UM-HSRI-77-36-2

FIRE DATA METHODOLOGY: VOLUME II

Estimation of Fire Incidents

Jairus D. Flora, Jr.
Lily Ch. Huang
Larry D. Roi
Peter Cooley

May 1977

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UM-HSRI-77-36-2		

4. Title and Subtitle	5. Report Date
Fire Data Methodology: Vol. II Estimation of Fire Incidents	May 1977
	6. Performing Organization Code
7. Author(s) Jairus D. Flora,Jr., Lily C. Huang, Larry D. Roi, Peter Cooley	8. Performing Organization Report No.
9. Performing Organization Name and Address Highway Safety Reserach Institute The University of Michigan Ann Arbor, Michigan 48109	10. Work Unit No. (TRAIS) 014759
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address U.S. Dept. of Commerce National Fire Prevention and Control Admin. Washington, D.C. 20230	13. Type of Report and Period Covered Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report suggests methods which may be used to extrapolate data on fire incidents from a few states to the nation. Caution is needed when attempting such extrapolations and attempts to obtain external validation are in order.

To illustrate the methodology, the report uses fire department data from the State of Michigan to develop a predictive model relating fire data to several census variables. The resulting model can be combined with census data to obtain projected fire rates for additional areas or for the U.S. Insurance claim records are used in an attempt to validate the model.

Detailed fire department data for the State of Michigan and conceptual problems with fire department reporting systems are summarized. These include non-reporting of some fires and some fire injuries, lack of detail, and data quality. Supplemental sampling is recommended to validate and to supplement the fire department data. Three national sampling plans are suggested as possible approaches to obtaining more timely, and detailed national data.

17. Key Words	18. Distribution Statement
Fires Fire Incident Data Analysis Statistical Modelling	

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
		48 plus Appendices	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

* U. S. GOVERNMENT PRINTING OFFICE : 1973 725-504/328

CONTENTS

The authors express their appreciation to the Office of the State Fire Marshal, State of Michigan, for its cooperation in use of the Michigan Fire Incident Reporting System data. Special thanks are due Captain William Rucinski and Lieutenant Myron Franks. Particular thanks are also due Dean Flesner, Gray Middleton, Leon Ingerham, Bob Kouts, and Jack Post, of the Michigan Division of the State Farm Fire and Casualty Company, for providing insurance claim data for each county in Michigan for use in modeling and validating fire department data.

EXECUTIVE SUMMARY

National data on fires in the United States have been incomplete. This has hampered efforts to develop programs in fire prevention, since the total magnitude of the problem has been unknown and data on the relative importance of causes of fires have been lacking. The Federal Fire Prevention and Control Administration's Fire Data Center was organized to develop the data needed to accurately assess the magnitude of the problem, assign priorities, and develop and implement effective countermeasures to reduce the property loss, injuries, and deaths caused by fires.

The National Fire Incident Reporting System (NFIRS) was developed to supply the needed data. This system works in cooperation with states. Each state collects and computerizes the data of fire department records within that state. These state data are then assembled in Washington and used to obtain a national picture of the fire problem.

The NFIRS, however, is still in the development stage. It is being implemented in phases, with groups of states being added each year. As a result, only a few states are currently supplying data to the NFIRS. Yet there is a current need for these data to assess the magnitude of the fire problem. This report deals with the problem of extrapolating data from a few states and localities to a national total. The data from Michigan are used to exemplify the methods suggested. However, it is expected that the models would differ somewhat if the total data set available to the NFPCA were used.

Modeling the National Fire Rate

Several methods of using the data from the U.S. census to aid in the extrapolation of fire rates to the U.S. as a whole were investigated. Previous work had suggested that data based on census tracts would be useful, but practical problems prevented this geographical detail of data from being used with statewide data sets. Instead, data were aggregated by counties. The differences in fire rates (fires per thousand persons or fires per thousand dwelling units) by county were modeled using county data from the U.S. census. The resulting model can be used with census data to predict a fire rate--and hence a total number of fires--for counties outside the states with centralized uniform fire data.

