THE UNIVERSITY OF CHICAGO

Epidemiology of Fall-Related Hospitalizations in Older Adults (July 2009 – July 2015): Exploring the national and regional associations with weather, demographic, and socioeconomic variables using machine learning and classic statistical approaches

By

Jade Benson

June 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts in Computational Social Science

Faculty Advisor: Diane Lauderdale

Preceptor: Shilin Jia

Table of Contents

Acknowledgements

It is bittersweet to finish this thesis and with it mark the end of such a transformative era at the University of Chicago. Six years, two degrees, two theses, at least one published paper – but these numbers do not even begin to capture what this experience has meant to me. I am most thankful for the incredible community that has supported me every step of the way. My thesis advisor, Diane Lauderdale, has been a constant source of inspiration to grow as a thoughtful and confident researcher and person. I could never fully express my gratitude for her guidance over the years, let alone her help creating this thesis on such a tight timeline. For his understanding and willingness to bounce around ideas, thank you to my preceptor, Shilin Jia. There have been so many meaningful friendships built over laughter, stories, and deep conversations that have defined me and this period of my life. I am particularly grateful for Egemen Pamukcu, Sabina Hartnett, and Franco Romaldo Mendes for being the best MACSS teammates and filling these last two years with such joy as we struggle through learning new disciplines together. Ellie Frank, Jamye Selby, and Josh Fonseca Rivera have shared so many memories and so much love with me, and I could not be luckier to have them as friends. Of course, my family has helped me create such a beautiful life through fanning my curiosity and empathy through the five L's, traveling the world together, and always encouraging me to ask why. I am so grateful for this step of my journey and am also so looking forward to beginning the next one. Thank you to everyone who made it possible.

Abstract

Falls are a leading cause of injury and death among older adults. There are few studies that use national longitudinal data and multiple weather metrics to explore the complex relationships of environmental risk factors. This thesis compares machine learning (random forests) and classic statistical (mixed-effects Poisson regressions) methods to examine the influence and spatial variation of weather, demographic, and socioeconomic characteristics on county fall-related hospitalization rates among adults 65 years and older in the United States. Three county-level data sources were combined: 1) monthly Medicare hospitalization claims from July 2009 – July 2015 with either primary diagnosis or external cause codes indicating an accidental fall (ICD-9-CM codes E880-E888, excluding E887) 2) weather variables scraped from NOAA's National Weather Service, and 3) sociodemographic characteristics from the 2010-2014 American Community Survey. National and regional modelling was conducted using both methods and each set of characteristics. Weather, demographic, and socioeconomic variables are all important predictors of fall-related hospitalization rates in older adults. Temperatures were the most influential weather variables at the national level and for all regions except the Northeast. The random forests identified that high fall-related hospitalization rates were associated with temperature variability. The mixed-effects Poisson models demonstrated a U-shaped relationship where both high maximum and low minimum temperatures are associated with greater risk at the national level and in the South and Midwest, controlling for relevant sociodemographics. The two modelling approaches are compared, and the mixed-effects Poisson regressions are consistently more accurate likely due to their adjustment for spatial variation and confounding.

Introduction

Falls are a leading cause of injury and death among older adults.[1] As the U.S. population ages, falls are an increasingly important public health problem.[2] In 2015, healthcare costs due to falls were estimated to cost approximately $50 billion, of which Medicare paid $28.9 billion.[3] Falls can cause immediate injuries like fractures or traumatic brain injuries, and they can also lead to indirect health consequences like restricted mobility, inability to perform daily living activities, social isolation, and depression.[1] Fall risk factors can be separated into individual and environmental factors.[1] Individual risk factors include demographic characteristics, comorbidities, and indicators of physical health (i.e. activity, balance, strength).[1,2,4,5] Home hazards are the most often studied environmental risk factors.[1,5-8] There have been mixed findings about how the natural environment influences fall risk for older adults.[9]

Many studies report higher fall-related injuries during winter months for both older adults living in the community[10-15] and residing in inpatient treatment.[16-18] However, it is debated whether this seasonal trend can be explained by other environmental factors like weather and geography. Stevens, et al. examined the rate of fatal and nonfatal falls among U.S. adults 65 years and older using national-level mortality and emergency department (ED) records in 2002 and found no seasonal variation.[19] Instead, geography and climate were more important factors, as there were higher fall incidences throughout the year for older adults living in colder climates.[19] These studies conducted in these colder climates have typically found that weather is more important to fall burden among older adults. In their study of daily incidence of hip fractures among women 45 years and older in Rochester, Minnesota from 1952 – 1989, Jacobsen, et al. demonstrated that increased winter risk for women aged 45 – 74 was attenuated after controlling for weather conditions.[20] However, for women 75 years and older, the increased winter risk of fracture did not meaningfully change when controlling for both weather conditions and season.[20] In a Montreal-based cohort of adults older than 65, Mondor et al. reported no seasonal variation to fall-related injuries, but higher burdens on days when a freezing rain warning was released.[21] During winters in Ontario, fall-related ED visits increased for up to a week following snowfall, but interestingly also increased during days with warmer temperatures across all age groups.[22] Boston-area hospitals also experienced increased rates of fall-related injuries following moderate snowfall, but this was primarily driven by adults under 65.[23] Many studies in Scandinavia have reported no seasonal variation in fall-related injuries.[24-26] Although the contribution of environmental risk factors to the increased winter fall rates is mixed, the relationship with temperature is more clear.

In their systematic review of the meteorological factors associated with fall risk, Chow et al. reported that lower temperatures and snow/ice cover were the most consistently demonstrated risk factors.[9] Among this literature, Giladi et al.'s study of wrist fractures among 2007 Medicare claims found that season, temperature, and weather conditions were all associated with increased risk.[14] Wrist fracture risk was positively associated with winter, freezing temperatures, snow, and a constructed measure of "slipperiness," which consolidated monthly temperature and weather metrics. Even after adjusting for season, Turner, et al. found that the incidence of fall-related hip fractures was associated with lower daily air temperatures for both men and women over 75 years of age in Sydney, Australia from 1998 – 2004.[27] Lin & Xirasagar also found significant associations between monthly fall rates and temperature after adjusting for season among older adults in Taiwan between 1997 – 2003, using auto-regressive integrated moving average (ARIMA) modeling.[28] However, Johnson, et al. found diverging patterns of fall-related injury risks by age, gender, and temperature in their study of wrist and hip fracture incidence among adult patients referred to their ED in the UK.[29] Lower minimum daily temperatures were associated with higher wrist fracture rates for men older than 50 years and for all women.[29] Hip fracture rates increased for older men with colder temperatures, but women 50 years or older experienced the largest hip fracture burden that did not vary with temperature.[29] A few studies have reported no significant results between temperature and fall risk, but typically have smaller sample sizes and shorter observation times.[9] Only one study of older adults found that injurious falls were more likely to occur in higher temperatures and this was among a small cohort of older adults with glaucoma near Baltimore, Maryland.[30]

Temperature is most often added into models as one continuous variable or categorized based on thresholds (typically below/above freezing).[9,14,19,31] Yeung et al. and Modarres et al. have both examined whether monthly minimum, average, and maximum temperatures along with other meteorological metrics are associated with falls in older adults.[13,15] Using Pearson correlation coefficients with adjusted monthly number of fall-related ED visits to a Hong Kong hospital in 2006, Yeung et al. found that all monthly temperature metrics were negatively correlated with falls.[13] Modarres et al. used ARIMA modelling to similarly demonstrate that all monthly temperatures were negatively correlated with hip fracture incidence from 1993 – 2004 in Montreal, Canada.[15] There are important relationships between temperature and falls that have yet to be explored, like temperature variability and whether both high and low temperatures present increased risk (U-shaped risk pattern).

In this thesis, I investigate the epidemiology of falls that caused hospitalizations among older adults in the United States, its association with weather metrics, the influence of demographic and socioeconomic variables, and how these relationships vary depending on machine learning (random forests) and classic statistical (mixed-effects Poisson regressions) modelling approaches at both the national and regional levels. As sociodemographic features are important individual fall risk factors and have been shown to influence environmental risk factors too, they are important to consider on their own and as confounders. I combine six years (July 2009 – July 2015) of monthly hospital admission claims related to falls from Medicare beneficiaries with weather data from the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) and county sociodemographic characteristics from the 5-year estimates of the 2010 – 2014 American Community Survey (ACS) to investigate multiple research questions. These include: 1) What is the effect of weather variables on monthly county fall-related hospitalization rates among adults aged 65 years and older? 2) What is the effect of county demographic variables on monthly county fall-related hospitalization rates? 3) What is the effect of socioeconomic characteristics on monthly county fall-related hospitalization rates? 4) What is the importance of all these variables on monthly county fall-related hospitalization rates when considered together? 5) Do these associations vary by region of the country? 6) How do different modeling approaches affect these relationships and their spatial patterns?

Methods

Data Sources

Medicare Falls Data

From the U.S. Centers for Medicare and Medicaid Services (CMS), aggregated county-level hospital admission claims related to falls of a 100 percent sample of fee-for-service beneficiaries from the Medicare Provider Analysis and Review (MedPAR) file from July 1st, 2009 – July 1st, 2015 were received.[32] Hospitalizations were considered related to falls if either the primary diagnosis or the primary external cause code fields reported an accidental fall (International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) external cause codes E880-E888, excluding E887)[1], following an algorithm developed by Kim et al. for identifying fall-related injuries in Medicare claims data.[32,33] Primary external cause codes (E-codes) detail the circumstances leading to injuries, which facilitates reimbursement and better study of injuries and their prevention.[34] For example, a patient is hospitalized following a slip and fall that resulted in a broken hip, the primary diagnosis would be the broken hip and the E-code would begin with E885 indicating the type of fall and that their health insurance would likely be liable for payment. The study window was selected as it was the longest period with a consistent coding schema. From 2005 – June 2009, the total number of fall-related hospitalizations reported (N = 131,790) was more than four times lower than the number of fall-related hospitalizations in a single year in the July 2009 – July 2015 period. Therefore, they were

[1] E880: accidental fall on or from stairs or steps; E881: accidental fall on or from ladders or scaffolding; E882: accidental fall from or out of building or other structure; E883: accidental fall into hole or other opening in surface; E884: other accidental falls from one level to another; E885: accidental fall on same level from slipping, tripping or stumbling; E886: fall on same level from collision, pushing, or shoving, by or with other person; E888: other and unspecified fall.

removed to avoid this meaningful reporting and coding inconsistency. The diagnosis codes were switched in the third quarter of 2015 to ICD-10, so these data were dropped to avoid any transition issues and the use of different methods for only a few months.[32,35] The Social Security Administration (SSA) county codes were converted to Federal Information Processing Series (FIPS) county codes using crosswalks released by CMS, resulting in 3,107 counties in the contiguous US.[36] Fall count data at the monthly level were received only for counties that reported more than 10 falls during the month. There were 72,081 monthly county-level fall records, which included 3,377,329 total falls from 1,790 counties from July 2009 – July 2015.

