Technical Note 1770

U.S. Department of Commerce
National Institute of Standards and Technology

Applied Economics Office
Engineering Laboratory
Gaithersburg, MD 20899

Evaluating Potential Bias in Non-Randomly Reported Fire Incident Data

David T. Butry
Douglas S. Thomas

Technical Note 1770

U.S. Department of Commerce
National Institute of Standards and Technology

Applied Economics Office
Engineering Laboratory
Gaithersburg, MD 20899

Evaluating Potential Bias in Non-Randomly Reported Fire Incident Data

David T. Butry
Douglas S. Thomas

Sponsored by:
National Institute of Standards and Technology
Engineering Laboratory

September 2012

Abstract

This analysis is part of an effort to develop statistics and uncertainty measures for characterizing, tracking, and better understanding the root causes of the total burden of fire in the United States. These measures will be used to develop performance metrics, enabling comparisons between the use of new fire mitigation technologies and their impact on the U.S. fire burden, with a particular focus on residential fires involving upholstered furniture. This portion of the analysis has the following objectives: (1) to develop a statistical approach for evaluating the 'representativeness' of fire incident data reported in the National Fire Incident Reporting System (NFIRS) to depict fire activity in non-reporting cities; (2) to test (statistically) for differences between reporting and non-reporting cities of those factors believed correlated with fire risk and NFIRS reporting status; and (3) to discuss how the findings could be used to weight NFIRS-based statistics to produce more accurate national statistics. Results show that factors believed correlated to fire risk occur at different rates between reporting and non-reporting cities. This suggests that detailed fire statistics derived from NFIRS data may not best represent the U.S. fire problem, as these factors are also correlated with NFIRS reporting status. However, a weighting scheme, based on propensity scores, may provide a mechanism to adjust NFIRS-based fire incident statistics to provide more accurate nationwide metrics.

Keywords: NFIRS; fire risk; statistics; propensity score matching; sample selection bias

Preface

This study was conducted by the Applied Economics Office in the Engineering Laboratory at the National Institute of Standards and Technology. The study provides a synopsis of available data depicting the U.S. fire burden. This analysis is part of an effort to develop statistics and uncertainty measures for characterizing, tracking, and better understanding the root causes of the total burden of fire in the United States. These measures will be used to develop performance metrics, enabling comparisons between the use of new fire mitigation technologies and their impact on the U.S. fire burden, with a particular focus on residential fires involving upholstered furniture.

Cover Photographs Credits

Microsoft Clip Art Gallery Images used in compliance with Microsoft Corporation’s non-commercial use policy.

Acknowledgements

The authors wish to thank all those who contributed so many excellent ideas and suggestions for this report. Special appreciation is extended to Joshua Kneifel, Applied Economics Office (AEO), and William Pitts, Fire Research Division, for their thorough reviews and many insights and to Carmen Pardo for her assistance in preparing the manuscript for review and publication. The authors also wish to thank Robert Chapman, AEO, and Nicos S. Martys, Materials and Structural Systems Division, for their reviews.

TABLE OF CONTENTS

1 Introduction

1.1 Background

This analysis is a part of an effort to develop statistics and uncertainty measures for characterizing, tracking, and better understanding the root causes of the total burden of fire in the United States. The purpose of these statistics will be to provide details on the circumstances, causes, and development of fires and the related deaths, injuries, and property damage by major fire incident category (residential structure, non-residential structure, natural vegetation, vehicle), as well as provide details on the costs related to fire protection and loss mitigation (e.g., fire protection of constructed facilities; standards and codes development, testing, and implementation; wildland fuel treatments, etc.). This information will be used to develop performance metrics, enabling comparisons between the use of new fire mitigation technologies and their impact on the U.S. fire burden, with a particular focus on residential fires involving upholstered furniture.

The U.S. Fire Administrations National Fire Incident Reporting System (NFIRS) provides detailed information on more than one million fire incidents each year (on average from 2002 to 2009). While NFIRS is the most comprehensive accounting of individual fire incidents in the U.S., it represents only a partial census. Many incidents are not reported. Many cities/jurisdictions do not report to NFIRS, as it is a voluntary system. Thus, using specific information contained in NFIRS to generalize to the U.S. may be misleading if the partial census is unrepresentative of the non-reporting collection of cities and states.

1.2 Purpose and Approach

This analysis has the following objectives: 1) to develop a statistical approach for evaluating the 'representativeness' of fire incident data reported in NFIRS to depict fire activity in non-reporting cities; (2) to test (statistically) for differences between reporting and non-reporting cities on those factors believed to be correlated with fire risk and the NFIRS reporting status; and (3) to discuss how the findings could be used to weight NFIRS-based statistics to produce more accurate national statistics.

Propensity score matching (PSM) techniques are presented as an approach to compare samples or subsets of data that are potentially afflicted by self- or sample-selection bias—i.e., when the data is not a random collection of observations, but rather, is influenced by other processes that affect the conclusions. PSM eliminates (or reduces) bias by conditioning comparisons using data that describes the underlying self- or sample-selection process. It is commonly used for program evaluation and impact analysis, in fields such as labor economics and epidemiology. Commonly, PSM is used to compare the effectiveness of a program or treatment on a selected outcome. For instance, it has been used to assess the impact a job training program (*treatment*) had on participants' post-training wages (*outcome*).

At this stage, the analysis will not yet focus on impact measurement, but rather to *develop PSM as a proof-of-concept* to compare characteristics believed correlated with both NFIRS reporting status and fire risk (e.g., population size and socio-economic factors) between NFIRS-reporting and non-reporting cities. If these characteristics are 'similar' between reporting and non-reporting groups, then NFIRS-based data can be used to produce statistics representative of the U.S. fire problem. If this is not the case, PSM offers a method to facilitate appropriate comparisons.

If reporting and non-reporting cities demonstrate significant differences in their composition of factors believed to affect reporting status and fire risk, then PSM may be used to weight incident data derived from NFIRS to produce representative fire risk statistics. The focus of future analysis will be on statistics related to fire fatalities, injuries, and property damage occurring from residential fires involving upholstered furniture.

