Building and occupant characteristics as predictors of temperature-related health hazards in American homes

Arfa N. Aijazia,b, , Stefano Schiavona\*, Duncan Callawayc

a Center for the Built Environment, University of California, Berkeley, CA 94720

b School of Architecture, University of Waterloo, Canada

c Energy and Resources Group, University of California, Berkeley, CA 94720

\* Corresponding author:

Email address: [schiavon@berkeley.edu](mailto:schiavon@berkeley.edu)

Address: Center for the Built Environment (CBE)

University of California, Berkeley

390 Wurster Hall #1839

Berkeley, CA 24720-1839

Keywords: Thermal resilience, Housing, Extreme heat, Extreme cold, Overheating, Overcooling, Public health, Climate change

Target publication: [Science of the Total Environment](https://www.sciencedirect.com/journal/science-of-the-total-environment)

# Abstract

Many cities and regions are making significant investments towards planning for extreme temperature and in particular extreme heat. Heat vulnerability indices (HVI) are used to track spatial variation in extreme temperature risk to target mitigation interventions. Most HVI focus on demographic characteristics, which generally relate to vulnerability, and lack information about the building stock, which mediate the occupant’s exposure to extreme temperatures. In this study, we use the Energy Information Administration’s (EIA) Residential Energy Consumption Survey (RECS) to estimate prevalence of temperature-related illness in the United States and develop machine learning models using climate, demographic, and building characteristics to predict them. Temperature-related illness affects approximately 2 million households annually, around 1% of the total population. The models we develop predict temperature-related illness with up to 85% accuracy. The most important feature is energy insecurity, which describes the household’s ability to maintain and operate heating, ventilation, and air conditioning (HVAC) systems. Our results offer guidance for municipalities to improve both 1) data collection, enabling them to better identify at-risk households and 2) interventions, such as by targeting factors that could mitigate temperature-related health hazards.

1. Introduction
   1. Background

Though less visibly destructive than floods, hurricanes, and other hazards, prolonged periods of extreme temperatures are the leading cause of weather-related deaths in the United States (Berko et al. 2014). Globally, extreme ambient temperature (either too hot or too cold) contributes to 6.5-10% of all deaths (Sera et al. 2019; Zhao et al. 2021). Nearly 90% of global deaths attributed to temperature are cold-related, which is consistent with findings in the United States (Berko et al. 2014). The public health impact of extreme temperatures is undercounted because it aggravates several underlying conditions, but may not be noted as a contributing cause on hospital records or death certificates (Ostro et al. 2009; Lane 2018). At least 17 chronic conditions such as heart disease, diabetes, kidney disease, and respiratory infection show a J-shaped relationship with temperature (Burkart et al. 2021), meaning that the disease prevalence increases in both extreme high and low ambient temperatures. In addition to these serious health impacts, extreme temperature exposure impairs labor productivity (Lai et al. 2023), sleep quality (Obradovich et al. 2017), and cognitive performance (Laurent et al. 2018).

Several macro trends are pushing policy makers to prioritize emergency planning and disaster mitigation with regards to extreme temperatures. The first is anthropogenic climate change, which is increasing frequency and intensity of extreme weather (IPCC 2021). Second, individuals may be exposed to hazardous temperatures during power outages, as was seen across the Northeastern United States following Hurricane Sandy in 2012 (Henry and Ramirez-Marquez 2016) and in Texas in February 2021 (King, Rhodes, and Zarnikau 2021). Major electrical grid failures have increased by more than 60% in recent years (Stone et al. 2021). Finally, aging global populations mean more individuals will be susceptible to extreme temperature stress. In 2019, less than a tenth of the global population was over the age of 65 and by 2050 this number will increase to 1 in 6 (United Nations 2020). Age is a well-documented risk factor for temperature-related illness and death (Oudin Åström, Bertil, and Joacim 2011), so an older population has greater vulnerability.

* 1. Heat Vulnerability Indices enable planning for extreme temperatures

Most cities face substantial variation in intra-city vulnerability to extreme temperatures. The discourse in public agencies and academic literature around thermal vulnerability focuses on extreme heat, even though the mortality rate from extreme cold is significantly higher than that of extreme heat (Berko et al. 2014). In principle, many of the socioeconomic vulnerabilities contributing to heat-related illness and death also apply to extreme cold. Many cities and other jurisdictions use heat vulnerability indices (HVI) to better allocate resources on emergency response to heat, such as the location of cooling centers (Reid et al. 2009; Rinner et al. 2010; Uejio et al. 2011; Nayak et al. 2018). An HVI is typically a weighted sum of variables related to heat vulnerability such as income or poverty level, age, social isolation, and land cover characteristics calculated at a localized level like census tract.

HVI seldom include variables related to the local building stock. However, building characteristics, including level of insulation, presence of HVAC, and air tightness, can exacerbate or mitigate occupant exposure to heat. Indoor exposure, particularly at home, accounts for a sizable portion, 38-85%, of heat-related deaths (Fouillet et al. 2006; Wheeler et al. 2013; Iverson et al. 2020). An individual will on average spend around 67% of their time in a residence (Klepeis et al. 2001). This proportion is even higher for vulnerable populations such as infants and the elderly, who on average spend 89% and 78% of their time in a residence respectively (Matz et al. 2014).

A review by Samuelson et al. (Samuelson et al. 2020), found that out of 20 HVI’s from different cities and regions, eight included the year of construction, nine included central air-conditioning (AC) ownership, three included floor of residence, and one included rooftop albedo and thermal mass. Many building characteristics that could potentially impact indoor heat exposure such as orientation, envelope properties, and construction type, such as detached single-family or high-rise multifamily, are not included. City-level tax assessor data typically records year of construction and the presence of central AC at the parcel level, so these variables are attractive proxies for the contribution of the built environment. However, there is no obvious link between these variables and heat vulnerability.

HVI for New York State (Nayak et al. 2018) and the cities of Toronto (Rinner et al. 2010) and Philadelphia (Uejio et al. 2011) all considered older homes to have a higher risk for heat exposure due to a presumption of lack of insulation, lower likelihood of AC, and correlation with other risk factors like poverty. However, several studies monitoring indoor temperatures in European residences found that older buildings had, in summer, significantly cooler temperatures than newer ones (Beizaee, Lomas, and Firth 2013; Maivel, Kurnitski, and Kalamees 2015; Pathan et al. 2017), perhaps because of the thermal mass properties of stone construction typical of older European homes. Temperature monitoring in American residential buildings did not find a strong correlation between construction age and measured indoor temperature during times the home was actively heated or cooled (Booten et al. 2017). A simulation-based comparison of representative housing models in Boston found older typologies had lower maximum indoor temperatures (Samuelson et al. 2020). These studies demonstrate that construction age alone cannot capture indoor heat exposure.

HVI also typically consider AC prevalence, particularly that of central AC systems. In homes where AC is present, the cost of operating and maintaining AC systems may prohibit their use in a way that sufficiently protects residents from the adverse effects of heat. In recent investigations of indoor heat deaths, the Maricopa County Department of Public Health (MCDPH) found that in 91% of cases, AC was present (MCDPH 2019), but the AC was either broken (87%), disconnected from electricity (5%), or functioning but not turned on (8%). Clearly, the sheer presence of AC alone is not a protective factor against overheating.

The primary barriers to including additional building level characteristics is data availability at sufficient scale and awareness of their importance. However, new methods of data acquisition are rapidly becoming available such as smart thermostat data (Ecobee 2021) and satellite and street-level imagery (New et al. 2020). Understanding the role of the building and other household characteristics will enable public agencies to target emergency planning efforts and resources in the short-term like locations of cooling and warming centers and the long-term such as social and building weatherization programs.

* 1. Research gaps and objectives

Several research gaps relate to the role of building characteristics on temperature-related illness and death.

First, there is a lack of empirical evidence that examines the link between building characteristics and predicting temperature-related illness and death. Studies assessing the sensitivity of overheating risk to building characteristics often use building performance simulations. These studies use simulation outputs such as maximum daily room temperature (Mavrogianni et al. 2012; Samuelson et al. 2020), percent of time in different U.S. Occupational Safety and Health Administration (OSHA) heat index (HI) risk categories (Sun, Specian, and Hong 2020), or the degree-hours the wet-bulb globe temperature (WBGT) index exceeded a threshold value (Baniassadi, Heusinger, and Sailor 2018) to model the risk of overheating.. While there are many thermal indices, as yet none of them are validated for personal exposure indoors, meaning the recommended thresholds are not based on empirical observations of temperature-related health hazards in this context (Kuras et al. 2017; Kenny et al. 2019).