Weather Variables

Weather variables were collected from NOAA's NWS for counties in the contiguous United States using the meteostat python library.[37] The latitude and longitude of every county's centroid point were assembled using GeoPandas[38] on the projected version (CRS code EPSG:4326) of the U.S. counties geojson file from the Python Graph Gallery.[39] The five closest reporting weather stations to the centroids were identified. Then to maximize data completeness, the closest station that began reporting daily weather records prior to 07/01/2009 and through 07/01/2015 was selected and scraped. There were 3,109 counties with 1,402 unique closest weather stations that reported data over the study period available through the meteostat library. For each station, daily weather variables attained were precipitation (mm), snow depth (mm), average wind speed (km/h), and minimum, maximum, and average temperature (°C). Summary statistics (minimum, mean, and maximum) were then calculated for each weather metric at the monthly level. For greater interpretability, six monthly summary weather variables were selected: average monthly temperature (average of all the daily average temperatures), minimum monthly temperature (the minimum of all minimum daily temperatures), maximum monthly temperature (the maximum of all maximum daily temperatures), maximum monthly precipitation (the maximum of all maximum daily precipitations), maximum monthly snow fall (the maximum of all maximum daily snow falls), and the maximum monthly windspeed (the maximum of all maximum daily windspeeds). These temperature metrics were selected to represent both the average monthly climate and the extremes of the temperature variability that may lead to increased fall risk. The maximum monthly precipitation and snow falls allow for any storm events to be captured, which have often been linked to fall-related hospitalizations and injuries.[9,22,23,31] Giladi, et. al incorporated maximum windspeeds into their slipperiness score, as this can be related to greater windchill and hazardous conditions.[14] By aggregating to the monthly level, some of the random variation in daily weather metrics are smoothed and it removes the need to lag variables. When daily records are available, some studies have used weather metrics on that day,[21,29] while others argue that weather variables require lags which range from 1 – 15 days after.[23,40,41] The monthly analysis can capture general trends between weather variables and fall rates, and has often been used as the unit of analysis in this literature.[9,13,24,28] However, examining these trends at the monthly level may not accurately capture the fall burden related to specific storms or extreme weather events.[21,23,31] Although including the maximum of the daily maximum values attempted to address this concern, it may not accurately capture the average climate conditions that affect day-to-day risk and could over-represent outlier weather days.

Demographic and Socioeconomic Variables

The 5-year 2010 – 2014 American Community Survey (ACS) data were used to estimate county-level population, demographic, and socioeconomic characteristics. Fall rates were calculated using the population 65 years of age and older in each county and scaled to be per 100,000. Additional potentially relevant demographic characteristics included the total county population density and the proportions of the over 65-year-old adults in each county that were: 85 years of age or older, female, and non-Hispanic white. County-level socioeconomic variables were median income, the Gini Index of Income Inequality, and the proportion of adults 65 years of age and older living below the poverty line. The Gini Index of Income Inequality is a summary statistic of the dispersion of income across the entire income distribution, ranging from 0 where everyone receives the same income to 1 where only one person receives all the income.[42]

Analysis

In this analysis, I examine to what extent the weather, demographic, and socioeconomic variables are associated with fall-related hospitalization rates in older adults and how different modeling approaches affect these results. Months that were either missing fall counts or any of these predicting variables were excluded, resulting in an analytic sample of 66,945 monthly fall rates (3,124,305 total falls) from 1,733 unique counties from July 2009 – July 2015 (73 months). Two rounds of modelling were performed – national and regional. The national round conducted models using the monthly fall rates and predictors for all US counties included in the analytic sample. The regional round conducted each model separately by region after coding monthly fall rates and predictors as belonging to counties in each of the four US census regions (West, Midwest, South, & Northeast).[43] Four models were conducted at the national level and for each region. Every model uses the monthly county-level fall rates per 100,00 adults aged 65 years and older as the outcome and the predictors for these models were 1) weather variables, 2) demographic characteristics, 3) socioeconomic factors, and 4) all metrics. Random forests and mixed-effects Poisson regressions were compared for each round of analysis.

Random forests combine an ensemble of decision trees and add randomness to the selection of the training set, node split, and features, which makes their predictions more robust and accurate.[44] Decision trees are inductive, non-parametric machine learning algorithms that split instances based on the feature that optimizes the information gain and recursively repeat this process until either homogenous nodes or pre-pruning stopping criteria are reached. [45-47] To better account for the time series structure, a 5-fold forward chaining strategy was used for cross-validation with 2014 – 2015 data as the test set (about 25% split).[48] Hyperparameter tunning was conducted using a grid with the maximum number of tree nodes (maximum depth) ranging from 2 – 5, the minimum number of samples per leaf as either 1, 5, or 10, and the number of trees in the forest as either 50, 100, or 200. Features were evaluated and ranked according to their relevance in predicting fall-related hospitalization rates using Mean Decrease Impurity (MDI).[2] A feature's MDI is calculated in each tree by averaging the reduction of the variance in fall-related hospitalization rates on nodes where the feature was used to split over all nodes, these scores are then averaged across all trees in the ensemble.[49] Features that are not used at all receive a zero value and those consistently closest to the root nodes have higher scores. The MDI values indicate which features were important for predicting the outcome, but not the

[2]MDI is also referred to as "Gini Importance."

directionality or relationship between all the variables in the trees. The scikit-learn python package was used to fit and evaluate all random forest regressors.[50,51]

The mixed-effects Poisson regressions included fixed effects for each state to account for state-to-state variation not explained by the included variables, random effects for each county to better account for overdispersion, and an offset term for log county population aged 65 years and older.[52] The data were again split into the 2009 – 2013 training and 2014 – 2015 testing sets and each of the variables was standardized. Since the mixed-effects Poisson models require no hyperparameter tunning nor are they explicitly forecasting models, no forward-chaining strategy was used.[53] Rather than ranking and selecting the features for prediction like the random forests, the mixed-effects Poisson regressions consider all of the features simultaneously. The estimates returned from these models can be interpreted as the factor change in the outcome for a unit increase in the predictor, controlled for the other variables in the model. Mixed-effects Poisson models were conducted in R using the lme4 package for generalized linear mixed-effects models,[54] tidyverse for data preparation,[55] and ML metrics for model evaluation.[56]

Mean absolute percentage error (MAPE) was the evaluation metric for all models, as it has been shown to be well-suited for forecasting applications, model comparison, and interpretation.[57] It is interpreted as the average percentage difference between the forecasted and actual values.

Results

Descriptive Results

The cumulative national fall rates suggested both temporal and spatial patterns. During winter months (December, January, February), fall rates displayed higher peaks (Figure 1). Similar patterns were observed in each of the four census regions (Figure 1a). The peaks in the Western region were lower than any of the valleys in the other three regions and the highest overall peak was in January 2015 in the Northeast region.

Figure 1. National cumulative monthly fall rates per 100,000 adults 65 and older for all available counties in the analytic sample, July 2009 – July 2015. January of each year is marked with a dashed line and winter months (December, January, February) are shaded.

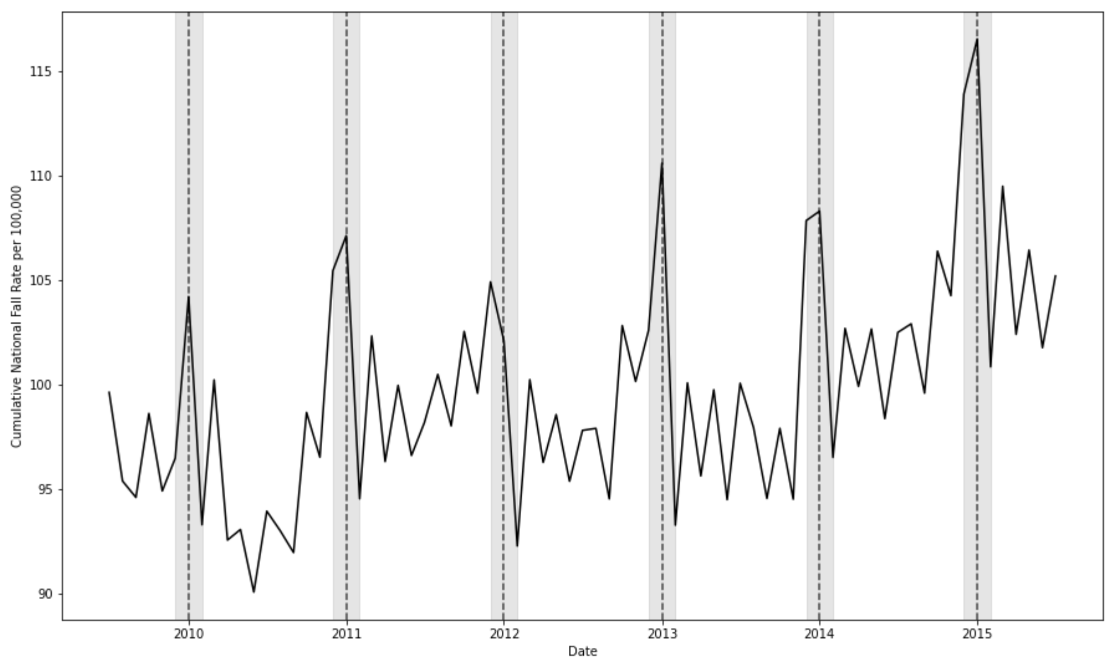
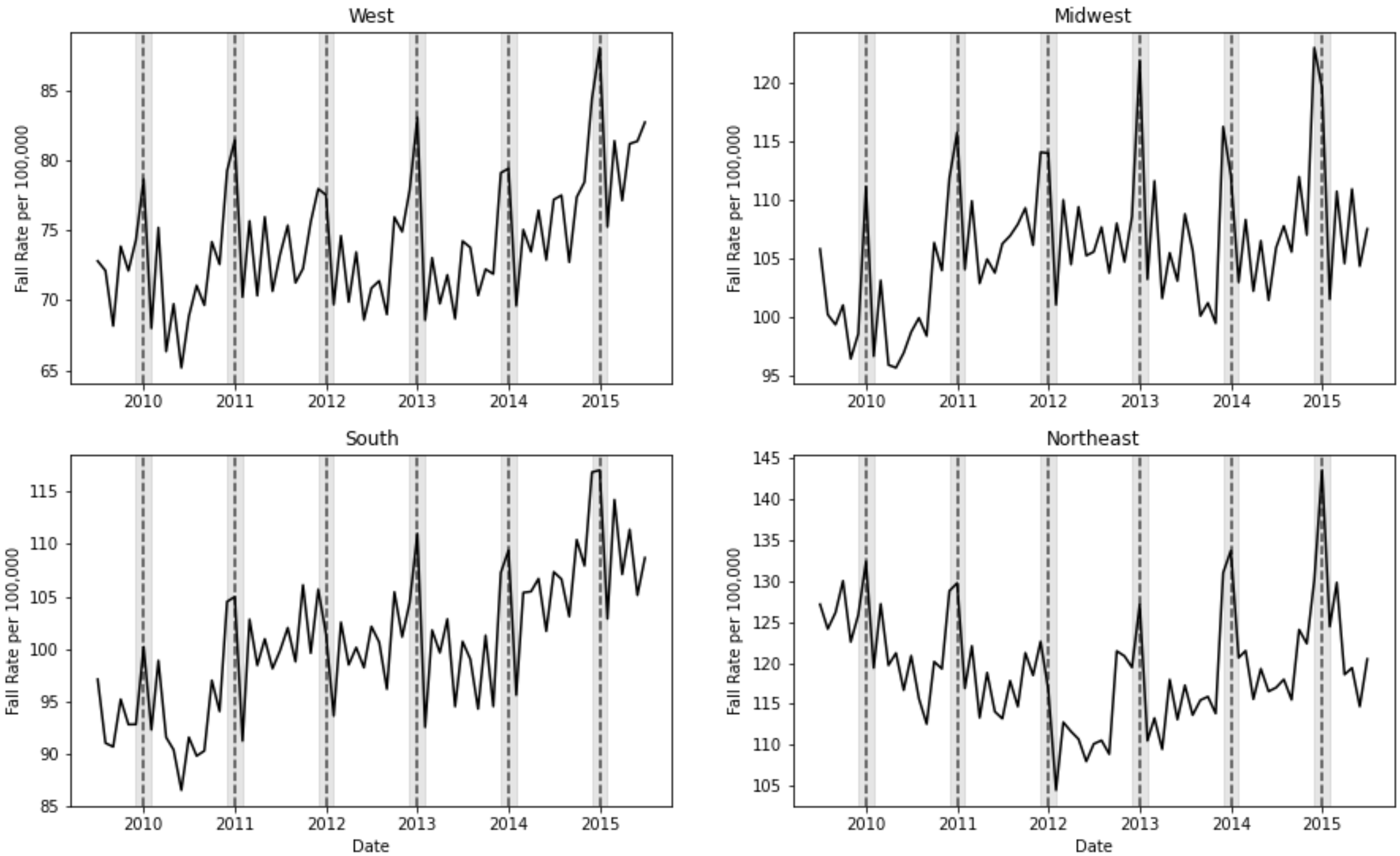


