Modeling Residential Fire Risk:

A Multilevel Approach

Kit Miller

Northeastern University

Research Practicum in Urban Informatics

Table of Contents

Modeling Residential Fire Risk:

A Multilevel Approach

Despite long term trends showing declining numbers of deaths, injuries, property damage, and overall incidents nationwide, residential building fires are still a leading cause of childhood injury and property damage in the United States. In 2014, 379,500 residential building fires caused nearly seven billion dollars in property damage, injured over 12,000 persons, and caused 2,765 deaths (U.S. Fire Administration, 2016).[1] For the same year, unintentional fire or burning was among the top ten leading causes of child injury and death (National Center for Injury Prevention and Control, 2016). Most of these injuries and fatalities occur at home: although residential building fires account for roughly 30 percent of all fires annually in the United States, they cause at least 75 percent of civilian fire fatalities and injuries each year (U.S. Fire Administration, 2016).

Even as residential fire continues to exact a heavy toll in the United States[2], we still lack a detailed understanding of the underlying spatial, social, and economic patterns of fire incidence. It is a well-established fact that residential fires are not uniformly distributed across geographies or socioeconomic groups, but there have been few rigorous statistical studies since the 1980s to reinforce or extend early research on demographic and building factors associated with greater fire risk. Studies from the late 1970s and early 1980s (Gilliam, 1985; Gunther, 1984; Karter & Donner, 1977; Schaenman, Hall, Schainblatt, Swartz, & Karter, 1977) all found that the incidence of fire was linked to higher rates of poverty and poorer housing quality. Since then, virtually every study of socioeconomic characteristics has shown that lower levels of income are

[1] 2014 is the latest year for which fire data is available from the U.S. Federal Emergency Management Agency.

[2] In 1979, the United States had the highest death rate from fire (34 deaths per million) of any industrialized country. That rate is now 10.7 deaths per million. Although drastically better, the United States still has the 10th highest death rate out of 24 industrialized countries.

either directly or indirectly tied to an increased risk of fire (USFA National Fire Data Center, 1997). Several studies also demonstrated a relationship between single-parent households with children and residential fire rates.

Nearly all of these studies on fire distribution have analyzed fire rates at spatial resolutions no greater than the neighborhood or census tract level. Some studies have compared fire rates between census tracts of a single city and other studies have compared fire rates between cities. Studies that have examined building characteristics have either looked at building factors like average age or percent vacant in the aggregate or at summary statistics of fire rates by building type. No known studies have looked at variations in fire rates by street segment, a level of spatial resolution that has become more common in the field of criminology over the last 20 years (D. O'Brien & Winship, 2016). Consequently, although we have some indications of variation in fire rates based on building age and vacancy rates, we know little about any other building-level factors or street-level factors that may be associated with higher rates of fire.

The current study seeks to address this research gap by using multiple logistic regression and multilevel mixed-effects models at the building level and census tract levels to assess fire risk for residential buildings in Boston, MA. The goals of the study are threefold. First, we want to build upon past research into socioeconomic and building factors at the census tract level and assess whether Boston fits the fire risk patterns identified in previous studies. Secondly, we want to determine if there are consistent building-level factors across the city that significantly increase the risk of fire in certain types of buildings. Such a determination would warrant further research into building-level risk assessments for other cities and would extend our knowledge of residential fire risk with greater specificity. Thirdly, we want to assess whether multi-level mixed effects models, by accounting for inter-neighborhood variance, can provide reasonably accurate assessments of building fire risk. Do multi-level mixed-effects models allow for reasonably accurate identification, for example, of a high-risk pool of buildings that has a significantly higher likelihood of experiencing a fire within a given time frame? If so, then further research again is warranted into the use of multi-level models for other cities, possibly with additional predictor variables and additional geographic or temporal groupings.

Thus, in attempting to answer these three questions, we hope to determine whether micro-level risk assessment and multi-level models show promise for the future of fire research and municipal approaches to fire risk assessment. Before presenting the data and analysis of the current study, the next two sections will 1) review the existing literature on the distribution of fire risk within cities and across socioeconomic groups and 2) show how the current study can extend our knowledge of fire risk from the conclusions reached in these previous studies.

Previous Research: Socioeconomic and Building Factors

The earliest known studies of fire incidence and its relationship to building stock criteria (Ridley, 1927; Simon, Shepard, & Sharp, 1943) looked at relationships between fire losses and the number and the assessed valuation of buildings at risk. Simon Shepard and Sharp (1943) found there was significant variation from census tract to census tract in the rate of fires, almost all of which could be attributed to single-family dwellings, meaning that the fire rates for multifamily dwellings were much more stable throughout the City (Jennings, 1996).

The first social approaches to fire risk were published in the 1970s. An early study by Schaenman, Hall, Schainblatt, Swartz and Karter (1977) used multiple regression analysis to study fire rates and census tract variables in five U.S. cities. The authors found three variables that explained the largest percentages of intra-tract variations in fire rates across all locations.

The first was parental presence, defined as the percentage of children under the age of 18 living with both parents. The second was poverty, defined as the percentage of persons whose incomes fell below the poverty line. And third was under-education, which is defined as the percentage of persons over the age of 25 who had fewer than eight years of schooling. Since Schaenman et al (1977), virtually every empirical study has found an association between poverty and fire rates. Karter and Donner (1978) looked at fire incidents and census data from five cities and identified poverty and parental presence as significant factors as well. Munson (1977), however, was the first to suggest different systems of causality at work in high and low-income neighborhoods. Based on regression models for fire rates in Charlotte, NC, Munson observed that population factors dominated in low-income tracts while housing factors were strongest in the high-income tracts (M. Munson, 1977)

Karter and Donner also found three significant housing factors associated with greater fire risk. The first was crowdedness, defined as the percentage of year-round housing units with at least 1.01 persons per room. The second was home ownership, defined as the percentage of year-round housing units that are owner-occupied. And the third was housing vacancy, defined as the percentage of year-round vacant housing units. A series of studies in the UK (Chandler, 1979; Chandler, Chapman, & Hallington, 1984) found significant correlations in fire incidence with the age of housing stock, homeownership, socioeconomic status, unemployment rates, and proportion of children in the population. Later studies (Fahy & Miller, 1989; Gunther, 1984) lent further support to the relationship between poverty and fire rates. Donnell's regression analysis of Syracuse, New York census tracts (Donnell, 1980) found the strongest correlations between structural fire incidence and abandoned structures, percent of families below the poverty line, percent black population, and median family income (negative).

