A Comparison of Three Modeling Approaches for Analyzing the Impact of Rhode Island Urban Heat Islands on Emergency Medical Services Utilization

By

Yiwen Liang

B.S., The Ohio State University, 2021

Thesis

Submitted in partial fulfillment of the requirements for the Degree of Master of Science in the Department of Biostatistics at Brown University

PROVIDENCE, RHODE ISLAND

May 2024

AUTHORIZATION TO LEND AND REPRODUCE THE THESIS

As the sole author of this thesis, I authorize Brown University to lend it to other institutions or individuals for the purpose of scholarly research.

Date ______________ ____________________________________
Yiwen Liang, Author

I further authorize Brown University to reproduce this thesis by photocopying or other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Date ______________ ____________________________________
Yiwen Liang, Author

This thesis by Yiwen Liang is accepted in its present form by the Department of Biostatistics as satisfying the thesis requirement for the degree of Master of Science.

Date ______________ ______________________________
Christopher H. Schmid, Advisor

Date ______________ ______________________________
Katelyn Moretti, Reader

Date ______________ ______________________________
Stavroula A. Chrysanthopoulou,
Director of the Master's
Graduate Program in Biostatistics

Approved by the Graduate Council

Date ______________ ______________________________
Thomas A. Lewis, Dean of the Graduate School

Acknowledgments

I am deeply grateful for the completion of this thesis, which would not have been possible without the unwavering support and guidance of the remarkable people around me. First and foremost, I extend my heartfelt gratitude to Dr.Christopher Schmid. As my thesis adviser, he has been consistently supportive and approachable. His vast experience and insights into teaching and statistics have been instrumental in inspiring and guiding me throughout this research journey. His willingness to share knowledge, offer constructive feedback, and provide guidance during challenging times have been invaluable. His mentorship has not only enriched my research but has also inspired me to strive for excellence in my future endeavors.

I also want to thank Dr.Katelyn Moretti, my reader, for her belief in my capabilities and continuous encouragement. It has been such a privilege to be part of her research group and to learn and discuss with so many talented and thoughtful professors and students. I'd especially like to thank Dr.Adam R. Aluisio from the Department of Emergency Medicine and John Matthew Nicklas from the Department of Earth, Environmental and Planetary Sciences, for their invaluable advice and assistance with data processing and model construction. Their support and expertise have been indispensable to this thesis.

Furthermore, I would like to thank all the professors, administrative staff, and fellow students in the Department of Biostatistics for fostering such a supportive and friendly environment throughout my academic journey here at Brown University. The approachability of the professors has been truly commendable, always willing to offer guidance, share their expertise, and engage in meaningful discussions. The staff members have been consistently supportive, ensuring that we students have the resources and assistance we need. The warm and collaborative atmosphere within the department has not only facilitated my research, but also made my time here incredibly rewarding and enjoyable. I feel fortunate to have been a part of such a welcoming community.

Lastly, I'd like to express my deepest gratitude to my family who have supported me throughout my studies abroad over the past seven years. Their unconditional love has been my anchor, sustaining me through every success and setback. I am so glad and blessed to have them being there for me in every success or failure. I also want to thank my friends both in the States and back in China. Their willingness to listen and support has been a source of strength and a ray of light during challenging times.

And my special thanks goes to McDonald's and all the fast-food places I ordered from, which spared me from cooking and sustained me in my busiest and most difficult moments.

Abstract of A Comparison of Three Modeling Approaches for Analyzing the Impact of Rhode Island Urban Heat Islands on Emergency Medical Services Utilization

By Yiwen Liang, M.S., Brown University, May 2024

Background: The escalating frequency, intensity, and duration of extreme heat events due to climate change pose significant threats to public health and strain healthcare systems. Previous research by Dr. Moretti and colleagues in Rhode Island (RI) found that increased summer temperatures were positively correlated with an increased use of emergency medical services (EMS). However, that study did not account for a potentially uneven geographical distribution of these adverse effects of heat on health that might arise from urban heat islands (UHI), small areas with temperatures elevated by their local built environment. UHIs are also frequently home to socially and economically disadvantaged individuals who live in regions with high values of the Area Deprivation Index (ADI). This research expands upon Dr. Moretti's research to estimate the relative impacts of hot days in summer, heatwaves, UHIs, and socioeconomic factors on EMS utilization across Rhode Island using data from June-August in the years from 2018 to 2021.

Methods: We leverage five existing datasets that describe EMS utilization, daily temperature, UHI indices, ADI scores, and population distribution in RI, to form a new comprehensive database. We utilize day-level and day-subgroup level quasi-Poisson regressions to model the number of daily EMS runs from residential areas with high ADI (7-10). The effects of heat on outcomes may arise from both current and lagged temperatures, which can be modeled either collectively or separately.

We employ quasi-Poisson regression models and develop three models with both day and day-subpopulation as the unit of analysis. Temperature is modeled in two ways. One approach uses a cross-basis matrix from a distributed lag nonlinear model (DLNM); the other uses functions of current and lagged temperatures. All models include other covariates to assess the effects of UHI in addition to demographic and other temporal effects.

Results: The primary influence on EMS utilization is the current daily mean temperature, showing a nearly positive linear association. Day-level analyses indicated that higher proportions of patients from UHI areas correlate with increased EMS utilization, and the day-subpopulation model suggests a positive UHI effect on EMS utilization when assuming consistent age and UHI distributions across the population and EMS patients. The day-level model reveals that areas with greater proportions of female residents and those under 65 have higher EMS call rates, and the day-subpopulation model suggests a greater EMS utilization for females and individuals younger than 65.

Conclusion: Rising ambient temperatures and temporal factors correlate with increased EMS use. Living in UHIs may also increase EMS use but conclusive determinations cannot be made from the data available. Although the results from all three models align qualitatively, the different units of analysis result in distinct interpretations of the findings. Therefore, we cannot draw a firm conclusion. More data are needed on individuals who do not use EMS services to provide a basis for informed policy-making in EMS planning and resource allocation.

Table of Contents

Acknowledgments . . . . **iv**

List of Tables . . . . **x**

List of Figures . . . . **xii**

1 Introduction **1**

2 Data **5**

2.1 Existing Datasets 5

2.1.1 The Rhode Island National Emergency Medical Services Information System (NEMSIS) V3 dataset 5

2.1.2 The National Centers for Environmental Information of the National Oceanic and Atmospheric Administration (NOAA) 6

2.1.3 Urban Heat Island Severity for U.S. Cities 6

2.1.4 2021 Area Deprivation Index v4.0.1 7

2.1.5 2020 Census Data 7

2.2 Database Formation 8

2.3 Descriptive Statistics of EMS Runs 10

2.3.1 Patient characteristics 11

2.3.2 Temperature 13

2.3.3 Time 14

2.4 Insights from the Exploratory Data Analysis 15

3 Setting Up Predictors **17**

3.1 Patient Characteristics 17

3.2 Temperature 18

3.3 Time 19

3.4 Confounding 19

4 Models **24**

4.1 Model 1: Day-Level with DLNM 24

4.2 Model 2: Day-level with Current and Lagged Temperature 27

4.3 Model 3: Day-subpopulation level with Current and Lagged Temperature 28

5 Results from Three Models **31**

5.1 Model 1: Day-level with DLNM 31

5.2 Model 2: Day-level with Current and Lagged Temperature 35

5.3 Model 3: Day-subpopulation level with Current and Lagged Temperature 39

6 Discussion **47**

6.1 Results Comparison 47

6.2 Models Evaluation 49

6.3 Limitations 50

7 Conclusion **53**

A Supplemental Figures **55**

References **55**

List of Tables

2.1 Degrees range for UHI category definitions 7

2.2 EMS response and patient characteristics from areas with different UHI severity index in Rhode Island, June–August 2018-2021 (% within UHI level). Entries are numbers of EMS runs with percentage within UHI level in parentheses. 12

2.3 Number and proportion of EMS runs from areas of different ADI scores in 2018 to 2021s. 14

2.4 Number and proportion of EMS runs at different day of the week. 15

3.1 Summary of five slopes of linear regressions of the daily number of EMS runs on the proportion of runs for UHI>0 areas (overall) 20

3.2 Summary of five slopes of linear regressions of the daily number of EMS runs on the proportion of runs for UHI>0 areas in (a) residential, high ADI, (b) non-residential, high ADI, (c) residential, low ADI, (b) non-residential, low ADI areas. 22

3.3 Summary of five slopes of linear regression in (a) residential, high ADI, (b) non-residential, high ADI, (c) residential, low ADI, (b) non-residential, low ADI areas. 22

4.1 Distribution of EMS runs and RI population in high ADI areas across gender, age, and UHI. 29

4.2 Formulas for population estimation give arbitrary P(UHI|age$\geq$65)=a and P(UHI|age<65)=b, assuming N individuals in total and x are $\geq$65 years old. 30

5.1 Summary of results of UHI_prop, weekday/weekend, year, age_prop, and female_prop from Model 1. 34
5.2 Summary of results of daily average temperature, UHI_prop, weekday/weekend, year, age_prop, and female_prop from Model 2. 37
5.3 Summary of estimation results of daily mean temperature, year, weekday/weekend, gender, and interaction terms when P(UHI|age$\geq$65)=0.5 and P(UHI|age<65)=0.5. 43

List of Figures

2.1 Database formation 9

2.2 Flow diagram for exclusion. 10

2.3 Distribution of the daily number of EMS runs. 11

2.4 Distribution of the daily temperature from June to August in 2018 to 2021. 13

2.5 Distribution of the daily EMS counts with respect to daily mean temperature. 14

3.1 Confounding illustration using linear association between daily EMS counts and proportion of runs from UHI>0 areas with boxplots representing marginal distributions. 20

3.2 Confounding illustration using linear association between daily EMS counts and proportion of runs from UHI>0 areas with boxplots representing marginal distributions, in (a) residential, high ADI, (b) non-residential, high ADI, (c) residential, low ADI, (b) non-residential, low ADI areas. . . . 21

5.1 Three-dimensional plot showing the estimated exposure-lag–response association between temperature and EMS in high ADI areas 32

5.2 Temperature-EMS association at lag=0. 33

5.3 Predicted vs. observed daily number of EMS runs from Model 1. 35

5.4 A: variation in predicted daily mean number of EMS runs. The five curves represent 5 functional forms of daily mean temperature, with the remaining factors in the model fixed. The five functional forms are (1) linear scale; (2) natural cubic splines with 2 degrees of freedom; (3) natural cubic splines with 3 internal knots place at 10, 75, 90 percentiles of the temperature distribution; (4) quadratic; and (5) cubic. Fixed factors are: year (2021), weekday/weekend (weekday), UHI_prop=0.5, age_prop=0.5, female_prop=0.5. B: the distribution of daily mean temperatures. 36
5.5 Predicted vs. observed daily number of EMS runs from Model 2. 38
5.6 A: variation in predicted probabilities of EMS runs when P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5. The five curves represent 5 functional forms of daily mean temperature, with the remaining factors in the model fixed. The five functional forms are (1) linear scale; (2) natural cubic splines with 2 degrees of freedom; (3) natural cubic splines with 3 internal knots place at 10, 75, 90 percentiles of the temperature distribution; (4) quadratic; and (5) cubic. Fixed factors are: year (2021), weekday/weekend (weekday), gender (female), age category (≥65), UHI. B: the distribution of daily mean temperatures. 40
5.7 Distribution of overall effect of *UHI* on across different values of P(UHI|age≥65) and P(UHI|age<65). 42
5.8 Predicted number of EMS runs in each subgroup (per 10,000) over four years, with P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5. 45
5.9 Predicted vs. observed daily number of EMS runs from Model 3 when P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5. 46

A.1 Predicted number of EMS runs in each subgroup (per 10,000) over four years, with P(UHI|age≥65)=0.2 and P(UHI|age<65)=0.2. 56

A.2 Predicted number of EMS runs in each subgroup (per 10,000) over four years, with P(UHI|age≥65)=0.8 and P(UHI|age<65)=0.8. 57

Chapter 1 Introduction

Climate change has been described as the greatest global health threat of the 21st century. One of its detrimental results is the increasing frequency, intensity, and duration of extreme heat events and heatwaves (Watts, 2020). Global climate changes have increased the ambient temperature and the frequency of heat waves, with recent research pointing out that over the past two decades, heat-related deaths among individuals aged over 65 have increased by 70% (WHO, 2023). In Charlottesville, Virginia, data from 2005–2016 showed a higher rate of emergency department admissions during heatwaves compared to non-heatwave periods (Davis, 2018). The increasing temperatures not only jeopardize individual health but also compound the strain on healthcare systems. A well-established correlation exists between escalating ambient temperatures and increased rates of morbidity and mortality, thereby intensifying the utilization and strain on healthcare infrastructure.

The adverse effects of heat on health are not uniformly distributed, exhibiting variations both temporally and spatially. In our study, we focus on one of the temporal effects known as the lag effect. The lag effect is an essential consideration in public health research, representing the time gap between an exposure and the manifestation of adverse health outcomes. Fatalities and hospitalizations resulting from heat can manifest swiftly, occurring on the same day as the exposure, or exhibit a lagged effect, becoming evident several days later (WHO, 2024). This delayed impact tends to exacerbate mortality or illness, especially among individuals with underlying chronic conditions and those already frail.

One significant instance of the uneven geographical distribution of the deleterious effects of heat is urban heat islands (UHIs), which refer to urbanized areas characterized by elevated temperatures compared to their surrounding areas. This phenomenon arises because built structures like buildings, roads, and other urban infrastructure absorb and subsequently re-emit more solar heat than natural landscapes such as forests and water bodies. The concentration of these heat-absorbing structures and the scarcity of green spaces in urban areas contribute to the formation of "islands" marked by elevated temperatures in contrast to their urban green or non-urban counterparts. Notably, the urban heat island effect leads to daytime temperatures in urban locales typically 1–7°F higher compared to outlying areas, while nighttime temperatures experience a rise of approximately 2–5°F (EPA, 2024). Climate change and heat islands interact in important ways. The Climate Science Special Report predicts that as temperatures steadily rise and urban areas continue to grow and alter in terms of their structure, spatial expanse, and population density, the heat island effect will worsen (EPA, 2022) (USGCRP, 2017).

Such environmentally induced temperature variations have the potential to disproportionately impact different sociodemographic and socioeconomic groups. Heat islands significantly elevate the risk of heat-related health issues, including hyperthermia, respiratory complications, heat cramps, and heat stroke. While extreme heat affects everyone, certain age groups face heightened vulnerability. Older adults are particularly susceptible to extreme heat events because of factors such as poor health, limited mobility, social isolation, and heightened sensitivity to high temperatures. Likewise, young children are more vulnerable due to their smaller size, higher breathing rates relative to body size, and ongoing respiratory system development. During heatwaves, elevated levels of ozone pollution and smog can exacerbate asthma and other respiratory conditions, especially in young children.

Individuals with lower socioeconomic status or marginalized populations may also face increased risks during extreme heat events due to factors such as poor housing conditions, lack of air conditioning, and extended outdoor working hours. Research has shown that people living below the poverty line are exposed to higher levels of surface urban heat island effects compared to those living above the poverty line (Angel Hsu, 2021). Such disparities in the impact of heat effects on the population raise concerns about potential health inequities that may impede achieving better health. Health equity, defined as the absence of systematic disparities in health between different social groups with varying levels of social advantage or disadvantage, is crucial in ensuring fair and just health outcomes (Braveman & Gruskin, 2003). Disparities in health status across various groups within a population are a global challenge, spanning differences in age, gender, socioeconomic status, ethnicity, disability, and geographical location (Reid & Robson, 2000).

Research carried out by Moretti and colleagues demonstrated that increased summer daily temperatures are associated with more frequent emergency medical services (EMS) encounters within Rhode Island (Moretti, 2021). However, their study did not account for the urban heat island effect, lagged temperature effects and potential disparities between population groups. This study aims to estimate the relative impacts of hot days in summer, heatwaves, UHIs, and socioeconomic factors on EMS utilization across Rhode Island using data from June-August in the years from 2018 to 2021. Prediction and management of EMS demand provides an opportunity to proactively address challenges to and pressures on the healthcare system posed by climate change.

In Chapter 2, we introduce five existing datasets that we use to form a comprehensive database to address these questions, and then detail the process of database formation, including the inclusion and variable selection criteria. We categorize the variables into three groups: patient characteristics, temperature, and time, and provide exploratory analyses corresponding to each in this chapter. Chapter 3 delves into how these three types of variables can be incorporated into the quasi-Poisson regression model, considering two different units of analysis. In Chapter 4, we introduce three models, each with different forms of variables for describing the association between heat and EMS utilization and explain the implementation of each. Chapter 5 presents the results obtained from these three models collectively. Chapter 6 offers a comparative analysis of the results from the three models, evaluating their respective performances. We interpret the findings, discuss their practical implications, and identify the study's limitations. Finally, in Chapter 7, we review the conclusions drawn from our study.