Figure 1a. Regional cumulative monthly fall rates per 100,000 adults 65 and older for all available counties in the analytic sample, July 2009 – July 2015. January of each year is marked with a dashed line and winter months (December, January, February) are shaded.



Only 58% of counties reported any monthly fall counts over the study period, due to the data protection restrictions (more than 10 reported falls during a month in the county). The cumulative county fall rates per 100,000 adults 65 years and older from July 2009 – July 2015 are displayed in Figure 2. Counties with missing data were primarily located in the middle of the United States and had smaller populations. The average county size for counties that did not report any monthly data within the study period was only 1,986 adults 65 years and older (IQR: [1,022, 2,746]), while the counties that did report data had an average older adult population size of 22,683 (IQR: [5,276, 19,277]). The county population of adults 65 years and older is displayed in Figure 2a, which also demonstrates that the majority of counties missing fall data were in the smallest quintile of population.

Figure 2. Cummulative county fall rates per 100,000 adults 65 years and older from July 2009 – July 2015 using monthly data and 2010 – 2014 ACS 5-year population estimates, in the analytic sample. The four census regions (West, Midwest, South, & Northeast) are outlined in black.

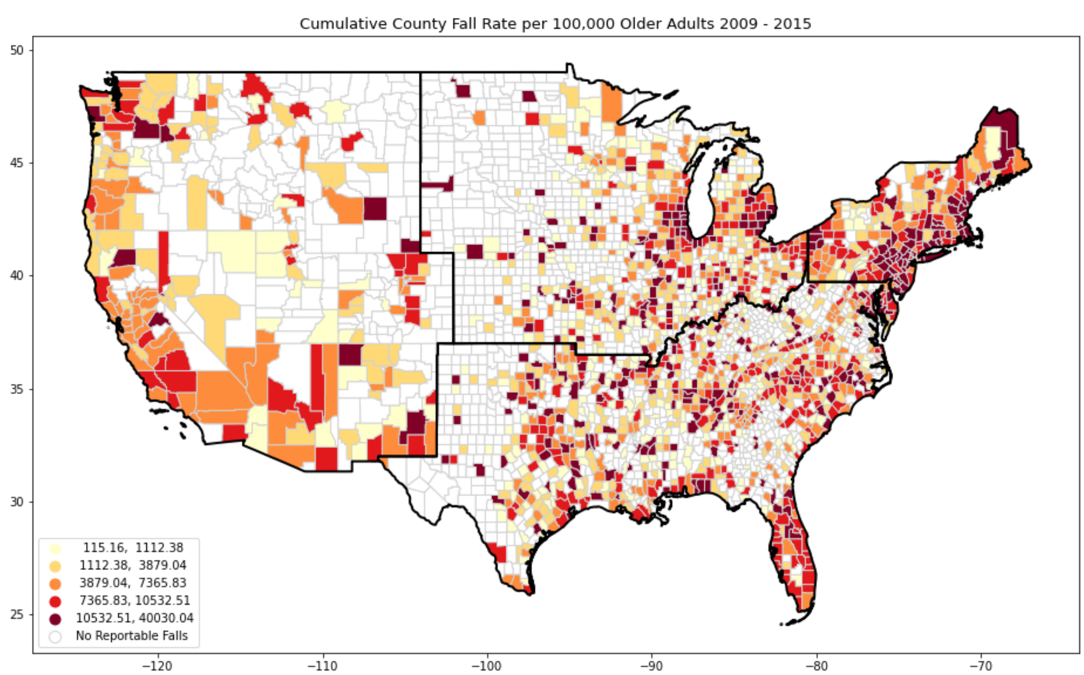
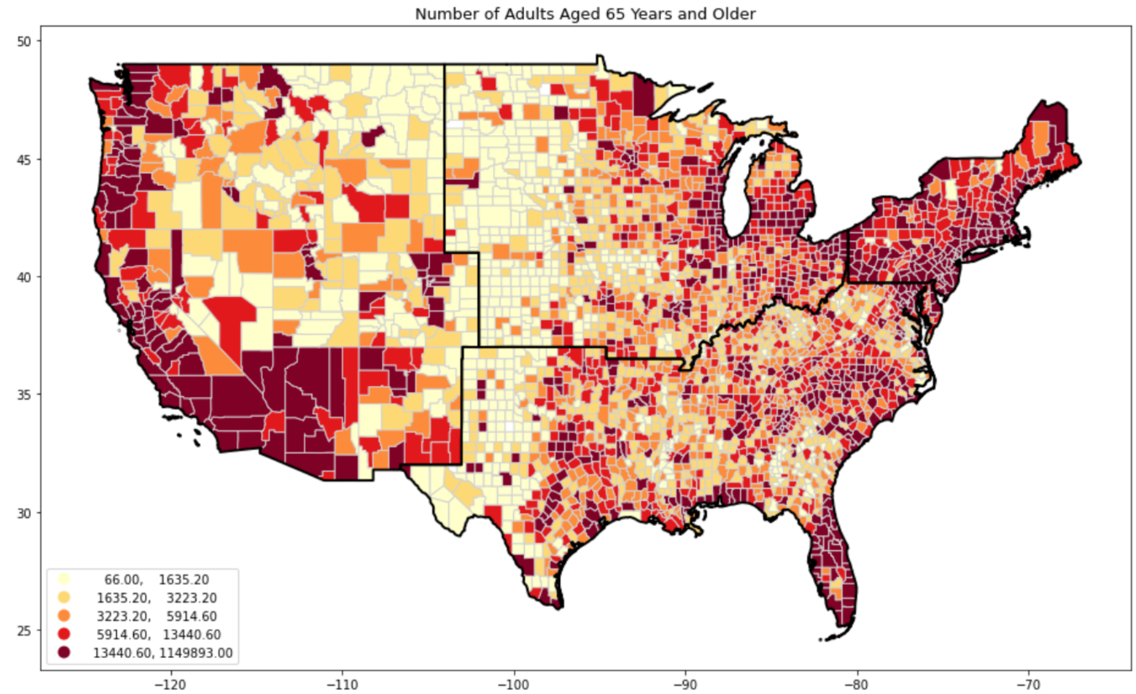


Figure 2a. Quintiles of county population of adults 65 years and older from 2010 – 2014 ACS 5-year population estimates. The four census regions (West, Midwest, South, & Northeast) are outlined in black.



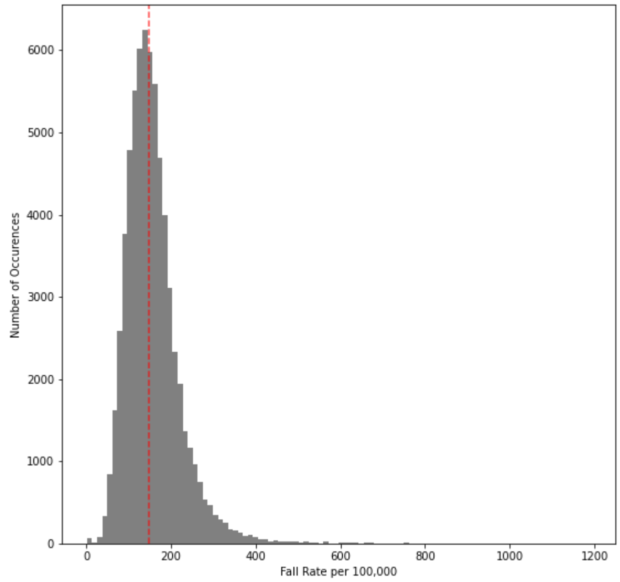
The analytic sample consisted of 66,945 monthly fall rates from 1,733 unique counties over the 73 months in the study period. Summary statistics of the analytic sample for each of the variables can be found in Table 1. The average monthly fall rate was 157.0 fall-related hospitalizations per 100,000 adults 65 years and older. The averages of monthly temperatures were 0.7°C (33.3°F) for the minimums, 13.7 °C (56.7°F) for the averages, and 27.2°C (81.0°F) for the maximums. On average, counties had more older women and had high proportions of non-Hispanic white older adults. The average proportion of older adults living below the poverty line was 0.08 and the average median income was $25,936.

Table 1. Summary statistics for outcome, weather, demographic, and socioeconomic variables in the analytic sample (N = 66,945).

	Variable	**Mean (IQR)**
Outcome	Fall-Related Hospitalization Rate per 100,000	157.0 (114.1, 185.6)
Weather Variables	Average Monthly Temperature (°C)	13.7 (6.6, 21.7)
	Minimum Monthly Temperature (°C)	0.7 (-6.6, 9.4)
	Maximum Monthly Temperature (°C)	27.2 (22.0, 33.3)
	Maximum Monthly Precipitation (mm)	11.7 (0.0, 17.3)
	Maximum Monthly Snowfall (mm)	7.8 (0.0, 0.0)
	Maximum Monthly Windspeed (km/h)	21.5 (16.2, 26.2)
Demographic Variables	Proportion 85 Years and Older	0.13 (0.11, 0.15)
	Proportion Female	0.56 (0.55, 0.57)
	Proportion Non-Hispanic White	0.86 (0.81, 0.96)
	Total County Population Density (people/miles^2)	48.4 (8.3, 36.2)
Socioeconomic Variables	Proportion Below Poverty Line	0.08 (0.07, 0.10)
	County Median Income ($)	25,936 (22,123, 28,169)
	Gini Index of Income Inequality	0.45 (0.43, 0.46)

The distribution of the monthly fall rates was right skewed with about 80% of the values equal to or below 200 falls per 100,000 adults aged 65 years and older (Figure 3). The highest fall rate was 1,190.5 falls per 100,000 older adults that occurred in January 2014 in the small county of Smith, Kansas (1,008 adults aged 65 years and older). Even among the analytic sample which excluded the smallest population counties, the distribution of the population of adults aged 65 years and older was also right skewed with half of the counties below 4,217 older adults and only 6% of the counties had more than 50,000 older adults.