Gilliam (1985) published what is likely the first study to analyze fire risk at multiple geographic levels. In his study of fire incidence in Highland Park, Michigan Gilliam looked at three levels of analysis: the census tract, census block group, and individual housing unit. Gilliam found different variables significant at each level of analysis. Among these were income, crowding, education, house value, presence of children, and age of residents (Gilliam, 1985; Jennings, 1999).

Jennings (1996) was the first to propose a conceptual model for fire incidence that made the analytic distinction between fire initiation and fire loss (Jennings, 2013). Jennings developed a four-variable model to explain variation in fire incidence (expressed as fires/capita) across census tracts in Memphis, TN over four years. These four variables were selected based on a review of the literature and consistent with the conceptual model. The variables used were 1) percent vacant housing units; 2) percent of the population under 16 or over 65; 3) household income, and 4) percent of households with children headed by a single parent. The model utilized a weighted least squares regression and accounted for 83 percent of the variation in fire rates between tracts (Jennings, 2013).

Thus, common risk factors that emerge from the literature include various indices of poverty, family structure, and vacant housing percentages. The inconsistency of other factors such as overcrowding, housing stock age, and race suggests variations in factors in different cities or perhaps multicollinearity in some of the early models. Nevertheless, the research established irrefutably that residential fire risk was not distributed evenly across the built environment and neighborhoods, and demographic groups. More recent studies in the last 15 years have shifted to GIS and spatial statistics as a primary means of analysis and have revealed new spatial and temporal patterns in the incidence of fire.

Current Study

The current study examines the distribution of fire incidents in Boston from 2010 through 2014 using fire incident data from the National Fire Incident Reporting System (NFIRS). In a process outlined in the next section, the fire incidents were geocoded to match addresses within the City of Boston's Master Address List and then linked to U.S. Census data and property data from the City of Boston's Tax Assessor's Database. This linkage of fire incidents with property and census data allows us to analyze the distribution of fire across all residential buildings in Boston as well as across local geographies such as census tracts and street segments.

As indicated, one of the goals of the study is to find consistent building-level factors that are associated with a greater probability of a fire. Drawing upon data from the City of Boston Tax Assessor's database, we analyzed all possible factors available for residential properties to find those with the greatest significant effect on fire probability and build a multiple logistic regression model. Drawing on previous work that utilized records from Boston's 311 database (O'Brien & Sampson, 2015), we include measurements of private neglect and emergency medical calls at the address level for 2012-2014 to determine if citizen-reported data can help assess building-level fire risk.

Under the first goal of extending past research, we aggregate the number of fires by census tract and then linear predictors of census tract fire rates. Next, to partition the variance in fire probabilities across differing levels of geographic specificity and to evaluate multilevel modeling for the assessment of urban fire risk, we apply multilevel mixed-effects logistic models that combine tract-level and building-level predictors. Finally, we analyze the predictive value of the multilevel model by cross-validating it with our sample data.

Methods and Data

Data Sources and Measures

The current study utilizes records from six data sets:

- City of Boston's 2015 Tax Assessor's Database
- City of Boston's Master Address List
- National Fire Incident Reporting System (NFIRS) records 2010-2014
- U.S. Census Bureau American Community Survey 5-Year Estimates 2010-2014
- City of Boston's archive of 311 service requests 2011-2013
- City of Boston's archive of 911 dispatches 2011-2013

The Tax Assessor's database contains 165,661 records for every uniquely identifiable property parcel in Boston. The data include variables related to parcel ownership, composition, valuation information, and location—including latitude and longitude coordinates, and 2010 Census block, group, and tract identifiers. Excluding individually owned units in a building, these 165,661parcels are distributed across 96,132 unique street-level addresses. Of these addresses, 88,905 represent residential or mixed-use commercial/residential buildings.

The Master Address List is based on the Tax Assessor's records and contains 118,918 records for every unique address in the city of Boston. Important data fields in the Master Address List include a unique location ID, address components, street segment ID, census tract and block group IDs, number of parcels, and land usage. Of these 118,918 records, 96,621 represent residential or mixed/use commercial/residential buildings. Thus, there are 7,716 more unique residential addresses in the Master Address List (96,621) that are in the Tax Assessor's Database (88,905).

The NFIRS database contains detailed annual records of fire and other emergency incidents as reported voluntarily by 23,000 fire departments across 50 states and the District of

Columbia (U.S. Fire Administration, 2015). Incidents are defined by NFIRS as the actual situation that emergency personnel found when they arrived and may include situations as diverse as a false alarm, rubbish fire, or building fire (U.S. Fire Administration, 2015). Incidents reported from Boston between 2010 and 2014 were filtered to include only residential building structure fires and to exclude fires outside of a main building or fuel burner, boiler, or incinerator fires. (Table 1).