Factor analysis was attempted to combine the county socio-economic and demographic variables into "factors" which might be more useful in predicting fire rates than the original variables. However, although five major factors were identified from the twenty-four candidate variables, these factors were no more successful in predicting fire rates than the variables themselves, so were not used.

Similarly, weighted least squares was used to determine if better predictions could be obtained than with ordinary least squares. Again, the result was not an improvement over ordinary least squares, so the latter was used.

The best model for predicting fire rate explained 53.4% of the variation of the fire rates among the counties in Michigan. Further, in the aggregate, it gave a fire rate within 6% of that calculated from the total number of fires. The final model was

$$F/C = -7.3824 + 1.445X_1 = 0.4X_2 + 115.5X_3 + 0.0216X_4, \text{ for low-density counties,}$$

$$F/C = 3.165 + 1.076X_1 = 0.588X_2 + 56.16X_3 + 0.548X_4, \text{ for high-density counties.}$$

In the model, F/C stands for fires per thousand persons, X_1 is the percent of the population over 65, X_2 is the percent of families with annual income of less than $3000, X_3 is the percent of the population which are Aid to Dependent Children Recipients, and X_4 is the percent of housing unit changes during the decade from 1960 to 1970. This model gave an estimated fire rate of 7.83 per thousand persons for the State of Michigan, compared with a rate of 7.40 calculated from the total number of fires reported to the State Fire Marshal's Office. The model predicts a fire rate of 7.96 for the State of Ohio, which could be compared with the fire rate derived from the NFIRS data.

The data from the fire department reports from the State of Michigan were summarized. Among the highlights noted were the following. Residential fires amounted to slightly over one-fourth of the fires reported by fire departments. However, residential fires resulted in over half of all fire injuries and nearly three-fourths of fire fatalities. Among residential fires, the leading causes were: heating (16% of the fires), cooking (12%), flammable liquids (11%), and smoking (10%). However, among the residential fires resulting in injuries or fatalities, the leading causes were: smoking (21% of the fires), cooking (15%), heating (14%), and flammable liquids (8%). Among all fires the leading causes were somewhat different: Open flames (17%) was the most frequent cause, followed by smoking (7.2%), heating (6.6%), and other equipment (6.4%).

Several conceptual problems as well as a number of practical problems were identified in the fire department data system. Certain types of fires are not covered in fire department reports. In addition, some injuries may occur in fire incidents which are not reported to fire departments. The problem of defining coverage suggests that a careful definition of what constitutes a "fire incident" is needed. Probably some severity threshold will be needed to ensure that all such incidents are consistently present in the data set. Similarly, definition of the severity or detailed information on the injury is needed. Otherwise, there is the risk of recording the

consequences of, say, a severe burn, as the consequences of a "fire injury" while estimating the number of such injuries based on the total number of all injuries, however minor.

Several reasons led to the recommendation that the NFIRS be augmented with a parallel and/or supplemental sampling system. This recommended national sample of fire incidents would serve three purposes: It would provide accurate data earlier than NFIRS, it would provide a means of checking the completeness and validity of the NFIRS data, and it would provide for the collection of detailed data relevant to current special topics, which would be impractical with the entire NFIRS.

Three different sampling approaches have been suggested. The first is the least expensive and would provide the least additional information. It consists of selecting a sample of fire departments in the U.S. and arranging for those fire departments to supply data on all incidents on a standard form to the NFPCA. These data would be collected and reported by existing fire department personnel, and supported by supplemental funds from the NFPCA. The second approach would utilize persons other than fire department personnel to collect and code the data from the sampled fire departments. The third approach would use regionally located special investigation teams who would conduct in-depth investigations of a sample of fires in their respective areas. The third approach is the most costly, but would provide detailed accurate and uniform data on fires and fire victims through a specially trained investigation team. These detailed data would be from a probability sample of fires in the U.S. A selection among the three types would depend on the detail of data desired and the resources to be committed to the sample effort.