2 Methods

2.1 Propensity Score Matching

Propensity score matching (PSM) is a technique used to evaluate the impact of programs (see Imbens and Wooldridge [2009] for an excellent review of the program evaluation literature). It is commonly used to compare the effect of a program ('treatment') on an 'outcome.' (Often the focus is to measure the average treatment effect on the treated). For instance, PSM has been used to measure the impact a job training program (the treatment) had on participants' wages (the outcome). A statistical challenge occurs because the analysts are using non-randomized observational data. (Again, in the case of job training, participants' wages might be compared to that earned by non-participants). Thus, the outcomes from program participants may be partly influenced by other, confounding factors, making treatment selection a non-randomized process. Essentially, job training participants may choose to participate based on factors (e.g., education) that also influence their wage potential. In such cases, simple comparisons made between participants and non-participants may be biased, as the presence of confounding factors mask the true impact from program participation.

The novelty of the PSM method resides in its ability to facilitate comparisons between treated and non-treated observations using data that describes the non-randomized treatment selection process. It is particularly useful when there is a large number of possible confounding factors. For instance, if only one confounder were suspected, matching treated and non-treated observations would be fairly straightforward. The treated and non-treated observations could be matched pairwise based on the same value of the confounder. However, when a large number of confounders are present, this becomes very difficult (if at all possible). In empirical applications, the propensity score is calculated as the probability of treatment selection, and estimated as a function of all potential confounders. Thus, the propensity score is a scalar value, making matching straightforward for matching purposes, but it is also useful as it contains all the variation in the confounding variables that influence the selection process.

2.2 Balancing Score

Successful implementation of the PSM method requires the propensity score to be a *balancing* score. That is, a balancing condition must be satisfied for the PSM method to produce unbiased treatment effect estimates. The balancing condition requires the confounders to be independent (uncorrelated) of treatment status (treated/non-treated) conditioned on the propensity score. The standard statistical test for assessing the balancing score requirement is a Student's t-test of the means of each confounder between the matched treated and non-treated samples. When the null hypothesis of the t-test (H0: the difference in the treated and non-treated sample means equal zero) cannot be rejected, for any of the confounders, the propensity score is said to balance the confounders.

For unmatched samples, the balancing test becomes a useful way to evaluate the similarities in the two sample groups. In this analysis, those factors believed to be correlated with cities' NFIRS reporting status and level of fire risk are tested for balance between reporting and non-reporting cities. If a subset of (or all) factors are found to be unbalanced (the null hypotheses are rejected), data taken from reporting cities may not do a very good job of representing the rest of the non-reporting U.S.—i.e., meaning the fire risks faced by non-reporting cities may be different.

In this analysis, a set of factors believed to influence cities' NFIRS reporting status and fire risk is tested for balance. For any unbalanced confounders, an estimated propensity score is used to achieve balance. The potential of the propensity score is to provide a mechanism to create weighted datasets using NFIRS incident data to describe fire risk for the rest of the non-reporting Nation. As the focus of this analysis is to develop PSM as a proof-of-concept for evaluating the representativeness of cities that report to NFIRS to other cities in the U.S., the set of confounders selected are not meant to be exhaustive; however, they are deemed a reasonable set to demonstrate the technique.

2.3 Implementation

The pre-written routines PSMATCH2 and PSTEST (see Leuven and Sianesi, 2003) were used to perform covariate (confounder) balance testing using Stata version 12.1. PSMATCH2 estimates a propensity score and matches the scores. PSTEST performs tests of covariate balance. In this analysis, PSMATCH2 was run to generate the propensity scores that were used for covariate balance testing in PSTEST. The propensity scores were estimated using a logit specification. Kernel matching was used to create matched comparisons (see Leuven and Sianesi [2003] for additional details).

3 Data

3.1 'NFIRS Cities'

Individual fire incidents reported in NFIRS from 2002 to 2009 were geocoded based on their ZIP code or city name. GIS data from ESRI's ArcGIS version 10 Data & Maps were used to match the incident data one of three ways (in preferred order): (1) on reported ZIP code, (2) on reported city name that matched the PO_NAME ('Post Office Name') found in the ESRI ZIP code GIS, or (3) on the city name that matched the NAME found in the cities GIS, which was later spatially joined with the ZIP code boundary file. Thus, each matched fire incident was geocoded to a city (Post Office Name—PO_NAME). All analyses were performed on aggregated city data.

After city name standardization and data cleansing were performed on the raw NFIRS incident records, the geocode match rate was 97.66 %. This is out of 8 333 134 reported fire incidents from 2002 to 2009. Of the 2.34 % (194 804) unmatched incidents, 59.34 % (113 605) did not report a ZIP code or city name. Thus, it was not possible to match these incidents. Arizona (13.39 %), South Dakota (11.37 %), and Florida (5.30 %) had the highest percentage of missing locational data. The remaining unmatched incidents (1.39 % overall) did contain a non-missing ZIP code or city name, but these could not be matched to the data in the GIS data. It appeared for most cases, the ZIP code number was invalid. Maryland (8.60 %), South Dakota (5.16 %), and Virginia (3.33 %) had the highest percentage of invalid or unmatchable locational data.

Table 3-1 summarizes reporting to NFIRS. Based on the matched data, 96.01 % of all cities reported at least one incident. However, only 41.90 % of 24 970 cities reported incidents each year from 2002 and 2009. It is evident that reporting has become more common over time. In 2009, nearly 82 % of all U.S. cities (as defined in this analysis) reported at least a fire incident to NFIRS.

Table 3-1. Count of cities and percent of total (n = 24 970) reporting to NFIRS.