Figure 3. Histogram of monthly fall rates per 100,000 older adults in the analytic sample (N = 66,945). The median fall rate (147 falls per 100,000) is indicated with a dashed red line.



The regional analysis aimed to better characterize the spatial variation in how weather, demographic, and socioeconomic variables were associated with fall-related hospitalization rates. Two-way ANOVAs were conducted to analyze the difference in average fall-related hospitalization rates and all predictive variables across the four regions. Descriptive statistics by region and results from the two-way ANOVAs can be found below in Table 2. The means of fall-related hospitalization rates and all other variables were significantly different across the regions. The average fall rate was highest in the Southern region (170.2 falls per 100,000 adults aged 65 years and older) and lowest in the Western region (112.5 falls per 100,000 adults aged 65 years and older). The Midwestern and Northeastern regions were similar in average monthly temperatures (minimum, mean, and maximum). The Northeastern region had a much higher average population density and the highest proportions of women and adults 85 years of age and older. The Southern region had the lowest mean county-level median income ($24,752) and the highest proportion of older adults living below the poverty line (0.10).

Table 2. Summary statistics for sample, outcome, weather, demographic, and socioeconomic characteristics by census region (West, Midwest, South & Northeast). Mean and IQR presented in parentheses when applicable. Two-way ANOVA F-statistics and p-values also presented.

	Variable	**West**	**Midwest**	**South**	**Northeast**	**ANOVA (F-statistic; p-value)**
Sample	States	'WA', 'OR', 'CA', 'MT', 'ID', 'WY', 'NV', 'UT', 'CO', 'AZ', 'NM'	'ND', 'SD', 'NE', 'KS', 'MN', 'IA', 'MO', 'WI', 'IL', 'IN', 'MI', 'OH'	'TX', 'OK', 'AR', 'LA', 'MS', 'AL', 'GA', 'FL', 'SC', 'NC', 'KY', 'TN', 'WV','MD', 'DE', 'VA', 'DC'	'PA', 'NY', 'NJ', 'VT', 'NH', 'MA', 'CT', 'RI', 'ME'	
	Number of Counties in the Analytic Sample)	193	506	838	196	
	Number of Monthly Rates	8,476	17,023	30,936	10,330	
Outcome	Fall-related Hospitalization Rate per 100,000	113.8 (80.7, 137.4)	161.6 (120.2, 191.7)	170.2 (122.8, 198.0)	145.4 (117.0, 171.5)	1830.2 p <0.001
Weather Variables	Minimum Monthly Temperature (°C)	0.5 (-5.0, 7.2)	-3.5 (-12.8, 6.7)	4.2 (-3.8, 13.0)	-2.7 (-10.6, 6.7)	2345.8 p <0.001
	Average Monthly Temperature (°C)	13.1 (7.2, 19.1)	10.4 (1.4, 19.8)	16.9 (10.0, 24.3)	10.2 (1.8, 18.8)	2537.0 p <0.001
	Maximum Monthly Temperature (°C)	27.3 (20.6, 34.4)	24.9 (17.8, 32.2)	29.6 (25.0, 34.4)	24.0 (17.0, 31.6)	1975.6 p <0.001
	Maximum Monthly Precipitation (mm)	5.6 (0.0, 6.4)	10.5 (0.0, 15.7)	14.0 (0.0, 22.1)	11.7 (0.0, 18.5)	233.3 p <0.001
	Maximum Monthly Snowfall (mm)	5.8 (0.0, 0.0)	12.5 (0.0, 0.0)	3.3 (0.0, 0.0)	14.7 (0.0, 0.0)	255.3 p <0.001
	Maximum Monthly Windspeed (km/h)	21.9 (16.0, 26.7)	25.0 (20.1, 29.5)	19.5 (14.8, 23.7)	21.7 (16.5, 26.4)	2342.3 p <0.001

Demographic Variables	Proportion 85 Years and Older	0.13 (0.12, 0.14)	0.14 (0.13, 0.16)	0.12 (0.10, 0.13)	0.15 (0.13, 0.16)	7663.2 $p < 0.001$
	Proportion Female	0.55 (0.53, 0.56)	0.56 (0.56, 0.57)	0.56 (0.55, 0.58)	0.57 (0.56, 0.58)	2848.5 $p < 0.001$
	Proportion Non-Hispanic White	0.81 (0.69, 0.94)	0.93 (0.93, 0.97)	0.82 (0.76, 0.93)	0.90 (0.88, 0.98)	3876.0 $p < 0.001$
	Total County Population Density (people/miles^2)	35.6 (3.1, 28.4)	37.5 (11.5, 35.9)	35.9 (7.5, 28.2)	114.2 (15.3, 72.4)	802.6 $p < 0.001$
Socioeconomic Variables	County Median Income ($)	26,180 (22,176, 29,643)	26,301 (23,194, 28,342)	24,752 (21,013, 26,728)	28,672 (24,331, 31,669)	1404.3 $p < 0.001$
	Gini Index	0.45 (0.43, 0.46)	0.43 (0.41, 0.45)	0.45 (0.43, 0.47)	0.44 (0.42, 0.46)	1609.9 $p < 0.001$
	Proportion Below Poverty Line	0.08 (0.07, 0.10)	0.07 (0.06, 0.08)	0.10 (0.07, 0.11)	0.08 (0.06, 0.08)	3153.4 $p < 0.001$

Model Results

National Models

National Random Forests

The national round of random forests used all the monthly fall-related hospitalization rates and the relevant variables for each of the four models without explicitly adjusting for spatial variation. The weather random forest ranked the monthly maximum and minimum temperatures as the most predictive features of monthly fall-related hospitalization rates (MDI values of 0.50 and 0.47, respectively). Precipitation, snow fall, and wind speed were not used to split any nodes across the forest (MDI = 0). Months with highly variable temperatures (both lower monthly minimums and higher monthly maximums) experienced the highest falls-related hospitalization rates. The testing MAPE was 40.8%, which means that with a random forest constructed only from temperature variables, the average predicted monthly fall rates for 2014 – 2015 were 40.8% different than the actual fall rates. The demographic random forest ranked population density and proportion female as the only important predictive features (each had 0.50 MDI). The predictive performance of this model for 2014 – 2015 fall-related hospitalization rates was improved by 2.6% from the weather model (test error of 38.2%). Rural counties and counties with a higher proportion of older women experienced the highest fall-related hospitalization rates. The socioeconomic random forest only considered median income in all its splits. The highest falls-related hospitalization rates were in the counties with the lowest median income and the lowest falls-related hospitalization rates were in the highest median income counties. The socioeconomic model performance fell between the weather and demographic models with a test MAPE of 39.9%. The full competitive random forest model with all features was the most accurate, with 4.6% improvement upon the best individual model (test MAPE 33.6%). The ranking of features for this prediction was: proportion female (MDI 0.35), population density (MDI 0.33), proportion non-Hispanic white (MDI 0.22), median income (MDI 0.09), and less than 0.01 MDI for all other variables. None of the weather variables contributed to the prediction. Table 3 includes the training and testing MAPE, the ranking of the important features with MDI values greater than 0.01, and the optimal hyperparameters for all the national random forests.

Table 3. Mean absolute percentage errors (test, train), ranking of the non-zero important features (MDI value), and optimal hyperparameters from the four national random forest models.

	Test (Train) MAPE	**Rank of Important Features (MDI)**	**Optimal Hyperparameters**
Weather Model	40.8% (34.0%)	Maximum temperature (0.50), minimum temperature (0.46), average temperature (0.04)	Max depth: 2, min samples per leaf: 5, number of trees: 100
Demographic Model	38.2% (32.2%)	Population density (0.50), proportion female (0.50)	Max depth: 2, min samples per leaf: 1, number of trees: 200
Socioeconomic Model	39.9% (33.3%)	Median income (1.0)	Max depth: 2, min samples per leaf: 10, number of trees: 50
Full Model	33.6% (32.1%)	Proportion female (0.35), population density (0.33), proportion non-Hispanic white (0.22), median income (0.08), Gini Index of Income Inequality (0.01), proportion poor (0.01)	Max depth: 4, min samples per leaf: 1, number of trees: 200

National Mixed-Effects Poisson Models

The same four models (weather, demographic, socioeconomic, and full) were conducted again at the national level using mixed-effects Poisson regression. The weather model had significant but small effects for monthly minimum, average, and maximum temperatures, and maximum windspeed (Table 4). For one standard deviation increase in the monthly average temperature, the estimated fall-related hospitalization rate decreases by a factor of 0.991 (95% CI: [0.986, 0.995]; $p < 0.001$), after controlling for all other weather variables. Both lower monthly minimum and higher monthly maximum temperatures were associated with higher monthly fall-related hospitalization rates. After controlling for all weather variables, higher windspeeds were associated with lower fall rates.

Table 4. Estimates and 95% confidence intervals from mixed-effects Poisson regression model using weather variables on training data (2009 – 2013) in the analytic sample.

	Estimate	95% Confidence Interval	P-value
Monthly Average Temperature	0.990	[0.986, 0.995]	< 0.001
Monthly Minimum Temperature	0.989	[0.986, 0.992]	< 0.001
Monthly Maximum Temperature	1.005	[1.002, 1.007]	0.002
Monthly Maximum Precipitation	1.000	[0.999, 1.001]	0.62
Monthly Maximum Snowfall	1.001	[1.000, 1.002]	0.054
Monthly Maximum Windspeed	0.999	[0.998, 1.000]	0.133

The demographic mixed-effects Poisson model had significant associations for every variable (Table 5). Counties with a standard deviation higher population density were associated with a 0.804 factor decrease in fall-related hospitalization rates (95% CI: [0.754, 0.857]; $p < 0.001$). After controlling for all other demographic variables, higher proportions of women 65 years and older were associated with a decrease in fall rates (coefficient estimate: 0.680; 95% CI: [0.631, 0.733]; $p < 0.001$). Among the adults 65 years and older, higher proportions of non-Hispanic white and those 85 years of age and older were associated with increased fall-related hospitalization rates.

Table 5. Estimates and 95% confidence intervals from mixed-effects Poisson regression using demographic variables on training data (2009 – 2013).