Incident Type Code	Description	Included in Study
111	**Building fire. Excludes confined fires (113–118).**	**Yes**
112	Fire in structure, other than in a building. Included are fires on or in piers, quays, or pilings: tunnels or underground connecting structures; bridges, trestles, or overhead elevated structures; transformers, power or utility vaults or equipment; fences; and tents.	No
113	Cooking fire involving the contents of a cooking vessel without fire extension beyond the vessel.	No
114	Chimney or flue fire originating in and confined to a chimney or flue. Excludes fires that extend beyond the chimney (111 or 112).	No
115	Incinerator overload or malfunction, but flames cause no damage outside the incinerator	No
116	Fuel burner/boiler, delayed ignition or malfunction, where flames cause no damage outside the firebox.	No
117	Commercial compactor fire, confined to contents of the compactor. Excluded are home trash compactors.	No
118	Trash or rubbish fire in a structure, with no flame damage to the structure or its contents.	No
100	**Other Structure Fire, includes fires out on arrival**	**Yes**

Table 1: *NFIRS Structure Fire Incident Codes*

There were 1,726 such residential building fires in the city of Boston between 2010 and 2014 in the NFIRS database. The fires were spread evenly across the five years, with an average of 343 fires, or twenty percent of the total, in each year. For those fires with more detailed descriptive data, the ignition causes and factors indicate that these residential fires can involve building-related and/or human-behavioral conditions (Table 2).

Cause / Factor	Building/Human Typology	Number of Fires	Percent
Equipment / Heat	Building	596	34.5%
Smoking	Human	127	7.4 %
Cooking	Human	104	6.0 %
Unsupervised Children	Human	79	4.6 %
No Smoke Detector	Human	75	4.3 %
Alcohol or Drugs	Human	37	2.1 %

Table 2: *Fires 2010-2014 by Selected Causes and Factors*

Less than half of these fires could be called serious in terms of injury or property damage (Tables 2-3). Only about forty percent caused more than $5,000 in damage. Only two percent caused any injury. Less than one percent resulted in a fatality. Nevertheless, the focus of this study is on the risk of fire, or, in terms of Jennings's conceptual model, on fire ignition and not fire loss. All fires that necessitate a response from a fire department can be considered potentially serious, and property damage, injury, and fatality are outcomes influenced by numerous post-ignition factors beyond the scope of the study. For these reasons, no threshold of fire severity applied to the data set other than the omission of self-contained fires as noted in Table 1.

Property Loss	Count	Percent
$0	109	6.3 %
$0 – $1,000	485	28.1%
$1,000 - $5,000	434	25.1%
$5000 - $20,000	439	14.8%
$20,000 - $1M	432	25.0%
$1M – 5M	6	0.3%

Table 3: *Fires 2010-2014 by Property Loss*

Injuries	Count	Percent
0	1691	98 %
1	20	1.2 %
2 – 15	15	0.8 %
Fatalities	**Count**	**Percent**
0	1718	99.5 %
1 -2	8	0.5 %

Table 4: *Fires 2010-2014 by Injuries and Fatalities*

These fire incident records were extracted from the NFIRS database and geocoded using the Google geocoding API to standardize address formats. The resulting geocoded address was algorithmically matched to properties and addresses from the Master Address List and Tax Assessor's Database.[3] Adjustments were made to this initial data set of residential buildings. Many records, particularly apartment and condo buildings, contained either missing or outlying data. To provide the key predictor data for as many buildings as possible, living area, building age, and assessed value variables were imputed based on neighborhood mean values for missing records in the data set. In addition, the analysis showed that tax-exempt buildings (code E) encompass too wide a variety of building types—including hospitals, college dormitories, and public housing—to be easily incorporated into the study. Mixed Use buildings (code RC) were

[3] Five fires were later removed from the final sample because they were matched to building types not in the study or to buildings for which tax records were incomplete or missing.

similarly problematic in terms of defining the residential and non-residential portions of each building. These land use designations were dropped from the study. The land use designations selected for the study included single-family, two, three, and four six-family homes, apartment buildings, condo buildings, and Tax Exempt 121A buildings (Table 5). Restricting our data to these land use types resulted in a final sample of 89,257 records.

Code	Description	Range of Units	Count	Percent
R1	1-Unit House	1	30,976	34.7 %
R2	2-Unit House	2	21,723	24.3 %
R3	3-Unit House	3	16,431	18.4 %
R4	4-6 Unit Building	4-6	3,603	4.1%
CM	Condo Building	3 - 373	10,818	12.1 %
A	Apartment Building	7 - 279	4,863	5.4 %
EA	Tax-Exempt 121A Section 10	1 - 670	843	0.9 %
TOTAL	-		89,257	100%

Table 5: *Building Types Included in Study*

Drawing on recent work (O'Brien & Sampson, 2015) we also utilized adjusted counts of private neglect and medical emergencies per address and per street segment for each record in our dataset. The Private Neglect count comprises reports to the city's Constituent Relationship Management (CRM) or 311 system related to housing issues (e.g., pests), the uncivil use of private space (e.g., illegal rooming houses or parking on lawns), and complaints about big buildings (i.e., condos) (O'Brien & Sampson, 2015). The Medical Emergency counts comprise 911 calls for medical emergencies at specific addresses. Data are available for 2011 to 2013, and counts were tabulated and adjusted based on land use type for each of these three years. [4] These adjusted scores were centered and scaled at the mean for each year and were then averaged for the entire three-year period.

[4] Adjusted counts control for street-level and census tract level infrastructure and environmental variances. For more information, see D. O'Brien & Winship, 2016.