	Count	Percent
Any Year	23 973	96.01
2002	14 593	58.45
2003	16 979	68.00
2004	18 058	72.32
2005	19 638	78.65
2006	20 038	80.25
2007	20 228	81.02
2008	20 225	81.00
2009	20 431	81.83
All Years	10 461	41.90

3.2 Assembled Confounders

Two data sets were utilized to assemble the confounders: the Census of Population and Housing (see U.S. Department of Commerce, 2004) and the National Fire Department Census (see U.S. Fire Administration, 2006). The Census of Population and Housing contains socio-economic variables about population such as income, sex, and race as well as housing items such as the status of a housing unit (occupied or unoccupied), median age of units, and median value. The data is parsed out by census block group. The National Fire Department Census contains basic information about fire departments listed with the U.S. Fire Administration. The variables that were drawn from these datasets were chosen to represent items that both impact fire reporting and the occurrence of fire. Again, as the focus of this analysis is to develop PSM as a proof-of-concept for evaluating the representativeness of cities that report to NFIRS to other cities in the U.S., the set of confounders selected are not meant to be exhaustive; however, they are deemed a reasonable set to demonstrate the technique. Each observation represents an NFIRS city as defined in the previous section.

The variables are grouped into four categories, and are meant to directly measure of proxy for the category titles: (1) fire station resources; (2) fire station responsibility; (3) community resources; and (4) community risk attributes. The fire station resources and responsibility variables were selected based on their expected impact on the level of fire prevention a station is able to engage in as well as their impact on the ability to expend resources on reporting fires in NFIRS. The variables in the community resource grouping and the community risk attributes grouping affect a community's attitude towards fire prevention. It is thought that the community's attitude impacts a fire station's ability and behavior regarding both the reporting of fire and fire prevention efforts. Shown below in Sections 3.2.1 through 3.2.4 are the variable names, the abbreviation used, and a short description of each confounder.

3.2.1 Fire Station Resources

Fire Stations (FD): Total number of fire stations within the city boundary from the National Fire Department Census

Staff (FD_STAFF): The total number of fire department staff, including career and volunteer fire fighters as well as non-firefighting staff (of those fire departments found within the city) listed in the National Fire Department Census

3.2.2 Fire Station Responsibility

Population (POP): The total population from the Census of Population and Housing

Residential Units (UNITS): Total number of residential units, including vacant, owner occupied, and renter occupied units, listed in the Census of Population and Housing

Urban Population (PCT_URBAN): Urban population divided by total population, as listed in the Census of Population and Housing

Fire Station Distance (FD_DIST): Distance (in meters) from an NFIRS city, as defined in Section 3.1, to the nearest fire station listed in the National Fire Department Census

Unit Age (AGE_OWNROCC): The number of years from the median year of construction for owner occupied housing, as listed in the Census of Population and Housing, to 2012

3.2.3 Community Resources

Income (INCOME): Aggregate income for the NFIRS city divided by the total population, both taken from the Census of Population and Housing

Poverty (PCT_POVERTY): Population in poverty divided by total population, both taken from the Census of Population and Housing

3.2.4 Community Risk Attributes

Disabled (PCT_DISABLED): Disabled population divided by total population, both taken from the Census of Population and Housing

Education (PCT_25_HS): Population that is aged 25 or older that has at minimum a high school diploma or equivalent divided by total population, both taken from the Census of Population and Housing

Gender (PCT_MALES): Population that is male divided by total population, both taken from the Census of Population and Housing

Male Youth (PCT_YMALES): Population that is male and 17 years or younger divided by total population, both taken from the Census of Population and Housing

Owner Occupied Units (PCT_OWNROCC): Units that are owner occupied divided by the total number of units, both taken from the Census of Population and Housing

Race (PCT_WHITE): White population divided by total population, both taken from the Census of Population and Housing

Unit Value (VAL_OWNROCC): The median value of units that are owner occupied, taken from the Census of Population and Housing

Vacancy (PCT_VACANT): Residential units that are vacant divided by the total number of units, both taken from the Census of Population and Housing

Veterans (PCT_VETERANS): Population aged 18 to 64 that are veterans divided by total population, both taken from the Census of Population and Housing

4 Results

4.1 Propensity Score—Full Sample

An estimated propensity score was generated by regressing NFIRS reporting status (1 = report; 0 = no report) on a set of covariates believed correlated with reporting status and fire risk. The regression included all 24 957 cities (full sample) for which data was available. Of these, 42 % reported each year from 2002 to 2009 (see Table 4-1).

Table 4-1. Number of cities by NFIRS report status for all years 2002 to 2009.

	Reported	Not Reported	Total
All Cities	10 460	14 497	24 957
Strata 1 (Pop. ≤ 10,000)	7138	12 835	19 973
Strata 2 (10,000 < Pop. ≤ 25,000)	1844	933	2777
Strata 3 (Pop. > 25,000)	729	1478	2207

The regression results are shown in Table 4-2. Nearly all included covariates were found to be statistically correlated (5 % level) with NFIRS report status. The exceptions included: FD_STAFF, INCOME, and PCT_DISABLED. (The label "_cons" denotes a constant [intercept] term.)

Table 4-2. Regression (logit) results from propensity score estimation on the full sample.

NFIRS	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
POP	-.0000467	5.13e-06	-9.12	0.000	-.0000568	-.0000367
FD	.0613524	.0078902	7.78	0.000	.0458878	.076817
FD_DIST	-.0001352	7.17e-06	-18.87	0.000	-.0001492	-.0001212
FD_STAFF	-.0001432	.0002059	-0.70	0.487	-.0005467	.0002603
UNITS	.0001204	.0000135	8.90	0.000	.0000939	.0001469
PCT_VACANT	-1.573478	.2135262	-7.37	0.000	-1.991982	-1.154975
PCT_OWNROCC	.84165	.2012925	4.18	0.000	.4471239	1.236176
VAL_OWNROCC	-.0025435	.0003852	-6.60	0.000	-.0032986	-.0017884
AGE_OWNROCC	-.002079	.0003419	-6.08	0.000	-.0027491	-.0014089
INCOME	.0048746	.0043154	1.13	0.259	-.0035834	.0133326
PCT_POVERTY	-4.536126	.369018	-12.29	0.000	-5.259388	-3.812864
PCT_URBAN	.9871813	.052924	18.65	0.000	.8834521	1.09091
PCT_MALES	-3.109848	.5890458	-5.28	0.000	-4.264357	-1.95534
PCT_YMALES	4.273011	.7145594	5.98	0.000	2.872501	5.673522
PCT_WHITE	.4768351	.1150638	4.14	0.000	.2513142	.7023561
PCT_25_HS	-1.067173	.2428898	-4.39	0.000	-1.543228	-.5911174
PCT_DISABLED	.1239005	.1994842	0.62	0.535	-.2670812	.5148823
PCT_VETERANS	2.969865	.3140909	9.46	0.000	2.354258	3.585471
_cons	.772505	.4651782	1.66	0.097	-.1392274	1.684237

