e0138765.
- 495 20. Lundberg SM and Lee SI. A Unified Approach to Interpreting Model Predictions.
496 *Adv Neur In*. 2017;30.
- 497 21. Curriero FC, Heiner KS, Samet JM, Zeger SL, Strug L and Patz JA. Temperature
498 and mortality in 11 cities of the eastern United States. *Am J Epidemiol*. 2002;155:80–
499 7.
- 500

Figure 1. States and communities participating in CARES

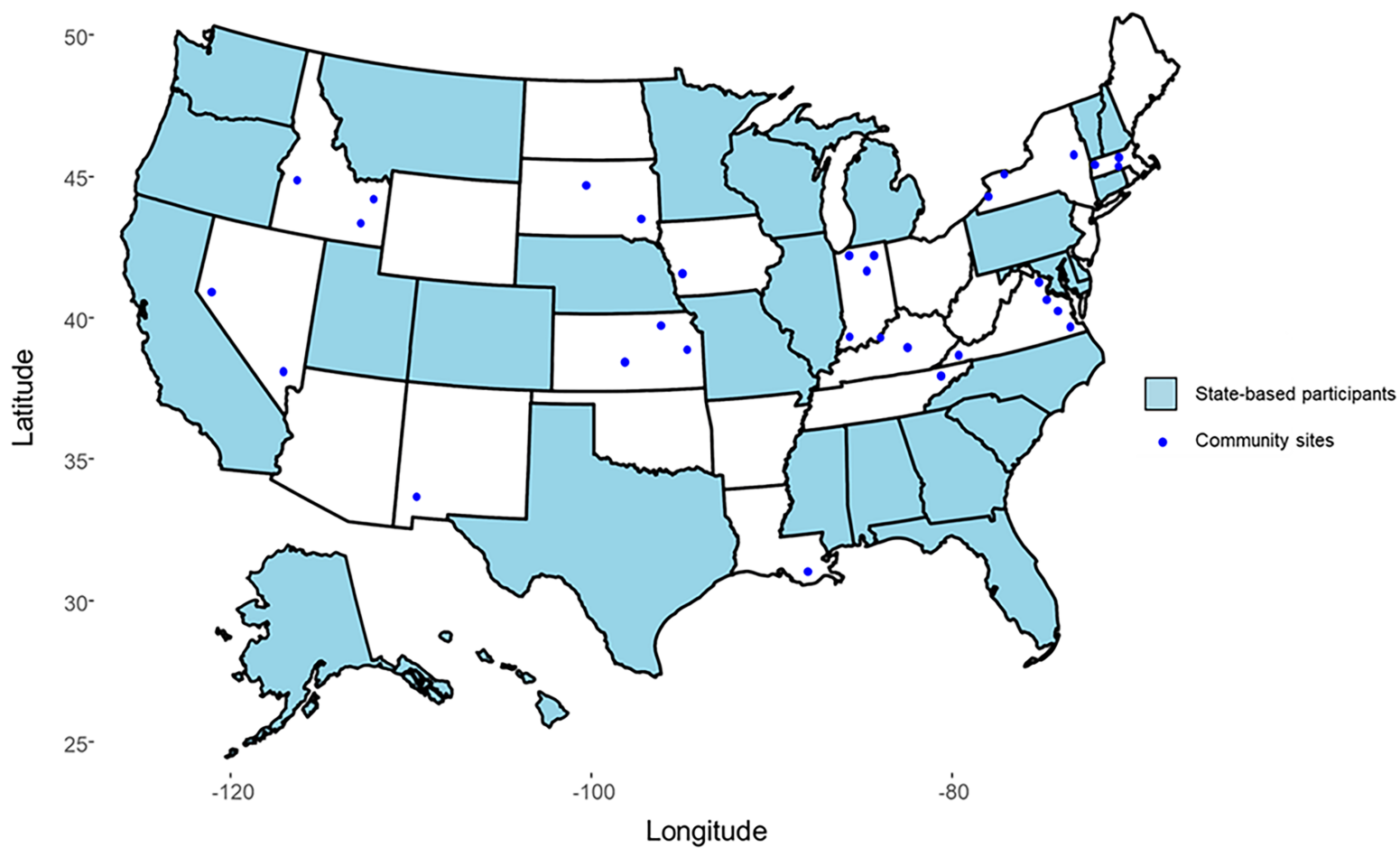


Figure 2. Incidence of out-of-hospital cardiac arrest by meteorological condition