Unit of Analysis

The current study utilizes two units of analysis: the building level and the U.S. Census Tract level. The building level was chosen in favor of the individual unit or parcel level because there is little significant data available for fire prediction at the unit level and because the unit-level location of a fire within a building is inherently less significant than the building location for the public policy and municipal planning purposes. Moreover, a fire in a building affects other units in the same building. Thus, parcels were condensed to 89,257 unique building addresses from records in the Tax Assessor's Database. Addresses are linked to street segments based on census TIGER Line data (n = 13,767 segments with addresses) and are defined as the undivided length of the street between two intersections or an intersection and a dead end. Finally, census tracts (n = 178 tracts from the 2010 census) are used as a proxy for the neighborhood (O'Brien & Winship, 2016).[5]

Analysis

We analyze fire risk for the aggregated five-year period as opposed to each individual year. With only about 1,700 fires for nearly 90,000 residential properties, fire is too rare an event to model reliably for individual years. In all cases, it is the occurrence or ignition of a fire that is measured as opposed to any outcome or damage estimates from the fires. The approach entailed four phases. First, we fit single and multiple linear regression models to identify the strongest and most significant predictors of higher or lower fire rates at the census tract level. In Phase 2, we fit single and multiple logistic regression models to identify the building level factors that have the greatest effect on building fire probabilities.

[5] See O'Brien and Winship (2016) for a complete description of how addresses, street segments, and tracts were linked.

In Phase 3, we fit generalized mixed models, using the same building-level variables from our logistic models in Phase 2 plus the addition of street and tract-level factors. Whereas the logistic models assume the same intercept for fire risk across the city, the multi-level models allow for varying intercepts for each census tract to account for variability in fire probabilities due to environmental factors. We assess this variance and how it changes with the addition of various predictors, drawing inferences on the interplay between tract-level and building-level factors. Finally, in Phase 4, we assess the predictive value of the multilevel model using a 10-fold cross-validation algorithm that divided the sample into test and training data.

Results

Census Tract Level Predictors and Linear Models

The data indicate that it is race more than poverty that is associated most with higher neighborhood fire rates. Of all individual tract-level variables, the percentage white or percentage black population has the highest r-squared values from single regressions. The data from 2011 to 2014 show that nearly twenty percent of tract variation in fire rates is explained by its white or black population. By contrast, although poverty rates and median household income are moderately correlated with a black population (ρ 0.47 and -0.55), neither of these variables appear significant as predictors. The data also indicate that predominately black neighborhoods have higher fire rates than predominately Hispanic neighborhoods with all other factors being equal. Hispanic and foreign-born percentages—both strongly correlated (ρ=0.64) to each other—accounted for only four and seven percent of the variation, respectively. Moreover, when controlling for median household income in multiple regression, the effect of the black population on tract fire rates is nearly unchanged.

Most other factors correlated with fire rates are all at least moderately correlated with the percentage black population. The percentage of female-headed households, useful as a proxy for single-parent families, is shown to correlate slightly with fire rates, albeit to a degree less than indicated by previous studies (Schaenman, Hall, Schainblatt, Swartz, & Karter, 1977). It should be noted that the percentage of female-headed households correlates very strongly with the percentage black population (ρ 0.78) in Boston. Similarly, the unemployment rate shows a slight correlation with increased fire rates and is itself highly correlated with the percentage black population (ρ 0.77). Other lesser correlates of the black population are measures indicating high percentages of children (percentage family households and population under eighteen) and percentage of families on public assistance are shown to correlate slightly with higher fire rates.

Supporting previous findings, median building age (Chandler, Chapman, & Hallington, 1984; Munson & Oates, 1983) and vacant unit percentage (Karter Jr. & Donner, 1977), neither of which correlate strongly with race or with any other socioeconomic variable, appear to have a significant positive correlation on neighborhood fire rates. Median building age is shown to account for nearly as much variation in fire rates as the white or black population, but the distribution of building age is highly left-skewed. Sixty-three percent of census tracts in Boston have a median building age of fifty-nine years. If we correct this skewness and examine the correlation between median building age and fire rates for the other 37% of tracts, the correlation appears even stronger ($\rho = 0.54$, $r^2 = 0.29$). The percentage of renters and the population density, however, had no effect on fire rates.

NAME	R	R^2	P
Percentage Black	0.48	0.23	0.0000 ***
Percentage White	0.46	0.21	0.0000 ***
Median Building Age [6]	0.35	0.12	0.0000 ***
Percentage Female Headed Households	0.32	0.10	0.0000 ***
Percentage Under 18	0.31	0.10	0.0000 ***
Percentage of Households that are Families	0.31	0.10	0.0000 ***
Percentage Households on Public Assistance	0.29	0.09	0.0001 ***
Unemployment Rate	0.28	0.08	0.0002 ***
Percentage Foreign Born	0.26	0.07	0.0005 ***
Vacant Unit Percentage	0.24	0.06	0.0015 **
Median Household Income	0.21	0.04	0.0071 **
Percentage Hispanic	0.18	0.03	0.0179 *
Percentage of Families Under Poverty Line	0.17	0.03	0.0243 *
Percentage Renters	0.04	0.00	0.5692
Population Density	0.03	0.00	0.7044
Percentage Over 65	0.03	0.00	0.6265

In broad terms, the Census Data suggest that the prototypical highest fire risk neighborhoods in Boston would be poorer than average, have more black residents or more foreign-born residents (or both), more children, more single-parent households, older buildings, and higher rates of vacancy than average. These findings are entirely consistent with previous research with the possible addition of foreign-born residents as a significant predictor. It is important to note, however, that the black population, poverty, children under eighteen, and single-parent households are all intercorrelated. The collinearity suggests that it may be more useful to define socioeconomic factors in fire risk in terms of a nexus of such variables that define higher-risk populations as opposed to concentrating on single variables.