The tests of individual covariate balance are shown in Table 4-3. For each variable, two means tests were run: between the (1) ‘unmatched’ and (2) ‘matched’ data. The unmatched means test is a means comparison between the two reporting groups (report to NFIRS; do not report to NFIRS). The matched means test is a comparison between reporting groups after matching on their propensity scores (matches were created based probability of reporting to NFIRS).

In addition to the Student’s t-test of the means, Table 4-3 shows the percent bias and percent bias reduction achieved from matching. The percent bias is defined as the “[percent] difference of the sample means in the treated and non-treated (full or matched) sub-samples as a percentage of the square root of the average of the sample variances in the treated and non-treated groups” (Leuven and Sianesi, 2003). (In this analysis, treated denotes reporting). The percent bias and percent bias reduction, as measured by PSTEST, are not a focus of this analysis.

Examining the unmatched results from Table 4-3, it can be seen that those variables that are statistically correlated to NFIRS reporting status (those significant variables shown in Table 4-2) also have statistically different means (5 % level) between the two reporting groups. The implication is those variables that affect the probability of NFIRS reporting occur at different rates between the reporting groups. Thus, any comparisons made on fire risk between the two groups could be problematic due to differences in their confounders. Matched comparisons, based on the propensity score, are meant to alleviate this statistical issue, but only if balance can be achieved.

Based on the reported results shown in Table 4-3, the propensity score achieves balance for a number of covariates; however, a number of covariates still fail to balance. These variables include: POP, UNITS, PCT_VACANT, PCT_OWNROCC, VAL_OWNROCC, INCOME, PCT_POVERTY, PCT_URBAN, and PCT_DISABLED. Thus, this particular propensity score is problematic, as it is not a balancing score. (The lack of balance for INCOME and PCT_DISABLED is not an issue, as neither was found significant [5 %

level] in the propensity score model). Commonly, higher-order terms (of the unbalanced covariates) are added to the propensity score model to create balance. While not shown, the use of higher-order terms failed to produce a balancing propensity score over the full sample of cities. However, stratifying the cities into three groups based on population size produced useful results.

Cities were placed into one of three groups based on population size: (1) population less than or equal to 10 000 people; (2) population greater than 10 000 people, but less than or equal to 25 000; and (3) population greater than 25 000. The number of cities by NFIRS reporting status is shown in Table 4-1. The population group thresholds were determined through trial and error, as an issue occurred with higher population cities. Specifically, for cities with very large populations there were relatively fewer non-reporting cities to pose as matches for reporting cities. But as will be shown below, the final grouping produced strata-specific propensity scores that created covariate balance.

Table 4-3. Balancing results on full sample.

Variable	Unmatched Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
POP	Unmatched	17503	7032.2	18.2		14.66	0.000
	Matched	17503	14762	4.8	73.8	2.71	0.007
FD	Unmatched	2.8132	1.3138	29.2		23.10	0.000
	Matched	2.8132	2.7561	1.1	96.2	0.50	0.617
FD_DIST	Unmatched	555.5	2539.2	-30.1		-22.82	0.000
	Matched	555.5	619.43	-1.0	96.8	-1.11	0.265
FD_STAFF	Unmatched	70.711	31.862	23.6		19.01	0.000
	Matched	70.711	71.075	-0.2	99.1	-0.11	0.914
UNITS	Unmatched	7283.2	2830.5	19.0		15.38	0.000
	Matched	7283.2	6119.8	5.0	73.9	2.77	0.006
PCT_VACANT	Unmatched	.11511	.17196	-44.7		-33.98	0.000
	Matched	.11511	.11108	3.2	92.9	2.88	0.004
PCT_OWNROCC	Unmatched	.68932	.64546	33.1		25.40	0.000
	Matched	.68932	.69339	-3.1	90.7	-2.43	0.015
VAL_OWNROCC	Unmatched	108.33	98.926	13.4		10.24	0.000
	Matched	108.33	111.46	-4.4	66.7	-3.27	0.001
AGE_OWNROCC	Unmatched	51.344	55.087	-6.5		-4.95	0.000
	Matched	51.344	52.268	-1.6	75.3	-1.55	0.121
INCOME	Unmatched	19.527	17.855	23.9		18.47	0.000
	Matched	19.527	19.782	-3.6	84.8	-2.58	0.010
PCT_POVERTY	Unmatched	.10338	.13162	-37.8		-28.82	0.000
	Matched	.10338	.10051	3.8	89.8	3.31	0.001
PCT_URBAN	Unmatched	.37768	.20769	46.0		36.14	0.000
	Matched	.37768	.39415	-4.5	90.3	-2.98	0.003
PCT_MALES	Unmatched	.49641	.50308	-21.5		-16.30	0.000
	Matched	.49641	.49604	1.2	94.5	1.07	0.285
PCT_YMALES	Unmatched	.13287	.13218	2.8		2.17	0.030
	Matched	.13287	.13242	1.8	34.8	1.46	0.143
PCT_WHITE	Unmatched	.89255	.86718	15.1		11.46	0.000
	Matched	.89255	.89316	-0.4	97.6	-0.30	0.762
PCT_25_HS	Unmatched	.84689	.82051	24.7		19.02	0.000
	Matched	.84689	.84803	-1.1	95.7	-0.85	0.398
PCT_DISABLED	Unmatched	.3317	.35509	-19.1		-14.63	0.000
	Matched	.3317	.32838	2.7	85.8	2.13	0.033
PCT_VETERANS	Unmatched	.17408	.16845	11.0		8.47	0.000
	Matched	.17408	.17403	0.1	99.2	0.07	0.946

4.2 Propensity Score—Strata 1 (Cities with Population ≤10 000)

Strata 1 is a subset of cities with a population of 10 000 people or less. This comprised 80 % of all cities analyzed. Of these, 7138 cities out of 19 973 (36 %) reported to NFIRS every year from 2002 to 2009 (see Table 4-1).