The second research gap is the limited understanding of how HVI models perform when built using building characteristics in conjunction with other markers of socioeconomic vulnerability such as income and age. Very few studies using building performance simulations review the interaction of building and occupant characteristics. Baniassadi et al. (Baniassadi et al. 2019) account for some effects of occupant income by modeling AC non-functionality and occupant age by using a conservative value for their overheating threshold.

To overcome these research gaps, this study trains and evaluates models that predict temperature-related illness based on a nationwide survey of building and household characteristics in American homes. This study revolves around two research questions:

1. Would a HVI with detailed information about the building be more accurate for predicting the risk of health hazards? If so, by how much?
2. Which building and occupant characteristics contribute the most to predicting the risk of health hazards?

To answer these questions, we leverage state of the art machine learning models and a train-validate-test pipeline to identify the best performing models and their hyperparameters. Note that our focus is on what data are most valuable for temperature-related illness *prediction*, rather than identifying causal relationships between variables and health outcomes.

More accurate predictions will allow public agencies to better identify at-risk households and strategize limited resources for short-term planning like locations of cooling and warming centers and long-term planning like building weatherization and social programs. Understanding the contributions of building and occupant characteristics can prioritize data collection efforts.

1. Materials and methods
   1. Residential energy consumption survey (RECS) data

The main source of data for this study is the Residential Energy Consumption Survey (RECS), which is administered by the U.S. Energy Information Administration (EIA) (EIA 2018; 2022). RECS is a periodic survey that has collected detailed energy characteristics, usage patterns, and demographics of American households since 1978. The primary objective of RECS is to estimate future energy demand and improve energy efficiency and building design.

Of relevance for this study, the three most recent cycles of RECS, 2009, 2015, 2020, ask respondents whether “in the last year, did anyone in your household needed medical attention because the home was too hot?” or “too cold?” (EIA 2016; 2020). This study treats an affirmative response to either question as a temperature-related illness. While the questions are self-reported and do not specify duration and severity of extreme temperatures and who in the household needed medical assistance, it provides a source of ground truth that the household experienced a hazardous interior thermal environment. We focus on the two most recent RECS surveys, namely the 2015 and 2020 RECS. Responses to our questions of interest are not available in the public data file for the 2009 RECS due to infrequent responses risking disclosure of sensitive and confidential household information.

Each RECS is an independent cross-sectional study of residential energy use, so each iteration of the survey is slightly different. Theoretically, it’s possible to select the same home twice, but it is highly unlikely and occurs rarely. The EIA selects samples to statistically represent all U.S. households occupied as a primary residence at the time of the survey. The most significant difference between the 2015 and 2020 survey cycles is the mode of execution. The 2015 survey cycle collected data through a combination of computer-assisted personal interviews, internet, and mailings. The 2020 survey cycle relied entirely on self-administered web and paper questionaries. Because there were no in-person interviews, the 2020 survey did not use a clustered sampling method like in 2015. The impact of this change is a three-fold increase in sample size – from 5,686 in 2015 to 18,496 in 2020. Sample size is inversely proportional to the standard error and, so larger samples generally result in narrower confidence intervals for population and subpopulation estimates. The two survey years also had minor differences in variable coding.

Table 1 shows counts of heat-, cold-, or any temperature-related illness in the 2015 and 2020 RECS. For the predictive model, we treat each sample as an independent observation. However, for population estimate, we reviewed the results of each year separately due to differences in sampling methods. For each year, RECS calculates the sample weight, which indicates the number of households in the population that observation represents. Inclusion of replicate weights allows for calculation of the sampling error. We followed the EIA’s procedure for calculating population estimates, standard errors, and confidence intervals in R programming language (EIA 2019; 2023).

Table . Observations of temperature-related illness in RECS

|  |  |  |  |
| --- | --- | --- | --- |
| **Temperature-related illness** | **2015** | **2020** | **Total** |
| Heat-related | 39 | 76 | 115 |
| Cold-related | 54 | 120 | 174 |
| Any temperature | 81 | 171 | 252 |
| None | 5,605 | 18,496 | 24,101 |

To explore patterns in households that experienced temperature-related illness we narrowed the over 750 household characteristics described in the RECS dataset to approximately 25 related to either vulnerability or exposure to extreme temperature. These variables fall under 3 categories: climate, demographics, and buildings. We describe these building and household characteristics in the subsequent sections. Table 2 provides an overall summary of all input variables.

By default, the RECS dataset encodes all variables as numerical quantities. We retained the numerical values for truly numerical household characteristics like the construction age. We transformed the variable into categorical for ordinal data such as the level of insulation, or binary variables such as presence of back-up generator or on-site solar. We merged categorical variables based on the literature on heat and cold-related vulnerability, for example non-white versus white instead of individual racial and ethnic categories. We also derived new variables of interest such as poverty, which combines the number of household members with income level and thermal mass, which combines insulation level and exterior wall or roof material.

Climate

Ground surface temperature is a climatic variable often reported in HVI (Uejio et al. 2011), because it represents local exposure to extreme temperatures. Due to the EIA’s objective to forecast energy demand, climatic variables in RECS are oriented towards HVAC system operation, such as cooling and heating design-temperatures.

Demographics

Epidemiological studies have investigated the relationship between different demographic and socioeconomic variables on heat-related mortality. Elderly age is a vulnerability factor, but there is some ambiguity around the cut-off for higher risk: 60, 65, 70, or 75 (Applegate et al. 1981; O’Neill, Zanobetti, and Schwartz 2003; Ballester et al. 1997; Centers for Disease Control and Prevention (CDC) 1995; Conti et al. 2005). The elderly may be more likely to have co-morbidities and take medication that affect thermal perception and regulation. They also may have limited mobility to access cooling centers or limit AC usage due to fixed income. Economic conditions and heat-related mortality are related. The economic situation was measured by poverty (Naughton et al. 2002; Curriero et al. 2002), unemployment (Nayak et al. 2018), renter status (Uejio et al. 2011; Wright et al. 2020), and utility payment (Wright et al. 2020). Klinenberg’s sociological analysis of the 1995 Chicago heat wave found a higher risk of death in individuals with limited social connections such as those living alone (Klinenberg 2015). These individuals may be at higher risk of not being checked on regularly during a heat emergency and they may have less help in coping with heat. On the other hand large households (7+ members) may also have elevated heat mortality risk (Uejio et al. 2011). The impact of race and ethnicity on heat-related mortality is mixed, with some studies finding a higher risk for African Americans or non-white racial and ethnic groups (O’Neill, Zanobetti, and Schwartz 2005; Schwartz 2005).

Table . Summary of household characteristics derived from the RECS dataset relevant to the household’s vulnerability or exposure to extreme temperature.

| **Category** | **Variable** | **Variable description** | **Type a** |
| --- | --- | --- | --- |
| Climate | Cooling design temperature | Dry bulb design temperature (°F) expected to be exceeded 1% of the time | N |
| Heating design temperature | Dry bulb design temperature (°F) expected to be exceeded 99% of the time | N |
| Demographic | Non-white | Householder (respondent) race is non-white, or ethnicity is of Spanish descent. | B |
| Older than 65 | Respondent or household member age is > 65 | B |
| Lives alone | Number of household members = 1 | B |
| Large households (7+ members) | Number of household members > 7 | B |
| Poverty | Calculated from gross income and number of household members based on U.S. Census Bureau definition for poverty threshold for that year | B |
| Unemployed | Respondent is unemployed or retired | B |
| Education level | Highest education attained is high school or equivalent |  |
| Renting | Household pays rent | B |
| Pays utility or for other fuel | Household pays for electricity, natural gas, propane, and/or fuel oil | B |
| Buildings: construction | Construction age | Estimated year when housing unit was built (taken as the maximum of the range in RECS response coding) | N |
| Apartment | Type of housing unit is low-rise or high-rise apartment | B |
|  | Mobile | Type of housing unit is mobile home | B |
| Buildings: envelope | Thermally massive wall | Estimated thermal mass based on exterior wall material and presence of insulation | N |
| Thermally massive roof | Estimated thermal mass based on exterior roof material and presence of insulation | N |
| Insulation | Level of insulation | N |
| Infiltration | Frequency of draft | N |
| Window per room | Number of windows per room as an approximation for window-to-wall ratio | N |
| Glazing type | Type of glass in most windows | N |
| Buildings: HVAC | AC type | Air conditioning equipment used | N |
| Heating type | Space heating equipment used | N |
| Energy insecurity | Household reported difficulty paying energy bills; repairing or replacing broken heating or cooling equipment; or that they had kept their home at unsafe temperatures because of cost concerns | B |
| Fans | Number of ceiling, floor, window, and/or table fans used | N |
| Off-grid | Home has back-up generator or on-site solar electricity generation | B |

a Type includes numerical (N) and binary (B)