1.0 INTRODUCTION

Our knowledge of the problem of fires in the United States is subject to considerable uncertainty. There is, however, general agreement that there are many preventable fires and that the annual toll of property loss, personal injuries, and deaths caused by fires could be reduced. In order to determine where to allocate efforts to reduce fires, and to determine the extent of potential fire reduction as well as reductions actually achieved, it is necessary to have valid and reliable information about the fires which occur in the U.S. To obtain the information needed for determining the size and scope of the problem, as well as for suggesting countermeasures, the National Fire Prevention and Control Administration (NFPCA) has begun implementation of a National Fire Incident Reporting System (NFIRS) which collects data from fire departments in each state at a central location. These state data are then combined by the NFPCA to estimate national statistics.

Since the NFIRS data are available from only a few of the first states participating in the NFIRS, complete and detailed data dealing with the fire problem in the U.S. is now limited. This report addresses the problem of methods which can be used with data from a limited number of states to arrive at the best available estimates of fires and their consequences to better understand the fire problem.

U.S. census data are used in Section 2.0 to develop models for extrapolating data from the few states to the NFIRS to the U.S. These methods are illustrated by developing a prediction model based on fire data from the State of Michigan. Fire insurance claim data from the State Farm Fire and Casualty Company are used to obtain a similar model, and the two are compared. An attempt at cross-validation is made, which suggests that fire department data may not include all incidents which come to the attention of insurance companies.

Section 3.0 analyzes fire data from the State of Michigan in some detail. Fire incidents are related to the various types of affected property, causes of fire, and injuries produced.

In Section 4.0, conceptual problems using a fire data system based on fire department reporting are considered. Some of the practical problems in obtaining reliable data from such a system are discussed, with particular emphasis on our experience with the data from the State of Michigan, with some non-coverage in terms of certain types of fires and fire injuries identified. The need for careful definition and documentation of fire size, fire severity, and resultant injuries is noted.

Section 5.0 suggests that a national sample of fire department data be collected to parallel the NFIRS. There are three basic needs for this sample. These are to provide: (1) better estimates before the NFIRS is fully operational, (2) a validation check on the NFIRS after it is operational, and (3) the opportunity for data collection with data in greater detail than is feasible with the NFIRS. Three types and levels of sampling effort are discussed. The selection of a preferred system depends on the resources to be committed to the sampling effort as well as the degree of detail desired in the basic data.

2. METHODS FOR EXTRAPOLATING FIRE INCIDENT DATA

2.0 Data Used

Three data sets were used in describing and illustrating the methods of analysis suggested to improve the national estimates of fire incidents and their consequences. The data on 24 demographic variables collected by the U.S. Census and tabulated by counties for the United States were used as predictor variables.

Data collected by the State of Michigan Fire Marshall's office through the Michigan Fire Incident Reporting System (MFIRS) provided the main data on fire incidents in the State of Michigan.

Data on fire insurance claims for each county in Michigan were provided to HSRI through the courtsey of State Farm Insurance Company. Two dependent variables were also used with the State Farm data. These were the commercial fire rate and the residential fire rate. State Farm Insurance Company provided a record of the total number of fire claims, the total incurred loss from these fire claims (for State Farm only, of course), the number of policies in force, and the total insurance liability for each county. These data were separated into home (personal) and commercial categories. The residential fire rate is defined as the number of fire claims per 1000 personal policies in force, and the commercial fire rate as the number of fire claims per 1000 commercial policies in force. The financial loss from these fire incidents was not analyzed.