A logit model was used to construct the propensity score for strata 1. The regression results are shown in Table 4-4. Most variables are statistically significant at the 5 % level. Note, however, the inclusion of an additional higher-order term: PCT_WHITE2. This represents a squared term. This was done to achieve covariate balance for PCT_WHITE (results without the higher order term are not shown). The significant variables include: POP, FD, FD_DIST, UNITS, PCT_VACANT, PCT_OWNROCC, AGE_OWNROCC, PCT_POVERTY, PCT_URBAN, PCT_YMALES, PCT_WHITE2, and PCT_VETERANS.

Table 4-4. Regression results from logit model used to generate a balancing score on Strata 1 (Pop. ≤ 10 000).

NFIRS	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
POP	.000109	.0000266	4.11	0.000	.000057	.0001611
FD	.0621099	.0168356	3.69	0.000	.0291127	.0951071
FD_DIST	-.0000677	7.02e-06	-9.64	0.000	-.0000814	-.0000539
FD_STAFF	-.0001999	.000698	-0.29	0.775	-.001568	.0011682
UNITS	.0003653	.0000574	6.36	0.000	.0002527	.0004778
PCT_VACANT	-.9798014	.2826211	-3.47	0.001	-1.533729	-.4258742
PCT_OWNROCC	1.507662	.260535	5.79	0.000	.9970228	2.018301
VAL_OWNROCC	-.0006733	.0004865	-1.38	0.166	-.0016269	.0002803
AGE_OWNROCC	-.0009804	.0003893	-2.52	0.012	-.0017434	-.0002174
INCOME	-.0084254	.0053548	-1.57	0.116	-.0189207	.0020698
PCT_POVERTY	-4.669032	.4428048	-10.54	0.000	-5.536913	-3.801151
PCT_URBAN	-.4188619	.0738062	-5.68	0.000	-.5635194	-.2742045
PCT_MALES	-1.256555	.663221	-1.89	0.058	-2.556444	.0433342
PCT_YMALES	3.730159	.819803	4.55	0.000	2.123374	5.336943
PCT_WHITE	-1.219834	.6529122	-1.87	0.062	-2.499518	.0598502
PCT_WHITE2	1.26062	.4573367	2.76	0.006	.3642568	2.156984
PCT_25_HS	-.2844151	.2801674	-1.02	0.310	-.8335332	.2647029
PCT_DISABLED	.3105467	.2260051	1.37	0.169	-.1324151	.7535085
PCT_VETERANS	2.277193	.3610494	6.31	0.000	1.569549	2.984837
_cons	-1.362596	.5831269	-2.34	0.019	-2.505503	-.2196878

Table 4-5 presents the results of the balancing tests. For all variables found significant in Table 4-4, the means are statistically different (5 %) in the unmatched tests. For instance, the average population is 3543 people for cities that report to NFIRS and 1940 people for cities that do not report. However, when conditioned on the propensity score (matched results), all variables are balanced, meaning there is not a statistical difference in the means of each variable between reporting groups. For instance, when using the propensity score to match 'like' cities, the population mean of the non-reporting cities increases to 3566, which is not statistically different (5 % level) than the population of the reporting cities. Thus, matched reporting and non-reporting cities have a similar distribution of populations.

Table 4-5. Balancing results on Strata 1 (Pop. ≤ 10 000).

Variable	Unmatched Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
POP	Unmatched	3542.6	1939.9	68.5		47.60	0.000
	Matched	3542.6	3565.8	-1.0	98.6	-0.52	0.600
FD	Unmatched	1.3779	.89121	34.5		23.86	0.000
	Matched	1.3779	1.4031	-1.8	94.8	-0.90	0.369
FD_DIST	Unmatched	765.64	2639.4	-26.0		-17.15	0.000
	Matched	765.64	838.42	-1.0	96.1	-0.87	0.382
FD_STAFF	Unmatched	32.968	20.614	37.7		26.10	0.000
	Matched	32.968	33.999	-3.1	91.7	-1.63	0.103
UNITS	Unmatched	1554.9	883.84	61.7		43.21	0.000
	Matched	1554.9	1567.7	-1.2	98.1	-0.62	0.532
PCT_VACANT	Unmatched	.13303	.18451	-38.6		-25.28	0.000
	Matched	.13303	.13224	0.6	98.5	0.42	0.677
PCT_OWNROCC	Unmatched	.69912	.64602	40.0		26.32	0.000
	Matched	.69912	.69818	0.7	98.2	0.48	0.633
VAL_OWNROCC	Unmatched	98.736	89.849	14.8		9.83	0.000
	Matched	98.736	99.912	-2.0	86.8	-1.18	0.238
AGE_OWNROCC	Unmatched	48.069	52.676	-8.4		-5.45	0.000
	Matched	48.069	48.802	-1.3	84.1	-1.03	0.301
INCOME	Unmatched	18.539	17.302	20.2		13.42	0.000
	Matched	18.539	18.63	-1.5	92.6	-0.89	0.373
PCT_POVERTY	Unmatched	.10509	.13238	-37.8		-24.76	0.000
	Matched	.10509	.10558	-0.7	98.2	-0.47	0.641
PCT_URBAN	Unmatched	.19718	.12772	23.9		16.43	0.000
	Matched	.19718	.20364	-2.2	90.7	-1.23	0.217
PCT_MALES	Unmatched	.49961	.50458	-15.6		-10.12	0.000
	Matched	.49961	.49931	0.9	94.1	0.68	0.496
PCT_YMALES	Unmatched	.13353	.13212	5.6		3.69	0.000
	Matched	.13353	.13296	2.3	59.9	1.56	0.120
PCT_WHITE	Unmatched	.91529	.88108	21.3		13.80	0.000
	Matched	.91529	.91135	2.5	88.5	1.79	0.074
PCT_WHITE2	Unmatched	.85461	.81082	20.0		13.05	0.000
	Matched	.85461	.84841	2.8	85.8	1.93	0.054
PCT_25_HS	Unmatched	.83982	.8192	19.4		12.91	0.000
	Matched	.83982	.83796	1.8	91.0	1.13	0.259
PCT_DISABLED	Unmatched	.33925	.35906	-15.7		-10.37	0.000
	Matched	.33925	.33967	-0.3	97.8	-0.22	0.826
PCT_VETERANS	Unmatched	.17705	.17128	11.5		7.62	0.000
	Matched	.17705	.17666	0.8	93.4	0.47	0.635