Buildings: construction

Building construction represents variables related to the building age and form. As mentioned in Section 1.2, several city and state-level HVI use construction age as a catch-all for other building characteristics that affect the indoor thermal environment (Rinner et al. 2010; Uejio et al. 2011; Nayak et al. 2018). A building performance simulation study of London dwellings found a significant impact of archetype, a combination of construction age and construction type on overheating risk (Mavrogianni et al. 2012). Samuelson et al. (Samuelson et al. 2020) suggests that detached buildings may be less vulnerable due to a greater potential for exposed walls to exchange heat and more opportunities for cross-ventilation. Similarly, Lomas (Lomas 2021) singles out flats or apartments because of more limited opportunities for natural ventilation. Mobile or manufactured homes may also increase heat or cold exposure due to poor energy efficiency (Harrison and Popke 2011), an issue common in even newer mobile homes (Hart et al. 2002).

Buildings: envelope

The building envelope represents the materials that separate the interior and exterior of the building. Building performance simulations find wall insulation reduces the overheating risk when applied to the exterior, but may increase overheating risk when applied to the interior (Mavrogianni et al. 2012; Porritt et al. 2012). Porritt et al. (Porritt et al. 2012) also found a correlation between roof and wall surface reflectivity i.e., inverse of solar absorptivity and overheating risk. Samuelson et al. (Samuelson et al. 2020) infers that other building envelope characteristics like infiltration, and window-to-wall ratio may also be significant. One HVI considered houses with thermally massive materials to have greater adaptive capacity (Inostroza, Palme, and de la Barrera 2016). Thermal mass describes building materials with high heat capacity, such as brick, stone, and concrete, which can buffer temperature fluctuations. For this strategy to work, the material must be exposed to both the interior and exterior i.e. limited insulation.

Buildings: HVAC

Building HVAC characteristics describe the presence (Curriero et al. 2002), type (O’Neill, Zanobetti, and Schwartz 2005), and functionality (Naughton et al. 2002; MCDPH 2019) of HVAC systems. Fan are cost-effective and energy efficient solutions to keep people comfortable indoor by increasing evaporation and convective heat losses (Jay et al. 2015). Finally, we also consider availability of alternate power sources such as a back-up generator or on-site solar panels as they may reduce interruptions to HVAC systems.

* 1. Machine learning

We used machine learning to predict the reporting of a temperature-related illness event, which we treated as a binary classification problem since the RECS survey is coded as a yes or no response. The input features for the machine learning model are described in Table 2. We focus on comparing the performance of models trained with and without building characteristics.

We know from Table 1 that there is an extreme imbalance in the RECS data – less than 1% of all households experienced temperature-related illness. This imbalance is problematic because a naïve model that always predicts the majority class, i.e. no temperature-related illness, will have a high accuracy, 99% in this case, but will fail to predict any observations in the minority class, i.e. occurrence of temperature-related illness. Imbalanced data is a common issue in other domains such as disease diagnosis, customer churn prediction, and fraud detection. As in our case, imbalanced data problems generally have a high cost associated with failure to predict the minority class. We employ several techniques in the machine learning model building process to address the imbalanced data (He and Garcia 2009; Kaur, Pannu, and Malhi 2019; Krawczyk 2016), we describe them below.

We first pre-processed the data set to remove variables with zero or near-zero variance. These variables can negatively impact model performance as they may become zero variance after the data is subdivided. This step removed the variable for large households and pays utility and/or fuel. We also checked for highly correlated variables (magnitude of Spearman’s correlation coefficient > 0.75) and linear combinations, but no variable met the threshold for removal. We then standardized input variables to have zero mean and unit variance. This step prevents variables with larger numerical quantities from having undue influence, particularly in regression-based modeling methods.

We then split the RECS dataset into training and test data, using 80% for training and holding 20% for testing, which prevents overfitting. We bootstrapped this process with 30 iterations to quantify the uncertainty in model performance due to the training data split. For each training and test split, we then used 5-fold cross validation repeated 5 times to further split the training data into training and validation sets for selecting machine learning model hyperparameters.

We compared performance from several machine learning algorithms, listed in Table 3. These algorithms vary in their underlying structure and assumptions about input features. We selected these algorithms because of their ability to accept class weights and availability in the R *caret* package. We applied an exhaustive grid search of 100 values to find the best performing hyperparameter settings for each machine learning algorithm.

Table . Summary of machine learning algorithms

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Hyperparameters (min, max)** | **R implementation** |
| Generalized linear model | None | glm (Kuhn et al. 2023) |
| Multivariate Adaptive Regression Spline | #Terms: (1, 100)  Product degree: 1 | earth (Milborrow 2023) |
| Penalized discriminant analysis | Shrinkage penalty coefficient: (0, 0.1) | pda (Hastie and Tibshirani 2023) |
| Penalized multinomial regression | Weight decay = (0, 0.1) | Multinom (Ripley and Venables 2023) |
| Bagged classification and regression tree | None | treebag (Peters et al. 2023; Wickham 2023; Meyer et al. 2023) |
| Stochastic gradient boosting | # Boosting iterations: (50, 500)  Max. tree depth: 1  Shrinkage: (510-3, 510-2)  Min. terminal node size: 10 | gbm (Greenwell et al. 2022; Wickham 2023) |
| Random forest | # Randomly selected predictors: (1, # of variables)  Splitting rule: Gini impurity, extremely randomized  Min. node size: (1, 5) | ranger (Meyer et al. 2023; Greenwell et al. 2022; Wickham et al. 2023) |
| Single layer neural network | # Hidden units: (1, # of variables)  Weight decay: (10-7, 10-1) | nnet (Ripley and Venables 2023) |

We employed the following strategies to address the inherent class imbalance in the RECS data set: 1) stratified sampling; 2) fewer cross-validation folds; 3) class weights; 4) sub-sampling; and 5) appropriate performance metrics. Stratified sampling means that any time we created divisions in the data set such as splitting the training and test data or subdividing the training data into cross-validation folds, we partitioned the data based on occurrence of temperature-related illness. This way each subset maintained the same proportion of the dependent variable as the original dataset. We also set 5 folds versus the common practice of 10 folds for cross-validation. This allowed us to hold more observations of temperature-related illness for the validation set when tuning hyperparameters. Thirdly we test the effect of class weights on model performance. Class weights impose a heavier cost on errors in the minority class. Fourth, we considered the effect of several sub-sampling techniques during cross-validation. Up-sampling randomly replicates instances of the minority class. We also tried two hybrid methods, the synthetic minority oversampling technique (SMOTE) and random oversampling examples (ROSE), which down-sample the majority class and synthesize new data points in the minority class. SMOTE draws artificial samples by choosing points on the line connecting minority class observations to its nearest neighbors in the feature space (Fernandez et al. 2018). ROSE uses smoothed bootstrapping to draw artificial samples from the feature space neighborhood around the minority class (Menardi and Torelli 2014).