	Estimate	95% Confidence Interval	P-value
Population Density	0.804	[0.754, 0.857]	< 0.001
Proportion Female	0.680	[0.631, 0.733]	< 0.001
Proportion non-Hispanic White	1.111	[1.027, 1.206]	0.009
Proportion 85 years and older	1.161	[1.082, 1.246]	< 0.001

The socioeconomic mixed-effects Poisson model also had highly significant associations for every variable (Table 6). A standard deviation higher median county income was associated with less than half the fall-related hospitalization rates (coefficient estimate: 0.467; 95% CI: [0.435, 0.501]; $p < 0.001$). Higher proportions of older adults under the poverty line were associated with increased fall-related hospitalization rates. After controlling for the other socioeconomic variables, higher values of the Gini Index of Income Inequality (more inequality) were associated with decreased fall-related hospitalization rates.

Table 6. Estimates and 95% confidence intervals from Poisson regression of Model 3 (socioeconomic variables) on training data (2009 – 2013).

	Estimate	95% Confidence Interval	P-value
Median Income	0.467	[0.435, 0.501]	< 0.001
Gini Index of Income Inequality	0.644	[0.609, 0.681]	< 0.001
Proportion under the Poverty Line	1.150	[1.087, 1.218]	< 0.001

In the full mixed-effects Poisson model, all variables significant in the prior models remained significant. The effect size of population density decreased in the full model and the effect of older adults under the poverty line increased. Unlike the random forest models which prioritize predictive accuracy and feature selection, the weather variables remained significant even though they continued to have small effect sizes.

Table 7. Estimates and 95% confidence intervals from mixed-effects Poisson regression using all variables on training data (2009 – 2013).

	Estimate	95% Confidence Interval	P-value
Monthly Average Temperature	0.991	[0.986, 0.995]	< 0.001
Monthly Minimum Temperature	0.989	[0.986, 0.992]	< 0.001
Monthly Maximum Temperature	1.005	[1.002, 1.007]	< 0.001
Monthly Maximum Precipitation	1.000	[0.999, 1.001]	0.604
Monthly Maximum Snowfall	1.001	[1.000, 1.001]	0.055
Monthly Maximum Windspeed	0.999	[0.998, 1.000]	0.146
Population Density	0.902	[0.854, 0.952]	< 0.001
Proportion Female	0.703	[0.661, 0.748]	< 0.001
Proportion non-Hispanic White	1.415	[1.315, 1.524]	< 0.001
Proportion 85 years and older	1.195	[1.127, 1.268]	< 0.001
Median Income	0.602	[0.4561, 0.646]	< 0.001
Gini Index of Income Inequality	0.751	[0.710, 0.794]	< 0.001
Proportion under the Poverty Line	1.515	[1.424, 1.611]	< 0.001

The mixed-effects Poisson regressions had better performance than their corresponding random forest models, with test errors consistently around 30.4% MAPE (about a 3.2% improvement from the best random forest). The training and testing MAPEs were also more consistent across all the mixed-effects Poisson models (Table 8).

Table 8. Mean absolute percentage error (MAPE) for all national mixed-effects Poisson regression models.

	Test Error	**Train Error**
Weather Model	30.38	29.60
Demographic Model	30.43	29.61
Socioeconomic Model	30.42	29.61
Full Model	30.44	29.60

Regional Models

Regional Random Forests

For the regional round of analysis, the same four random forests using each set of variables (weather, demographic, socioeconomic, and full) were constructed for each of the 4 census regions (16 total models). There were different patterns across the models and regions in what and how features predicted fall-related hospitalization rates. The uneven amounts of unique observations meaningfully affected model performance, as demonstrated by the differences between the high error in the Western region (smallest number of monthly fall rates) and the relatively comparable accuracies for all other regions. The weather models alone consistently had the worst predictive accuracies. In full competitive models, the demographic characteristics were particularly important. The random forests constructed at the regional level were more complicated than the national models, typically with the highest maximum depth and number of estimators provided through the hyperparameter grid tunning. The testing and training MAPE, ranking of features with MDI values greater than 0.01, and the optimal hyperparameters for all regional random forests can be found in Table 9.

The Western region had much smaller fall-related hospitalization rates than the other regions but followed similar patterns as the national models. The Western weather random forest had the highest feature importance for the minimum monthly temperature and the maximum windspeed. However, the incredibly high error (100% MAPE) makes this random forest unusable for description or prediction. The demographic random forest performed much better than the weather model, but still had a much higher error percentage than any of the national models

(53.2% MAPE). The same pattern as the national demographic model was observed with population density and proportion female continuing to be the most important features for fall-related hospitalizations predictions (higher rates for more rural and female counties). The socioeconomic random forest had a similar error as the demographic model and considered all the features in its predictions (median income MDI 0.45, proportion under the poverty line 0.30, and Gini Index of Income Inequality 0.25), as opposed to the national model that only used median income. In the full model, population density and proportion female were the most important features and weather variables again were rarely used for prediction in the entire random forest (MDI < 0.01).

The Midwestern weather random forest included more features in its predictions than the national model. There was about an even contribution of monthly average temperature and maximum windspeed (MDI 0.28), followed by maximum precipitation and minimum temperature (MDI ~0.17). The use of multiple weather features in prediction combined with the complicated random forest structure (highest values of all hyperparameters provided), suggests that there is a complex relationship in the Midwestern region between weather and fall-related hospitalization rates. The demographic model improved accuracy and continued to demonstrate the pattern of first selecting on population density (MDI 0.72) and then proportion female (MDI 0.16), with small contributions of the other two features. The socioeconomic model again selected median income as the most important feature, but it also had the highest feature importance for the Gini Index of Income Inequality (MDI 0.29). The full Midwestern model had high predictive accuracy (24.5% MAPE) and again primarily split on demographics (population density and proportion female).

The Southern random forests displayed slightly different patterns than the comparable national or other regional models. Although temperature variables were again the most important features in the Southern weather model, it was primarily monthly averages and minimums that predicted fall-related hospitalization rates (0.47 and 0.44 MDI, respectively). Overall cooler temperatures presented the highest risk in the South as opposed to variable monthly temperatures or high monthly maximum temperatures. The ranking of the feature importance values from the Southern demographic model also displayed a different order. The most important feature was again population density (MDI 0.55); however, this was followed by the proportion of non-Hispanic white older adults (MDI 0.22), and then the proportion of adults 85 years and older (MDI 0.15). Counties that were more rural, old, and white had higher fall-related hospitalization rates. The Southern socioeconomic random forest followed the same pattern as the national model with only median income as the predictive feature (MDI 1.0). Between the highest and lowest median income counties, the fall-related hospitalization rate more than doubled. The full Southern model continued to rank the demographic features as the most important, but again had a different order than the national or other regional models.

The Northeastern random forests also followed different patterns. In the Northeastern weather model, the monthly maximum windspeed was the most important feature (MDI 0.31) followed by the monthly minimum temperature (MDI 0.20) and a more even split between precipitation (MDI 0.15), maximum temperature (MDI 0.12), and snow (MDI 0.10). This was again the most complicated random forest, suggesting a complex relationship between multiple weather variables and fall-related hospitalization rates. The Northeastern demographic model was also different in its ranking of predictive features with the proportion of adults 85 years of age and older (MDI 0.37) as the most important followed by proportion female (MDI 0.32). Population density was the least important. The Northeastern socioeconomic random forest almost evenly ranked the importance of the proportion under the poverty line (MDI 0.46) and median income (MDI 0.42). The full Northeastern model was the most accurate of all the regional models with an MAPE of 23.7%. This model again prioritized demographic features in its predictions of fall-related hospitalization rates, in the same order as the demographics model. It also ranked socioeconomic variables higher in feature importance than other regional models (MDI 0.17 for median income and MDI 0.10 for proportion under the poverty line).

Table 9. Results from the regional random forests: testing (training) MAPE, ranking of features (MDI values), and optimal hyperparameters. Only features with MDI values greater than 0.01 are presented.

		West	**Midwest**	**South**	**Northeast**
Weather Model	*Testing (Training) MAPE*	100% (84.4%)	32.9% (31.6%)	31.7% (34.8%)	29.6% (26.6%)
	Feature Rank (MDI)	Minimum temperature (0.51), maximum windspeed (0.23), maximum temperature (0.12), average temperature (0.09), maximum precipitation (0.04)	Maximum temperature (0.28), maximum windspeed (0.28), maximum precipitation (0.17), minimum temperature (0.17), average temperature (0.08), maximum snow (0.02)	Average temperature (0.46), minimum temperature (0.44), maximum precipitation (0.05), maximum temperature (0.04), maximum windspeed (0.01)	Maximum windspeed (0.31), minimum temperature (0.22), precipitation maximum (0.15), maximum temperature (0.12), maximum snow (0.10), average temperature (0.09)
	Optimal Hyper-parameters	Max depth: 5, min samples per leaf: 10, number of trees: 200	Max depth: 5, min samples per leaf: 10, number of trees: 200	Max depth: 2, min samples per leaf: 1, number of trees: 50	Max depth: 5, min samples per leaf: 1, number of trees: 200
Demographic Model	*Testing (Training) MAPE*	53.2% (45.2%)	24.7% (22.8%)	26.2% (30.2%)	24.8% (22.5%)
	Feature Rank (MDI)	Population density (0.61), proportion female (0.26), proportion	Population density (0.72), proportion female (0.16), proportion non-	Population density (0.54), proportion non-Hispanic white	Proportion 85 years and older (0.37), proportion female (0.32), proportion

		non-Hispanic white (0.07), proportion 85 years and older (0.06)	Hispanic white (0.09), proportion 85 years and older (0.03)	(0.22), proportion 85 years and older (0.15), proportion female (0.08),	non-Hispanic white (0.17), population density (0.15)
	Optimal Hyper-parameters	Max depth: 5, min samples per leaf: 10, number of trees: 200	Max depth: 5, min samples per leaf: 5, number of trees: 50	Max depth: 4, min samples per leaf: 1, number of trees: 200	Max depth: 5, min samples per leaf: 10, number of trees: 50
Socioeconomic Model	*Testing (Training) MAPE*	55.5% (49.7%)	28.1% (26.4%)	29.7% (31.4%)	24.1% (22.5%)
	Feature Rank (MDI)	Median income (0.45), proportion poor (0.30), Gini Index of Income Inequality (0.25)	Median income (0.54), Gini Index of Income Inequality (0.29), proportion poor (0.17)	Median income (1.0)	Proportion poor (0.46), median income (0.42), Gini Index of Income Inequality (0.12)
	Optimal Hyper-parameters	Max depth: 5, min samples per leaf: 10, number of trees: 200	Max depth: 5, min samples per leaf: 10, number of trees: 200	Max depth: 2, min samples per leaf: 1, number of trees: 200	Max depth: 5, min samples per leaf: 10, number of trees: 50
Full Model	*Testing (Training) MAPE*	53.2% (44.3%)	24.5% (22.6%)	26.3% (29.8%)	23.7% (26.6%)
	Feature Rank (MDI)	Population density (0.55), proportion female (0.22), Gini Index of Income Inequality (0.08), median income (0.06), proportion non-Hispanic white (0.03), proportion poor (0.02), proportion 85 years and older (0.02)	Population density (0.69), proportion female (0.14), proportion non-Hispanic white (0.05), median income (0.04), Gini Index of Income Inequality (0.03), proportion 85 years and older (0.03), proportion poor (0.02)	Population density (0.46), proportion non-Hispanic white (0.19), proportion 85 years and older (0.14), median income (0.10), proportion poor (0.06), Gini Index of Income Inequality (0.03), proportion female (0.02),	Proportion 85 years and older (0.30), proportion female (0.22), median income (0.17), proportion poor (0.10), population density (0.09), Gini Index of Income Inequality (0.07), proportion non-Hispanic white (0.03)
	Optimal Hyper-parameters	Max depth: 5, min samples per leaf: 10, number of trees: 200	Max depth: 5, min samples per leaf: 10, number of trees: 50	Max depth: 4, min samples per leaf: 10, number of trees: 200	Max depth: 5, min samples per leaf: 10, number of trees: 50