4.3 Propensity Score—Strata 2 (Cities with 10 000 < Population ≤ 25 000)

Strata 2 is a subset of cities with a population greater than 10 000, but less than or equal to 25 000. This comprised 11 % of all cities analyzed. Of these, 1844 cities out of 2777 (66 %) reported to NFIRS every year from 2002 to 2009 (see Table 4-1). Of the three population groups analyzed, Strata 2 had the best reporting rate.

A logit model was used to construct the propensity score for strata 2. The regression results are shown in Table 4-6. Fewer than half of the variables are statistically significant at the 5 % level. The significant variables include: FD_DIST, VAL_OWNROCC, AGE_OWNROCC, PCT_URBAN, PCT_YMALE, PCT_WHITE, PCT_25_HS, and PCT_DISABLED.

Table 4-6. Regression results from logit model used to generate a balancing score on Strata 2 (10 000 < Pop. ≤ 25 000).

NFIRS	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
POP	.0000366	.0000341	1.07	0.282	-.0000302	.0001034
FD	-.0278099	.0157079	-1.77	0.077	-.0585968	.002977
FD_DIST	-.0002859	.0000456	-6.27	0.000	-.0003752	-.0001965
FD_STAFF	-.0000576	.0001984	-0.29	0.772	-.0004464	.0003313
UNITS	2.73e-06	.0000796	0.03	0.973	-.0001534	.0001588
PCT_VACANT	-.0959396	1.158332	-0.08	0.934	-2.366228	2.174348
PCT_OWNROCC	-.2681101	.581523	-0.46	0.645	-1.407874	.871654
VAL_OWNROCC	-.0041966	.0010696	-3.92	0.000	-.0062929	-.0021003
AGE_OWNROCC	-.0028531	.0008738	-3.27	0.001	-.0045657	-.0011404
INCOME	.017607	.0117991	1.49	0.136	-.0055188	.0407327
PCT_POVERTY	-1.347311	1.237029	-1.09	0.276	-3.771843	1.077221
PCT_URBAN	-.616943	.2281261	-2.70	0.007	-1.064062	-.1698241
PCT_MALES	2.754427	2.100399	1.31	0.190	-1.36228	6.871133
PCT_YMALES	7.799151	2.614047	2.98	0.003	2.675713	12.92259
PCT_WHITE	1.620091	.342858	4.73	0.000	.9481015	2.29208
PCT_25_HS	2.075133	.8951543	2.32	0.020	.3206633	3.829604
PCT_DISABLED	1.821125	.8673618	2.10	0.036	.1211267	3.521122
PCT_VETERANS	.9443208	1.139133	0.83	0.407	-1.288338	3.17698
_cons	-4.851424	1.724398	-2.81	0.005	-8.231182	-1.471665

Table 4-7 presents results of the balancing tests. Of the eight statistically significant (5 % level) variables from the logit model (see Table 4-6), six are not balanced in the unmatched tests (see Table 4-7). All six are balanced by the propensity score matching procedure, however.

Table 4-7. Balancing results on Strata 2 (10 000 < Pop. ≤ 25 000).

Variable	Unmatched Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
POP	Unmatched	15898	15697	4.8		1.19	0.234
	Matched	15898	15848	1.2	75.3	0.36	0.721
FD	Unmatched	3.1969	2.8692	11.0		2.76	0.006
	Matched	3.1969	3.2678	-2.4	78.3	-0.75	0.455
FD_DIST	Unmatched	117.69	1146.3	-33.5		-10.06	0.000
	Matched	117.69	122.7	-0.2	99.5	-0.21	0.837
FD_STAFF	Unmatched	82.082	76.36	3.2		0.70	0.481
	Matched	82.082	87.187	-2.8	10.8	-0.86	0.387
UNITS	Unmatched	6574.9	6368.6	10.0		2.52	0.012
	Matched	6574.9	6562.1	0.6	93.8	0.19	0.847
PCT_VACANT	Unmatched	.08538	.08629	-1.2		-0.31	0.759
	Matched	.08538	.08542	-0.0	96.4	-0.01	0.989
PCT_OWNROCC	Unmatched	.68762	.66523	18.2		4.71	0.000
	Matched	.68762	.68802	-0.3	98.2	-0.11	0.915
VAL_OWNROCC	Unmatched	127.64	153.04	-25.4		-6.79	0.000
	Matched	127.64	128.15	-0.5	98.0	-0.20	0.840
AGE_OWNROCC	Unmatched	52.169	68.075	-23.6		-6.56	0.000
	Matched	52.169	52.891	-1.1	95.5	-0.52	0.605
INCOME	Unmatched	21.314	22.326	-9.8		-2.60	0.009
	Matched	21.314	21.389	-0.7	92.6	-0.27	0.786
PCT_POVERTY	Unmatched	.09707	.11343	-19.7		-5.31	0.000
	Matched	.09707	.09559	1.8	91.0	0.72	0.474
PCT_URBAN	Unmatched	.67753	.74207	-25.3		-6.33	0.000
	Matched	.67753	.67793	-0.2	99.4	-0.05	0.962
PCT_MALES	Unmatched	.49131	.49335	-6.7		-1.79	0.074
	Matched	.49131	.49112	0.6	90.8	0.25	0.805
PCT_YMALES	Unmatched	.13199	.13096	4.5		1.17	0.243
	Matched	.13199	.13097	4.5	1.2	1.54	0.123
PCT_WHITE	Unmatched	.87225	.81276	36.2		9.51	0.000
	Matched	.87225	.87242	-0.1	99.7	-0.04	0.970
PCT_25_HS	Unmatched	.85546	.83819	15.5		4.01	0.000
	Matched	.85546	.85552	-0.1	99.7	-0.02	0.985
PCT_DISABLED	Unmatched	.32161	.31977	1.7		0.44	0.662
	Matched	.32161	.31889	2.5	-47.7	0.81	0.416
PCT_VETERANS	Unmatched	.16968	.15425	31.0		7.78	0.000
	Matched	.16968	.16872	1.9	93.8	0.62	0.535