Between 2010 and 2014, the mean fire rate for Boston census tracts was 2.08 fires per thousand persons, with a standard deviation of 1.3 and a range from 0 to 8.5. Looking at the

[6] Median Building Age is not normally distributed.

twelve census tracts with the highest fire rates from 2011 to 2014 (Table X), we can see that these tracts have higher-than-average values for most of the aforementioned significant variables. There are some exceptions. Some tracts have lower-than-average vacancy rates. One tract in East Boston has a lower-than-average black population but a much higher-than-average population of foreign-born residents. Other tracts in Mattapan and Roxbury have lower-than-average percentages of foreign-born residents but higher percentages of black residents. Only one tract in Roxbury has lower than average percentages of families and children under eighteen. However, it has a higher-than-average vacancy rate.

Tract Number	**Neighborhood**	**Fires per 1000** [7]	**Median Build. Age Z**	**Vacant Unit Z**	**Black Z**	**For. Born Z**	**Female Headed Z**	**Family Z**	**Age Under 18 Z**
918	South Dorchester	6.82	0.58	0.42	1.21	1.03	1.10	1.20	1.02
1003	Mattapan	6.45	0.58	0.93	2.25	-0.03	1.34	1.38	1.11
922	South Dorchester	6.34	0.58	2.19	0.79	1.22	0.54	0.98	0.14
916	South Dorchester	5.37	0.58	0.67	0.97	0.71	0.16	1.05	0.78
903	Roxbury	4.93	0.58	-0.02	1.63	-0.12	2.22	1.46	1.73
917	South Dorchester	4.74	0.58	-0.27*	1.14	1.31	0.72	1.32	1.73
506	East Boston	4.61	0.58	-0.23*	-0.83*	3.37	0.24	0.32	0.63
1002	Mattapan	4.40	0.58	1.39	2.33	-0.40*	1.16	1.46	1.31
906	Roxbury	4.29	0.58	1.32	1.55	1.22	1.11	1.20	0.99
911	North Dorchester	4.15	0.58	0.44	-0.67*	0.63	-0.11	0.8	1.02
819	Roxbury	4.14	0.28	0.40	1.95	-0.79*	1.04	0.61	0.67
814	Roxbury	4.11	0.58	1.37	0.96	-0.83*	0.15	-0.18*	-0.27*

Table 5: *Z-Scores of Key Variables for Census Tracts with Highest Fire Rates 2010-2014. * denotes anomalies where variables are significantly below the mean.*

In developing a multiple regression model for fire rates with available census data, one must contend with the aforementioned multicollinearity of several of the socioeconomic variables. A stepwise algorithm finds a best-fit model with four predictor variables: black

[7] Mean Fires Per 1000 for all tracts is 2.1 with a standard deviation of 1.3.

population, white population, vacant unit percentage, and median year built. This model explains about 40% of the variance in fire rates across census tracts. A similar model with the addition of the percentage of foreign-born performs virtually the same (Table 6). Regardless, it does not appear that there is a model with common census tract variables that can explain more than half of the variance in fire rates. Lacking more specific survey data, it is difficult to model fire rates with any greater accuracy. In addition to accuracy concerns, these linear models may lack enough geographic specificity to be of substantial use for cities. It might be possible to model fire risk within smaller areas, but the rareness of residential fires may pose problems here.

Variable	Estimate	P	Adjusted R^2
Percentage Black	2.20	0.0409 *	
Percentage White	-0.68	0.5366	
Percentage Foreign Born	1.90	0.2065	
Vacant Unit Percentage	5.08	0.0305	
Median Building Age	0.04	0.000 ***	
MODEL	-	0.000 ***	0.4042

Table 6: *Multiple Regression Model for Census Tract Fire Rates*

Building Level Predictors and Logistic Models

For individual buildings, three characteristics—owner occupancy, building type, and a number of units—have the greatest effect on fire probabilities in Boston. Each unit increase in the size of a building increases the probability of fire by roughly 2%. Compared to owner-occupied buildings, non-owner-occupied buildings have an over 100% greater probability of a fire. Building type is closely related to each of these variables, but still has a significant effect on fire probabilities even when controlling for the other two factors. Taken in isolation, the raw probabilities of each type of building (Table 6) reflect the effect of renter occupancy and the number of units. Apartments have the greatest average number of units, nearly the lowest percentage of owner-occupancy, and the highest probability of fire. Conversely, single-family homes have the lowest probability and nearly the highest percentage of owner-occupancy.

Building Type	% Fully Renter Occupied	Median Units	Properties having at least 1 fires	Probability of Fire
Apartment	97 %	12	284	0.06
Tax-Exempt 121A	100 %	1	46	0.06
4-6 Unit House	81 %	4	113	0.03
3-Unit House	57 %	3	427	0.03
Condominium	13 %	4	249	0.02
2-Unit House	40 %	2	300	0.01
1-Unit House	19 %	1	212	0.01

Table 7: *Fire Probabilities for Building Types 2010 - 2014*

If we look at individual logistic predictors (Table 8), apartments and 121A buildings (including subsidized housing) have the highest odds ratio of fires, followed by renter-occupancy. Other multi-unit buildings types and condo buildings increase odds ratios significantly. The number of floors and number of units, each co-related, have a similar positive effect on fire probability. Another parameter, years since remodeling, was computed based on a combination of the year remodeled and year built variable from the tax data. For buildings that have not been remodeled, the year remodeled was set to the year built. This value was then subtracted from 2014 to get the final value. Thus, years since remodeled may capture more fully the state of the building and would capture any fire mitigating effect of post-remodeling inspections. Nevertheless, this parameter appears to have a negligible effect on fire probability.