Regional Mixed-Effects Poisson Models

The same four mixed-effects Poisson models with each set of variables (weather, demographic, socioeconomic, and full) were conducted for each of the four regions. Almost all the weather variables were significant in the Western, Midwestern, and Southern weather models. In the Northeast, only monthly maximum precipitation and maximum temperature were significant after controlling for all other weather variables. In the Western weather model, the temperature variables had the largest effect sizes. A standard deviation increase in average monthly temperatures was associated with a 1.021 (95% CI: [1.007, 1.034]; $p < 0.01$) increase in the fall-related hospitalization rates. Increased monthly minimum and maximum temperatures were associated with decreased fall-related hospitalization rates. A standard deviation increase in monthly snowfall and precipitation was also associated with a small increase in fall-related hospitalization rates. In the Midwest, higher monthly average and minimum temperatures were associated with decreased fall-related hospitalization rates (around 0.98). Higher monthly maximum temperatures, precipitation, and snow were associated with increased rates. The Southern region followed similar patterns as the Midwestern region but experienced slightly decreased fall-related hospitalizations with a standard deviation increase in snowfall (estimate: 0.998; 95% CI: [0.997, 0.999]; $p < 0.001$). In the Northeastern region, only two weather variables (maximum temperature and precipitation) were associated with fall-related hospitalization rates. Table 10 displays the results from the regional weather mixed-effects Poisson regressions.

Table 10. Estimates [95% confidence intervals] for regional mixed-effects Poisson regressions with weather variables, conducted on the 2009 – 2013 training data.

	West	Midwest	South	Northeast
Monthly Average Temperature	1.021 [1.007, 1.034] **	0.988 [0.979, 0.997] **	0.982 [0.976, 0.987] ***	0.995 [0.983, 1.006]
Monthly Minimum Temperature	0.979 [0.970, 0.988] ***	0.981 [0.975, 0.988] ***	0.995 [0.991, 0.999] *	0.994 [0.986, 1.002]
Monthly Maximum Temperature	0.981 [0.973, 0.990] ***	1.015 [1.010, 1.020] ***	1.011 [1.008, 1.014] ***	0.993 [0.987, 0.998] *
Monthly Maximum Precipitation	1.004 [1.000, 1.006] *	1.002 [1.000, 1.004] *	0.999 [0.998, 1.000]	1.006 [1.004, 1.009] ***
Monthly Maximum Snowfall	1.004 [1.001, 1.007] **	1.003 [1.002, 1.005] ***	0.998 [0.997, 0.999] ***	1.001 [0.999, 1.004]
Monthly Maximum Windspeed	0.997 [0.994, 1.000]	0.997 [0.995, 0.999] **	1.001 [0.999, 1.002]	1.000 [0.997, 1.004]

* $p < 0.05$
**$p < 0.01$
***$p < 0.001$

In the regional demographic mixed-effect Poisson models, the proportion of adults 65 years and older that are female was consistently associated with decreased fall-related hospitalization rates across all regions after controlling for other demographics. Population density and proportion of older adults 85 years and older were associated with fall-related hospitalization rates in all but the Northeastern region. A standard deviation increase in the proportion non-Hispanic white adults was associated with an increase in fall-related hospitalization rates only in the Midwest and Northeast. The results from the regional demographic mixed-effects Poisson models can be found in Table 11.

Table 11. Estimates [95% confidence intervals] for regional mixed-effects Poisson regressions with demographic variables, conducted on the 2009 – 2013 training data.

	West	Midwest	South	Northeast
Proportion 85 years and older	1.366 [1.126, 1.658] **	1.193 [1.092, 1.304] ***	1.084 [0.990, 1.186] *	1.032 [0.889, 1.198]
Proportion Female	0.411 [0.340, 0.500] ***	0.866 [0.789, 0.951] **	0.873 [0.784, 0.972] *	0.508 [0.424, 0.608] ***
Proportion non-Hispanic White	0.779 [0.600, 1.012]	1.395 [1.190, 1.634] ***	1.031 [0.934, 1.138]	1.372 [1.110, 1.700] **
Population Density	0.675 [0.540, 0.843] ***	0.537 [0.455, 0.635] ***	0.582 [0.511, 0.665] ***	1.043 [0.898, 1.211]

* p < 0.05
**p< 0.01
***p <0.001

In the regional socioeconomic mixed-effects Poisson models, a standard deviation higher median county income was associated with halved fall-related hospitalization rates. A higher Gini Index of Income Inequality (more unequal) was also associated with decreased fall-related hospitalization rates in all regions, after controlling for the other socioeconomic variables. An increased proportion of older adults under the poverty line had different associations by region – there was no significant association in the Western region, a decreased association in the Northeast, and increased associations in South and Midwest. Detailed results from these regressions are presented below in Table 12.

Table 12. Estimates [95% confidence intervals] for regional mixed-effects Poisson regressions with socioeconomic variables, conducted on the 2009 – 2013 training data.

	West	**Midwest**	**South**	**Northeast**
Median Income	0.391 [0.319, 0.480] ***	0.515 [0.462, 0.574] ***	0.488 [0.438, 0.543] ***	0.583 [0.490, 0.693] ***
Gini Index of Income Inequality	0.703 [0.586, 0.844] ***	0.551 [0.503, 0.604] ***	0.721 [0.665, 0.782] ***	0.651 [0.559, 0.758] ***
Proportion under the Poverty Line	0.999 [0.835, 1.196]	1.128 [1.026, 1.239] *	1.237 [1.146, 1.334] ***	0.832 [0.709, 0.977] *

* $p < 0.05$
**$p < 0.01$
***$p < 0.001$

The full regional mixed-effects Poisson regression models followed similar patterns as their component models. Notably, the same associations with the weather variables remained after controlling for the relevant sociodemographics. Higher monthly minimum temperatures were associated with a decrease in fall-related hospitalization rates in the West, Midwest, and South. The association between higher monthly maximum temperatures and fall-related hospitalization rates varied by region – higher in the Midwest and South, but lower in the West and Northeast. Higher snowfall was associated with increased fall-related hospitalization rates in the West and Midwest, but lower rates in the South. Higher maximum precipitation was associated with higher fall-related hospitalization rates in the Northeast and Midwest. The associations with population density were mostly attenuated in the full regional models, except in the Midwest where more rural counties were associated with about a 50% increase in fall-related hospitalization rates. As opposed to the random forests, higher proportions of women were associated with decreased fall-related hospitalization rates after controlling for the sociodemographics. Median county income continued to have a large association with fall-related hospitalization rates in all regions. The results of the regional mixed-effects Poisson regression models with all variables are in Table 13.

Table 13. Estimates [95% confidence intervals] for regional mixed-effects Poisson regressions with all variables, conducted on the 2009 – 2013 training data.

	West	Midwest	South	Northeast
Monthly Average Temperature	1.021 [1.007, 1.034] **	0.988 [0.979, 0.997] **	0.982 [0.976, 0.987] ***	0.995 [0.983, 1.006]
Monthly Minimum Temperature	0.979 [0.970, 0.988] ***	0.981 [0.975, 0.988] ***	0.995 [0.991, 0.999] *	0.994 [0.986, 1.002]
Monthly Maximum Temperature	0.981 [0.973, 0.990] ***	1.015 [1.010, 1.020] ***	1.011 [1.008, 1.014] ***	0.992 [0.987, 0.998] *
Monthly Maximum Precipitation	1.003 [1.000, 1.006]	1.002 [1.000, 1.004] *	0.999 [0.998, 1.000]	1.006 [1.004, 1.009] ***
Monthly Maximum Snowfall	1.004 [1.001, 1.007] **	1.003 [1.002, 1.005] ***	0.998 [0.997, 0.999] ***	1.001 [0.999, 1.004]
Monthly Maximum Windspeed	0.997 [0.994, 1.001]	0.997 [0.995, 0.999] **	1.001 [0.999, 1.002]	1.000 [0.997, 1.004]
Proportion 85 years and older	1.271 [1.068, 1.513] **	1.166 [1.080, 1.257] ***	1.110 [1.026, 1.202] **	1.045 [0.914, 1.195]
Proportion Female	0.551 [0.457, 0.666] ***	0.857 [0.790, 0.929] ***	0.786 [0.719, 0.860] ***	0.538 [0.458, 0.631] ***
Proportion non-Hispanic White	1.413 [1.032, 1.934] *	1.453 [1.265, 1.668] ***	1.335 [1.216, 1.465] ***	1.546 [1.242, 1.875] ***
Population Density	0.899 [0.722, 1.120]	0.676 [0.583, 0.784] ***	0.961 [0.850, 1.087]	1.046 [0.899, 1.218]
Median Income	0.619 [0.496, 0.774] ***	0.725 [0.655, 0.802] ***	0.580 [0.516, 0.652] ***	0.744 [0.638, 0.867] ***
Gini Index of Income Inequality	0.817 [0.691, 0.968] *	0.769 [0.704, 0.839] ***	0.790 [0.726, 0.861] ***	0.894 [0.787, 1.017]
Proportion under the Poverty Line	1.466 [1.179, 1.824] ***	1.326 [1.223, 1.437]	1.510 [1.388, 1.643] ***	1.216 [1.014, 1.459] *

* $p < 0.05$
**$p < 0.01$
***$p < 0.001$

The descriptive and predictive performance of the regional mixed-effects Poisson regressions was more consistent across models and regions compared to the random forests. The Western models actually had the lowest testing MAPE, when this had been the region with the highest error rates in the random forest models. The highest accuracy for predicting 2014 – 2015 fall-related hospitalization rates for the regional random forest models was the Northeastern full model with an MAPE of 23.7%, which was comparable to the Western mixed-effects Poisson models around 24.2% MAPE. The MAPE values for the regional mixed-effects Poisson models are below in Table 14.

Table 14. Test (train) MAPE values for regional mixed-effects Poisson models.