4.4 Propensity Score—Strata 3 (Cities with Population>25 000)

Strata 3 is a subset of cities with a population greater than 25 000. This comprised the remaining 9 % of cities analyzed. Of these, 729 cities out of 2207 (33 %) reported to NFIRS every year from 2002 to 2009 (see Table 4-1).

A logit model was used to construct the propensity score for strata 3. The regression results are shown in Table 4-8. Note, however, the inclusion of two additional higher-order terms: VAL_OWNROCC2 and INCOME2. These represent squared terms. This was done to achieve covariate balance for VAL_OWNROCC and INCOME (results without higher order terms are not shown).

Nine of the 20 variables are statistically significant at the 5 % level. The significant variables include: FD_DIST, PCT_VACANT, PCT_OWNROCC, VAL_OWNROCC, VAL_OWNROCC2, INCOME, PCT_URBAN, PCT_WHITE, and PCT_VETERANS.

Table 4-8. Regression results from logit model used to generate a balancing score on Strata 3 (Pop. > 25 000).

NFIRS	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
POP	-4.37e-06	4.05e-06	-1.08	0.281	-.0000123	3.57e-06
FD	-.005469	.0087939	-0.62	0.534	-.0227047	.0117668
FD_DIST	-.0003709	.00006	-6.18	0.000	-.0004886	-.0002532
FD_STAFF	.000157	.0003237	0.49	0.628	-.0004774	.0007915
UNITS	.000012	.0000102	1.18	0.239	-7.99e-06	.000032
PCT_VACANT	-4.113417	1.580257	-2.60	0.009	-7.210664	-1.01617
PCT_OWNROCC	-2.387989	.6880056	-3.47	0.001	-3.736456	-1.039523
VAL_OWNROCC	-.0280108	.0029056	-9.64	0.000	-.0337058	-.0223159
VAL_OWNROCC2	.0000158	4.92e-06	3.21	0.001	6.15e-06	.0000254
AGE_OWNROCC	-.0025255	.0013451	-1.88	0.060	-.0051617	.0001108
INCOME	.1879741	.049743	3.78	0.000	.0904795	.2854687
INCOME2	-.0008456	.0006513	-1.30	0.194	-.0021221	.0004309
PCT_POVERTY	-.9395408	1.600866	-0.59	0.557	-4.07718	2.198098
PCT_URBAN	-1.055298	.495994	-2.13	0.033	-2.027429	-.0831678
PCT_MALES	2.075407	4.354222	0.48	0.634	-6.458711	10.60953
PCT_YMALES	6.977429	3.792087	1.84	0.066	-.4549258	14.40978
PCT_WHITE	1.982503	.4055677	4.89	0.000	1.187605	2.777401
PCT_25_HS	2.022126	1.307303	1.55	0.122	-.5401411	4.584393
PCT_DISABLED	-1.738697	1.275423	-1.36	0.173	-4.238479	.7610854
PCT_VETERANS	-3.462222	1.466424	-2.36	0.018	-6.336361	-.5880829
_cons	-.0676141	3.111953	-0.02	0.983	-6.166929	6.031701

Table 4-9 presents results of the balancing tests. Of the nine statistically significant (5 % level) variables from the logit model (see Table 4-8), eight are not balanced in the unmatched tests (see Table 4-9). (INCOME is statistically significant in the logit model, but not for the unmatched means test). All eight are balanced by the propensity score matching procedure.

Table 4-9. Balancing results on Strata 3 (Pop. > 25 000).