Chapter 2 Data

Developing a statistical model to thoroughly investigate the impact of summer heat, heatwaves, urban heat islands (UHIs), and socioeconomic variables on healthcare delivery requires data on EMS utilization, daily temperature, a measure of UHI, and a measure of socioeconomic status (SES). Given that a single dataset typically cannot encompass all of this information simultaneously, our study leverages five existing datasets to form a new comprehensive database.

2.1 Existing Datasets

2.1.1 The Rhode Island National Emergency Medical Services Information System (NEMSIS) V3 dataset

The emergency medical services (EMS) utilization data are obtained from the Rhode Island National Emergency Medical Services Information System (NEMSIS) V3 database, managed by the Center of Emergency Medical Services within the State of Rhode Island Department of Health. It serves as a repository for essential ambulance run data gathered from all agencies operating within Rhode Island. This database comprises patient-level information, encompassing demographics such as age and gender, along with specific details including the date, time, and address of the location where each patient is pickedup (including the 5-digit zip code) for each EMS call.

2.1.2 The National Centers for Environmental Information of the National Oceanic and Atmospheric Administration (NOAA)

Daily weather data are sourced from the National Centers for Environmental Information of the National Oceanic and Atmospheric Administration (NOAA). This database offers public access to historical weather data and information throughout the United States. Specifically, we extracted daily minimum, mean and maximum temperatures from the weather station at Theodore Francis Green State Airport, Warwick, RI, for the specified periods of interest: June to August from 2018 to 2021. The daily temperatures are measured and recorded in degrees Fahrenheit.

2.1.3 Urban Heat Island Severity for U.S. Cities

The urban heat island severity data are derived from the urban heat island severity layer within the map generated by the Trust for Public Land's ParkServe ® (The Trust for Public Land, n.d.). This raster graphic, with a resolution of 30 meters, is produced from ground-level thermal sensor data (Landsat 8 imagery band 10) obtained from the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA) during the summers of 2018 and 2019. The map provides an assessment of relative heat severity for every pixel across cities in the United States. The Urban Heat Island index is categorized on a scale of 0 to 5 using the Jenks Natural Breaks classification method (Jenks, 1967). A rating of 0 indicates a non-heat island (at or below the mean temperature for the city), while 1 denotes relatively mild heat areas (slightly above the city's mean temperature), and 5 represents severe heat areas (significantly above the average temperature for the city). The range of degrees above the mean temperature of each UHI category is presented in Table 2.1.

UHI Index	0	1	2	3	4	5
Severity	Non-UHI	Very Low	Low	Moderate	High	Very High
Temperature above mean (°F)	Below mean	1	2-3	4-5	6-7	8-12

Table 2.1: Degrees range for UHI category definitions

2.1.4 2021 Area Deprivation Index v4.0.1

We use the Area Deprivation Index (ADI), a measure established by the Health Resources Services Administration (HRSA) to characterize the relative socioeconomic conditions of neighborhoods, as the indicator of SES. The ADI scores utilized in this study are derived from the 2021 Area Deprivation Index, collected and refined by the research team at the Center for Health Disparities Research within the School of Medicine and Public Health at the University of Wisconsin-Madison (Kind, 2018). These scores represent decile rankings, ranging from 1 to 10, of census block groups based on socioeconomic disadvantage at the state level. Group 1 denotes the least disadvantaged, while 10 represents the most disadvantaged block groups. The ADI incorporates various factors including income, education, employment, and housing quality. ADI scores are assigned to each census block group, allowing for the quantification of socioeconomic disadvantage across Rhode Island. Each census block group is referenced by a unique Federal Information Processing Series (FIPS) code.

2.1.5 2020 Census Data

The population by sex and age for each census block group in Rhode Island are obtained from 2020 Census data. Census tracts, which are small, permanent statistical subdivisions of a county, typically have a population size ranging from 1,200 and 8,000 inhabitants. Within these tracts, Block Groups (BGs) are further statistical divisions, designed to contain between 600 and 3,000 people. Each block group provides the population distribution by gender and age category in each census block group, identified by its 12-digit Federal Information Processing Series (FIPS) code. Age from the Census are organized in five-year increments and we group them into three commonly used categories given varying known risk to heat exposure: under 5 years old, 5 to 64 years old, and 65 years old and above.

2.2 Database Formation

The process begins by combining the patient-level data on daily statewide EMS runs from the NEMSIS database, with NOAA daily temperature data by linking on the date of the EMS encounter. The addresses in this combined dataset are complete, comprising street number and name, city, state, and 5-digit zip code. Subsequently, we incorporate the Urban Heat Island Severity indices associated with each encounter by merging on the address associated with the runs.

Merging with the ADI data requires associating the address associated with each run to the appropriate 12-digit Federal Information Processing Series (FIPS) code associated with the ADI scores assigned to each census block. Each census block group is assigned a unique FIPS code, and multiple 9-digit zip code areas are wholly contained within a census block group. The first step is therefore to associate each address in the combined database with its 9-digit zip code in the U.S. Postal Service (USPS) server and therefore with a FIPS code. The ADI scores can then be merged into the combined database by linking on the FIPS code. The FIPS code then serves as the identifier to merge with the Census data. Population data by gender and age groups for each block group are incorporated into the expanded dataset by 12-digit FIPS code.

The process of database formation is outlined in Figure 2.1. This comprehensive dataset now includes all EMS runs conducted across Rhode Island during the summer months (June through August) from 2018 to 2021. It incorporates patient-level information, the

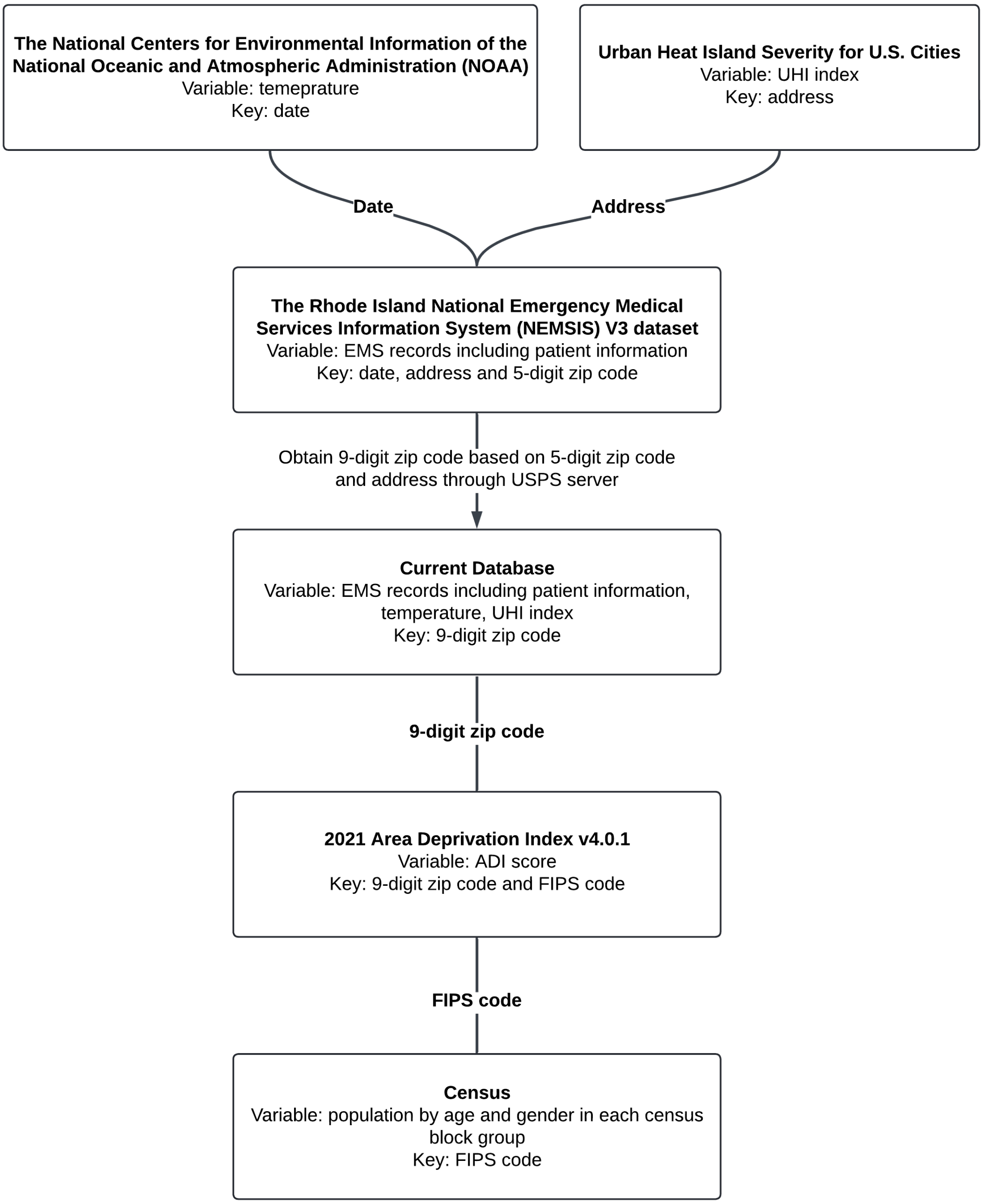


Figure 2.1: Database formation

socioeconomic status of the pick-up location, and population data categorized by age and gender for each census block group in Rhode Island.

Upon assembly, the database has three main categories of variables available to study the impact of various factors on the daily EMS counts. The first is patient characteristics, which includes patient's age, sex, UHI index and ADI score for patient's pick-up location. The second is daily temperature. The last type of variable is the day and date of each run. Since the data used in this study spans the years 2018 to 2021, encompassing the period of the COVID-19 pandemic, calendar year may be an important determinant of risk variation. Other time-related variables available include the day of the week. Subsequently, we will summarize the dataset according to these three categories.

2.3 Descriptive Statistics of EMS Runs

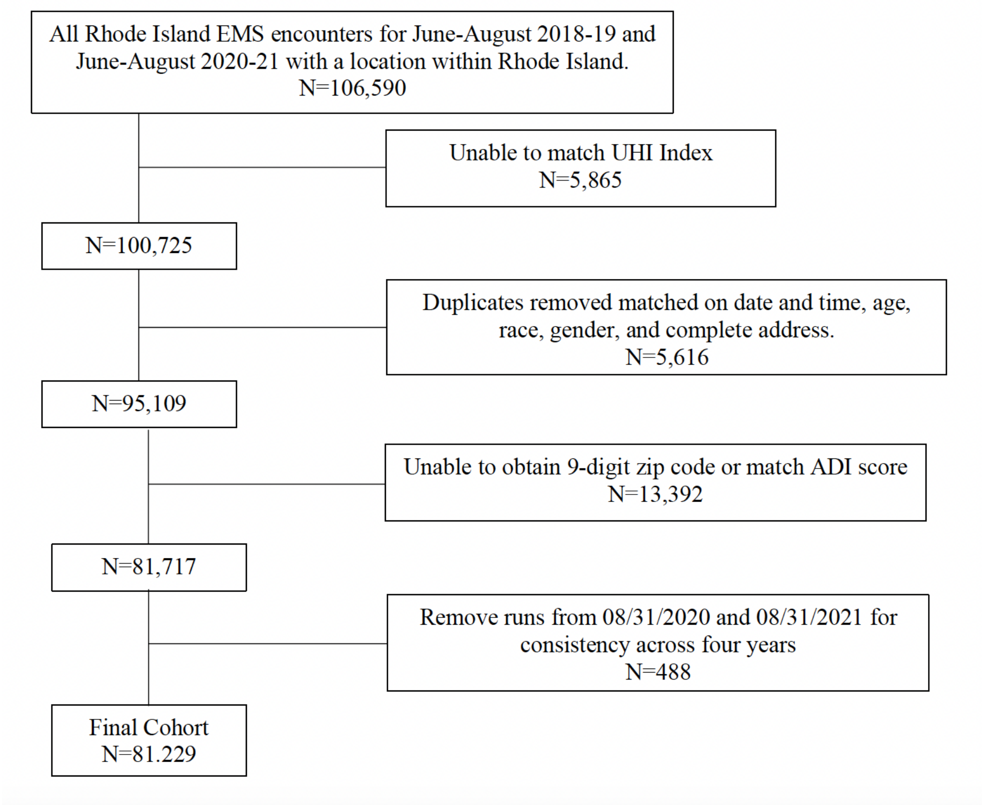


Figure 2.2: Flow diagram for exclusion.

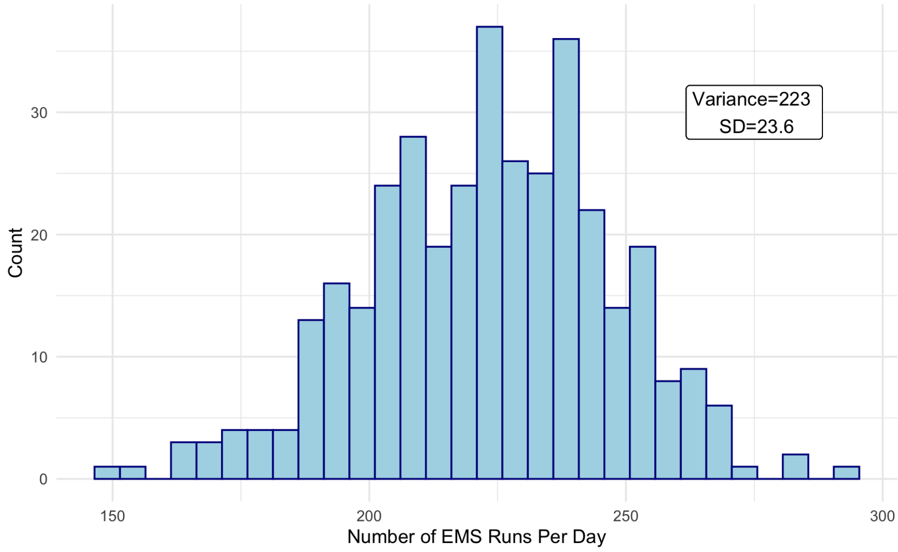


Figure 2.3: Distribution of the daily number of EMS runs.

In total, 106,590 encounters were included from June to August for the years 2018 to 2021. After removing duplicates and records that couldn't be matched with UHI indices or ADI scores, the final cohort comprised 81,717 EMS encounters (Figure 2.2). To maintain data consistency across the four years, we excluded EMS records for 08/31/2020 and 08/31/2021 since there was no corresponding data for the same dates in 2018 and 2019.

Figure 2.3 depicts the overall distribution of the response variable, the daily EMS counts. The mean daily number of EMS runs is 223 with SD=23.6.

2.3.1 Patient characteristics

Table 2.2 presents the descriptive characteristics of patients by areas with varying levels of UHI severity within the cohort. More than half of the encounters originated from a UHI region with severity 0, with only 15% of EMS runs occurring in areas with a UHI severity greater than 2. The average daily number of EMS encounters ranged from 123 to 1 across UHI indices from 0 to 5, respectively. The mean age of patients in the database was 58.05 years old, with patients from UHI=0 (mean age: 59.87) regions exhibiting higher mean ages compared to other regions. We also summarized the EMS distribution across three age categories, <5, 5-64, and ≥65. The observed EMS records of children younger than 5 years old only accounts for approximately 1% of the runs in each UHI index, so we can consider combining the two categories <65. Additionally, there was a greater proportion of female patients in the database, with higher representation in all UHI levels except for UHI=4 levels.