	West	**Midwest**	**South**	**Northeast**
Weather Model	24.06 (22.41)	30.89 (30.47)	32.53 (31.65)	28.39 (27.92)
Demographic Model	24.18 (22.41)	30.79 (30.47)	32.57 (31.65)	28.49 (27.92)
Socioeconomic Model	24.25 (22.41)	30.86 (30.47)	32.54 (31.65)	28.44 (27.92)
Full Model	24.29 (22.41)	30.76 (30.47)	32.58 (31.65)	28.48 (27.92)

Discussion

Weather, demographic, and socioeconomic variables are all important predictors of fall-related hospitalization rates in older adults. However, these associations varied both by region and modeling approach. Overall, temperature variables were the most influential weather variables at the national level and for all regions except the Northeast. The random forests identified months with variable temperatures (both high maximums and low minimums) as having the highest fall-related hospitalization rates. While the mixed-effects Poisson regressions suggested a more U-shaped temperature relationship, where even holding constant for all temperature and weather variables, both high monthly maximum temperatures and low minimum temperatures were associated with increased fall-related hospitalization rates at the national level and in the Midwestern and Southern regions. The mixed-effects Poisson models allowed the sociodemographic confounders to be controlled for while estimating the small effects of the weather variables on fall-related hospitalization rates, whereas the competitive random forests dropped these variables in favor of the more predictive social factors. In the random forests, population density and the proportion of women 65 years and older were consistently the most predictive demographic features (more rural and female counties presented the highest risk). However, after controlling for all variables in the mixed-effects Poisson models, only rural Midwestern counties experienced increased fall-related hospitalization rates and the directionality of the relationship with the proportion of older women flipped. Median county income was the most predictive socioeconomic variable across most models. The predictive accuracy of the random forest and the mixed-effects Poisson regressions were relatively similar, however the regional mixed-effects Poisson regression models with all variables were more consistent and had mean absolute percentage errors ranging from 24.3% to 32.6%. These models allow for better understanding of the epidemiology of fall-related hospitalization rates and could also be useful for public health planning.

The finding that lower average and minimum temperatures are associated with increased fall risk is well-supported in the literature.[9,14,27,28] However, in the full national mixed-effects Poisson regression models, higher monthly maximum temperatures were associated with increased fall-related hospitalization rates which differs from two studies that used the same weather metrics.[13,15] When separating by region, this U-shaped relationship between increased fall-related hospitalizations and both higher maximum and lower minimum monthly temperatures only exists in the Midwest and South. The study population and methods may contribute to these different findings as one was conducted in Hong Kong using Pearson correlation coefficients[13] and the other was conducted in Montreal, Canada using ARIMA modelling.[15] It was surprising that snowfall was not significantly associated with monthly fall-related hospitalizations at the national level, as this has been one of the most consistently demonstrated associations,[9] but its effect was seen regionally. Stevens et al. also found that regional variation was an important predictor of fall rates, but took a different approach to grouping states based on whether their average January temperature was below freezing.[19] They found that fatal fall rates were higher in these cooler states regardless of season, where this thesis demonstrated a more complicated relationship between climate and fall-related hospitalizations.[19] Although the seasonality of falls is a popular and contentious topic among this literature, [9,15,16,20,28,58] this was not explicitly explored here due to varying definitions of what constitutes a season based on region, climate, and year; however, it may be worth considering in future analysis. The machine learning applications in this field are primarily used to predict whether a patient will fall based on individual characteristics, in order to allocate additional resources for fall prevention and older adult care.[59-61] Young et al. & Liu et al. both found that random forests were the best performing models for that prediction task.[59,60] There was one study, Crowson et al, that modeled fall-related death rates using multiple machine learning methods but used healthcare utilization data as predictors.[62] However, the modeling techniques and metrics used here have not yet been applied to these research questions.

Conducting this analysis with both random forests and mixed-effects Poisson regressions allowed for an interesting opportunity to compare and contrast these approaches. Random forests can accurately predict complicated nonlinear relationships as the constitutive trees consider at each node what variable split maximizes information gain and, in doing so, create final nodes that operate as conditional statements[3] that are then averaged across in the final forest. They are also relatively robust to noise, errors, and missing data.[44] Lastly, they rank the most important variables for prediction which can be useful for feature selection and possibly for prioritizing interventions. However, random forests on their own do not adjust for confounding like classic statistical methods and improving this is an active area of machine learning research.[63,64] Some confounding adjustment methods for machine learning include sample matching, inverse probability weighting, penalized learners, and restricted permutations.[63] Another approach could include adjusting the rates for geography, age and other confounders, which is how Crowson et. al handled this problem.[62] These strengths and limitations were demonstrated through this project. The conditional decision-making structure identified that months with variable temperature (both high maximums and low minimums) had the highest fall-related

[3] For example, a final node could represent that in months where the minimum temperature is below -4°C *and* the maximum is above 19°C, the average fall-related hospitalization rate is 167 falls per 100,000 older adults.

hospitalization rates. In mixed-effects Poisson regression, this cannot be easily identified without creating additional features, as it holds all other variables constant when estimating one feature's effect. However, without applying additional confounding adjustment techniques for random forests, some important problems arose. The sociodemographic confounding adjustment is necessary in this context as they may influence the associations between weather variables and fall-related hospitalization rates. When these sociodemographic characteristics were added to the full models, the decision-making criteria of the random forests completely ignored the weather variables due to their small effects. There was also a Simpson's paradox where the proportion of older women was positively associated with fall-related hospitalization rates in the random forests, but after adjusting for other demographic characteristics in the mixed-effects Poisson regressions this association flipped.

There are also strengths and limitations to the mixed-effects Poisson regression modeling. They are designed for repeated count data and can more easily adjust for spatial variation and confounders, but they cannot account for interactions without explicit model specification and it does not perform optimally if dispersion is present.[65] The mixed-effects Poisson regression models accounted for spatial variation and the skewed county populations along with all other variables by including fixed state and random county effects and population offset terms. To avoid random forests dictated by individual county rates, spatial variation was only explicitly considered by separating observations into regions. However, it may be useful in future random forest modelling to adjust the rates geographically or explicitly include spatial units to identify areas with a disproportionate burden. The mixed-effects Poisson models also do not allow for as much flexibility in their predictions as the random forests. There may be important interactions that were not specified and should be explored in additional sensitivity analyses. Although there are opportunities to improve both approaches through further model specification and feature engineering, overall, most of the models provided accurate descriptions and predictions that are useful for understanding and predicting the epidemiology of fall-related hospitalization among older adults.

This thesis contributes to the field by combining disparate data sources and modeling approaches. Various weather, demographic, and socioeconomic variables were scraped from publicly available, government sources and combined with six years of monthly county-level fall-related Medicare hospitalization records from across the United States. This time frame allows for larger trends to be captured and greater generalizability. The selection of the monthly weather metrics aimed to capture both average climate and more extreme events at a stable unit of analysis that did not require lagging. The two different modelling approaches also revealed a complicated geospatial relationship with these multiple monthly weather metrics and fall-related hospitalization rates. As discussed, the random forest models and the mixed-effects Poisson models presented different strengths and limitations. However, the predictive accuracy could be useful across modelling approaches. Particularly with the geospatial and confounding adjustment of the mixed-effects Poisson regressions, some regional models could predict monthly county fall rates with less than 25% error.

There are also important limitations to acknowledge. This thesis uses aggregated county-level variables for all analyses, and these associations should not be interpreted at the individual level. These associations may vary or not exist at the individual level, and individual sociodemographics were not released. Additionally, the county-level sociodemographics may not be the same as the individuals who were hospitalized due to falls. There was also meaningful missingness that may affect the generalizability of these findings. Due to the strict data protection protocol, about 40% of the counties in the United States did not report any monthly fall counts over the entire study period (July 2009 – July 2015). These counties tended to be more rural and potentially may have different associations than the ones that did report data. The monthly weather metrics may not accurately capture general climate trends or the fall-related hospitalization burden of weather events (particularly extreme ones). The improved performance of the mixed-effects Poisson models could also mean that there are meaningful state and county level characteristics that affect fall-related hospitalization rates that are not accounted for here.

Weather, demographic, and socioeconomic variables were important for predicting monthly county fall-related hospitalization rates across the United States from July 2009 – July 2015. Temperature variables tend to be the most important weather variables, with variation and extreme minimums and maximums presenting the largest risk but differing by region. Demographic variables are the most influential characteristics, but their associations differ depending on confounding adjustment through the modelling approaches. Socioeconomic variables are also relevant, mainly the county's median income. Machine learning and classical statistics approaches were compared, with the mixed-effects Poisson models more consistently outperforming the random forests, likely due to its ability to account for spatial variation and confounding. There are meaningful spatial trends to the epidemiology of falls in the United States that are influenced by individual and environmental risk factors.

References

1. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. *Maturitas*. May 2013;75(1):51-61. doi:10.1016/j.maturitas.2013.02.009
2. Dipietro L, Campbell WW, Buchner DM, et al. Physical Activity, Injurious Falls, and Physical Function in Aging: An Umbrella Review. *Med Sci Sports Exerc*. 06 2019;51(6):1303-1313. doi:10.1249/MSS.0000000000001942
3. Florence CS, Bergen G, Atherly A, Burns E, Stevens J, Drake C. Medical Costs of Fatal and Nonfatal Falls in Older Adults. *J Am Geriatr Soc*. 04 2018;66(4):693-698. doi:10.1111/jgs.15304
4. Masud T, Morris RO. Epidemiology of falls. *Age Ageing*. Nov 2001;30 Suppl 4:3-7. doi:10.1093/ageing/30.suppl_4.3
5. Feldman F, Chaudhury H. Falls and the physical environment: a review and a new multifactorial falls-risk conceptual framework. *Can J Occup Ther*. Apr 2008;75(2):82-95. doi:10.1177/000841740807500204
6. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. *Cochrane Database Syst Rev*. Sep 12 2012;(9):CD007146. doi:10.1002/14651858.CD007146.pub3
7. Lord SR, Menz HB, Sherrington C. Home environment risk factors for falls in older people and the efficacy of home modifications. *Age Ageing*. Sep 2006;35 Suppl 2:ii55-ii59. doi:10.1093/ageing/afl088
8. Stevens JA. Falls among older adults--risk factors and prevention strategies. *J Safety Res*. 2005;36(4):409-11. doi:10.1016/j.jsr.2005.08.001
9. Chow KP, Fong DYT, Wang MP, Wong JYH, Chau PH. Meteorological factors to fall: a systematic review. *Int J Biometeorol*. Dec 2018;62(12):2073-2088. doi:10.1007/s00484-018-1627-y
10. Kakara RS, Moreland BL, Haddad YK, Shakya I, Bergen G. Seasonal variation in fall-related emergency department visits by location of fall - United States, 2015. *J Safety Res*. 12 2021;79:38-44. doi:10.1016/j.jsr.2021.08.002
11. Wareham K, Johansen A, Stone MD, Saunders J, Jones S, Lyons RA. Seasonal variation in the incidence of wrist and forearm fractures, and its consequences. *Injury*. Mar 2003;34(3):219-22. doi:10.1016/s0020-1383(02)00212-7
12. Qian XX, Chau PH, Kwan CW, et al. Seasonal pattern of single falls and recurrent falls amongst community-dwelling older adults first applying for long-term care services in Hong Kong. *Age Ageing*. 12 01 2019;49(1):125-129. doi:10.1093/ageing/afz139
13. Yeung P-Y, Chau P-H, Woo J, Wai-Ting Yim V, Rainer TH. Higher incidence of falls in winter among older people in Hong Kong,. *Journal of Clinical Gerontology and Geriatrics*. 2011;2(1):13-16. doi:https://doi.org/10.1016/j.jcgg.2011.02.002.
14. Giladi AM, Shauver MJ, Ho A, Zhong L, Kim HM, Chung KC. Variation in the incidence of distal radius fractures in the U.S. elderly as related to slippery weather conditions. *Plast Reconstr Surg*. Feb 2014;133(2):321-332. doi:10.1097/01.prs.0000436796.74305.38
15. Modarres R, Ouarda TB, Vanasse A, Orzanco MG, Gosselin P. Modeling seasonal variation of hip fracture in Montreal, Canada. *Bone*. Apr 2012;50(4):909-16. doi:10.1016/j.bone.2012.01.004