Variable	Unmatched Matched	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
POP	Unmatched	86939	85651	0.7		0.17	0.867
	Matched	86939	74430	7.3	-871.3	2.24	0.025
FD	Unmatched	9.2673	6.7459	16.0		3.81	0.000
	Matched	9.2673	8.6517	3.9	75.6	1.32	0.187
FD_DIST	Unmatched	86.744	2517	-53.4		-14.46	0.000
	Matched	86.744	70.504	0.4	99.3	0.82	0.414
FD_STAFF	Unmatched	238.83	172.5	14.4		3.33	0.001
	Matched	238.83	247.47	-1.9	87.0	-0.52	0.604
UNITS	Unmatched	35836	32596	4.7		1.04	0.301
	Matched	35836	30820	7.2	-54.8	2.12	0.034
PCT_VACANT	Unmatched	.06568	.06061	11.2		2.55	0.011
	Matched	.06568	.06657	-2.0	82.4	-0.56	0.576
PCT_OWNROCC	Unmatched	.64412	.6104	25.7		5.87	0.000
	Matched	.64412	.64742	-2.5	90.2	-0.77	0.444
VAL_OWNROCC	Unmatched	130.56	189.7	-60.3		-14.83	0.000
	Matched	130.56	133.24	-2.7	95.5	-1.16	0.246
VAL_OWNROCC2	Unmatched	21214	51007	-49.6		-12.87	0.000
	Matched	21214	21479	-0.4	99.1	-0.27	0.784
AGE_OWNROCC	Unmatched	66.133	78.256	-21.5		-5.01	0.000
	Matched	66.133	66.343	-0.4	98.3	-0.12	0.903
INCOME	Unmatched	22.069	21.856	2.4		0.57	0.566
	Matched	22.069	22.46	-4.5	-83.8	-1.51	0.131
INCOME2	Unmatched	536.11	581.44	-8.4		-1.99	0.046
	Matched	536.11	554.38	-3.4	59.7	-1.19	0.235
PCT_POVERTY	Unmatched	.10295	.14142	-38.6		-9.53	0.000
	Matched	.10295	.09989	3.1	92.0	1.31	0.191
PCT_URBAN	Unmatched	.87539	.9325	-43.5		-9.15	0.000
	Matched	.87539	.88339	-6.1	86.0	-1.50	0.134
PCT_MALES	Unmatched	.48733	.48913	-10.8		-2.53	0.011
	Matched	.48733	.4855	11.0	-2.1	3.22	0.001
PCT_YMALES	Unmatched	.13083	.13474	-16.9		-3.84	0.000
	Matched	.13083	.12826	11.1	34.0	3.29	0.001
PCT_WHITE	Unmatched	.80805	.692	61.6		14.33	0.000
	Matched	.80805	.80882	-0.4	99.3	-0.13	0.895
PCT_25_HS	Unmatched	.87034	.82096	43.4		10.33	0.000
	Matched	.87034	.86976	0.5	98.8	0.19	0.851
PCT_DISABLED	Unmatched	.30787	.33032	-23.6		-5.32	0.000
	Matched	.30787	.309	-1.2	94.9	-0.36	0.721
PCT_VETERANS	Unmatched	.16522	.13674	51.7		11.68	0.000
	Matched	.16522	.1624	5.1	90.1	1.57	0.117

5 Summary and Future Research

The results demonstrate there are differences in socioeconomic and fire department characteristics between cities that report to NFIRS and those that do not. Should these factors also affect fire risk, which are correlated to NFIRS reporting status, then generalizations made about fire safety and risk based on NFIRS data (only) will not apply to non-reporting regions of the United States.

For 'Strata 1' cities (those with populations less than or equal to 10 000), non-reporting cities have smaller populations (POP), fewer fire departments (FD), fewer housing units (UNITS) than do reporting cities. The proportion of owner occupied housing (PCT_OWNROCC) and proportion of city classified as urban (PCT_URBAN) are lower in non-reporting cities, as well. In addition, the proportion of the population that are young male (PCT_YMALES), White (PCT_WHITE) (PCT_WHITE is correlated with NFIRS reporting status in the propensity score model without higher order terms), and veteran (PCT_VETERANS) are lower in non-reporting cities. The distance to closest fire department (FD_DIST), proportion of housing units classified as vacant (PCT_VACANT), age of owner occupied housing (AGE_OWNROCC), and proportion of population classified as living below the poverty level (PCT_POVERTY) are greater in non-reporting cities.

For 'Strata 2' cities (those with populations greater than 10 000, but less than or equal to 25 000), non-reporting cities have distance to closest fire department (FD_DIST), value of owner occupied unit (VAL_OWNROCC), age of owner occupied unit (AGE_OWNROCC), and proportion of city classified as urban (PCT_URBAN) greater than reporting cities. The proportion of population classified young male (PCT_YMALE), proportion of population classified as White (PCT_WHITE), proportion of population age 25 or older with a high school education (PCT_25_HS), and proportion of population classified as disabled (PCT_DISABLED) are lower in non-reporting cities.

For ‘Strata 3’ cities (those with populations greater than 25 000), non-reporting cities have lower proportion of housing units classified as vacant (PCT_VACANT), proportion of units that are owner occupied (PCT_OWNROCC), proportion of population classified as White (PCT_WHITE), and proportion of population classified as veterans (PCT_VETERANS) than reporting cities. Non-reporting cities also have lower incomes (INCOME) than do reporting cities. The distance to closest fire department (FD_DIST), value of owner occupied units (VAL_OWNROCC), and proportion of city classified as urban (PCT_URBAN) are greater in non-reporting cities than in reporting cities.

While fire incident data from non-reporting cities are not available, the matching approach developed here presents a possible path forward toward producing detailed NFIRS-based statistics that are more consistent with the current U.S. fire problem. Risk comparisons (e.g., the effect residential fire sprinklers has on occupant safety) made on matched reported incident data would ensure those factors that are correlated with reporting status and fire risk are balanced. Thus, the results would be generalizable to the rest of the U.S., as differences between reporting and non-reporting cities have been taken into account. The next phase of the larger analysis is to use this proof-of-concept to generate a better understanding of the risks associated with residential upholstered furniture in home fires. This will include revisiting the relevant confounders (covariates) needed.

References

Dehejia, R.H., Wahba, S., 2002. Propensity score-matching methods for nonexperimental causal studies. The Review of Economics and Statistics 84, 151-161.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature 47, 5-86.

Leuven, E., Sianesi, B., 2003. PSMATCH2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Version 3.0.0, http://ideas.repec.org/c/boc/bocode/s432001.html, Last accessed on 12 January 2009.

U.S. Dept. of Commerce, Bureau of the Census, and Inter-university Consortium for Political and Social Research. Census of population and housing, 2000 [United States]: Block group subset from summary file 3 [Computer file]. ICPSR ed. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census, and Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producers], 2004. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004. doi:10.3886/ICPSR13576

U.S. Fire Administration, National Fire Data Center, 2006. National fire incident reporting system 5.0. FEMA, Washington, DC.