	UHI Index						**p-value**
	0	**1**	**2**	**3**	**4**	**5**	
Total EMS runs	44,847	11,903	12,004	9,737	2,414	324	
Mean daily EMS runs	123	33	33	27	7	1	
Year							<0.001
2018	11,357 (25%)	3,025 (25%)	3,237 (27%)	2,798 (29%)	582 (24%)	67 (21%)	
2019	11,506 (26%)	3,043 (26%)	3,158 (26%)	2,865 (29%)	627 (26%)	74 (23%)	
2020	10,477 (23%)	2,759 (23%)	2,515 (21%)	1,944 (20%)	535 (22%)	82 (25%)	
2021	11,507 (26%)	3,076 (26%)	3,094 (26%)	2,130 (22%)	670 (28%)	101 (31%)	
Age (Mean (SD))	59.87 (23.40)	57.34 (23.42)	55.83 (23.15)	54.92 (22.44)	51.35 (20.47)	57.66 (20.80)	<0.001
Age Category							<0.001
<5	492 (1.1%)	145 (1.2%)	140 (1.2%)	98 (1.0%)	25 (1.0%)	1 (0.3%)	
5-65	22,361 (50%)	6,687 (56%)	7,144 (60%)	6,125 (63%)	1,750 (73%)	187 (58%)	
>65	21,966 (49%)	5,059 (43%)	4,712 (39%)	3,507 (36%)	638 (26%)	136 (42%)	
Gender							<0.001
Female	23,789 (53%)	6,355 (53%)	6,252 (52%)	4,998 (51%)	1,138 (47%)	180 (56%)	
Male	21,025 (47%)	5,537 (47%)	5,747 (48%)	4,735 (49%)	1,272 (53%)	144 (44%)	
Unknown (Unable to Determine)	33 (<0.1%)	11 (<0.1%)	5 (<0.1%)	4 (<0.1%)	4 (<0.1%)	0 (0%)	
ADI							<0.001
1	5,084 (11%)	335 (2.8%)	319 (2.7%)	471 (4.8%)	146 (6.0%)	5 (1.5%)	
2	6,171 (14%)	386 (3.2%)	292 (2.4%)	397 (4.1%)	392 (16%)	22 (6.8%)	
3	6,222 (14%)	1,102 (9.3%)	605 (5.0%)	387 (4.0%)	130 (5.4%)	46 (14%)	
4	6,152 (14%)	1,344 (11%)	1,033 (8.6%)	1,147 (12%)	145 (6.0%)	10 (3.1%)	
5	3,772 (8.4%)	1,068 (9.0%)	1,500 (12%)	974 (10%)	291 (12%)	51 (16%)	
6	5,453 (12%)	976 (8.2%)	990 (8.2%)	570 (5.9%)	173 (7.2%)	10 (3.1%)	
7	3,515 (7.8%)	1,706 (14%)	1,644 (14%)	1,209 (12%)	256 (11%)	25 (7.7%)	
8	2,950 (6.6%)	1,608 (14%)	1,689 (14%)	1,240 (13%)	125 (5.2%)	36 (11%)	
9	2,533 (5.6%)	1,638 (14%)	1,567 (13%)	1,104 (11%)	148 (6.1%)	1 (0.3%)	
10	2,577 (5.7%)	1,624 (14%)	2,105 (18%)	1,772 (18%)	546 (23%)	116 (36%)	
GQ[1]	359 (0.8%)	116 (1.0%)	256 (2.1%)	308 (3.2%)	60 (2.5%)	1 (0.3%)	
PH[2]	0 (0%)	0 (0%)	0 (0%)	158 (1.6%)	2 (<0.1%)	1 (0.3%)	
U[3]	59 (0.1%)	0 (0%)	4 (<0.1%)	0 (0%)	0 (0%)	0 (0%)	

Runs(% within UHI level); Mean(SD)
Pearson's Chi-squared test; Wilcoxon rank sum test
[1] GQ: suppression due to a high group quarters population.
[2] PH: suppression due to low population and/or housing.
[3] U: a unique zip code often representing businesses or large entities with significant mail volume, and thus omitted from the American Community Survey.

Table 2.2: EMS response and patient characteristics from areas with different UHI severity index in Rhode Island, June–August 2018-2021 (% within UHI level). Entries are numbers of EMS runs with percentage within UHI level in parentheses.

Table 2.2 also includes the two-way cross-tabulation of EMS runs between UHI indices and ADI scores. The distribution splits neatly on UHI=0 vs UHI>0 by ADI score at a threshold of 7. In areas with UHI=0, the greater proportion of EMS runs are from areas with ADI <7, but when UHI>0, the greater proportion are from areas with

ADI≥7. Overall, statistically significant differences were observed in patient characteristics including age, gender, and ADI scores among those from different UHI severity regions.

2.3.2 Temperature

Figure 2.4 displays the maximum, minimum, and average daily temperatures from June to August for the years 2018 to 2021. Temperatures are recorded in degrees Fahrenheit for this analysis. The blue line represents the average temperature, while the shaded area is bounded by the minimum and maximum temperatures. The temperature trends across the four years exhibit a relatively consistent pattern.

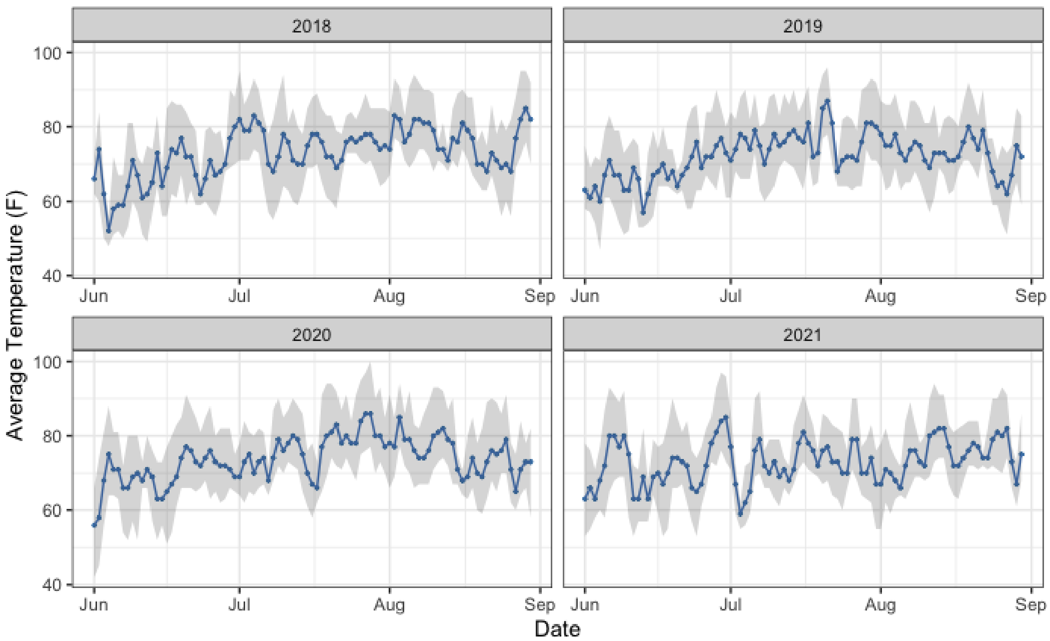


Figure 2.4: Distribution of the daily temperature from June to August in 2018 to 2021.

Figure 2.5 illustrates the relationship between the daily average temperature and the daily number of EMS runs. The general trend indicates that the daily EMS count rises with increasing temperatures, and this association is approximately linear.

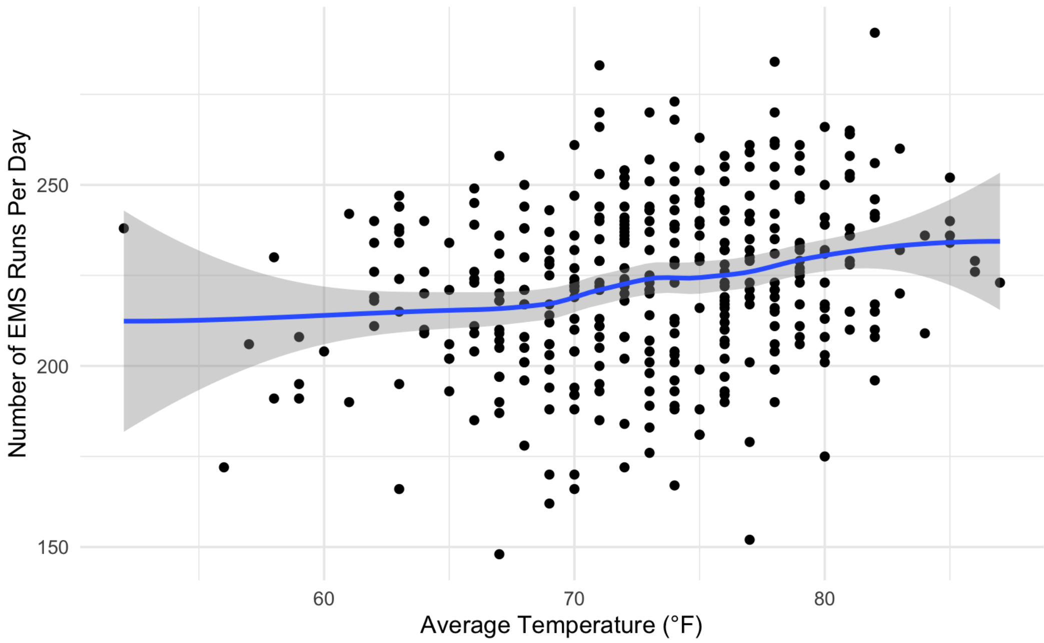


Figure 2.5: Distribution of the daily EMS counts with respect to daily mean temperature.

2.3.3 Time

As summarized in Table 2.2, the distribution of records from areas with different UHI severity levels remained similar for 2018 and 2019, with a notable decrease in total EMS runs in 2020 during the COVID-19 pandemic, which led to significant changes in people's lifestyles and work patterns. EMS encounters rebounded in 2021, with a higher number of calls from high UHI regions compared to 2018 and 2019.

Year	ADI Score										Total
	1	2	3	4	5	6	7	8	9	10	
2018	1591 (8%)	2073 (10%)	2265 (11%)	2563 (12%)	1956 (10%)	2162 (10%)	2125 (10%)	1913 (9%)	1771 (9%)	2254 (11%)	21,066
2019	1775 (9%)	2062 (10%)	2306 (11%)	2627 (13%)	2039 (9%)	2056 (10%)	2051 (10%)	1949 (9%)	1699 (8%)	2307 (11%)	21,273
2020	1445 (8%)	1644 (9%)	1958 (11%)	2195 (12%)	1785 (10%)	1853 (10%)	1940 (11%)	1792 (10%)	1542 (8%)	1912 (11%)	18,312
2021	1549 (8%)	1,881 (9%)	1,963 (10%)	2,446 (12%)	1,876 (9%)	2120 (10%)	2239 (11%)	1994 (10%)	1979 (10%)	2267 (11%)	20,578

Table 2.3: Number and proportion of EMS runs from areas of different ADI scores in 2018 to 2021s.

The distribution of EMS records from areas with different ADI scores remained fairly constant across years, however (Table 2.3).

Year	**Day of week**						
	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**
2018	3271 (16%)	3080 (15%)	3006 (14%)	3008 (14%)	2989 (14%)	2785 (13%)	2927 (14%)
2019	3213 (15%)	3020 (14%)	3159 (15%)	3114 (15%)	3028 (14%)	2877 (14%)	2862 (13%)
2020	2806 (15%)	2676 (15%)	2530 (14%)	2646 (14%)	2667 (15%)	2528 (14%)	2459 (13%)
2021	3006 (15%)	3042 (15%)	3157 (15%)	2954 (14%)	2921 (14%)	2787 (14%)	2711 (13%)

Table 2.4: Number and proportion of EMS runs at different day of the week.

Another variable related to time is day of the week. Table 2.4 shows the distribution of runs by calendar year and day of the week. While the total number of runs is lower in 2020 for all days of the week, the relative proportion of runs by day is stable across years. However, fewer runs occur on Saturday and Sunday compared with Monday-Friday.

2.4 Insights from the Exploratory Data Analysis

For the purpose of simplifying subsequent modeling and interpretation, we dichotomize the urban heat island index. As shown in Table 2.2, the distributions of EMS runs across age categories and ADI scores in UHI=0 areas appear different between areas when UHI = 0 and UHI>0. Categorizing based on zero and non-zero UHI indices also distinguishes between regions that are not affected by the heat island effect and those where the heat island effect is present. Hence a binary variable is created to indicate whether an area exhibits UHI conditions or not. Urban heat island indices of 0 will be classified as non-UHI, while indices of 1 to 5 will be classified as UHI.

We also dichotomize the ADI scores, creating a binary variable to indicate whether an area is disadvantaged or not. Based on the cross-tabulation of UHI and ADI, as shown in Table 2.2, the distribution of EMS runs by UHI differs notably between ADI scores below and at or above 7. Consequently, census block groups with ADI scores lower than 7 will be classified as low ADI areas, indicating less socioeconomically disadvantage. Conversely, ADI scores ranging from 7 to 10 will be classified as high ADI areas, indicating areas that are the most disadvantaged.

Previous studies have highlighted variations in emergency medical services (EMS) demand (Cantwell, 2015) and emergency department (ED) utilization (Hitzek, 2022) across different days of the week, distinguishing between weekdays and weekends. Therefore, we decide to include weekday/weekend indicator instead of day of the week in our model, and will set weekday as reference.

Given the objective to examine the relative effects of various factors on EMS utilization across Rhode Island with the response variable being the daily EMS counts, Poisson regression is indicated. However, the variance of the daily number of EMS calls (557) is more than twice as large as the mean (223). This indicates that the outcome variable is overdispersed, violating a key assumption of the Poisson model. To address this issue, we use quasi-Poisson regression in subsequent modeling. Quasi-Poisson regression relaxes the assumption that the mean and the variance are equal by introducing a scale parameter that allows for the modeling of overdispersed count data.

Chapter 3 Setting Up Predictors

We can form the quasi-Poisson regression models with two different units of analysis: runs per day or runs per day in subpopulations defined by patient characteristics in order to be able to use variables like age, sex, ADI and UHI in modeling at a patient level. The first approach requires aggregating these characteristics into day-level variables.

3.1 Patient Characteristics

Patient age and sex, as well as the UHI and ADI for each patient's EMS pick-up location comprise the patient characteristics for each EMS run. Thus, for the first modeling approach, the number of runs can be associated with the proportion of individuals from different age, sex, ADI, and UHI categories. Intuitively, if daily run counts increase with these proportions, then these variables may be risk factors. Hence, we used the proportion of EMS calls that were from females (*female_prop*), the proportion of patients utilizing EMS that were older than 65 (*age_prop*), and the proportion of runs from areas with UHI>0 (*UHI_prop*) as the aggregating representation of patient characteristics. For the subpopulation with day units, the response is the number of runs per day within each subpopulation defined by the combinations of age (< 65, ≥ 65), sex (male, female), ADI (1-6, 7-10) and UHI (0, 1-5).

3.2 Temperature

The impacts of heat on the outcome may arise from both current temperatures and lagged temperatures. These effects might not be linear and can be modeled either collectively or separately to capture their unique contributions to the response.

Through literature review, we discovered that the distributed lag non-linear model (DLNM), particularly its cross-basis term, may be helpful for modeling temperature and its lag effect simultaneously. DLNMs are developed to estimate health impact projections under climate change scenarios, and the developers of DLNM have utilized temperature-related mortality projections as a practical example to illustrate the modeling framework of DLNMs. We believe that the same methodological framework should also apply to our research, allowing us to explore the complex relationships between temperature and EMS utilization (Vicedo-Cabrera, 2019). The detailed introduction of DLNM will be provided in Section 4.1.

The cross-basis matrix in the DLNM effectively models the temperature and time using different functions and combines them without explicitly including current and lagged temperatures in the model. Theoretically, we should be able to generate a cross-basis matrix for each subgroup and find a way to incorporate them into both day-level and day-subpopulation level models. However, due to limitations in the programming settings of the DLNM in R (Gasparrini, 2011), when the unit of analysis is not the same as that on which temperature is measured (day), the cross-basis matrix cannot be constructed properly. Consequently, we are only able to incorporate the cross-basis matrix of the mean temperature into the day-level analysis.

On the other hand, modeling the impacts of current and lagged temperatures separately is applicable for both day-level and day-subpopulation analyses. We can utilize linear, polynomial, or spline forms for each, with and without interactions.

Maximum, mean, and minimum ambient temperatures have been employed in relevant studies by various researchers. Previous research suggests that no single temperature measure significantly outperforms the others when using temperature as a predictor of mortality (Barnetta, 2010). Moretti's earlier study (Moretti, 2021) examined the association of all three temperatures with EMS utilization and found no difference in their effects. Therefore, in this study, we focus on using the daily mean temperature as a representative measure of daily temperature.

3.3 Time

Both year and day of the week are coded as categorical variables in the data. They are day-level variables which can be easily incorporated in both analyses. Based on the exploratory analysis presented in Chapter 2, the distribution of EMS runs in 2020 differs from the other years. Therefore, when including calendar year as a categorical variable in our models, we will set 2020 as the reference year. As noted earlier, day of the week is included as a binary variable weekend/weekday with weekday as the reference.

3.4 Confounding

While conducting the exploratory data analysis, we observed differences in the numbers of EMS encounters across years, which leads us to further investigate the potential confounding factors in this study. We compare the crude (dark blue line) and stratified (year-specific) measures of association between the proportion of runs each day from UHI>0 areas and the daily number of EMS runs to determine the presence and extent of confounding (Figure 3.1). Figure 3.1 demonstrates differences between the stratified measures of associations and the crude measure, along with the marginal distribution displayed in boxplots. The slope of each line is displayed in Table 3.1. Table 3.1 shows that the number of EMS runs varies less with the proportion of runs from UHI>0 areas in all year-specific models compared to the association in the overall population. Since the year-specific estimates of the slope differ from the unadjusted (overall) estimate by more than 10%, we may conclude that there are potential confounding effects by calendar year (Pourhoseingholi, 2012).