A promising finding for the future of fire risk assessment is the significance of assessed value and medical and private neglect scores. Assessed value per square foot was computed for all buildings and then scaled to z-scores based on building type, thus controlling for variation in assessed values for different building types but not for the significant variations in assessed values by neighborhood.[8] Another metric was computed by scaling to z-scores for neighborhood and building type, thus controlling for both. In the first case, we see a substantial decrease in fire probability (odds ratio of 0.77) with each standard deviation increase in assessed value relative to the mean per building type. In the second case, we see how greater or lesser assessed value for a building relative to its neighborhood increases or decreases its probability by a still significant but more modest amount (odds ratio 0.91). These findings suggest an approximate method—compromised, to be sure, by inconsistencies and inaccuracies in building assessments—to model the socioeconomic status of household residents in estimating building fire risk.

Variable	**Unit**	**Intercept Value**	**Estimate (Odds Ratio)**	**P**
Apartment	binomial 0 or 1	0	3.82	0.000 ***
Tax-Exempt 121A Building	binomial 0 or 1	0	3.16	0.000 ***
Fully Renter Occupied	binomial 0 or 1	0	1.96	0.000 ***
4-6 Unit House	binomial 0 or 1	0	1.79	0.000 ***
3-Unit House	binomial 0 or 1	0	1.59	0.000 ***
Condo Building	binomial 0 or 1	0	1.31	0.000 ***
Medical Residuals 2011-2013	z-score	Mean	1.18	0.000 ***
Private Neglect Residuals 2011-2013	z-score	Mean	1.09	0.000 ***
Number of Floors	1 floor	1	1.02	0.000 ***
Number of Units	1 unit	1	1.02	0.000 ***
Years Since Remodeled	1 year	0 (remodeled in 2014)	1.00	0.000 ***
Building Age	10 years	0 (newly built in 2014)	0.99	0.204
Assessed Value Per Square Foot 1	z-score by building type	Mean value per building type	0.77	0.000 ***

[8] Assessed value per square foot (per building type) correlates with percentage white population with a coefficient of 0.49 and median household income with a coefficient of 0.37.

Assessed Value Per Square Foot 2	z-score by building type and neighborhood	Mean value per building type and neighborhood	0.91	0.000 ***
2-Unit House	binomial 0 or 1	0	0.70	0.000 ***
1-Unit House	binomial 0 or 1	0	0.28	0.000 ***

Table 8: *Logistic Coefficients for Building Level Variables*

In a similar vein, the private neglect scores and medical scores attempt to measure micro-level conditions normally inaccessible to researchers without extensive surveying. The private neglect scores act as a general measure of building conditions and quality of maintenance. The medical scores serve as a generalized measure of the health and stability of residents. Each standard deviation above the mean in medical reports raises the probability of fire by about 18%, and each standard deviation increase in private neglect scores increases the probability by about 9%. It should be noted, however, that the study design does not enforce a temporal sequence on the medical and neglect reports. These medical and private neglect scores are averaged from 2011 to 2013 yearly scores and thus can encompass reports preceding, concurrent with, or following the actual fire incident.

Notable is the lack of any significant effect of building age, challenging both the results of our linear models as well as previous research (Chandler, Chapman, & Hallington, 1984; M. J. Munson & Oates, 1977). If anything, the data suggest that fire probabilities may decrease slightly with greater building age in Boston. Buildings built in the last fifty years have a higher probability of fire (0.023) than buildings built before 1964 (0.018), although the slight difference and the p-value of the logistic regression coefficient suggest that this could be due to pure chance. Collinearity can largely be ruled out as a factor in the linear correlation of fire rates and median building age as building age is not correlated with any other known census tract variables. And, as noted previously, if we correct for the non-normal distribution of median building age in census tracts by dropping the cluster of census tracts that have the oldest housing stock, we still find a significant correlation between fire rate and building age.

In fact, the discrepancy can be explained by the fact that there is enough variation in building ages within Boston census tracts to smooth out the difference. The housing stock is old in Boston compared to other American cities, but there were enough fires in relatively newer buildings within older neighborhoods such that the mean building age for buildings that had a fire was slightly lower (95 years) compared to the citywide mean (96 years). As an example, we can see that in a nearly normal distribution in building ages, there is a large cluster of newer buildings in census tracts with fires rates above the 90th percentile (Figure 1), suggesting a sufficient pool to distribute the occurrence of fires across the full range of building ages within these tracts. Thus, despite fire occurring in higher rates in older neighborhoods, individual building age does not appear to be a factor in raising the probability of fire in an individual building.

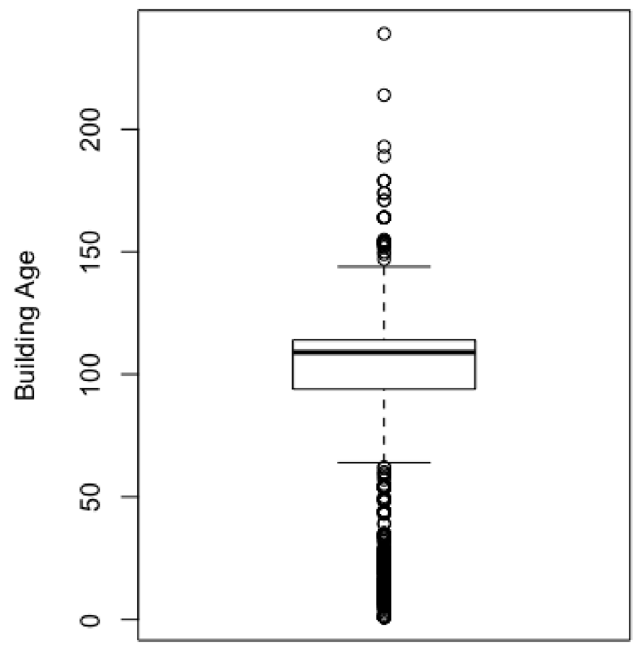

Figure 1: *Building Ages for Census Tracts with the Highest 10% of Fire Rates*

If we include all these variables in a multiple logistic regression model (Table 9), we can then control for the fixed effects of the number of units and owner occupancy and see more realistic coefficients for individual building types. Included here is the first metric for assessed building value scaled to building type (not to the neighborhood). Although it adds nothing to the log-likelihood of the model, we can include building age for reference purposes, as it also does not degrade the model significantly. The effect of the years since remodeled parameter appears to have no effect, but adding it improves the model's log-likelihood slightly. The number of floors and number of units is so closely correlated that one diminishes the effect of the other. The number of floors should be dropped from the multilevel models because its effect is lesser.