2.1 Direct Population Expansion

The simplest method of extrapolating from a subset of the U.S. to a national total is to multiply the total from the subset by the ratio of the U.S. population to the population of the subset. This expansion is appropriate for expanding totals to the U.S. It assumes

that the subset is similar in all regards to the U.S. population. In terms of rates--such as fires per capita--the expansion method merely means that the U.S. is assumed to have the same rate as the subset. No expansion is necessary. Thus, to estimate the total number of fire incidents in the U.S. for 1975, one might take the total for the State of Michigan, 74,970, and multiply by 23.23 (the U.S. population divided by the Michigan population) to get an estimate for the U.S. of 1,741,553. To similarly estimate a national fire rate, say per 1000 persons, the estimate would be simply to take the Michigan rate, 7.4/1000, as the estimate of the national fire rate.

The simple expansion method has the advantage of simplicity, and is reasonably informative in some instances, for example, in projecting the total number of fire fatalities. In that case, direct expansion gives an estimate of 7,225 fatalities nationally (for 1972) compared with 7320 estimated from the vital statistics using death certificates. However, if the subset of the U.S. is of a special nature, direct expansion can lead to substantial errors. If there is doubt whether the subset can be expanded to give valid national estimates, then caution is in order and some other method may be preferable.

2.2 Use of Census Data in Modeling

If the subset for which data are available differs substantially from a random sample of the U.S., then some attempt to correct for the differences before extrapolating national estimates from such a subset. The most direct method of correcting for such differences is to base the correction on national census data and the relationship of the census variables in the subset to the national experience.

The general approach to be used is to determine a model which relates the dependent variable of interest, fires per capita, say, to census variables of a demographic nature. If such a model can be found which fits the data well, then the census data available for the entire U.S. may be used together with the model to project a national estimate.

The assumptions underlying this approach, while still important, are less restrictive than assuming that the U.S. is just like the subset, only larger. The basic assumption here is that the relationship of the phenomenon to the census variables is consistent throughout the U.S., so that this relationship can be used to provide improved national estimates.

2.2.1 Variable Selection

Two types of variables must be considered in variable selection: Dependent and independent variables. The dependent variable is the variable which is being predicted. The objectives to be achieved in terms of desired estimates must be carefully defined. In this study, two different dependent variables were considered using fire data from the State of Michigan. These were the number of fires occurring per 1000 persons, and the number of fires occurring per 1000 housing units. The former expresses a fire rate on a per capita basis, while the latter expresses a fire rate on a per household basis. Fire rates rather than frequencies were chosen as the dependent variables, because frequencies would be largely predicted by population size. The effect of population size might swamp any other prediction attempts if frequencies rather than rates were used as the dependent variable.

Independent variables are those used in making the prediction. Typically these will be socioeconomic or demographic variables usually based on census data. The independent, or predictor, variables are assumed to have some association with the dependent variable which can be used to improve the estimates of the dependent variable.

Twenty-four socioeconomic and demographic variables were chosen as candidate independent variables to predict the fire rate. The 24 variables are listed in Table 2-1 and are more fully displayed in the 1972 County and City Data Book published in the U.S. Census.* Table 2-1 also gives the predictive power (R^2, or proportion of variation explained) of each of the individual variables.

---

*Bureau of the Census. County and City Data Book 1972 (A statistical abstract supplement). U.S. Government Printing Office, Washington, D.C. 1973.