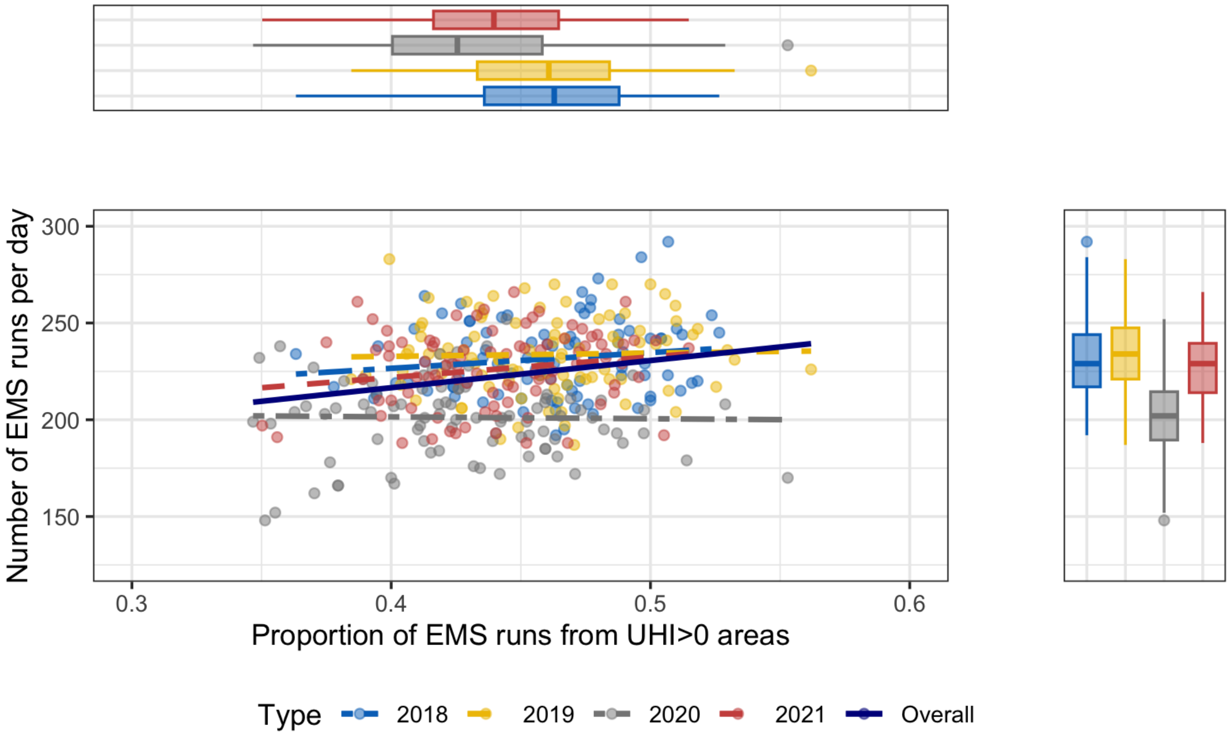
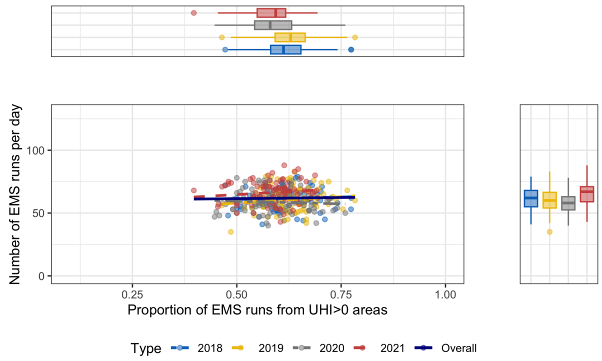


Figure 3.1: Confounding illustration using linear association between daily EMS counts and proportion of runs from UHI>0 areas with boxplots representing marginal distributions.

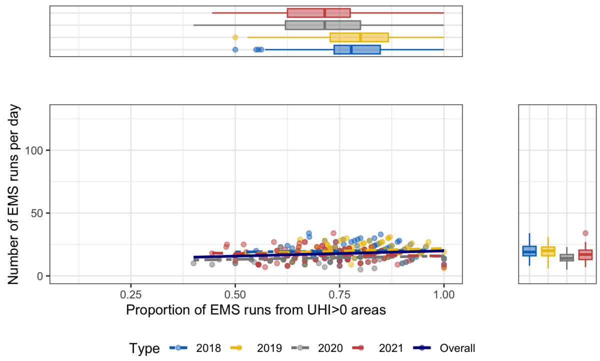
	Slope (95% CI)	**p-value**
2018	81.1 (-34.1, 196.4)	0.17
2019	16.8 (-97.5, 131.1)	0.77
2020	-9.7 (-109.4, 90.1)	0.85
2021	106.2 (-10.2 222.5)	0.08
Overall	140.4 (80.2, 200.6)	<0.001

Table 3.1: Summary of five slopes of linear regressions of the daily number of EMS runs on the proportion of runs for UHI>0 areas (overall)

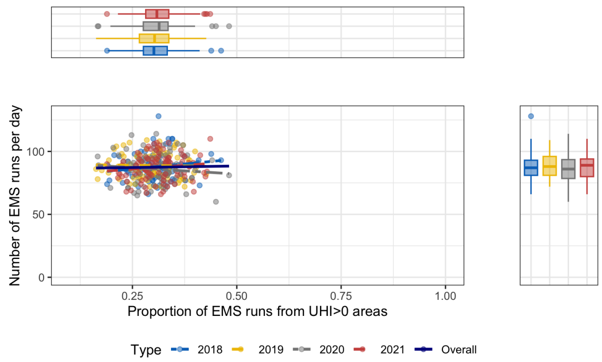
The cause of the confounding can be gleaned by noting that the number of runs and



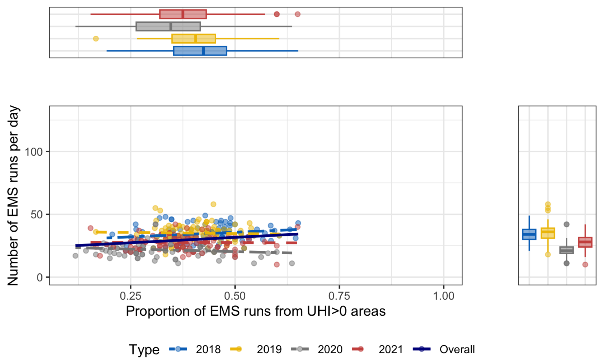
(a) Residential, high ADI areas.



(b) Non-residential, high ADI areas.



(c) Residential, low ADI areas.



(d) Non-residential, low ADI areas.

Figure 3.2: Confounding illustration using linear association between daily EMS counts and proportion of runs from UHI>0 areas with boxplots representing marginal distributions, in (a) residential, high ADI, (b) non-residential, high ADI, (c) residential, low ADI, (b) non-residential, low ADI areas.

the proportion of runs from high UHI areas were lower during the pandemic in 2020 than in other years. The overall slope of the association ignoring year then confounds this time trend with the within-year associations.

We also created similar figures and tables across residential/non-residential and low/high ADI areas, shown in Figure 3.2 and Table 3.2. First, we note that the daily number of EMS calls are generally greater from residential areas, and the proportions of runs from high ADI areas are higher in UHI>0 regions (Table 2.2). Table 3.2 indicates that the confounding effects by calendar year disappear in residential areas but still exist in non-residential areas.

We also note that the 95% confidence intervals are wider in the residential analyses

Year	(a) Residential, high ADI		(b) Non-residential, high ADI		(c) Residential, low ADI		(d) Non-residential, low ADI	
	Slope (95% CI)	p-value	Slope (95% CI)	p-value	Slope (95% CI)	p-value	Slope (95% CI)	p-value
2018	7.66 (-21.9,37.3)	0.61	2.50 (-9.13,14.12)	0.67	31.59 (-11.73, 74.90)	0.16	14.64 (1.16,28.11)	0.04
2019	20.26 (-11.6,52.1)	0.22	8.08 (-2.57,18.73)	0.14	-4.44 (-40.23,31.35)	0.81	-3.45 (-24.35,17.46)	0.75
2020	-8.85 (-33.4,15.7)	0.48	5.40 (-0.83,11.63)	0.09	-20.50 (-62.99,21.99)	0.35	-8.04 (-18.01,1.93)	0.12
2021	18.91 (-11.3,49.1)	0.22	-4.29 (-14.04,5.46)	0.39	24.34 (-16.18,64.85)	0.24	-1.05 (-13.82,11.72)	0.87
Overall	3.44 (-11,17.9)	0.64	8.37 (3.55,13.19)	<0.001	4.80 (-15.40,25.0)	0.64	17.21 (8.81, 25.6)	<0.001

Table 3.2: Summary of five slopes of linear regressions of the daily number of EMS runs on the proportion of runs for UHI>0 areas in (a) residential, high ADI, (b) non-residential, high ADI, (c) residential, low ADI, (b) non-residential, low ADI areas.

than in non-residential areas, despite the greater number of runs in residential areas. Table 3.3 displays the mean and standard deviation of the daily number of EMS runs by year and overall in each of the four residential/non-residential and low/high ADI areas. The standard deviations of the runs are larger in the residential areas, leading to wider confidence intervals in these areas due to the greater variances.

Year	(a) Residential, high ADI	(b) Non-residential, high ADI	(c) Residential, low ADI	(d) Non-residential, low ADI
2018	62 (9)	20 (6)	88 (11)	35 (7)
2019	61 (9)	20 (5)	89 (10)	35 (7)
2020	59 (8)	14 (4)	86 (12)	22 (5)
2021	66 (9)	17 (5)	87 (10)	28 (6)
Overall	62 (9)	18 (6)	87 (11)	30 (8)

Table 3.3: Summary of five slopes of linear regression in (a) residential, high ADI, (b) non-residential, high ADI, (c) residential, low ADI, (b) non-residential, low ADI areas.

To control for confounding and effects of pandemic in our analysis, we specifically focus on EMS runs from residential areas. During the pandemic, most people worked from home and spent most of their time within residential areas. Moreover, to focus on the most vulnerable populations most likely to lack adaptive strategies such as air conditioning, we further restrict the analysis to runs from regions with ADI scores greater than or equal to 7. Individuals residing in these communities are typically of lower socioeconomic status, less likely to have access to air conditioning, or more likely to be employed in jobs requiring physical or manual labor. These factors make them more vulnerable to the direct effects of extreme heat exposure and necessitate specialized attention. Therefore, for all the models in the subsequent chapters, we focus only on runs from residential areas with high ADI scores.

Chapter 4 Models

We developed three quasi-Poisson models, integrating the modeling approaches for different types of variables discussed in Chapter 3. Section 4.1 introduces the day-level modeling with the cross-basis matrix from the DLNM for mean temperature, while section 4.2 outlines the day-level modeling with common functional forms for mean temperature for comparison purposes. Finally, section 4.3 delves into the day-subpopulation level modeling using standard functional forms for mean temperature. By comparing the first and third models with the second one, we can determine whether DLNM is helpful in this study, and the difference in results between day and day-subpopulation level analyses.

4.1 Model 1: Day-Level with DLNM

This section describes the quasi-Poisson model with day as the unit of analysis and the DLNM formulation of temperature. The response is the daily count of EMS runs, with each record representing a day alongside all other day-level variables, and the denominator of the daily EMS call rate is the entire Rhode Island population. We assume that the Rhode Island population remains static over the 4-year period (2018-2021), so we do not need to include an offset term for the number at risk on each day.

As discussed in Chapter 1, heat exposure may have lagged effects on individual health, and the impact of the exposure, along with its potential lagged effects on the response, may also be non-linear. Thus, it is crucial to estimate the exposure-lag-response for health outcome assessments. Since risk depends on both the intensity and timing of prior exposures, the primary challenge in modeling lies in incorporating the additional temporal dimension required to describe this association. The distributed lag non-linear model (DLNM) proposed by Gasparrini and colleagues is designed to simultaneously account for non-linear relationships between the exposure and outcome variables, as well as delayed effects over time (Gasparrini, 2010). The DLNMs describe the intricate non-linear and lagged dependencies between the response variable y_{t0} measured at time t_0 and the lagged exposure variables x_{ti} through a cross-basis function. l_0 and L are the lower and upper limits of the range of time lags, l, between an exposure X_{t-l} and the response y_t at time t, i.e., $l \in [l_0, L]$. The time between $t - L$ and $t - l_0$ represents the lag period during which an exposure to x is assumed to affect the risk at time t. The association of interest is represented by the cross-basis function s, a two-dimensional space of functions created by combining the basis functions of the exposure intensity at exposure time and its lagged effects. The general form of the cross-basis function s can be expressed as:

$$s(x,t) = \int_{l_0}^{L} f(x_{t-l}) \cdot w(l)\ dl \approx \sum_{l=l_0}^{L} f(x_{t-l}) \cdot w(l).$$

The bi-dimensional function $f(x_{t-l}) \cdot w(l)$ is referred to as the exposure-lag-response function, where $f(x_{t-l})$ models the nonlinear exposure-response curve at a given time $t - l$, and $w(l)$ denotes the weight given to the exposure at lag l (time $t - l$) (Gasparrini, 2014). DLNM is able to use an arbitrary basis function in each space. The estimates obtained through cross-basis and corresponding interpretation will be explained in Section 5.1 along with the results.

Because the database includes records only from the summer months (June, July, and August), it comprises four disjoint time series The *group* argument in the *crossbasis()* function identifies the variable that defines these multiple time series. To address this issue, we specify argument *group* as calendar year while generating the cross-basis matrices.

This enables the function to segment the data at the end of each year. If we set the maximum lag day to L, then the values in the first L rows of each time series in the cross-basis matrix are automatically set to missing, as the cross-basis matrix for each series is constructed separately.

In this study, the environmental exposure and the outcome correspond to historical series of daily mean temperature and daily EMS utilization counts. The EMS-temperature association is modeled by using the cross-basis matrix from the DLNMs to describe non-linear and lagged dependencies. We modeled the exposure-response curve f with a natural cubic spline with three internal knots placed at the 10th, 75th and 90th percentiles of the temperature distributions, and the lag-exposure curve w with a natural cubic spline with three internal knots placed at equally spaced values in log scale. An extra intercept is always included in the basis for the space of lags. The minimum lag l_0 is set to 0 and the maximum lag day L set to 5, according to previous studies. Therefore, we ended up with a cross-basis matrix with $4 \times 5 = 20$ columns, and the values in the first five rows of each time series are set to missing. The general form of this model can be written as:

$$\begin{aligned}\log(\mathrm{E}(Y_t)) = &\beta_0 + C\boldsymbol{\beta_1} + \beta_2 \mathrm{UHI\_prop} + \beta_3 \mathrm{age\_prop} + \beta_4 \mathrm{female\_prop} + \beta_5 \mathrm{weekend} \\ &+ \beta_6 \mathrm{year} + \mathrm{Interaction\ terms},\end{aligned}$$

where Y_t is the observed daily number of EMS runs, β_0 is the intercept, C is the cross-basis matrix for daily mean temperature generated by the *dlnm* package (Gasparrini, 2011). The three aggregated patient characteristics variables are defined in Section 3.1, and the two time-related variables are defined in Section 3.3. The test of all two-way interactions is conducted using the Chi-squared test, comparing models with and without the interaction term. At a significance level of $\alpha = 0.05$, a p-value<0.05 implies that the inclusion of the interaction term leads to a significant improvement in model performance, and hence we should incorporate this interaction term. Consequently, the final model will include all main effects and significant interactions.

The *dlnm* package (Gasparrini, 2011) was used in the analysis for fitting DLNMs.

4.2 Model 2: Day-level with Current and Lagged Temperature

For the day-level analysis, our second approach utilizes common functional forms for temperature, serving as an intermediate model. We can then compare this model to the day-level analysis using DLNM to model the EMS-temperature association, as well as to the day-subpopulation level analysis which also employs common functional forms for temperature. The response is the daily count of EMS runs, and by assuming that the state population remains static over four years, no offset term is needed. The general form of the day-level quasi-Poisson model can be expressed as:

$$\begin{aligned}\log(E(Y_t)) = &\beta_0 + \boldsymbol{T}\beta_1 + \beta_2\text{UHI\_prop} + \beta_3\text{age\_prop} + \beta_4\text{female\_prop} + \beta_5\text{weekend} \\ &+ \beta_6\text{year} + \text{Interaction terms},\end{aligned}$$

where Y_t is the observed daily number of EMS runs, β_0 is the intercept, and T composed of T_0, T_1 and $T_0 * T_1$ where T_0 are the current temperatures and T_1 are lagged temperatures. We examine five functional forms for the temperatures: 1) linear scale; 2) natural cubic spline with 2 degrees of freedom; 3) natural cubic spline with three internal knots placed at the 10th, 75th, and 90th percentiles of the temperature distribution; 4) second-order polynomials; and 5) third-order polynomials. Additionally, we explored up to 5 days of lagged temperatures. We tested these five functional forms for both current and lagged temperatures. T is hence formed by choosing the same functional form for all temperatures at different times. Interactions among temperatures are also tested to account for the potential multiplicative effects of consecutive heat exposure, such as heatwaves. Other covariates are defined in Chapter 3, and variables with suffix *_prop* are aggregated individual-level patient characteristics. The test of the all the two-way interactions is conducted with the Chi-squared test, comparing models with and without the interaction term, and the final model will include all the main effects and significant interactions, with significance level $\alpha = 0.05$.

4.3 Model 3: Day-subpopulation level with Current and Lagged Temperature

An alternative approach is to use a day-subpopulation level analysis. The quasi-Poisson regression model with data grouped by subpopulation offers the advantage of incorporating individual-level covariates into the model, provided we have information on the population in each subgroup.