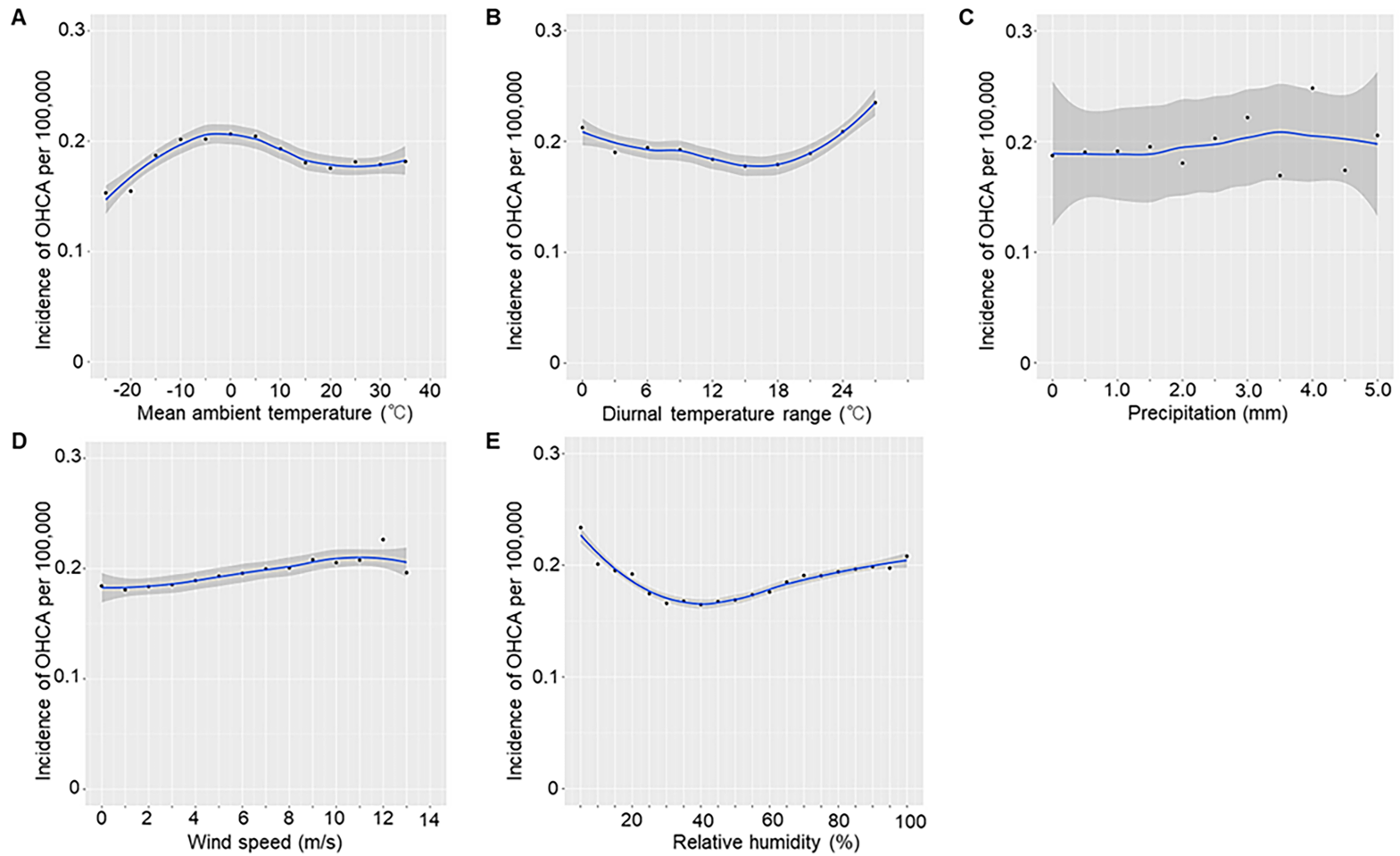


Figure 3. Observed versus predicted incidence of out-of-hospital cardiac arrest



Figure 4. Predictive importance