Table 2-1. PREDICTIVE POWER (R^2) OF CENSUS VARIABLES

Variable	Insurance Data (Highly insured counties only) M=43		Fire Marshall Report M=83	
	Residential	Commercial	# fires/ 1000 persons	# fires/1000 housing units
% units built prior to 1950	.537*	.018	.025	.006
% units change 1960-1970	.218*	.000	.1164*	.041
% older than 65	.036	.156*	.195*	.031
% under 18	.036	.054	.044	.000
% in social sec.	.097	.032	.156*	.071
% in aid to old	.096	.040	.184*	.022
% on ADC	.050	.032	.177*	.043
Median Income	.045	.084	.024	.172*
% family income less than $3000	.017	.071	.069	.148*
% family income $3000-$5000	.009	.124	.024	.145*
% persons in low income	.089	.038	.065	.125*
% female, as head of household	.002	.019	.019	.039
% unemployed	.001	.113	.001	.273*
Black population	.041	.016	.000	.059
% change in black population	.045	.038	.004	.043
Population density	.117	.032	.000	.018
Total housing units	.114	.030	.000	.017
Net population migration	.198*	.002	.028	.006
% unit owner occupant	.040	.037	.074	.003
% owner occup. vacant	.008	.172*	.089	.164*
% rent unit vacant	.032	.259*	.036	.164*
% unit with 1.01 or more persons/ room	.022	.023	.010	.077
% of family in low income	.025	.059	.048	.152*
% less than 5 years education	.004	.000	.000	.009

*significant at .01 level.

A method of factor analysis was used to reduce the number of independent variables from 24 to a smaller and more manageable number. Factor analysis also reduces the difficulty in modeling caused by inter-dependencies among the candidate variables. This approach resulted in the selection of five factors felt to adequately represent the 24 original explanatory variables. Each of these five resultant factors is basically a particular linear combination of the original 24 variables, with the special feature that the factors are mutually independent. When possible an attempt is made to identify the factors with some meaningful interpretation.

In the present case, the first factor may be identified with economic variables, heavily influenced by such variables as percent of low-income families, percent of vacant housing units, percent of persons older than 65, and percent of persons receiving social security payments. The second factor may be called stability. It consists of primarily of the net population migration from 1960 to 1970, the housing unit changes from 1960 to 1970, and the percent of housing units built prior to 1950. The third factor, called population density, includes the number of persons per square mile, the total number of housing units, and the number of Aid to Dependent Children recipients. The fourth factor, crowdedness, primarily consists of the percent of housing units with an occupancy of more than 1.01 persons per room. The fifth factor had no identifiable primary characteristic. Table 2-2 gives the factor loadings of the five main factors. The rescaled loadings are the coefficients of each variable in the linear combination of the 24 variables that comprise each factor. These five factors were the only ones with eigenvalues greater than, or equal to, unity (factor five had an eigenvalue of only 0.96). Taken together they account for 72% of the variability among the 24 independent variables.

Table 2-2

Factor Loadings

VARIABLE	INITIAL VALUES	STEP= 5 COMMUNAL.	RESCALED FACTOR LOADINGS (1)	(2)	(3)	(4)	(5)
3.POPDEN	.98656	.94723	-.39044	.35890	.70540	.36666 -1	-.40878
4.DEIMIG	.78661	.78320	-.39630 -1	-.71765	.42673	-.24040	-.16348
5.MEDAGE	.64761	.58727	-.45996	.25165	.48157	.86398 -3	.28367
6.CHNG%	.61632	.70271	-.47165	.23557	.31882	-.26901	.50074
7.HSE18	.75857	.71324	.46270	.33541	.10378	-.61296	-.12613 -1
8.HSE65	.94771	.85265	.90217	.10135	.12344 -1	-.22791 -1	-.16672
9.EDUC5	.59375	.32964	.35782	.37203	.43924 -1	.24701	-.15970 -1
10.UNEMPL%	.72088	.59128	.75708	.53552 -1	.66504 -1	.70695 -1	.76278 -1
11.FEMHEM%	.68549	.67428	.11371	.73771	.29929	.10677	.13042
12.INCME13	.98526	.90599	.94612	.31643 -1	.96849 -1	-.63782 -3	.35045 -1
13.INCMST05	.96860	.83960	.88581	.12888	.73855 -3	-.19554	.98419 -2
14.MEDINCM	.96739	.92215	-.94967	-.69133 -1	.11380	.27354 -1	-.50568 -1
15.FAMLOWIN	.95937	.89345	.92908	.97843 -1	.10861	.54102 -2	.96591 -1
16.PERLOWIN	.97806	.94004	.91966	.15184	.12401	-.34892 -1	.12444
17.JOLSEL	.95730	.88463	.92893	.28718 -3	.91843 -1	.13474 -1	-.11444
18.OLDHSE	.80281	.66299	.77623	.13977	.78166 -1	.18653	.43215 -2
19.HDC	.53824	.43405	.16039	.25973	.43697	.25520	.25055
20.TOTOWN13	.98696	.96206	-.38013	.38175	.71432	.24287 -1	-.40124
21.CHNGEM70	.79564	.80675	.35395	-.67556	.45990	-.96751 -1	.65019 -1
22.PRTORSH	.82883	.82811	.13162	.62834	-.55441	-.61750 -2	-.20778
23.VCHGTOWN	.78016	.71407	.74838	-.32898	.20855	-.38402 -1	-.28366 -1
24.VCHGYRM1	.71656	.48136	.61266	-.22152	.22686	-.67028 -1	-.31308 -1
25.OWNOCCUP	.60249	.29587	.30362	-.16361	-.12021	.34480	-.20875
26.MII.M1	.66882	.59941	.97380 -1	-.29699	.18722	.66579	.15298
	EIGENVALUE		9.4244	3.0475	2.5476	1.3252	.95784
	% VARIANCE		35.3	52.0	62.6	68.1	72.1