To implement this model, we select subpopulations defined by gender (male/female), age groups (< 65, ≥ 65), and UHI (non-UHI/UHI). The response variable of the analysis now becomes the number of EMS encounters from high ADI residential areas each day within each age, gender, and UHI subgroup, such as the number of runs from high ADI residential areas that are males less than 65 years old, residing in UHI areas. Therefore, for each day, we have $2 \times 2 \times 2 = 8$ run counts representing the different subgroups.

The general form of the quasi-Poisson model can be expressed as:

$$\begin{aligned}\log(\mathrm{Y}) = {} & \beta_0 + \boldsymbol{T}\beta_1 + \beta_2\mathrm{year} + \beta_3\mathrm{weekend} + \beta_4\mathrm{gender} + \beta_5\mathrm{age} + \beta_6\mathrm{UHI} \\ & + \mathrm{Interaction\ terms} + \log(\mathrm{Est.Pop}),\end{aligned}$$

where Y is the observed daily number of EMS runs in each subgroup, and β_0 is the intercept. T denotes the temperatures. Similar to Model 2, we evaluated five functional forms: 1) linear scale; 2) natural cubic spline with 2 degrees of freedom; 3) natural cubic spline with three internal knots placed at the 10th, 75th, and 90th percentiles of the temperature distribution; 4) second-order polynomials; and 5) third-order polynomials for temperatures and examined up to 5 days of lagged temperatures. T is determined by using the same functional form for all temperatures at different times. Interactions among temperatures were tested to account for multiplicative effects like heatwaves. Other covariates are defined in Chapter 3. All two-way interactions among covariates were also tested, and those with p-values less than 0.05 were kept. *log(Est.Pop)* is the offset term, serving as the denominator of the rate of the EMS calls for each subgroup.

To obtain the offset for the Poisson models, information on how the Rhode Island population is distributed across each subgroup is required. Population distribution by age group and gender from high ADI areas is available from census block group population data. However, information on the proportion of the population residing in UHI areas is missing.

<table>
<tr><th rowspan="2">Gender</th><th rowspan="2">Age Category</th><th colspan="3">EMS Runs</th><th rowspan="2">RI Population</th></tr>
<tr><th>non-UHI</th><th>UHI</th><th>Total</th></tr>
<tr><td rowspan="2">Male</td><td><65</td><td>3,456 (33%)</td><td>7,036 (67%)</td><td>10,492</td><td>170,981</td></tr>
<tr><td>≥65</td><td>1,888 (42%)</td><td>2,611 (58%)</td><td>4,499</td><td>24,041</td></tr>
<tr><td rowspan="2">Female</td><td><65</td><td>3,526 (34%)</td><td>6.796 (66%)</td><td>10,322</td><td>168,953</td></tr>
<tr><td>≥65</td><td>2,689 (42%)</td><td>3,681 (58%)</td><td>6,370</td><td>33,122</td></tr>
</table>

Table 4.1: Distribution of EMS runs and RI population in high ADI areas across gender, age, and UHI.

EMS run records provide the distribution of the population into these subgroups for individuals taken to the hospital. Table 4.1 summarizes the distribution of EMS runs and RI population from high ADI areas, categorized by gender, age groups and dichotomized UHI. Notably, for EMS encounters from high ADI areas, the probability of also being in a UHI appears consistent across genders, but varies among age groups. According to Table 4.1, among those under 65, about 67% of the EMS runs are from UHI areas compared with 58% among those over 65, regardless of gender. Some prior studies conducted in

Rhode Island and other countries have also indicated that the association between ambient temperatures and all-cause emergency department (ED) admissions, as well as mortality and morbidity, tends to be most pronounced among older age groups, while the impacts appear to be more similar across sex (Kingsley, 2016; Ragettli, 2019; Stafoggia, 2006).

It seems reasonable based on Table 4.1 and insights from previous research, to assume that in high ADI areas, the distribution of population between UHI and non-UHI regions is independent of gender within each age group. Thus we only need to determine the distribution of population by UHI group, conditional on age groups, i.e. P(UHI|age).

	Age <65	**Age ≥65**	**Total**
non-UHI	$(1-b)(N-x)$	$(1-a)x$	$N(1-b)+(b-a)x$
UHI	$b(N-x)$	ax	$N \cdot b+(a-b)x$
Total	$N-x$	x	N

Table 4.2: Formulas for population estimation give arbitrary P(UHI|age≥65)=a and P(UHI|age<65)=b, assuming N individuals in total and x are ≥65 years old.

Assuming the conditional probabilities of being in UHI areas given age are known such that P(UHI|age≥65)=a and P(UHI|age<65)=b, we can derive the population in each age-UHI subgroup as in Table 4.2. From census data, the Rhode Island population living in high ADI areas is $N = 297,097$, with $x = 57,163$ individuals aged≥65 years old.

Chapter 5 Results from Three Models

Three quasi-Poisson models were constructed in Chapter 4. Model 1 conducts day-level analysis using the cross-basis matrix from the DLNM for temperature and its lagged effects. Model 2 also operates at the day-level but employs common functional forms for both current and lagged temperatures. Meanwhile, Model 3 performs day-subpopulation level analysis with common functional forms for current and lagged temperatures. The subsequent sections present the three final models fitted and their results. The variables included and the specific forms of temperatures used in Models 2 and 3 are also explained.

5.1 Model 1: Day-level with DLNM

After testing all potential two-way interactions, two interactions associated with year, namely year:age_prop and year:female_prop, were found to have significant contributions to the model's performance. Hence, the final form of Model 1 can be written as:

$$\begin{aligned}\log(\mathrm{E}(Y_t)) = &\beta_0 + C\boldsymbol{\beta_1} + \beta_2\text{UHI\_prop} + \beta_3\text{age\_prop} + \beta_4\text{female\_prop} + \beta_5\text{weekend} \\ &+ \beta_6\text{year} + \beta_7\text{year:age\_prop} + \beta_8\text{year:female\_prop}\end{aligned}$$

Model 1 estimates the risk of EMS events at different temperatures relative to the reference temperature given by the minimum observed temperature, which is 52 degrees Fahrenheit.

The three-dimensional plot in Figure 5.1 represents a grid of relative risks for each combination of lag and potential temperature values. Each point on this surface signifies the individual contribution of risk for the specified lag day and temperature. For instance, to summarize the effect of temperature across all five previous days, sum the effect of each temperature at its corresponding lag from the surface in Figure 5.1. This yields the additive effects of temperature on the outcome. For example, if temperatures on the previous five days were 75, 80, 85, 80, 70, then the total effect of temperature on the current day would be the sum of the effects of temperatures of 75 degrees at lag 4, 80 at lag 3, 85 at lag 2, 80 at lag 1 and 70 at lag 0.

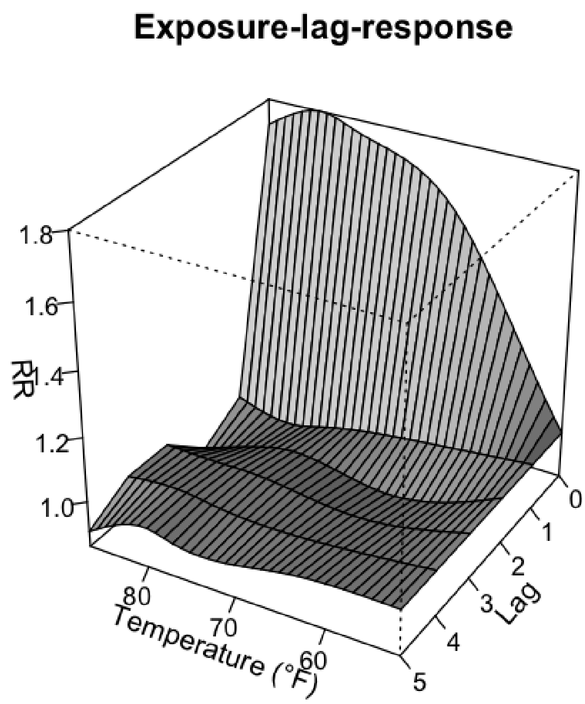


Figure 5.1: Three-dimensional plot showing the estimated exposure-lag–response association between temperature and EMS in high ADI areas

Note that the majority of the contribution to the effect of temperature occurs at lag=0; the effects at other lags are much smaller. This suggests that temperature may not exhibit a lagged effect on EMS runs. Figure 5.2 shows a slice of Figure 5.1 at lag=0. The relative risk (RR) of 1.89 at 80 degrees Fahrenheit, for example, indicates that the number of EMS runs on a day with a temperature of 80 degrees would be 1.89 times higher on average than that on a day with a temperature of 52 degrees.

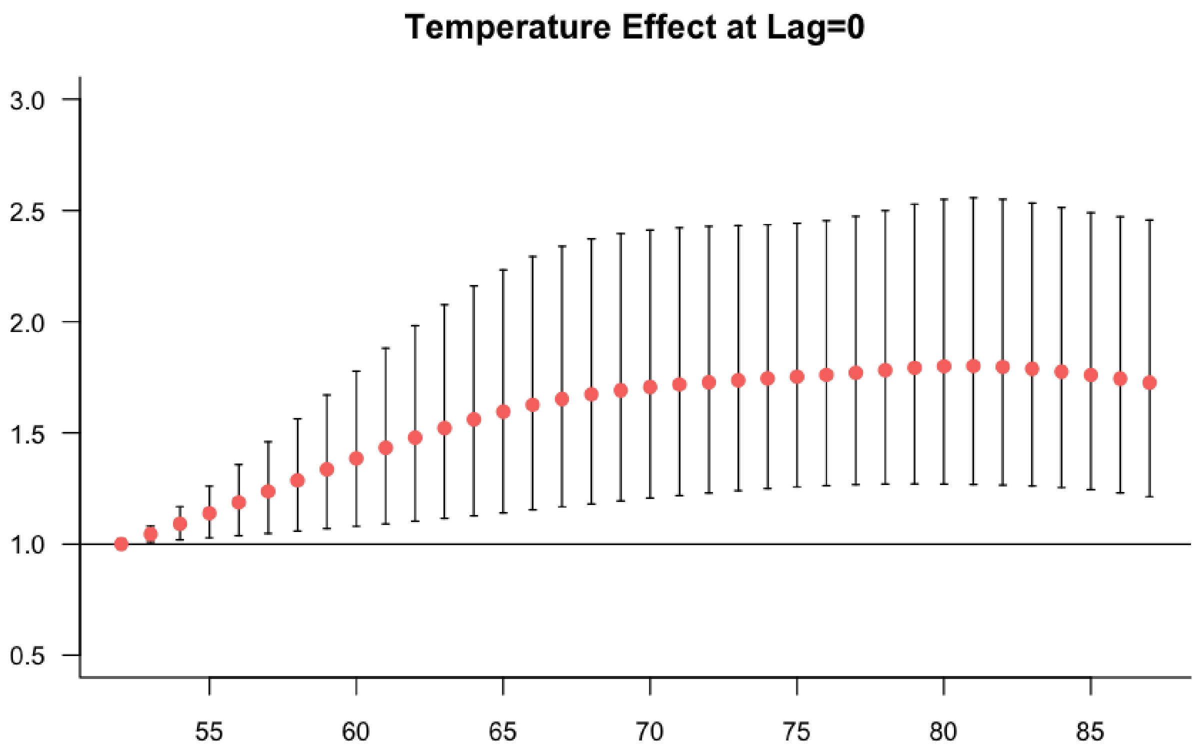


Figure 5.2: Temperature-EMS association at lag=0.

Table 5.1 presents the estimated impacts of covariates other than temperatures, along with their standard errors, and p-values. The effect of the proportion of EMS runs from UHI areas is 0.182 (RR: 1.20, 95% CI [0.94, 1.53]). Areas with a greater proportion of inhabitants who live in urban heat islands are estimated to have more runs, although this effect is not significant. The estimated coefficient of *weekend* of -0.047 suggests that the daily EMS call rate on weekends is 0.95 (95% CI [0.923, 0.985]) times that on a weekday. In general, the number of EMS encounters is significantly higher in 2018, 2019, and 2021 compared to 2020. The effects of the other two aggregated patient characteristics, proportion of old patients and female patients, vary across each year. In 2020 and 2021, areas with more residents aged older than 65 years are likely to have more runs, while areas with fewer residents aged older than 65 years are likely to have more runs in 2018 and 2019, although the impacts are minor. In 2019 and 2020, areas with more female residents are likely to have more EMS encounters, while areas with less female residents are likely to have more encounters in 2021, and almost no effect was found in 2018.

The predicted versus observed number of EMS runs is plotted in Figure 5.3, with mean squared error (MSE) of 61.2.

Variable	Coefficient	Standard Error	p-value
UHI_prop	0.182	0.123	0.141
age_prop≥65	0.342	0.232	0.143
female_prop	0.557	0.252	0.028
Weekend	-0.047	0.017	0.005
year2018	0.727	0.216	0.001
year2019	0.571	0.220	0.010
year2020	0	-	-
year2021	0.621	0.236	0.009
year2018 : age_prop≥65	-1.054	0.341	0.002
year2019 : age_prop≥65	-0.929	0.332	0.005
year2021 : age_prop≥65	0.015	0.313	0.961
year2018 : female_prop	-0.555	0.348	0.112
year2019 : female_prop	-0.367	0.343	0.286
year2021 : female_prop	-0.955	0.367	0.010

Table 5.1: Summary of results of UHI_prop, weekday/weekend, year, age_prop, and female_prop from Model 1.

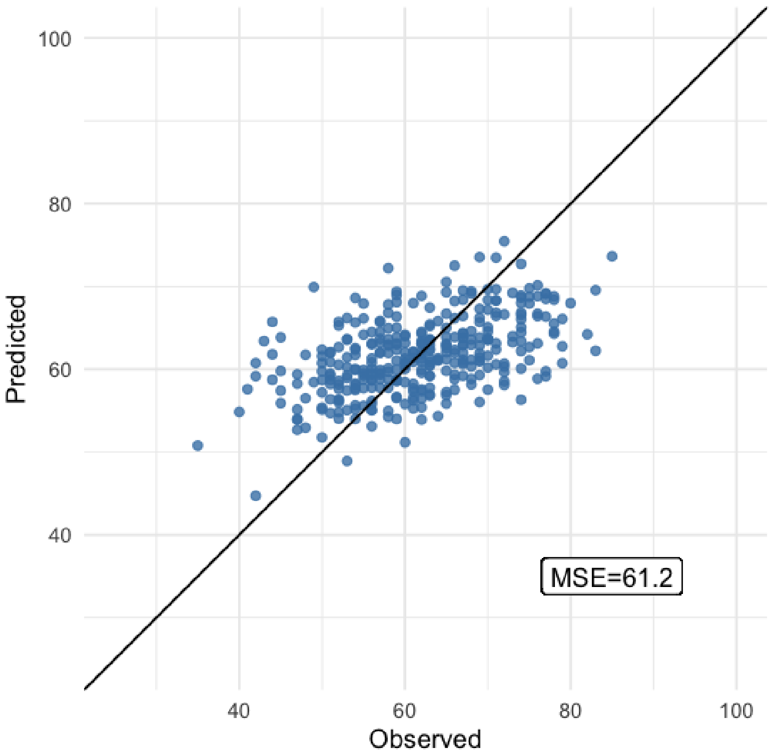


Figure 5.3: Predicted vs. observed daily number of EMS runs from Model 1.