16. Magota C, Sawatari H, Ando SI, et al. Seasonal ambient changes influence inpatient falls. *Age Ageing*. 05 01 2017;46(3):513-517. doi:10.1093/ageing/afw254
17. López-Soto PJ, Smolensky MH, Sackett-Lundeen LL, et al. Temporal Patterns of In-Hospital Falls of Elderly Patients. *Nurs Res*. 2016 Nov/Dec 2016;65(6):435-445. doi:10.1097/NNR.0000000000000184
18. Launonen AP, Lepola V, Saranko A, Flinkkilä T, Laitinen M, Mattila VM. Epidemiology of proximal humerus fractures. *Arch Osteoporos*. 2015;10:209. doi:10.1007/s11657-015-0209-4
19. Stevens JA, Thomas KE, Sogolow ED. Seasonal patterns of fatal and nonfatal falls among older adults in the U.S. *Accid Anal Prev*. Nov 2007;39(6):1239-44. doi:10.1016/j.aap.2007.03.011
20. Jacobsen SJ, Sargent DJ, Atkinson EJ, O'Fallon WM, Melton LJ. Population-based study of the contribution of weather to hip fracture seasonality. *Am J Epidemiol*. Jan 01 1995;141(1):79-83. doi:10.1093/oxfordjournals.aje.a117348
21. Mondor L, Charland K, Verma A, Buckeridge DL. Weather warnings predict fall-related injuries among older adults. *Age Ageing*. May 2015;44(3):403-8. doi:10.1093/ageing/afu199
22. Huynh D, Tracy C, Thompson W, et al. Associations between meteorological factors and emergency department visits for unintentional falls during Ontario winters. *Health Promot Chronic Dis Prev Can*. Dec 2021;41(12):401-412. doi:10.24095/hpcdp.41.12.01
23. Bobb JF, Ho KK, Yeh RW, et al. Time-Course of Cause-Specific Hospital Admissions During Snowstorms: An Analysis of Electronic Medical Records From Major Hospitals in Boston, Massachusetts. *Am J Epidemiol*. 02 15 2017;185(4):283-294. doi:10.1093/aje/kww219
24. Bergström U, Björnstig U, Stenlund H, Jonsson H, Svensson O. Fracture mechanisms and fracture pattern in men and women aged 50 years and older: a study of a 12-year population-based injury register, Umeå, Sweden. *Osteoporos Int*. Sep 2008;19(9):1267-73. doi:10.1007/s00198-007-0549-z
25. Saari P, Heikkinen E, Sakari-Rantala R, Rantanen T. Fall-related injuries among initially 75- and 80-year old people during a 10-year follow-up. *Arch Gerontol Geriatr*. 2007 Sep-Oct 2007;45(2):207-15. doi:10.1016/j.archger.2006.10.012
26. Falch JA, Ilebekk A, Slungaard U. Epidemiology of hip fractures in Norway. *Acta Orthop Scand*. Feb 1985;56(1):12-6. doi:10.3109/17453678508992970
27. Turner RM, Hayen A, Dunsmuir WT, Finch CF. Air temperature and the incidence of fall-related hip fracture hospitalisations in older people. *Osteoporos Int*. Apr 2011;22(4):1183-9. doi:10.1007/s00198-010-1306-2
28. Lin HC, Xiraxagar S. Seasonality of hip fractures and estimates of season-attributable effects: a multivariate ARIMA analysis of population-based data. *Osteoporos Int*. 2006;17(6):795-806. doi:10.1007/s00198-005-0060-3
29. Johnson NA, Stirling E, Alexander M, Dias JJ. The relationship between temperature and hip and wrist fracture incidence. *Ann R Coll Surg Engl*. May 2020;102(5):348-354. doi:10.1308/rcsann.2020.0030
30. Vongsachang H, Mihailovic A, E JY, et al. The Impact of Weather and Seasons on Falls and Physical Activity among Older Adults with Glaucoma: A Longitudinal Prospective Cohort Study. *Sensors (Basel)*. May 14 2021;21(10)doi:10.3390/s21103415

31. Gevitz K, Madera R, Newbern C, Lojo J, Johnson CC. Risk of Fall-Related Injury due to Adverse Weather Events, Philadelphia, Pennsylvania, 2006-2011. *Public Health Rep*. 2017 Jul/Aug 2017;132(1_suppl):53S-58S. doi:10.1177/0033354917706968
32. Sanghavi P, Pan S, Caudry D. Assessment of nursing home reporting of major injury falls for quality measurement on nursing home compare. *Health Serv Res*. 04 2020;55(2):201-210. doi:10.1111/1475-6773.13247
33. Kim SB, Zingmond DS, Keeler EB, et al. Development of an algorithm to identify fall-related injuries and costs in Medicare data. *Inj Epidemiol*. Dec 2016;3(1):1. doi:10.1186/s40621-015-0066-z
34. What are E-codes and why are they important? 2014. https://ncdetect.org/wp-content/uploads/sites/13428/2016/12/CCHI_E_CodeFactSheetJan2014.pdf
35. & CfM, Services M. ICD-10. Centers for Medicare & Medicaid Services. Accessed March 16, 2022, 2022. https://www.cms.gov/Medicare/Coding/ICD10
36. Research NBoE. SSA to Federal Information Processing Series (FIPS) State and County Crosswalk. National Bureau of Economic Research. 2022. https://www.nber.org/research/data/ssa-federal-information-processing-series-fips-state-and-county-crosswalk
37. *Meteostat*. 2022. https://dev.meteostat.net/guide.html#about-meteostat
38. *GeoPandas*. Version v0.8.1. Zenodo; 2020. https : // doi . org / 10.5281 / zenodo .3946761
39. Yan H. Data from: US-counties. 2021. *The Python Graph Gallery*.
40. Gasparrini A, Guo Y, Hashizume M, et al. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *Lancet*. Jul 25 2015;386(9991):369-75. doi:10.1016/S0140-6736(14)62114-0
41. Liang M, Ding X, Yao Z, Duan L, Xing X, Sun Y. Effects of ambient temperature and fall-related injuries in Ma'anshan, Anhui Province, China: a distributed lag nonlinear analysis. *Environ Sci Pollut Res Int*. Nov 2021;28(41):58092-58103. doi:10.1007/s11356-021-14663-2
42. Bureau USC. Gini Index. United States Census Bureau. Accessed April 25, 2022, 2022. https://www.census.gov/topics/income-poverty/income-inequality/about/metrics/gini-index.html
43. Bureau USC. 2010 Census Regions and Divisions of the United States. Accessed 4/1/2022, 2022. https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf
44. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32.
45. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *J Med Syst*. Oct 2002;26(5):445-63. doi:10.1023/a:1016409317640
46. Kingsford C, Salzberg SL. What are decision trees? *Nat Biotechnol*. Sep 2008;26(9):1011-3. doi:10.1038/nbt0908-1011
47. Singh S, Gupta P. COMPARATIVE STUDY ID3, CART AND C4.5 DECISION TREE ALGORITHM: A SURVEY. *International Journal of Advanced Information Science and Technology*. 2014;27(27):97 - 103.
48. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*. 2000;16(4):437 - 450. doi:https://doi.org/10.1016/S0169-2070(00)00065-0

49. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One*. Sep 28 2010;5(9)doi:10.1371/journal.pone.0012776
50. Breiman L. *Classification and regression trees*. Wadsworth statistics/probability series. Wadsworth International Group; 1984:x, 358 p.
51. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit - learn : Machine Learning in P ython. *Journal of Machine Learning Research*. 2011;12:2825 -- 2830.
52. Li D, Gaynor SM, Quick C, et al. Identifying US County-level characteristics associated with high COVID-19 burden. *BMC Public Health*. 05 28 2021;21(1):1007. doi:10.1186/s12889-021-11060-9
53. Svetunkov I. Forecasts Evaluation. In: OpenForecast M, ed. *Forecasting and Analytics with ADAM*. OpenForecast; 2022:chap 2. https://www.openforecast.org/adam/
54. Bates D, Machler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*. 2015;67(1):1 - 48. doi:10.18637/jss.v067.i01
55. Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. *Journal of Open Source Software*. 2019;4(43):1686. doi:10.21105/joss.01686
56. *"MLmetrics: Machine Learning Evaluation Metrics"*. Version 1.1.1. CRAN; 2016.
57. Myttenaere Ad, Golden B, Le Grand B, Rossi F. Mean Absolute Percentage Error for regression models. *Neurocomputing*. 2016;192:38-48. doi:https://doi.org/10.1016/j.neucom.2015.12.114
58. Mazzucchelli R, Crespí-Villarías N, Pérez-Fernández E, et al. Weather conditions and their effect on seasonality of incident osteoporotic hip fracture. *Arch Osteoporos*. Mar 15 2018;13(1):28. doi:10.1007/s11657-018-0438-4
59. Young AJ, Hare A, Subramanian M, Weaver JL, Kaufman E, Sims C. Using Machine Learning to Make Predictions in Patients Who Fall. *J Surg Res*. 01 2021;257:118-127. doi:10.1016/j.jss.2020.07.047
60. Liu CH, Hu YH, Lin YH. A Machine Learning-Based Fall Risk Assessment Model for Inpatients. *Comput Inform Nurs*. Apr 26 2021;39(8):450-459. doi:10.1097/CIN.0000000000000727
61. Usmani S, Saboor A, Haris M, Khan MA, Park H. Latest Research Trends in Fall Detection and Prevention Using Machine Learning: A Systematic Review. *Sensors (Basel)*. Jul 29 2021;21(15)doi:10.3390/s21155134
62. Crowson MG, Beyea JA, Cottrell J, et al. The predictive power of geographic health care utilization for unintentional fatal fall rates. *BMC Public Health*. Feb 16 2022;22(1):328. doi:10.1186/s12889-022-12731-x
63. Neto EC, Pratap A, Perumal TM, et al. A Permutation Approach to Assess Confounding in Machine Learning Applications for Digital Health. *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019:54 - 64. doi:10.1145/3292500.3330903
64. Bica I, Alaa AM, Lambert C, van der Schaar M. From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges. *Clin Pharmacol Ther*. 01 2021;109(1):87-100. doi:10.1002/cpt.1907

65. Yirga AA, Melesse SF, Mwambi HG, Ayele DG. Negative binomial mixed models for analyzing longitudinal CD4 count data. *Sci Rep*. 10 07 2020;10(1):16742. doi:10.1038/s41598-020-73883-7