Finally, we considered the class imbalance in our choice of performance metric. As illustrated earlier, the model’s overall accuracy (ratio of correct classifications to total observations) can be biased for heavily imbalanced classes. The routine choice for binary classification problems is the Receiver Operating Characteristic (ROC) curve. To understand this metric, we define a positive and negative class – the two outcomes of the predictive model. In our imbalanced data set, the positive class is the minority class and the negative class is the majority class. The ROC curve plots the true positive rate, also called the sensitivity (True Positives / (True Positives + False Negatives)), versus the false positive rate, (False Positives / (False Positives + True Negatives)) or 1 – specificity, (specificity is also known as the true negative rate), with different discrimination thresholds. The area under the receiver operator curve summarizes the ROC curve into a single metric that represents the prediction accuracy of the model. This metric can be misleading for imbalanced data because the false positive rate becomes very small when the number of negatives is very large. (Davis and Goadrich 2006; Fawcett 2006). The Precision-Recall (PR) curve, on the other hand, plots the precision, defined as the number of correct positive predictions divided by the total number of positive predictions (True Positives / (True Positives + False Positives)) by the recall, which quantifies that number of correct positive predictions by total number of positives (True Positives / (True Positives + False Negatives)) i.e. the same as the true positive rate in the ROC curve.

The PR curve is better suited for imbalanced data sets training because it is not concerned with negative class predictions i.e. the majority class. As with the ROC curve, the area under the PR curve summarizes the curve into a single metric, which we use to select the best hyperparameter values during cross-validation. In the test set, we will evaluate the model along three performance metrics: 1) balanced accuracy, 2) recall, and 3) precision. Balanced accuracy is defined as the average accuracy on either class, or, in other words the arithmetic mean of the sensitivity and specificity. For a naïve model that always predicts the majority class the sensitivity is 0, the specificity is 1, and so the balanced accuracy is 0.5. This serves as benchmark as the minimum performance value. Recall and precision are of interest because of the high-cost of not only temperature-related health hazards but also preventive measures.

For statistical analysis, we will use a paired t-test by bootstrap iteration i.e. the same training and test data split to compare models trained with different groups of input features. For results with statistical significance, p < 0.05, we will use Cohen’s d to quantify the effect size. We will interpret Cohen’s d as follows: 0.4 < |d| < 1.15 for recommended minimum practical effect, 1.15 < |d| < 2.70 for moderate effect, and |d| > 2.70 for strong effect (Ferguson 2009).

We used the statistical software R (R Core Team 2022) and its associated integrated development environment RStudio (RStudio Team 2021) to build and analyze all machine learning models. In particular, we used the *tidyverse* package (Wickham and RStudio 2023) for reading, manipulating, and visualizing data and the *caret* package (Kuhn et al. 2023) as a wrapper to conduct data pre-processing, resampling, and cross-validation as well as interface with the different machine learning algorithms.

1. Results
   1. Prevalence of temperature-related illness in population

We estimated the prevalence of temperature-related illness in U.S. households using sample weights provided by IEA. Figure 1 compares the inferred number of households affected by heat-related, cold-related, or any temperature-related illness in 2015 and 2020. Like the global and national trends discussed in Section 1.1, we find that cold-related hazards were more prevalent than heat-related ones. While overall the number of households with any temperature-related illness represents less than 1% of the total population, this still means that nearly 2 million households report needing medical attention for temperature related illness annually in the United States.

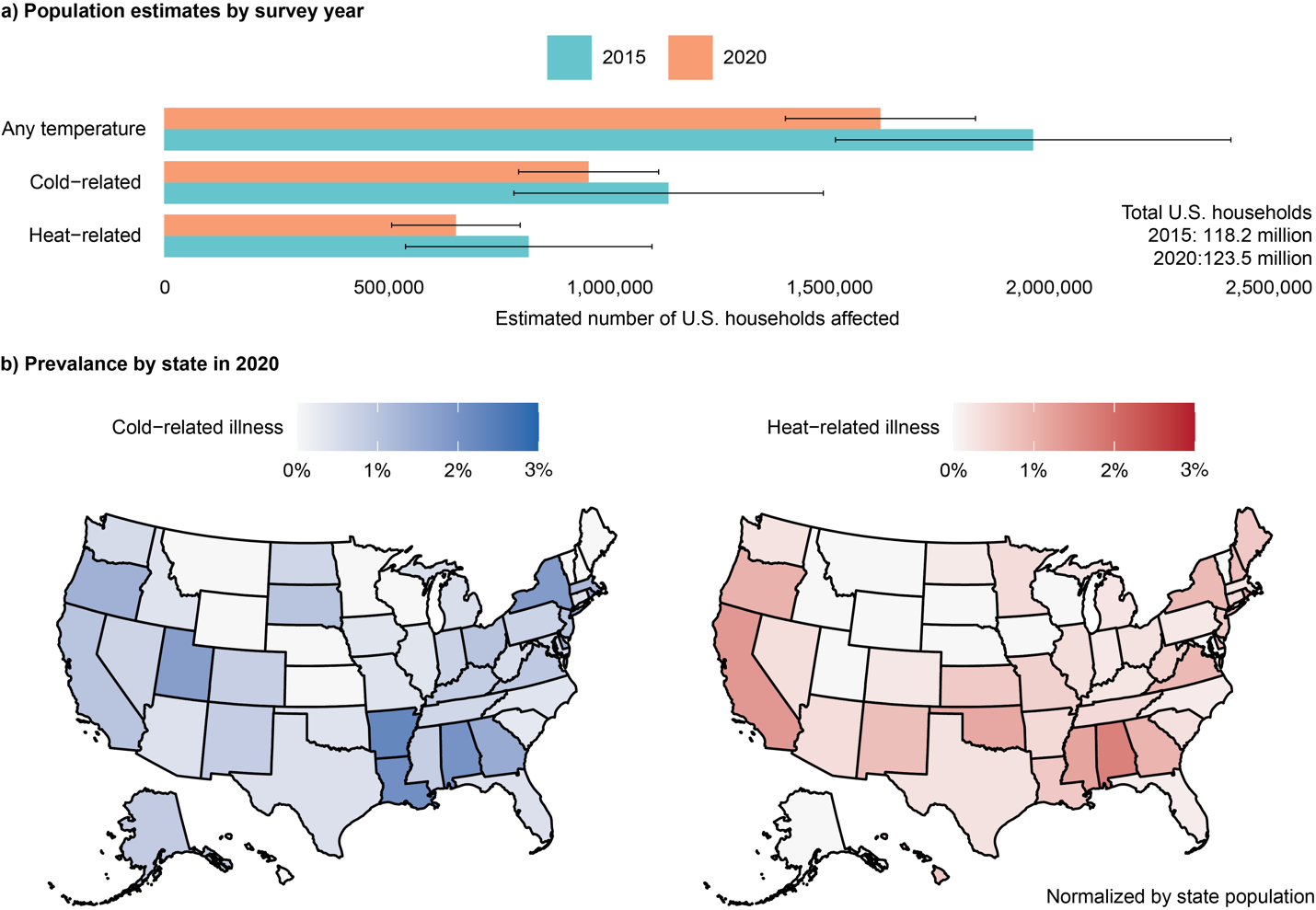


Figure : Prevalence of temperature-related illness in U.S. households based on the 2015 and 2020 Residential Energy Consumption Survey conducted by the U.S. Energy Information Administration. We calculated population estimates and standard errors from sample weights and replicate weights as recommended by the EIA (EIA 2019; 2023). Error bars represent the 95% confidence interval.

* 1. Predicting temperature-related illness

We constructed machine learning models to predict temperature-related illness. Figure 2a) shows the performance of all model iterations along three performance metrics: balanced accuracy, recall, and precision. Each bar represents machine learning models trained from the same set of input features, class imbalance scheme, and machine learning algorithm, a total of 80 models. The error bars represent the 95% confidence interval, which we calculated from 30 bootstrapped sample iterations, each with a different training and test data split. Generally, about half of the machine learning models performed better than a naïve model. Many of the poor performing models did not converge during model training. For well-performing models, the balanced accuracy and recall range from 70 to 85%. In comparison, the model precision is quite low, around 5%. This means that the models produce many false positives – households that we incorrectly predicted would have temperature-related illness.

Figure 2b) compares the best model performance from each input group. For the Climate + Demographics model the best machine learning algorithm was penalized discriminant analysis with SMOTE sub-sampling. For the “+ Buildings” model, the best machine learning algorithm was stochastic gradient boosting with up-sampling. We find a 13% increase in balanced accuracy, 15% increase in recall, and 3% increase in precision. These results are statistically significant with a p < 0.001 and have a moderate to strong effect size.