5.2 Model 2: Day-level with Current and Lagged Temperature

Considering the potential lagged effects of temperature on EMS encounters, we initially investigated whether to include lagged temperatures and determined the appropriate number of lags to use. For both current and lagged temperatures, we examined them using the five proposed functional forms. By including temperatures lagged up to five days and examining interactions among temperature variables, we found that the temperature's impact is significant only when the temperature of the current day is considered without lags. The predicted EMS call rate relative to daily mean temperatures, using the five different functional forms of temperature, along with a histogram of the temperature distribution, is displayed in Figure 5.4. To calculate the predicted counts from the model, other covariates were fixed at as UHI_prop=0.5, weekday/weekend (weekday), and year

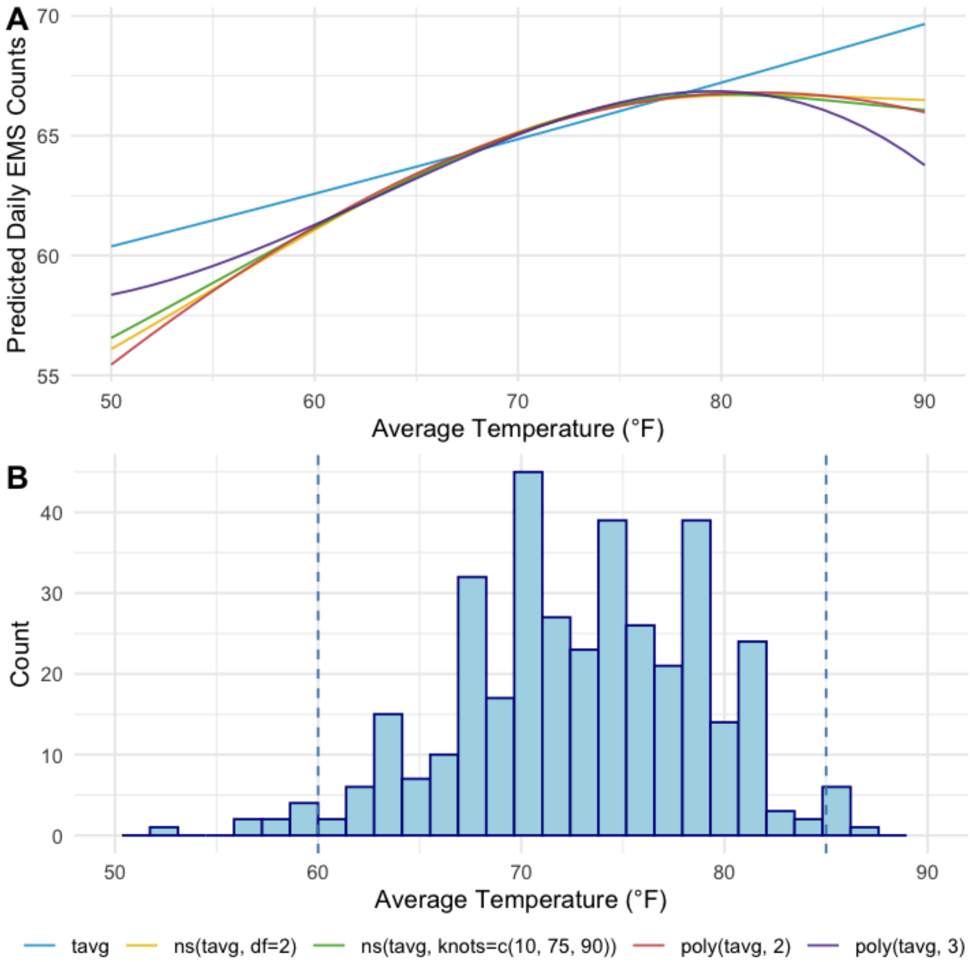


Figure 5.4: A: variation in predicted daily mean number of EMS runs. The five curves represent 5 functional forms of daily mean temperature, with the remaining factors in the model fixed. The five functional forms are (1) linear scale; (2) natural cubic splines with 2 degrees of freedom; (3) natural cubic splines with 3 internal knots place at 10, 75, 90 percentiles of the temperature distribution; (4) quadratic; and (5) cubic. Fixed factors are: year (2021), weekday/weekend (weekday), UHI_prop=0.5, age_prop=0.5, female_prop=0.5. B: the distribution of daily mean temperatures.

(2021) as the effect of temperature was modeled independently of the other variables. Within the range of the most concentrated daily mean temperatures (62-82), the predicted counts are similar for all five functional forms, varying by no more than 2 runs per day. We also tested all potential two-way interactions, and only one interaction between *year* and *age_prop* was significant. Therefore, we opted to include only the linear daily mean temperature (*tavg*) without any lagged temperatures and one interaction term in our model. Model 2 can be expressed as:

$$\begin{aligned}\log(E(Y_t)) =& \beta_0 + \beta_1 \text{tavg} + \beta_2 \text{UHI\_prop} + \beta_3 \text{age\_prop} + \beta_4 \text{female\_prop} + \beta_5 \text{weekend} \\ &+ \beta_6 \text{year} + \beta_7 \text{year:age\_prop}\end{aligned}$$

Variable	Coefficient	Standard Error	p-value
Daily mean temperature	0.004	0.001	0.003
UHI_prop	0.152	0.117	0.195
age_prop≥65	0.348	0.226	0.125
female_prop	0.098	0.121	0.421
weekend	-0.044	0.016	0.007
year2018	0.438	0.122	<0.001
year2019	0.355	0.123	0.004
year2020	0	-	-
year2021	0.072	0.117	0.537
year2018 : age_prop≥65	-1.053	0.326	0.001
year2019 : age_prop≥65	-0.861	0.322	0.008
year2021 : age_prop≥65	0.113	0.308	0.714

Table 5.2: Summary of results of daily average temperature, UHI_prop, weekday/weekend, year, age_prop, and female_prop from Model 2.

The results from Model 2 are presented in Table 5.2, including the estimated coefficients, standard errors, and p-values for daily mean temperature, proportion of EMS runs from UHI areas, proportion of patients aged over 65 years old, proportion of female patients, weekday/weekend, calendar year, and interaction between year and age. Higher temperatures increase the daily EMS call rate; a 10 degree increase is associated with an increase in the EMS call rate of 4%. Regression coefficients for other variables are similar to those from Model 1, except for the interaction between year and proportion of females. EMS encounter rates are elevated in 2018, 2019, and 2021 relative to 2020. EMS encounters on weekends are significantly less by about 5 percent compared to weekdays. Areas with higher proportions of urban heat islands and more females have non-significantly higher EMS run rates. As with Model 1, Model 2 suggests that areas with fewer residents aged older than 65 years are likely to have more runs in 2018 and 2019, while areas with more residents aged older than 65 years are expected to have more runs in 2020 and 2021.

The predicted versus observed number of EMS runs is plotted in Figure 5.5, with MSE=68.7.

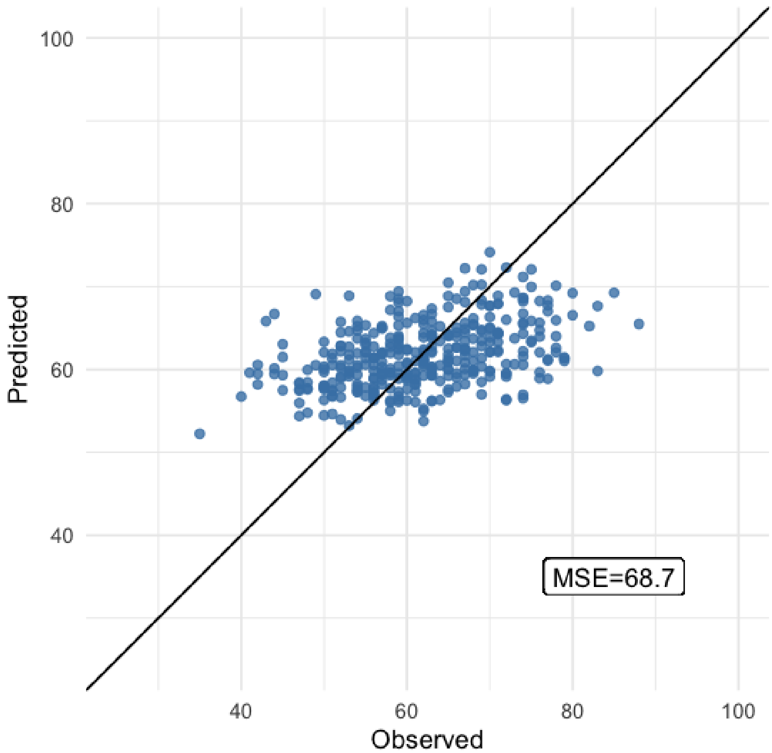


Figure 5.5: Predicted vs. observed daily number of EMS runs from Model 2.

5.3 Model 3: Day-subpopulation level with Current and Lagged Temperature

Similar to Model 2, we started with examining whether to include lagged temperatures, determining the appropriate number of lags to use, and the best functional forms to use for temperatures. By including temperatures lagged up to five days and examining interactions between temperature variables, we found that the temperature's impact is significant only when the temperature of the current day is considered without lags. The predicted number of EMS runs per 10,000 population in the given subgroup relative to daily mean temperatures, using the five different functional forms of temperature, is displayed in Figure 5.6. The curves have shapes very similar to those for temperature in Model 2. Here, we assume P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5. Other covariates in Figure 5.6 were set as follows: UHI, weekday, year 2021, female, ≥65. Within the range of the most concentrated daily mean temperatures, between 62 and 82 degrees, the predicted counts are similar for all five functional forms, varying by less than 1 run per 10,000 people per day in the given subgroup.

After testing all two-way interactions among daily mean temperature, calendar year, weekday/weekend, sex, age, and UHI, we found that the interactions year:UHI, age:UHI, year:age, and weekday:age were significant at $\alpha = 0.05$ and should be included in the model. Given that no interaction term associated with temperature was found, the shape of the curve in Figure 5.6 will not change if we set different values for these fixed factors. Therefore, we opted to include only the linear daily mean temperature without any lagged temperatures, and four interactions in our model. Thus, Model 3 can be written as:

$$\begin{aligned}\log(\mathrm{Y}) = {} & \beta_0 + \beta_1\mathrm{tavg} + \beta_2\mathrm{year} + \beta_3\mathrm{weekend} + \beta_4\mathrm{gender} + \beta_5\mathrm{age} + \beta_6\mathrm{UHI} \\ & + \beta_7\mathrm{year{:}UHI} + \beta_8\mathrm{age{:}UHI} + \beta_9\mathrm{year{:}age} + \beta_{10}\mathrm{weekend{:}age} + \log(\mathrm{Est.Pop}).\end{aligned}$$

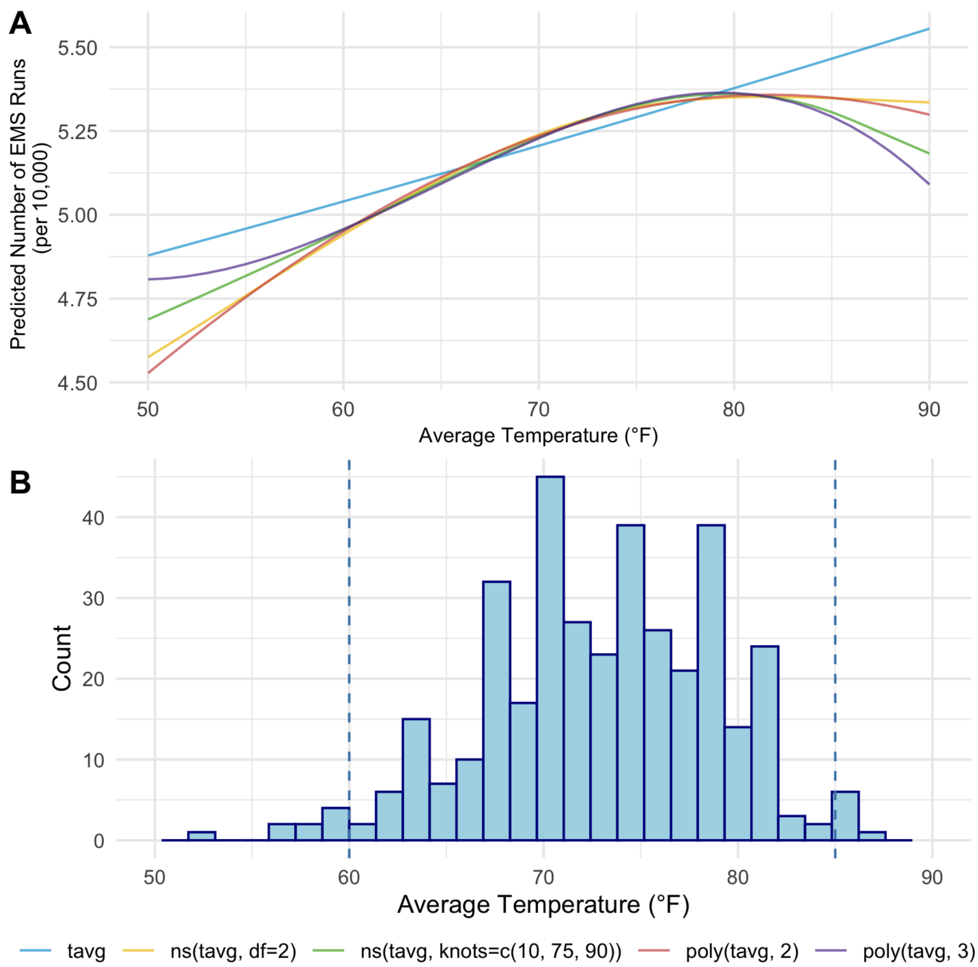


Figure 5.6: A: variation in predicted probabilities of EMS runs when P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5. The five curves represent 5 functional forms of daily mean temperature, with the remaining factors in the model fixed. The five functional forms are (1) linear scale; (2) natural cubic splines with 2 degrees of freedom; (3) natural cubic splines with 3 internal knots place at 10, 75, 90 percentiles of the temperature distribution; (4) quadratic; and (5) cubic. Fixed factors are: year (2021), weekday/weekend (weekday), gender (female), age category (≥65), UHI. B: the distribution of daily mean temperatures.

As discussed in Section 4.3, the subpopulation estimates for the 8 age-sex-UHI subgroup offsets in the quasi-Poisson model required assumptions about P(UHI|age≥65) and P(UHI|age<65). To estimate the population, we varied both probabilities from 0.05 to 0.95 in increments of 0.05 (19 values for each), then ran quasi-Poisson regressions for each setting. This process resulted in 361 simulations with different combinations of P(UHI|age≥65) and P(UHI|age<65). We observed that only the coefficients of age, UHI, and age:UHI were sensitive to the varying P(UHI|age≥65) and P(UHI|age<65).

The overall effect of UHI (sum of *UHI*, *age:UHI* and *year:UHI*) on individuals are of our main interest. Specifically, for those aged 65 years and older, the choice of P(UHI|Age<65) is irrelevant. Conversely, when focusing on those younger than 65 years, P(UHI|Age≥ 65) does not matter as the model predicts for only one of these groups. Thus, for each subgroup, only one set of probabilities is relevant, while the other is not. Figure 5.7 shows the distribution of the overall effect of UHI on individuals in two age groups over four years, varying the values of P(UHI|age).

In general, for individuals younger than 65 years old, the threshold separating positive and negative effects of UHI falls between a P(UHI|age<65) of 0.6 and 0.7. If the actual proportion of the younger population living in UHI areas is lower than this threshold, then living in a UHI increases the daily number of EMS runs. Conversely, if the actual P(UHI|age<65) exceeds the threshold, living in a UHI decreases the number of runs. For those aged 65 years and older, the threshold for is between P(UHI|age≥65) of 0.5 and 0.6. If the actual proportion of the elderly living in UHI areas is below this threshold, living in a UHI increases the daily number of EMS runs, while if the actual P(UHI|age<65) is greater than the threshold, then living in a UHI decreases the number of runs.

Aside from the coefficients of *UHI*, *age* and their interaction, the effects of other covariates in the model remained constant across the different combinations of P(UHI|age). Table 5.3 displays the estimated coefficients, standard errors, and p-values for daily mean temperature, year, weekday/weekend, gender, and all interaction terms when

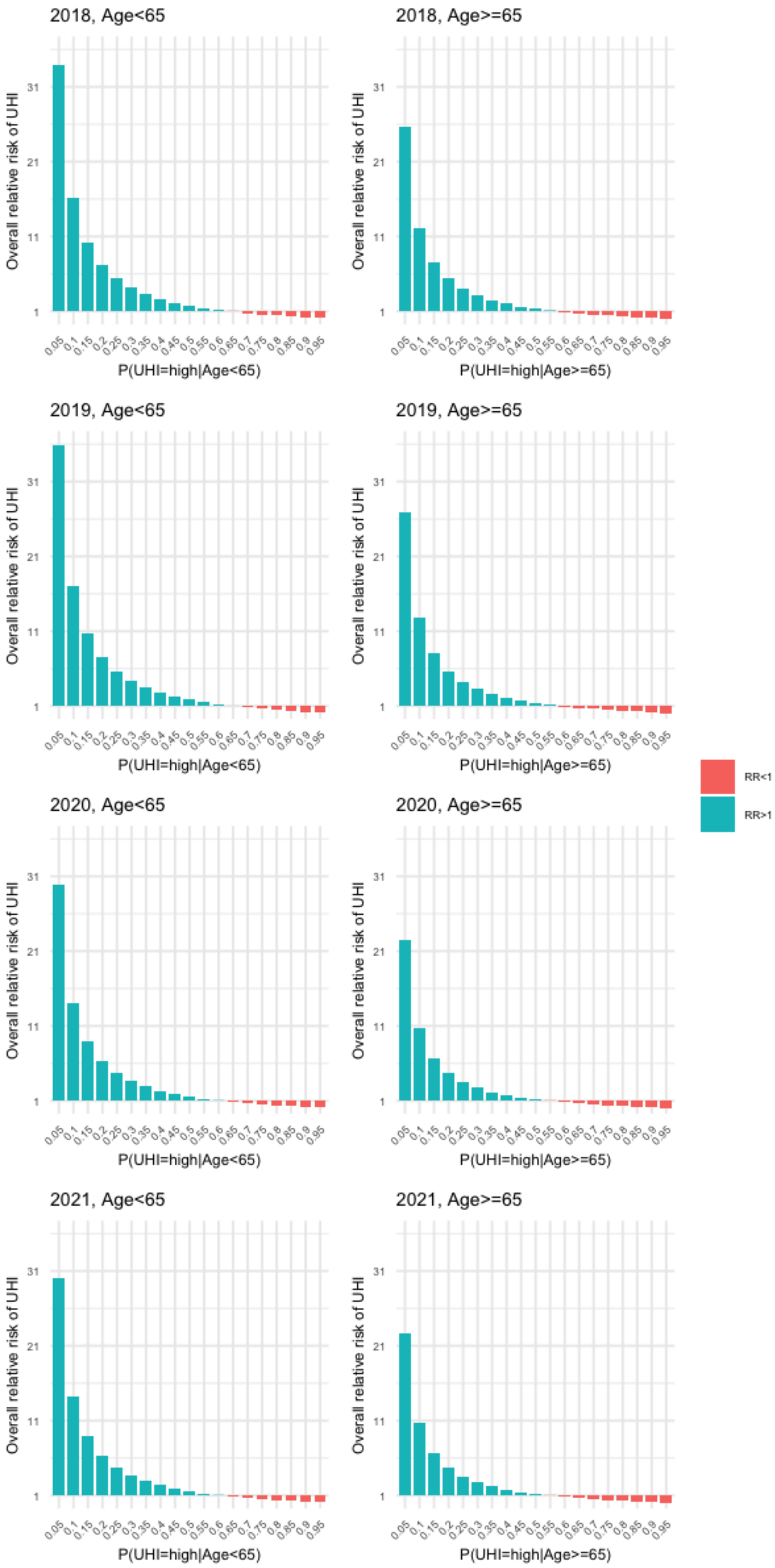


Figure 5.7: Distribution of overall effect of *UHI* on across different values of P(UHI|age≥65) and P(UHI|age<65).