Variable	**Estimate (Log Odds)**	**Odds Ratio**	**Std. Error**	**z value**	**Pr(>|z|)**
Apartment	1.85	6.37	0.11	17.31	0.0000 ***

Tax-Exempt 121A Building	1.60	4.95	0.19	8.37	0.0000 ***
4-6 Unit House	1.38	3.99	0.13	11.03	0.0000 ***
3-Unit House	1.28	3.58	0.09	14.16	0.0000 ***
Condo Building	0.92	2.50	0.10	8.90	0.0000 ***
2-Unit House	0.68	1.98	0.09	7.40	0.0000 ***
Fully Renter Occupied	0.16	1.17	0.06	2.51	0.0122 *
Medical Residuals 2011-2013	0.15	1.16	0.02	8.84	0.0000 ***
Private Neglect Residuals 2011-2013	0.07	1.08	0.02	4.71	0.0000 ***
Building Age (in decades)	0.02	1.02	0.01	1.50	0.1332
Number of Units	0.01	1.01	0.00	10.76	0.0000 ***
Number of Floors	0.00	1.00	0.00	2.60	0.0092 **
Years Since Remodeled	0.00	1.00	0.00	-4.53	0.0000 ***
Assessed Value Per Square Foot 1	-0.27	0.76	0.04	-7.55	0.0000 ***
(Intercept)	-4.91	0.01	0.11	-42.82	0.0000 ***

Table 9: *Multiple Logistic Regression Model for Building Level Fire Risk*

In this multiple regression model, the intercept can be interpreted as representing a single-floor, owner occupied, single-family home with an average assessed value and average medical and private neglect residuals. As with our tract-level analysis in Phase 1, the data suggest a nexus of factors comprising building-level fire risk. In broad terms, low-value apartment buildings with signs of private neglect and a disproportionate number of 911-calls will have the highest risk of a fire. Such findings do little to challenge common sense, but the ability to model and quantify these factors is the crucial point. Putting our findings from Phase 1 and Phase 2 together, the data indicate that such apartments may have an even greater risk of fire if located in neighborhoods with higher black or foreign-born populations, greater numbers of children, more single-parent households, and higher rates of vacancy. Some of these variables mixed together, however, have confounding and distorting effects when modeled. That is one reason why we turn to multilevel models.

Multilevel Models

It is established that fire rates vary by census tract in Boston. The variation is not dramatic, but it is large enough to be modeled and significant enough to be of statistical and practical interest. In terms of probabilities, if we fit an intercept-only 2-level model with buildings at level one and census tracts at level two, odds ratios for the census tract intercepts range from 0.28 to 5.92, with a standard deviation of about 2.0. If we fit the same type of model but with street segments substituted for census tracts, we see a larger variance in odds ratios, with odds ratios ranging from 0.5 to 58.4 and with a standard deviation of 4.5.

With multilevel mixed effects models, we can account for tract-level and street-level variance and isolate the so-called fixed building effects that we modeled in the previous section. We can also determine how much of the group-level variance is explained by these building-level predictors and limit the so-called random effects at these geographic levels. Of all the building level factors, the building type accounts for the largest percent of census tract variance—more than half. The number of units and owner occupancy also exert a sizable reduction in the tract variance, followed by assessed value and the 311 and 911 residual scores. Building age is shown to slightly increase the census tract variance.

Variable	Tract Level Std. Deviation	Tract Variance	Decrease in Variance by Percent
Building Type	0.462	0.213	56.0 %
Number of Units (z-score)	0.615	0.378	22.0 %
Fully Renter Occupied	0.640	0.410	15.4 %
Assessed Value Per Square Foot 1	0.671	0.450	7.0 %
Medical Residuals 2011-2013	0.688	0.474	2.2 %
Private Neglect Residuals 2011-2013	0.690	0.476	1.8 %
Years Since Remodeled	0.695	0.483	0.2 %
Baseline Census Tract Variance	0.696	0.484	0.0 %
Building Age (in decades)	0.697	0.485	-0.2 %

__Table 10:__ Percentage of Census Tract Variance Diminished by the Addition of Individual Building Factors to a Multilevel Model

Consistent with our findings, accounting for correlation and covariance among our variables, a two-level model with all the factors above (except building age) as building-level effects and a random census tract intercept accounts for 71% of the variance in fire probabilities across census tracts. We can take things further and fit a fuller model that adds additional predictors at the tract level, incorporating three of the key tract level predictors from Phase 1: percentage white population (stronger effect here than black population), percentage foreign-born, and percent of vacant housing units. This model accounts for 83% of the tract-level variance (Table 11). Because of the correlation between the percentage of white persons and the assessed building value, the second assessed value per square foot variable was selected to localize its effects at the building level. This model also showed a better log-likelihood than a model which used the other assessed value metric which is scaled only by building type.