Figure 2c) compares the magnitude of variable coefficients for the best regression model from each input group. Regression models allow for clearer interpretability of variable contributions, so even though this is not the best performing model for either input features group, its performance is within the 95% confidence interval. For both input feature groups, the best regression model happens to be penalized multinomial regression (Nibbering and Hastie 2022). This model type performs regularization, i.e. aims to reduce the number of input features by forcing coefficients of insignificant variables towards 0. For the Climate + Demographics input features group the best multinomial model used SMOTE sub-sampling and that of the best + Buildings input features group used ROSE sub-sampling. Also here the error bars represent the 95% confidence interval, which we calculated from 30 bootstrapped sample iterations, each with a different training and test data split. We greyed out points where the 95% confidence interval included 0. Variable names that are crossed out were eliminated during pre-processing due to zero or near-zero variance. We find that in the Climate + Demographics model, the variables with the largest magnitude are (in decreasing order): poverty, non-white, renting, employment, and older than 65, and lives alone. For the + Buildings model, the variables with the largest magnitude are (in decreasing order): energy insecurity, non-white, poverty, infiltration, and renting, AC type, employment, and thermally massive roof. What is most striking is the high magnitude of the coefficient for energy insecurity, in comparison to that of any other variable. When comparing the input groups, we see that the model selects almost the same demographics variables, however the magnitude of the coefficient is higher for the same variable in the Climate + Demographics model.

* 1. Machine learning modeling

Finally, we review the impact of two key machine learning modeling parameters: the algorithm and the class imbalance handling scheme along two performance metrics: balanced accuracy as a proxy for model performance and computation time. Figure 3 plots the interaction between the two machine learning modeling parameters and both performance metrics.

A screenshot of a graph

Description automatically generated

Figure : a) Overall machine learning model performance across all the 80 iterations along three metrics: balanced accuracy, recall, and precision. Each bar represents a machine learning model trained with the same input features, class imbalance handling scheme, and algorithm. The error bars represent the 95% confidence interval calculated from 30 bootstrapped samples, each with a different training and test data split. b) Shows the performance for the best machine learning model from each input features group. We calculated statistical significance using a paired t-test by bootstrap iteration i.e. the same training and test data split, and the effect size from Cohen’s d. We interpreted Cohen’s d as follows: 0.4 < |d| < 1.15 for recommended minimum practical effect, 1.15 < |d| < 2.70 for moderate effect, and |d| > 2.70 for strong effect (Ferguson 2009). c) Shows the variable contribution from the best regression

A chart of different colors

Description automatically generated with medium confidence

Figure : Interaction between two machine learning modeling parameters: the algorithm (on y-axis) and the class imbalance handling scheme (on x-axis). Results are shown for one input features group (Climate + Demographics + Buildings), but the trends are applicable to all groups. We binned the computation time for clarity. Computation time is the time it takes for model training included cross-validation and is binned as follows for clarity: fast (< 30 s), medium (30 s to 2 min) and slow (> 2 min).

The color of the point represents balanced accuracy, and the shape of the point represents the computation time binned into three groups, fast (< 20 s), medium (20 s to 1.5 min) and slow (> 1.5 min). From the plot, we can see that across all machine learning algorithms, no class imbalance handling scheme and class-weights had balanced accuracies near 50%, like a naïve model. We can also see that two machine learning algorithms, classification and regression trees and random forest had balanced accuracies near 50%, like a naïve model, across all class imbalance handling schemes except ROSE. Some machine learning algorithms like multivariate adaptive regression spline and random forest had consistently slow computation times across all class imbalance handling schemes. In some machine learning algorithms like penalized discriminant analysis and neural network, we find that up-sampling and SMOTE increase computation time relative to the other class imbalance handling schemes. Generalized linear models have the fastest computation times and models with up-sampling and SMOTE have balanced accuracies near the maximum of 85%. Generalized linear models with ROSE have balanced accuracy significantly worse than a naïve model.

1. Discussion
   1. Prevalence of temperature-related illness in population

The population estimates from RECS provide new information about self-reported prevalence of heat, cold, and any-temperature related illness in the United States. Although there is some U.S. national data on heat and heat related illness, namely the [Center for Disease Control’s Heat & Health Tracker](https://ephtracking.cdc.gov/Applications/heatTracker/), these sources often rely on data from hospital records or emergency room visits, which have been criticized for their ability to properly count temperature-related issues (XXX). While the RECS survey questions related to temperature-related illness are limited, it does explicitly indicate indoor extreme temperature exposure at home but it is based on self-reported assessment. RECS excludes vacant, seasonal or vacation homes, and group quarters such as prisons, military barracks, dormitories, and nursing homes. The exclusion of nursing homes is particularly relevant because they generally house a population with higher vulnerability. To our knowledge there are also no national statistics tracking cold-related illness or death, which our results and others show are a higher proportion of temperature-related health hazards.

* 1. Predicting temperature-related illness

Our results demonstrate the feasibility of machine learning modeling to predict temperature-related illness. Top-performing models can correctly categorize up to 85% of households (based on balanced accuracy) and identify up to 85% of households that reported temperature related illness (based on recall). However, these models generally have poor precision, around 5%, meaning that we assign more false positive classifications (classifying safe building as building with temperature related-illness). Given that temperature-related illness is generally underreported (xxx) providing public health interventions to these households may still be worthwhile. Moreover, from a policy perspective it is more important to have a high recall so people that are in need are found. Having a low precision leads to a more expensive, and therefore less cost-effective, policy intervention.

Our analysis of variable coefficients in the top-performing regression models indicates the importance of reported energy insecurity among all the considered variables. This finding is in line with investigations of indoor heat deaths in Maricopa County, Arizona (MCDPH 2019) and confirms that the simple presence of HVAC systems is not sufficient to protect against temperature-related health hazards. The HVAC must work and the building occupant should be able to afford to run it. Inclusion of energy insecurity and a handful of other building characteristics like infiltration rate resulted in significant performance improvement compared to a model with only demographic variables. This indicates the interaction of vulnerability (mostly captured in demographic variables) and exposure (mostly captured in detailed building variables) in the risk of temperature-related health hazards. While demographic information about households is generally available, our findings suggest value in collecting detailed data on energy insecurity. Our results can also be used to design more effective interventions to combat extreme-temperature related health hazards. Interventions that only provide HVAC units, such as a $10 million program launched by the Canadian province of British Columbia to install 8,000 portable AC units in vulnerable households over the next 3 years (Ministry of Health 2023), could be limited in their effectiveness if they also do not provide a plan for system maintenance (e.g., fix or replace when broken) and support for its operation (e.g. financial help to pay utility bills).

While RECS uniquely provides detailed information about the household’s demographic and building characteristics, survey responses are ultimately self-reported by a resident of the household. This means that building attributes that are not easy to see, such as insulation level and infiltration, may not be accurately reported. Further research validating RECS survey responses with on-site investigation or documentation can help clarify this gap.

* 1. Machine learning modeling

Imbalanced data sets are prevalent in many domains, so insight from the modeling parameters in this study may be useful in other contexts. Our results show an interaction between the machine learning algorithm and class imbalance handling scheme that cannot be predicted a priori. We find that, just as it is important to test a variety of machine learning algorithms, it is also important to compare multiple class imbalance handling schemes. These two modeling parameters have a computational cost associated with them, but that may not necessarily correlate to better model performance.

# Conclusions

This study finds that temperature-related illness affects at least 2 million households in the United States annually. We identified households who experienced temperature-related illness with 85% accuracy, but this requires detailed information about building characteristics, namely energy insecurity as it relates to the household’s ability to maintain and operate HVAC systems, which can safeguard against extreme temperature exposure. This finding is significant because it gives municipalities a pathway towards better data collection to identify at-risk households and better public health programming aimed at preventing in-home extreme temperature health hazards.