P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5. The effect of temperature on the daily EMS call counts is positive, suggesting that a 10-degree Fahrenheit increase in temperature leads to a 3% rise in the predicted number of EMS runs in each subgroup per 10,000 population. Given the presence of interactions among other variables, we present predictions for 64 scenarios (4 years × 2 weekday/weekend × 2 gender × 2 age categories × 2 UHI) in Figure 5.8.

Variable	Coefficient	Standard Error	p-value
Daily mean temperature	0.003	0.001	0.004
year2018	-0.046	0.035	0.184
year2019	-0.111	0.035	0.002
year2020	0	-	-
year2021	0.087	0.033	0.009
Weekend	-0.020	0.017	0.246
Female	0.090	0.014	<0.001
year2018 : UHI	0.130	0.039	0.001
year2019 : UHI	0.183	0.040	<0.001
year2021 : UHI	0.008	0.038	0.842
year2018 : age≥65	0.046	0.040	0.246
year2019 : age≥65	0.098	0.040	0.014
year2021 : age≥65	0.076	0.039	0.052
Weekend : age≥65	-0.082	0.028	0.004

Table 5.3: Summary of estimation results of daily mean temperature, year, weekday/weekend, gender, and interaction terms when P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5.

From Figure 5.8, we can observe small differences in the predicted number of EMS runs per 10,000 population between UHI and non-UHI areas, across age groups, and over the years. Individuals aged older than 65 years are expected to have more runs. The distinction between weekdays and weekends has a relatively minor impact, with fewer runs on weekends, especially among those 65 and older. Additionally, there is minimal variation in the predicted EMS runs between male and female individuals, with slightly more runs from females.

Similar prediction plots are available in Appendix A for scenarios where 1) P(UHI|age$\geq$65)=0.2 and P(UHI|age<65)=0.2 (Figure A.1), and 2) P(UHI|age$\geq$65)=0.8 and P(UHI|age<65)=0.8 (Figure A.2). When both probabilities are low, corresponding to the lower region of the plots in Figure 5.7, UHI increases the number of runs substantially more for both age groups, and the predicted number of EMS runs are greater for UHI areas than non-UHI areas and compared to the number in UHI areas displayed in Figure 5.8. Conversely, when both probabilities are high, placing us in the right region of the grid plot, UHI decreases the number of runs for both age groups, with predicted EMS runs being much higher in non-UHI areas compared to UHI and compared to the number in non-UHI areas in Figure 5.8.

The prediction directly obtained from Model 3 is the daily number of EMS runs in each subgroup. We sum across the 8 subgroups for each day to obtain the predicted number of EMS per day. The predicted versus observed daily number of EMS runs, when assuming P(UHI|age$\geq$65)=0.5 and P(UHI|age<65)=0.5, is plotted in Figure 5.9, with an MSE of 72.2.

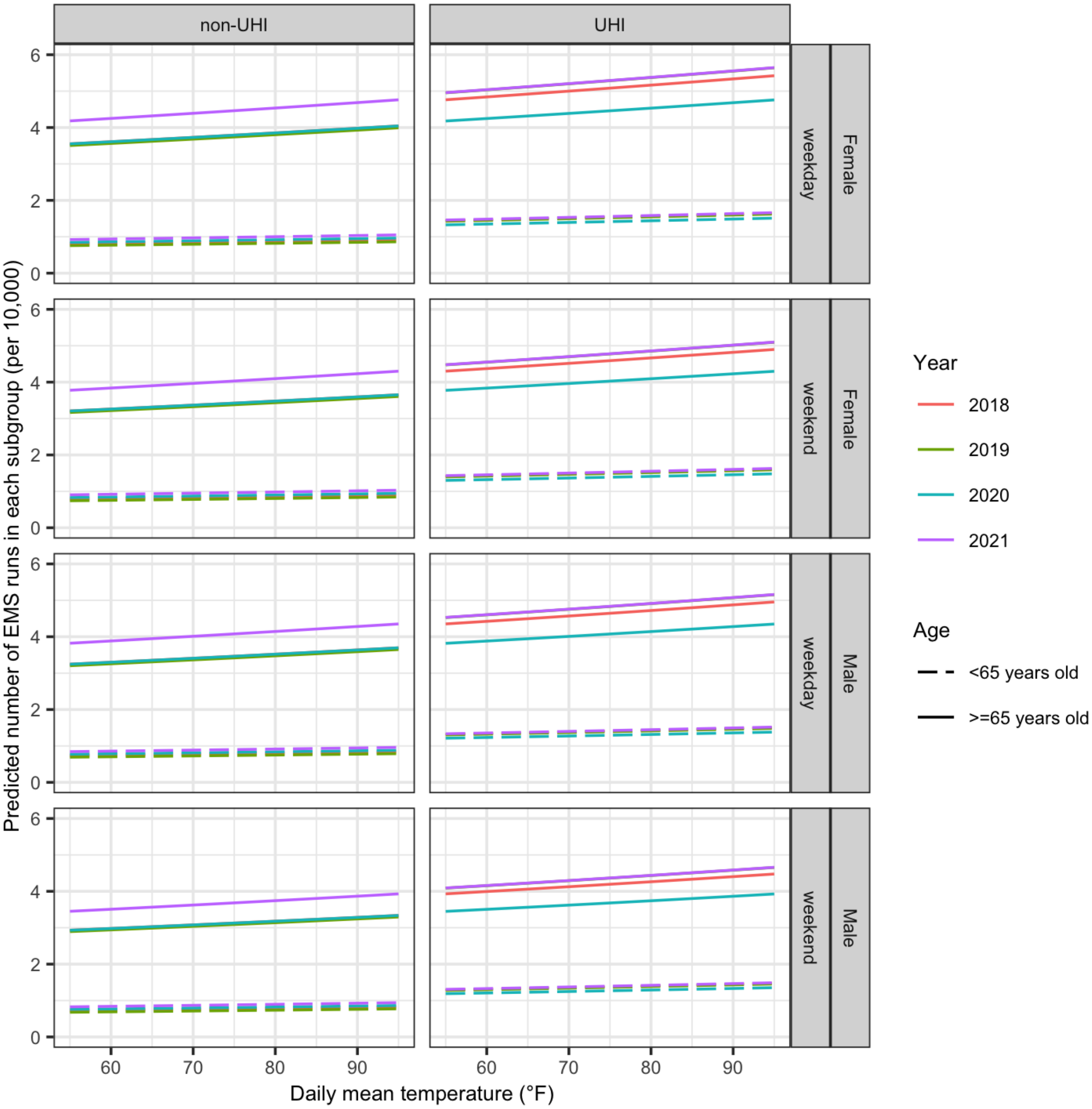


Figure 5.8: Predicted number of EMS runs in each subgroup (per 10,000) over four years, with P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5.

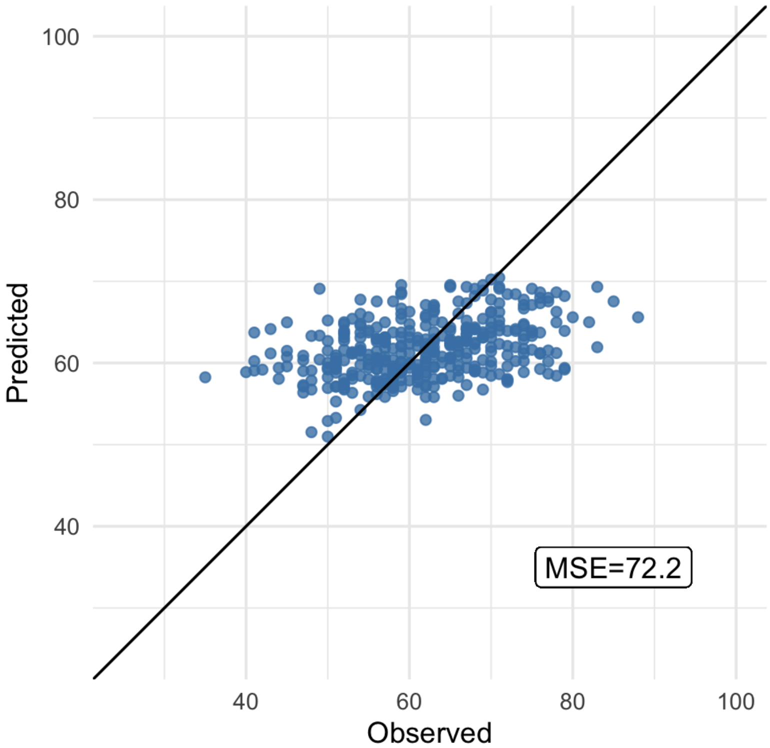


Figure 5.9: Predicted vs. observed daily number of EMS runs from Model 3 when P(UHI|age≥65)=0.5 and P(UHI|age<65)=0.5.

Chapter 6 Discussion

6.1 Results Comparison

We employed quasi-Poisson regression models on both day-level and day-subpopulation level data, utilizing various forms of temperature to assess the relative impacts of hot summer days, urban heat islands (UHIs), and socioeconomic factors on EMS utilization across Rhode Island from 2018 to 2021. Sections 4.1 and 5.1 focus on the day-level analysis, Model 1, leveraging the cross-basis matrix from DLNM to capture potential non-linear associations between heat exposure and response, including lagged effects. Model 2, presented in Sections 4.2 and 5.2, also conducts day-level analysis but employs separate functional forms for current and lagged temperature effects. In Sections 4.3 and 5.3, we transition to the day-subpopulation as the unit of analysis, facilitating the inclusion of additional patient-level variables in our analysis.

All three models indicate that the daily number of EMS runs increases with rising mean temperatures. Both Model 2 and Model 3 use a linear daily mean temperature without lag, and the estimated relative risks of temperature are similar: a 10-degree increase is associated with an increase in the EMS call rate of 3 to 4%. In Model 1, which utilizes the cross-basis matrix, the interpretation of the effect of a given temperature differs, with the reference being the temperature associated with the minimum EMS runs. All three models agree that the number of EMS runs on weekends is expected to be fewer than on weekdays. Both Model 1 and Model 2 suggest that the number of EMS runs on weekends is about 95% of that on weekdays. Model 3 further indicates that the fewer number of EMS runs on weekends is limited to individuals older than 65 years. Additionally, the three models yield consistent results regarding the differences in EMS utilization between years; specifically, EMS call rates are higher in 2018, 2019, and 2021 compared to 2020.

Patient characteristics—UHI, sex, and age—are incorporated into the models in two ways. In Models 1 and 2, which are day-level analyses, the variables are included as the proportion of patients from different UHI, sex, and age groups. Both models indicate that areas with higher proportions of the population residing in urban heat islands have non-significantly higher EMS call rates. Model 1 and Model 2 include interactions between year and the proportion of individuals aged over 65, yielding similar results: in 2018 and 2019, areas with fewer residents aged older than 65 years are likely to have more EMS encounters, while in 2020 and 2021, areas with more residents aged older than 65 years are expected to have more encounters. Models 2 implies that areas with more female residents are likely to have more EMS encounters, while Model 1 indicates that the effect of gender varies across years, with stronger impacts in 2019 and 2020.

In Model 3, we shift to day-subpopulation analysis and can directly incorporate patient characteristic variables into the model. The EMS call rate is significantly higher for females than males. Interactions between year and UHI and between year and age are also found to be significant. However, due to the assumptions made to estimate the subgroup population, the effects of UHI and age are highly sensitive to our choices of the probabilities of being in UHI areas given age. When assuming $P(UHI|age \geq 65)=0.5$ and $P(UHI|age<65)=0.5$, which approximately aligns with the distribution of EMS runs, the results suggest that the EMS call rate is higher for individuals living in UHI areas than for those who don't. The effect of UHI is more pronounced on younger people than on the elderly.

We also created prediction plots and computed the mean squared errors (MSE) of the predictions from the three models. It turns out that the three models have close predictive accuracy, with MSEs of 61.2, 68.8, and 72.2, respectively, despite employing two distinct approaches to model the effect of temperature and different units of analysis. The day-level model leveraging the cross-basis matrix for temperature has a better performance in predicting EMS utilization. Predicted versus observed outcome plots in Figures 5.3, 5.5, and 5.9 suggest that all three models produce narrow ranges of predicted values relative to the observed, which may be a sign that additional terms are required.

6.2 Models Evaluation

The potential weakness of the DLNM approach lies in its assumption that effects of lagged temperatures are additive. This does not incorporate potential non-additive effects that might arise from the accumulation of effects of a series of temperatures on consecutive days, such as with heatwaves. As depicted in Figure 5.1, DLNM estimates the exposure-lag-response association as the relative risk given the temperature and lag. However, when considering the characteristics of heatwaves, the effect of a temperature of 70 degrees at lag=1 should vary depending on the temperature today (lag=0), whether it is 70 degrees or 90 degrees, for example. To put it another way, we believe that these estimates should differ if we split the data into two subsets: one where the temperature at lag=1 is greater than at lag=0, and the other where the temperature at lag=1 is less than lag 0. However, in the 3D representation in Figure 5.1, we observe only one estimate for the effect of 70 degrees at lag=1. To summarize, DLNM marginalizes across all temperatures at a given lag to produce estimates, and it is unable to isolate the compounding effects of consecutive heatwaves from the additive model.

Having noted the limitations of the day-level analysis, it is also important to recognize that use of day-subgroup level models also did not provide a conclusive answer to the research question. The population distribution available from the Census Bureau only distinguishes between age and sex, with no information on the proportion of the population living in UHI areas. Hence, we had to make the assumption that the probability P(UHI|age) does not differ across genders. Subsequently, we estimated the population distribution of gender, age, and UHI in high ADI areas by applying probabilities of being in UHIs given age. However, the true value of the probabilities cannot be determined from the data we have and therefore it is not possible to know conclusively which estimate of the effect of UHI is correct.

The only information we have about these probabilities comes from the runs data and their addresses. From Table 4.1, we observe that for individuals$<$65 years old, the probability of EMS calls from UHIs is 0.66, whereas for those $\geq$65, it is 0.58. When locating these probabilities ($P(UHI|age<65)=0.66$ and $P(UHI|age\geq 65)=0.58$) in Figure 5.7, we find them at the boundary between positive and negative effects of UHIs, compared to non-UHIs. The probabilities of 0.66 and 0.58 are the proportions of the population from UHI areas among those utilizing EMS. But we need the proportion among all residents.

Unfortunately, it is not possible to disentangle the calculation of these probabilities from the effect of UHI itself. If living in a UHI increases the chance of needing EMS, then those using EMS would be more likely to live in a UHI than those not using EMS. This would imply that the overall probability of living in a UHI would be less than the probability among those using EMS, and so the correct probabilities in Figure 5.5 would be less than 0.66 and 0.58. In other words, the effect of UHI would be increased. But this is, of course, a circular argument, since this is what we have just assumed! The same circularity holds if living in a UHI decreases EMS use because then the models would produce UHI estimates that are associated with less EMS use.

6.3 Limitations

The data available for this analysis have several limitations related to the capture of weather data and the years in which the data were captured. These limitations hinder the acquisition of more precise local temperature measurements for individual runs and introduce additional variability into the study.

Given that there is only one weather station in Rhode Island, we must use a single temperature measure across the entire state for each day in our analyses. This assumes that this temperature applies uniformly to the entire state, with the primary variability in temperature on any given day being whether an individual lives in a heat island or not. While Rhode Island is a small state and so it might be reasonable to assume that variation in temperatures across non-built environments throughout the state may not be extreme, summertime temperatures in coastal areas (e.g., Newport) are typically cooler than those from inland regions.