Variable	Estimate (Log Odds)	Odds Ratio	Std. Error	z value	Pr(>|z|)
Apartment	1.72	5.56	0.11	15.37	0.0000 ***
Tax-Exempt 121A Building	1.44	4.22	0.19	7.62	0.0000 ***
4-6 Unit House	1.38	3.98	0.13	10.89	0.0000 ***
3-Unit House	1.08	2.94	0.09	11.53	0.0000 ***
Condo Building	1.01	2.75	0.11	9.30	0.0000 ***
2-Unit House	0.60	1.82	0.09	6.42	0.0000 ***
Medical Residuals 2011-2013	0.15	1.16	0.02	8.75	0.0000 ***
Number of Units (z-score)	0.14	1.15	0.01	11.91	0.0000 ***
Fully Renter Occupied	0.13	1.13	0.06	2.00	0.0455 *
Percentage Foreign Born	0.10	1.10	0.04	2.29	0.0219 *
Private Neglect Residuals 2011-2013	0.07	1.07	0.02	4.48	0.0000 ***

Vacant Unit Percentage	0.05	1.06	0.04	1.52	0.1280
Years Since Remodeled	0.00	1.00	0.00	-4.15	0.0000 ***
Assessed Value Per Square Foot 2 (scaled to neighborhood and building type)	-0.07	0.94	0.02	-2.83	0.0047 **
Percentage White	-0.23	0.79	0.05	-4.99	0.0000 ***
(Intercept)	-4.81	0.01	0.09	-51.16	0.0000 ***

Table 11: *Full Multilevel Model – Fixed Effect Estimates*

Predictive Value of the Models

The overall probability of fire for any given building within the five-year period of 2010 to 2014 was 0.018. Cross-validation results suggest that it would be possible to define a small high-risk pool of buildings in Boston and accurately predict fires for 25% of these buildings within a five-year time span (Table 12). By comparison, as expected based on the overall probability, a random selection of a similarly sized small pool of buildings accurately predicts only about 2% of buildings. Although these models have not been applied to actual future events and there are temporal complications based on the sequencing of private neglect and medical reports, these results validate the overall multilevel approach. It may turn out that private neglect and medical reports are still effective as predictor variables for fire events five years hence. Should more data be available for buildings—particularly housing inspection data—the model could be strengthened and the precision could be significantly improved.

Iteration	Test Sample	Predicted Positives	True Positives	False Positives	True Negatives	False Negatives	AUC	Precision
1	8,926	31	6	25	8729	166	0.70	0.19
2	8,926	18	4	14	8757	151	0.69	0.22
3	8,925	25	8	17	8737	163	0.71	0.32
4	8,926	30	10	20	8744	152	0.70	0.33
5	8,925	22	5	17	8757	146	0.72	0.23
6	8,925	18	4	14	8739	168	0.70	0.22
7	8,926	26	4	22	8739	161	0.70	0.15
8	8,926	16	5	11	8765	145	0.71	0.31
TOTALS	**71,405**	**186**	**46**	**140**	**69967**	**1252**	**0.71**	**0.25**

Table 12: Cross Validation Results

References

Chandler, S., Chapman, A., & Hallington, S. (1984). Fire incidence, housing, and social conditions—the Urban Situation in Britain. *Fire Prevention*, *19*(172).

Donnell, R. P. (1980). *Fire in the City: Spatial Perspectives on Urban Structural Fire Problems; Syracuse, New York*. Geography - Dissertations.

Fahy, R., & Miller, A. (1989). How Being Poor Affects Fire Risk. *Fire Journal*, *83*, 28–30.

Gilliam, G. A. (1985). *An analysis of residential fire patterns in Highland Park, Michigan, 1970 and 1977 using individual unit, block, and census track data*.

Gunther, P. (1984). Fire-cause patterns for different socioeconomic neighborhoods in Toledo. *Fire Journal*, *75*, 52–58.

Jennings, C. R. (1996). *Urban Residential Fires: An Empirical Analysis of Building Stock and Socioeconomic Characteristics for Memphis, Tennessee*. Cornell University.

Jennings, C. R. (1999). Socioeconomic characteristics and their relationship to fire incidence: a review of the literature. *Fire Technology*, *35*(1), 7–34.

Jennings, C. R. (2013). Social and economic characteristics as determinants of residential fire risk in urban neighborhoods: A review of the literature. *Fire Safety Journal*, *62*(PART A), 13–19. http://doi.org/10.1016/j.firesaf.2013.07.002

Karter, M., & Donner, A. (1977). *Fire rates and census characteristics. An analytical approach*. National Fire Protection Association.

Lindeman, R. H., Merenda, P. F., & Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis*. Scott, Foresman.

Munson, M. (1977). *Residential Fires and Urban Neighborhoods: An Empirical Analysis of Charlotte North Carolina*.

Munson, M. J., & Oates, W. E. (1977). *Community characteristics and the incidence of fire: an empirical analysis*. National Fire Prevention and Control Administration.

O'Brien, D. T., & Sampson, R. J. (2015). Public and Private Spheres of Neighborhood Disorder: Assessing Pathways to Violence Using Large-scale Digital Records. *Journal of Research in Crime and Delinquency*, *52*(4), 486–510. http://doi.org/10.1177/0022427815577835

O'Brien, D., & Winship, C. (2016). Journal of Quantitative Criminology What is Gained by Greater Granularity ? The Presence and Persistence of Problem Properties in Urban Neighborhoods Not applicable. *Journal of Quantitative Criminology*.

Ridley, C. E. (1927). *Measuring municipal government*. Municipal administration Service.

Schaenman, P. S., Hall, P., Schainblatt, A., Swartz, A., & Karter, M. (1977). *Procedures for improving the measurement of local fire protection effectiveness*. Urban Institute.

Simon, H. A., Shepard, R. W., & Sharp, F. W. (1943). *Fire losses and fire risks*. Bureau of public administration, University of California.

U.S. Fire Administration. (2016). *Fire Risk in 2014*. *Topical Fire Report Series* (Vol. 17).

USFA National Fire Data Center, F. (1997). Socioeconomic Factors and the Incidence of Fire, (June), 1–35.