# Reference

* 1. Applegate, William B., John W. Runyan, Linda Brasfield, Mary Lynn Williams, Charles Konigsberg, and Cheryl Fouche. 1981. “Analysis of the 1980 Heat Wave in Memphis\*.” *Journal of the American Geriatrics Society* 29 (8): 337–42. https://doi.org/10.1111/j.1532-5415.1981.tb01238.x.
  2. Ballester, F, D Corella, S Pérez-Hoyos, M Sáez, and A Hervás. 1997. “Mortality as a Function of Temperature. A Study in Valencia, Spain, 1991-1993.” *International Journal of Epidemiology* 26 (3): 551–61. https://doi.org/10.1093/ije/26.3.551.
  3. Baniassadi, Amir, Jannik Heusinger, and David J. Sailor. 2018. “Energy Efficiency vs Resiliency to Extreme Heat and Power Outages: The Role of Evolving Building Energy Codes.” *Building and Environment* 139 (July): 86–94. https://doi.org/10.1016/j.buildenv.2018.05.024.
  4. Baniassadi, Amir, David J. Sailor, Cassandra R. O’Lenick, Olga V. Wilhelmi, Peter J. Crank, Mikhail V. Chester, and Agami T. Reddy. 2019. “Effectiveness of Mechanical Air Conditioning as a Protective Factor Against Indoor Exposure to Heat Among the Elderly.” *ASME Journal of Engineering for Sustainable Buildings and Cities* 1 (1). https://doi.org/10.1115/1.4045678.
  5. Beizaee, A., K. J. Lomas, and S. K. Firth. 2013. “National Survey of Summertime Temperatures and Overheating Risk in English Homes.” *Building and Environment* 65 (July): 1–17. https://doi.org/10.1016/j.buildenv.2013.03.011.
  6. Berko, Jeffrey, Deborah D. Ingram, Shubhayu Saha, and Jennifer D. Parker. 2014. “Deaths Attributed to Heat, Cold, and Other Weather Events in the United States, 2006-2010.” *National Health Statistics Reports*, no. 76 (July): 1–15.
  7. Booten, Chuck, Joseph Robertson, Dane Christensen, Mike Heaney, David Brown, Paul Norton, and Chris Smith. 2017. “Residential Indoor Temperature Study.” NREL/TP--5500-68019, 1351449. https://doi.org/10.2172/1351449.
  8. Burkart, Katrin G., Michael Brauer, Aleksandr Y. Aravkin, William W. Godwin, Simon I. Hay, Jiawei He, Vincent C. Iannucci, et al. 2021. “Estimating the Cause-Specific Relative Risks of Non-Optimal Temperature on Daily Mortality: A Two-Part Modelling Approach Applied to the Global Burden of Disease Study.” *The Lancet* 398 (10301): 685–97. https://doi.org/10.1016/S0140-6736(21)01700-1.
  9. Centers for Disease Control and Prevention (CDC). 1995. “Heat-Related Mortality--Chicago, July 1995.” *MMWR. Morbidity and Mortality Weekly Report* 44 (31): 577–79.
  10. Conti, Susanna, Paola Meli, Giada Minelli, Renata Solimini, Virgilia Toccaceli, Monica Vichi, Carmen Beltrano, and Luigi Perini. 2005. “Epidemiologic Study of Mortality during the Summer 2003 Heat Wave in Italy.” *Environmental Research* 98 (3): 390–99. https://doi.org/10.1016/j.envres.2004.10.009.
  11. Curriero, Frank C., Karlyn S. Heiner, Jonathan M. Samet, Scott L. Zeger, Lisa Strug, and Jonathan A. Patz. 2002. “Temperature and Mortality in 11 Cities of the Eastern United States.” *American Journal of Epidemiology* 155 (1): 80–87. https://doi.org/10.1093/aje/155.1.80.
  12. Davis, Jesse, and Mark Goadrich. 2006. “The Relationship between Precision-Recall and ROC Curves.” In *Proceedings of the 23rd International Conference on Machine Learning - ICML ’06*, 233–40. Pittsburgh, Pennsylvania: ACM Press. https://doi.org/10.1145/1143844.1143874.
  13. Ecobee. 2021. “Donate Your Data Smart Wi-Fi Thermostats by Ecobee.” 2021. https://www.ecobee.com/donate-your-data/.
  14. EIA. 2016. “Residential Energy Consumption Survey: A Nationwide Study of Energy Use in American Homes.” Washington, DC: U.S. Department of Energy. https://www.eia.gov/survey/form/eia\_457/2015\_EIA-475A\_paper.pdf.
  15. ———. 2018. “Residential Energy Consumption Survey (RECS): 2015 Household Characteristics Technical Documentation Summary.” Washington, DC: U.S. Department of Energy.
  16. ———. 2019. “Residential Energy Consumption Survey (RECS): Using the 2015 Microdata File to Compute Estimates and Standard Errors (RSEs).” Washington, D.C.: U.S. Energy Information Administration. https://www.eia.gov/consumption/residential/data/2015/pdf/microdata\_v3.pdf.
  17. ———. 2020. “Residential Energy Consumption Survey (RECS) Form EIA-457A 2020 Household Questionnaire.” Washington, D.C.: U.S. Department of Energy. https://www.eia.gov/survey/form/eia\_457/2020\_RECS-457A.pdf.
  18. ———. 2022. “2020 RECS Survey Data.” May 2022. https://www.eia.gov/consumption/residential/data/2020/.
  19. ———. 2023. “2020 Residential Energy Consumption Survey: Using the Microdata File to Compute Estimates and Relative Standard Errors (RSEs).” Washington, D.C.: U.S. Energy Information Administration. https://www.eia.gov/consumption/residential/data/2020/pdf/microdata-guide.pdf.
  20. Fawcett, Tom. 2006. “An Introduction to ROC Analysis.” *Pattern Recognition Letters* 27 (8): 861–74. https://doi.org/10.1016/j.patrec.2005.10.010.
  21. Ferguson, Christopher J. 2009. “An Effect Size Primer: A Guide for Clinicians and Researchers.” *Professional Psychology: Research and Practice* 40 (5): 532–38. https://doi.org/10.1037/a0015808.
  22. Fernandez, Alberto, Salvador Garcia, Francisco Herrera, and Nitesh V. Chawla. 2018. “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary.” *Journal of Artificial Intelligence Research* 61 (April): 863–905. https://doi.org/10.1613/jair.1.11192.
  23. Fouillet, A., G. Rey, F. Laurent, G. Pavillon, S. Bellec, C. Guihenneuc-jouyaux, J. Clavel, E. Jougla, and Denis Hémon. 2006. “Excess Mortality Related to the August 2003 Heat Wave in France,” October, 16–24. https://doi.org/10.1007/s00420-006-0089-4.
  24. Greenwell, Brandon, Bradley Boehmke, Jay Cunningham, and G. B. M. Developers (https://github.com/gbm-developers). 2022. “Gbm: Generalized Boosted Regression Models.” https://cran.r-project.org/web/packages/gbm/index.html.
  25. Harrison, Conor, and Jeff Popke. 2011. “‘Because You Got to Have Heat’: The Networked Assemblage of Energy Poverty in Eastern North Carolina.” *Annals of the Association of American Geographers* 101 (4): 949–61. https://www.jstor.org/stable/27980241.
  26. Hart, John Fraser, Michelle J. Rhodes, John T. Morgan, and John T. Morgan. 2002. *The Unknown World of the Mobile Home*. Baltimore, UNITED STATES: Johns Hopkins University Press. http://ebookcentral.proquest.com/lib/berkeley-ebooks/detail.action?docID=3318195.
  27. Hastie, Trevor, and Robert Tibshirani. 2023. “Mda: Mixture and Flexible Discriminant Analysis.” https://cran.r-project.org/web/packages/mda/index.html.
  28. He, Haibo, and Edwardo A. Garcia. 2009. “Learning from Imbalanced Data.” *IEEE Transactions on Knowledge and Data Engineering* 21 (9): 1263–84. https://doi.org/10.1109/TKDE.2008.239.
  29. Henry, Devanandham, and Jose Emmanuel Ramirez-Marquez. 2016. “On the Impacts of Power Outages during Hurricane Sandy—A Resilience-Based Analysis.” *Systems Engineering* 19 (1): 59–75. https://doi.org/10.1002/sys.21338.
  30. Inostroza, Luis, Massimo Palme, and Francisco de la Barrera. 2016. “A Heat Vulnerability Index: Spatial Patterns of Exposure, Sensitivity and Adaptive Capacity for Santiago de Chile.” Edited by Jeffrey Shaman. *PLOS ONE* 11 (9): e0162464. https://doi.org/10.1371/journal.pone.0162464.