It is also essential to note that the dataset under study spans the years 2018 to 2021, encompassing the period of the COVID-19 pandemic. This unprecedented event brought about significant changes in people's lifestyles and work patterns. Specifically, during the pandemic, remote work became prevalent, leading to a decrease in the number of people commuting to their offices, many of which were situated in areas with high Urban Heat Island (UHI) indices. Consequently, we anticipate a reduction in the number of EMS calls from regions characterized by severe heat islands. Conversely, residential areas, typically associated with lower UHI indices, may have experienced an increase in EMS calls due to the greater presence of individuals at home. However, people's reduced willingness to visit hospitals in fear of COVID-19 may have offset this effect, resulting in relatively unchanged EMS call rates from less severe heat islands. The impact of the pandemic on people's behavior and health-seeking tendencies may therefore have influenced the proportion of runs from UHI areas as reflected in the aggregated variable *UHI_prop* over the study period.

Indeed, the impact of the pandemic on various aspects of life may extend beyond the short term and have lasting effects. This includes how we understand and analyze urban heat indices (UHIs) using historical data. It is imperative to consider the shifts in work patterns, lifestyles, and human behavior brought by the pandemic when conducting analyses and interpreting results related to UHIs. Adapting our analytical methods to account for the evolving societal dynamics post-pandemic will be crucial in ensuring the accuracy and relevance of our findings.

Chapter 7 Conclusion

In this study, we employed quasi-Poisson regression models with both day and day-subpopulation as the unit of analysis, using two approaches for modeling daily temperature, using a cross-basis matrix from a DLNM and using common forms of current and lagged temperatures, to assess the impacts of heat as well as UHIs, and socioeconomic factors on EMS utilization across Rhode Island using data from June-August in the years from 2018 to 2021.

The three models yielded consistent findings: the primary influence on EMS utilization was the current daily mean temperature, exhibiting a nearly linear positive association. Day-level analyses revealed that higher proportions of patients from UHI areas and a higher proportion of females were linked to increased EMS utilization. Additionally, EMS encounters on weekends were significantly lower than on weekdays. The effect of age was found to be different in 2018 and 2019 compared to 2020 and 2021.

The day-subpopulation model indicated that living in a UHI was associated with elevated EMS utilization, assuming a population distribution across age and UHI similar to the distribution observed for EMS utilization across age and UHI in Rhode Island. Moreover, this model suggested that females and those under 65 have higher EMS call rates, mirroring the effects observed for sex and age proportions in the day-level analyses. Though the results from all three models qualitatively align, the different units of analysis result in distinct interpretations of the findings.

The cross-basis matrix provided some improvement in prediction compared to the linear temperature approach, although it did not allow for estimates of effect for individual patients, but instead used proportions of individuals with different age, sex and UHI phenotypes. The day-level and day-subpopulation level analyses showed minimal disparity in predictive accuracy. However, the day-subpopulation level analysis relies on two assumptions: that the population distribution of UHI given age does not differ across genders and that we know the proportion of the population of different age groups residing in UHI areas. While we can test these assumptions by comparing the relative fits of models, we cannot draw firm conclusions without better information on these population distributions.

Preparation and mitigation strategies for population health in response to climate change necessitate a comprehensive understanding of its impacts on healthcare delivery. Our results will aid in identifying the most vulnerable populations geographically and will inform EMS planning and resource allocation, particularly in response to summer heat and heatwaves. By quantifying the impact of heat islands on EMS utilization, adjustments can be made to EMS staffing locations, urban planning, and mitigation strategies accordingly.

As discussed in Chapter 4, the estimation of subpopulations in Model 3 relies on assumptions due to the lack of specific information regarding the proportion of the population residing in UHI areas. One approach to acquire the population distribution across UHI areas is to overlay the UHI map with the census block database and estimate the proportion of the population residing in high UHI locations within each census block. However, due to the extensive workload and time constraints associated with this task, we opted to estimate the population distribution in this study. As a future step, we may consider randomly sampling census blocks, estimating the proportion of the population residing in UHI areas, and then aggregating the population distribution across the state to achieve more precise population distributions.

Appendix A Supplemental Figures

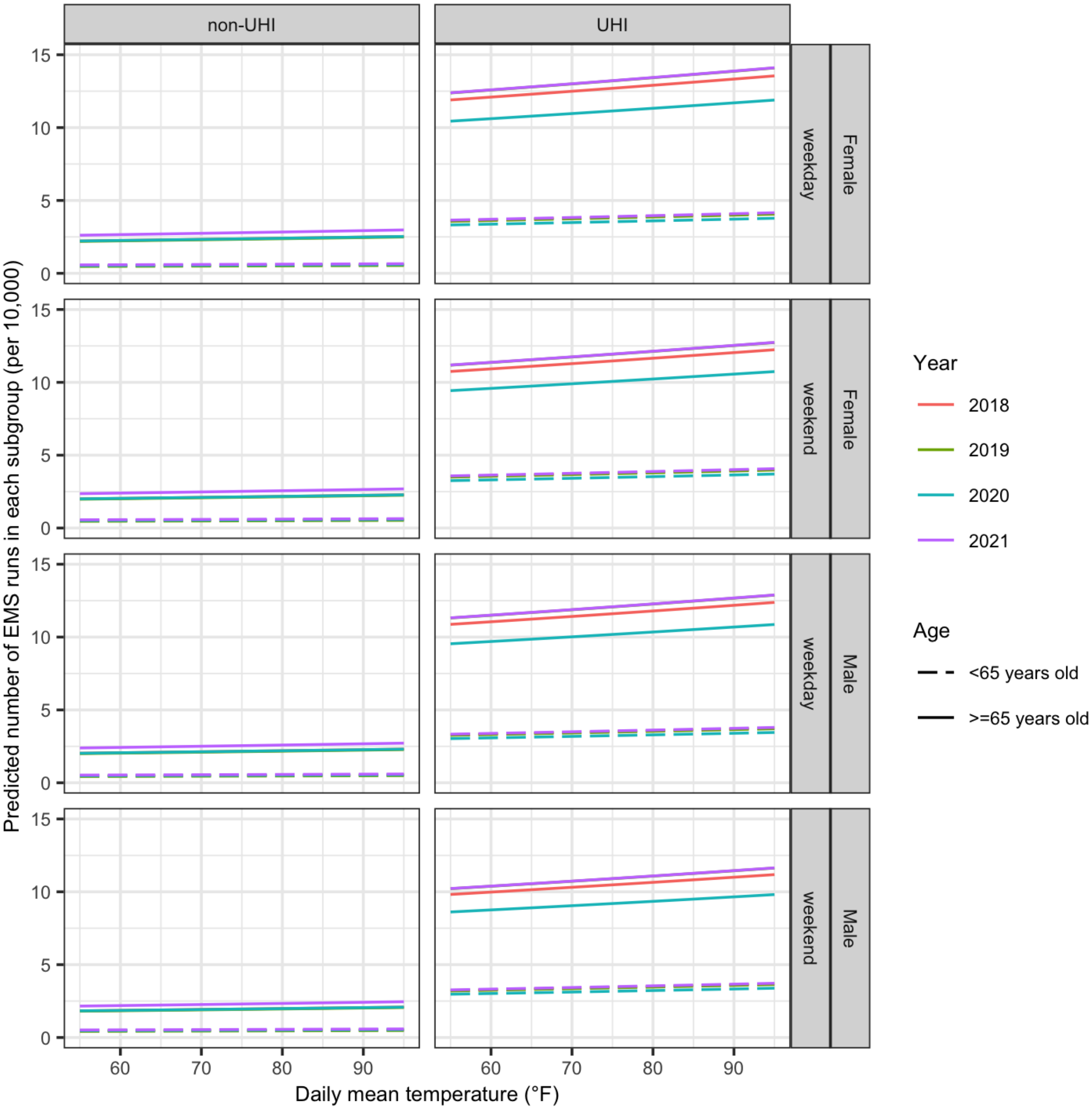


Figure A.1: Predicted number of EMS runs in each subgroup (per 10,000) over four years, with P(UHI|age≥65)=0.2 and P(UHI|age<65)=0.2.

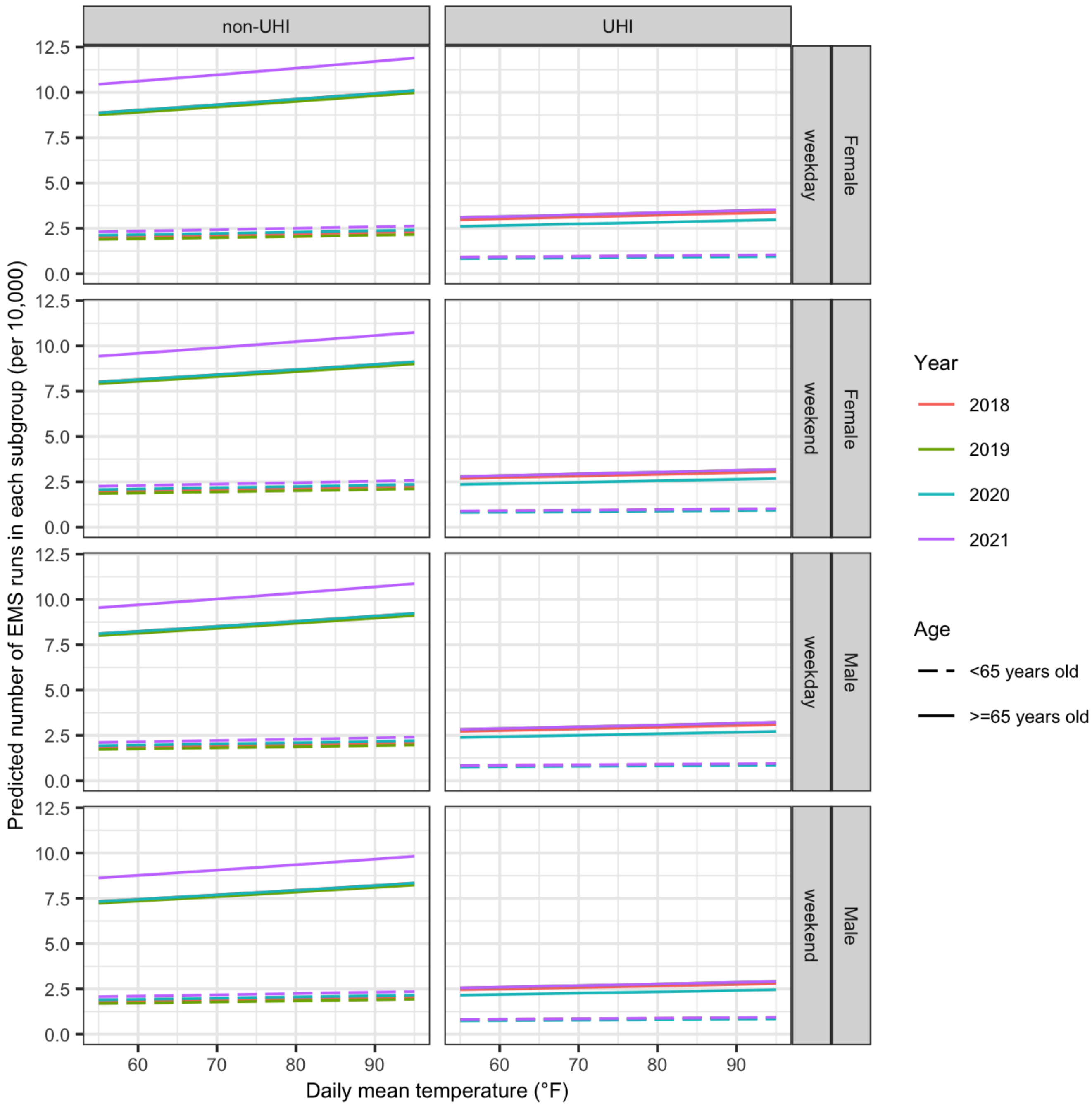


Figure A.2: Predicted number of EMS runs in each subgroup (per 10,000) over four years, with P(UHI|age≥65)=0.8 and P(UHI|age<65)=0.8.

References

Angel Hsu, T. C. D. M., Glenn Sheriff. (2021). Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications 12*.

Barnetta, S. A., A.G. (2010). What measure of temperature is the best predictor of mortality? *Environmental Research*, *110*, 604-611.

Braveman, P., & Gruskin, S. (2003, 4). Defining equity in health. *J. Epidemiol. Community Health*, *57*(4), 254–258.

Cantwell, A. M. K. S. M. L. T. S. P. D., Kate. (2015). Time of Day and Day of Week Trends in EMS Demand. *PREHOSPITAL EMERGENCY CARE*, *19*(3), 425–431. doi: doi: 10.3109/10903127.2014.995843

Davis, N. W. M., Robert E. (2018). The impact of heat waves on emergency department admissions in charlottesville, virginia, u.s.a. *International Journal of Environmental Research and Public Health*, *15*(7). Retrieved from `https://www.mdpi.com/1660-4601/15/7/1436` doi: doi: 10.3390/ijerph15071436

EPA. (2022). *Climate change indicators: U.s. and global temperature.* https://www.epa.gov/climate-indicators/climate-change-indicators-us-and-global-temperature. (Accessed: 2024-01-30)

EPA. (2024). *Heat island effect.* https://www.epa.gov/heatislands. (Accessed: 2024-01-30)

Gasparrini, A. (2010). Distributed lag non-linear models. *Statistics in medicine*, *29*(21), 2224–2234. doi: doi: https://doi.org/10.1002/sim.3940

Gasparrini, A. (2011). Distributed lag linear and non-linear models in R: the package dlnm. *Journal of Statistical Software*, *43*(8), 1–20. Retrieved from `https://doi.org/10.18637/jss.v043.i08` doi: doi: 10.18637/jss.v043.i08

Gasparrini, A. (2014). Modeling exposure–lag–response associations with distributed lag non-linear models. *Statistics in medicine*, *33*(5), 881–899. doi: doi: https://doi.org/10.1002/sim.5963

Hitzek, A. F.-R. M. M. S. L. K. A. S., Jennifer. (2022). Influence of Weekday and Seasonal Trends on Urgency and In-hospital Mortality of Emergency Department Patients. *Frontiers in public health*, *10*(711235). doi: doi: https://doi.org/10.3389/fpubh.2022.711235

Jenks, G. F. (1967). The Data Model Concept in Statistical Mapping. *International Yearbook of Cartography*, *7*, 186-190.

Kind, B. W., AJH. (2018). Making Neighborhood Disadvantage Metrics Accessible: The Neighborhood Atlas. *New England Journal of Medicine*, *378*, 2456-2458. (University of Wisconsin School of Medicine Public Health. 2021 Area Deprivation Index v4.0.1. Downloaded from https://www.neighborhoodatlas.medicine.wisc.edu/ Nov, 2023.) doi: doi: 10.1056/NEJMp1802313

Kingsley, J. G. R. R. V. G. A. W., Samantha L. Melissa N. Eliot. (2016). Current and Projected Heat-Related Morbidity and Mortality in Rhode Island. *Environmental Health Perspectives*, *124*(4), 460–467. doi: doi: http://dx.doi.org/10.1289/ehp.1408826

Moretti, G. M. B. S. L. B. A. N. . A. A. R., K. (2021). Increased Temperatures Are Associated with Increased Utilization of Emergency Medical Services in Rhode Island. *Rhode Island medical journal*, *104*(9), 24-28.

Pourhoseingholi, A. R. B. M. V., M.A. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology from Bed to Bench*, *5*(2), 79-83.

Ragettli, B. F. M. R., Martina S. Ana M. Vicedo-Cabrera. (2019). Impact of the warm summer 2015 on emergency hospital admissions in Switzerland. *Environmental Health*, *18*(66). doi: doi: https://doi.org/10.1186/s12940-019-0507-1

Reid, P., & Robson, B. (2000). Understanding health inequities. *Hauora: Māori Standards of Health IV. A study of the years*, *2005*, 3–10.

Stafoggia, F. A. D. B. A. B. L. C. E. C. N. d. F. D. L. S. D. M. M. M. P. M. R. P. P. P. S. R. M. R. A. S. C. P Massimo; Forastiere. (2006). Vulnerability to Heat-Related Mortality: A Multicity, Population-Based, Case-Crossover Analysis. *Epidemiology*, *17*(3), 315–323. doi: doi: 10.1097/01.ede.0000208477.36665.34

The Trust for Public Land, U., Descartes Labs. (n.d.). *Urban heat island severity for U.S. cities.* https://www.arcgis.com/apps/webappviewer/index.html?id=1b6cad6dd5854d2aa3d215a39a4d372d (Accessed: 2024-04-24)

USGCRP. (2017). *Climate science special report: Fourth national climate assessment, volume i* (Tech. Rep.). U.S. Global Change Research Program. doi: doi: doi: 10.7930/J0J964J6

Vicedo-Cabrera, F. S. A. G., Ana M. (2019). Hands-on Tutorial on a Modeling Framework for Projections of Climate Change Impacts on Health. *Epidemiology (Cambridge, Mass.)*, *30*(3), 321–329. doi: doi: https://doi.org/10.1097/EDE.0000000000000982

Watts, N. e. a. (2020). The 2020 report of The Lancet Countdown on health and climate change: responding to converging crises. *Lancet (London, England)*, *397*, 129-170.

WHO. (2023). *Climate change.* https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health. (Accessed: 2024-01-30)

WHO. (2024). *Heat waves.* https://www.who.int/india/heat-waves: :text=Deaths(Accessed: 2024-01-30)