  31. IPCC. 2021. “Sixth Assessment Report — IPCC.” 2021. https://www.ipcc.ch/assessment-report/ar6/.
  32. Iverson, Sally Ann, Aaron Gettel, Carla P. Bezold, Kate Goodin, Benita McKinney, Rebecca Sunenshine, and Vjollca Berisha. 2020. “Heat-Associated Mortality in a Hot Climate: Maricopa County, Arizona, 2006-2016.” *Public Health Reports* 135 (5): 631–39. https://doi.org/10.1177/0033354920938006.
  33. Jay, Ollie, Matthew N. Cramer, Nicholas M. Ravanelli, and Simon G. Hodder. 2015. “Should Electric Fans Be Used during a Heat Wave?” *Applied Ergonomics* 46 Pt A (January): 137–43. https://doi.org/10.1016/j.apergo.2014.07.013.
  34. Kaur, Harsurinder, Husanbir Singh Pannu, and Avleen Kaur Malhi. 2019. “A Systematic Review on Imbalanced Data Challenges in Machine Learning: Applications and Solutions.” *ACM Computing Surveys* 52 (4): 79:1-79:36. https://doi.org/10.1145/3343440.
  35. Kenny, Glen P., Andreas D. Flouris, Abderrahmane Yagouti, and Sean R. Notley. 2019. “Towards Establishing Evidence-Based Guidelines on Maximum Indoor Temperatures during Hot Weather in Temperate Continental Climates.” *Temperature (Austin, Tex.)* 6 (1): 11–36. https://doi.org/10.1080/23328940.2018.1456257.
  36. King, Carey, Josh Rhodes, and Jay Zarnikau. 2021. “The Timeline and Events of the February 2021 Texas Electric Grid Blackouts.” University of Texas at Austin. https://energy.utexas.edu/sites/default/files/UTAustin%20%282021%29%20EventsFebruary2021TexasBlackout%2020210714.pdf.
  37. Klepeis, Neil E., William C. Nelson, Wayne R. Ott, John P. Robinson, Andy M. Tsang, Paul Switzer, Joseph V. Behar, Stephen C. Hern, and William H. Engelmann. 2001. “The National Human Activity Pattern Survey (NHAPS): A Resource for Assessing Exposure to Environmental Pollutants.” *Journal of Exposure Science & Environmental Epidemiology* 11 (3): 231–52. https://doi.org/10.1038/sj.jea.7500165.
  38. Klinenberg, Eric. 2015. *Heat Wave: A Social Autopsy of Disaster in Chicago*. 2nd ed. Chicago: University of Chicago Press.
  39. Krawczyk, Bartosz. 2016. “Learning from Imbalanced Data: Open Challenges and Future Directions.” *Progress in Artificial Intelligence* 5 (4): 221–32. https://doi.org/10.1007/s13748-016-0094-0.
  40. Kuhn, Max, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, et al. 2023. “Caret: Classification and Regression Training.” https://cran.r-project.org/web/packages/caret/index.html.
  41. Kuras, Evan R., Molly B. Richardson, Miriam M. Calkins, Kristie L. Ebi, Jeremy J. Hess, Kristina W. Kintziger, Meredith A. Jagger, Ariane Middel, Anna A. Scott, and June T. Spector. 2017. “Opportunities and Challenges for Personal Heat Exposure Research.” *Environmental Health Perspectives* 125 (8): 085001.
  42. Lai, Wangyang, Yun Qiu, Qu Tang, Chen Xi, and Peng Zhang. 2023. “The Effects of Temperature on Labor Productivity,” June.
  43. Lane, Kathryn. 2018. “The Dangers of Cold Weather.” Public Health Post. November 14, 2018. https://www.publichealthpost.org/research/counting-cold-related-deaths-new-york-city/.
  44. Laurent, Jose Guillermo Cedeño, Augusta Williams, Youssef Oulhote, Antonella Zanobetti, Joseph G. Allen, and John D. Spengler. 2018. “Reduced Cognitive Function during a Heat Wave among Residents of Non-Air-Conditioned Buildings: An Observational Study of Young Adults in the Summer of 2016.” *PLOS Medicine* 15 (7): e1002605. https://doi.org/10.1371/journal.pmed.1002605.
  45. Lomas, Kevin J. 2021. “Summertime Overheating in Dwellings in Temperate Climates.” *Buildings and Cities* 2 (1): 487–94. https://doi.org/10.5334/bc.128.
  46. Maivel, Mikk, Jarek Kurnitski, and Targo Kalamees. 2015. “Field Survey of Overheating Problems in Estonian Apartment Buildings.” *Architectural Science Review* 58 (1): 1–10. https://doi.org/10.1080/00038628.2014.970610.
  47. Matz, Carlyn J., David M. Stieb, Karelyn Davis, Marika Egyed, Andreas Rose, Benedito Chou, and Orly Brion. 2014. “Effects of Age, Season, Gender and Urban-Rural Status on Time-Activity: Canadian Human Activity Pattern Survey 2 (CHAPS 2).” *International Journal of Environmental Research and Public Health* 11 (2): 2108–24. https://doi.org/10.3390/ijerph110202108.
  48. Mavrogianni, Anna, Paul Wilkinson, Michael Davies, Phillip Biddulph, and Eleni Oikonomou. 2012. “Building Characteristics as Determinants of Propensity to High Indoor Summer Temperatures in London Dwellings.” *Building and Environment*, Implications of a Changing Climate for Buildings, 55 (September): 117–30. https://doi.org/10.1016/j.buildenv.2011.12.003.
  49. MCDPH. 2019. “Heat-Associated Deaths in Maricopa County, AZ, Final Report for 2019.” Maricopa County, AZ: Maricopa County Department of Public Health. https://www.maricopa.gov/ArchiveCenter/ViewFile/Item/4959.
  50. Menardi, Giovanna, and Nicola Torelli. 2014. “Training and Assessing Classification Rules with Imbalanced Data.” *Data Mining and Knowledge Discovery* 28 (1): 92–122. https://doi.org/10.1007/s10618-012-0295-5.
  51. Meyer, David, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, Friedrich Leisch, Chih-Chung Chang (libsvm C++-code), and Chih-Chen Lin (libsvm C++-code). 2023. “E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien.” https://cran.r-project.org/web/packages/e1071/index.html.
  52. Milborrow, Stephen. 2023. “Earth: Multivariate Adaptive Regression Splines.” https://cran.r-project.org/web/packages/earth/index.html.
  53. Ministry of Health. 2023. “Province Launches New Initiative to Protect People during Extreme Heat Emergencies | BC Gov News.” The official website of the Government of British Columbia. BC Gov News. June 27, 2023. https://news.gov.bc.ca/releases/2023HLTH0095-001044.
  54. Naughton, Mary P, Alden Henderson, Maria C Mirabelli, Reinhard Kaiser, John L Wilhelm, Stephanie M Kieszak, Carol H Rubin, and Michael A McGeehin. 2002. “Heat-Related Mortality during a 1999 Heat Wave in Chicago1 1The Full Text of This Article Is Available via AJPM Online at Www.Ajpm-Online.Net.” *American Journal of Preventive Medicine* 22 (4): 221–27. https://doi.org/10.1016/S0749-3797(02)00421-X.
  55. Nayak, S. G., S. Shrestha, P. L. Kinney, Z. Ross, S. C. Sheridan, C. I. Pantea, W. H. Hsu, N. Muscatiello, and S. A. Hwang. 2018. “Development of a Heat Vulnerability Index for New York State.” *Public Health*, Special issue on Health and high temperatures, 161 (August): 127–37. https://doi.org/10.1016/j.puhe.2017.09.006.
  56. New, Joshua, Mark Adams, Eric Garrison, Brett Bass, and Tianjing Guo. 2020. “Urban-Scale Energy Modeling: Scaling Beyond Tax Assessor Data.” In , 7. 2020: ASHRAE and IBPSA-USA.
  57. Nibbering, Didier, and Trevor J. Hastie. 2022. “Multiclass-Penalized Logistic Regression.” *Computational Statistics & Data Analysis* 169 (May): 107414. https://doi.org/10.1016/j.csda.2021.107414.
  58. Obradovich, Nick, Robyn Migliorini, Sara C. Mednick, and James H. Fowler. 2017. “Nighttime Temperature and Human Sleep Loss in a Changing Climate.” *Science Advances* 3 (5): e1601555. https://doi.org/10.1126/sciadv.1601555.
  59. O’Neill, Marie S., Antonella Zanobetti, and Joel Schwartz. 2003. “Modifiers of the Temperature and Mortality Association in Seven US Cities.” *American Journal of Epidemiology* 157 (12): 1074–82. https://doi.org/10.1093/aje/kwg096.
  60. ———. 2005. “Disparities by Race in Heat-Related Mortality in Four US Cities: The Role of Air Conditioning Prevalence.” *Journal of Urban Health* 82 (2): 191–97. https://doi.org/10.1093/jurban/jti043.
  61. Ostro, Bart D., Lindsey A. Roth, Rochelle S. Green, and Rupa Basu. 2009. “Estimating the Mortality Effect of the July 2006 California Heat Wave.” *Environmental Research* 109 (5): 614–19. https://doi.org/10.1016/j.envres.2009.03.010.
  62. Oudin Åström, Daniel, Forsberg Bertil, and Rocklöv Joacim. 2011. “Heat Wave Impact on Morbidity and Mortality in the Elderly Population: A Review of Recent Studies.” *Maturitas* 69 (2): 99–105. https://doi.org/10.1016/j.maturitas.2011.03.008.
  63. Pathan, A., A. Mavrogianni, A. Summerfield, T. Oreszczyn, and M. Davies. 2017. “Monitoring Summer Indoor Overheating in the London Housing Stock.” *Energy and Buildings* 141 (April): 361–78. https://doi.org/10.1016/j.enbuild.2017.02.049.
  64. Peters, Andrea, Torsten Hothorn, Brian D. Ripley, Terry Therneau, and Beth Atkinson. 2023. “Ipred: Improved Predictors.” https://cran.r-project.org/web/packages/ipred/index.html.
  65. Porritt, S. M., P. C. Cropper, L. Shao, and C. I. Goodier. 2012. “Ranking of Interventions to Reduce Dwelling Overheating during Heat Waves.” *Energy and Buildings*, Cool Roofs, Cool Pavements, Cool Cities, and Cool World, 55 (December): 16–27. https://doi.org/10.1016/j.enbuild.2012.01.043.
  66. R Core Team. 2022. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
  67. Reid, Colleen E., Marie O’Neill, Gronlund, Carina J., Brines Shannon J., Brown Daniel G., Diez-Roux Ana V., and Schwartz Joel. 2009. “Mapping Community Determinants of Heat Vulnerability.” *Environmental Health Perspectives* 117 (11): 1730–36. https://doi.org/10.1289/ehp.0900683.
  68. Rinner, Claus, Dianne Patychuk, Kate Bassil, Shiraz Nasr, Stephanie Gower, and Monica Campbell. 2010. “The Role of Maps in Neighborhood-Level Heat Vulnerability Assessment for the City of Toronto.” *Cartography and Geographic Information Science* 37 (1): 31–44. https://doi.org/10.1559/152304010790588089.
  69. Ripley, Brian, and William Venables. 2023. “Nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models.” https://cran.r-project.org/web/packages/nnet/index.html.
  70. RStudio Team. 2021. “RStudio: Integrated Development Environment for R.” Boston, MA: RStudio, PBC. http://www.rstudio.com/.
  71. Samuelson, Holly, Amir Baniassadi, Anne Lin, Pablo Izaga González, Thomas Brawley, and Tushar Narula. 2020. “Housing as a Critical Determinant of Heat Vulnerability and Health.” *Science of The Total Environment* 720 (June): 137296. https://doi.org/10.1016/j.scitotenv.2020.137296.
  72. Schwartz, Joel. 2005. “Who Is Sensitive to Extremes of Temperature? A Case-Only Analysis.” *Epidemiology* 16 (1): 67–72. https://www.jstor.org/stable/20486001.
  73. Sera, Francesco, Ben Armstrong, Aurelio Tobias, Ana Maria Vicedo-Cabrera, Christofer Åström, Michelle L Bell, Bing-Yu Chen, et al. 2019. “How Urban Characteristics Affect Vulnerability to Heat and Cold: A Multi-Country Analysis.” *International Journal of Epidemiology* 48 (4): 1101–12. https://doi.org/10.1093/ije/dyz008.
  74. Stone, Brian, Evan Mallen, Mayuri Rajput, Carina J. Gronlund, Ashley M. Broadbent, E. Scott Krayenhoff, Godfried Augenbroe, Marie S. O’Neill, and Matei Georgescu. 2021. “Compound Climate and Infrastructure Events: How Electrical Grid Failure Alters Heat Wave Risk.” *Environmental Science & Technology* 55 (10): 6957–64. https://doi.org/10.1021/acs.est.1c00024.
  75. Sun, Kaiyu, Michael Specian, and Tianzhen Hong. 2020. “Nexus of Thermal Resilience and Energy Efficiency in Buildings: A Case Study of a Nursing Home.” *Building and Environment* 177: 106842. https://doi.org/10.1016/j.buildenv.2020.106842.
  76. Uejio, Christopher K., Olga V. Wilhelmi, Jay S. Golden, David M. Mills, Sam P. Gulino, and Jason P. Samenow. 2011. “Intra-Urban Societal Vulnerability to Extreme Heat: The Role of Heat Exposure and the Built Environment, Socioeconomics, and Neighborhood Stability.” *Health & Place*, Geographies of Care, 17 (2): 498–507. https://doi.org/10.1016/j.healthplace.2010.12.005.
  77. United Nations. 2020. “World Population Ageing, 2019 Highlights.” UN.
  78. Wheeler, Katherine, Kathryn Lane, Sarah Walters, and Thomas Matte. 2013. “Heat Illness and Deaths — New York City, 2000–2011.” *Morbidity and Mortality Weekly Report* 62 (31): 617–21. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4604987/.
  79. Wickham, Hadley. 2023. “Plyr: Tools for Splitting, Applying and Combining Data.” https://cran.r-project.org/web/packages/plyr/index.html.
  80. Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, Davis Vaughan, Posit Software, and PBC. 2023. “Dplyr: A Grammar of Data Manipulation.” https://cran.r-project.org/web/packages/dplyr/index.html.
  81. Wickham, Hadley, and RStudio. 2023. “Tidyverse: Easily Install and Load the ‘Tidyverse.’” R. https://CRAN.R-project.org/package=tidyverse.
  82. Wright, Mary K., David M. Hondula, Paul M. Chakalian, Liza C. Kurtz, Lance Watkins, Carina J. Gronlund, Larissa Larsen, Evan Mallen, and Sharon L. Harlan. 2020. “Social and Behavioral Determinants of Indoor Temperatures in Air-Conditioned Homes.” *Building and Environment* 183 (October): 107187. https://doi.org/10.1016/j.buildenv.2020.107187.
  83. Zhao, Qi, Yuming Guo, Tingting Ye, Antonio Gasparrini, Shilu Tong, Ala Overcenco, Aleš Urban, et al. 2021. “Global, Regional, and National Burden of Mortality Associated with Non-Optimal Ambient Temperatures from 2000 to 2019: A Three-Stage Modelling Study.” *The Lancet Planetary Health* 5 (7): e415–25. https://doi.org/10.1016/S2542-5196(21)00081-4.

# CRediT authorship contribution statement

**Arfa Aijazi:** Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation Writing–Original Draft, Visualization. **Stefano Schiavon**: Supervision, Validation, Writing–Review and Editing. **Duncan Callaway:** Methodology, Writing–Review and Editing

# Declaration of competing interest

The Center for the Built Environment at the University of California, Berkeley – with which some of the authors are affiliated, is advised by and funded in part by many partners that represent a diversity of organizations from the building industry – including manufacturers, building owners, facility managers, contractors, architects, engineers, government agencies, and utilities.

# Acknowledgements

# At the time of the study, Arfa Aijazi was supported by a Doctoral Completion Fellowship through the Graduate Division at the University of California, Berkeley. This research was also in part funded by the Center for the Built Environment (CBE) at University of California, Berkeley. CBE with which some of the authors are affiliated. This research used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). The authors also thank Dr. Matias Quintana with the Singapore-ETH Center for his assistance with developing machine learning modeling methodology.

# Data availability

All data and analysis code is provided on GitHub at: <https://github.com/anaijazi/RECSThermalMorbidity>