These five factors--economic, stability, density, crowdedness, and others--were considered in various combinations as candidate explanatory variables. The original 24 variables were also considered singly and in various combinations in selecting a prediction model. These resulting models are described in Section 2.4. Although the original set of 24 independent variables was reduced to five factors, which explained 72% of the variability among the independent variables, and which could even be given interpretations, it turned out that combinations of the original variables provided models with no better predictive power than the models based on the derived factors. Thus, the determination of the factors was not found to be useful in improving prediction.

2.2.2 Size of Reporting Unit

Previous studies have used regression techniques to model fire rates within cities, based on census tract information. Reasonably good predictive models based on census tracts were found, but attempts to use cities as reporting units to model differences in fire rates among cities were not successful. Results suggested that census variables could be used to predict fire rates based on census tracts, but that the same variables did not work well if used on larger units--i.e., cities.

Unfortunately, a census tract is not a common geographic concept. Indeed, for most states, census tracts are defined only for urban areas, so some other unit or additional units must of necessity be used in order to include non-urban areas. In addition, although census tract was a variable on the Michigan Fire Incident Report Form, that information was consistently missing. Thus it was not possible to use census tracts as the unit of reporting with Michigan data, and it seems unlikely that census tract data can be used for any statewide data. It should be possible to use census tracts as the reporting units if attention is restricted to urban areas in the

sense that those areas are divided into census tracts. However, there will still be practical difficulties in obtaining the data for each census tract from fire department reports. Thus, except for special urban studies such as UFIRS cities, it seems that practical difficulties preclude the use of census tracts as the reporting units at the present time. The same conclusion holds for smaller units, such as blocks. They may be useful concepts in urban areas with exceptional data capabilities, but do not appear to be useful in the near future for large-scale data.

Information was readily available by county for the State of Michigan, both from the State Farm Insurance data and from the Michigan Fire Incidence Reports. In addition, census data for counties are readily obtainable. Thus, the county was utilized as the reporting unit for the predictive models based on the Michigan fire data. The optimum reporting unit has still not be resolved. However, use of counties as the reporting units has several practical advantages: It reduces the amount of effort needed in the analysis and prediction (there are about 3000 counties in the entire U.S.). Census data are also readily available by county. Generally, fire data can easily and reliably be obtained by county. Finally, it was found that with the Michigan data, prediction based on counties utilizing ordinary least squares worked as well as prediction using weighted least squares. This provides some indication that assumptions for the least squares models are not seriously violated when using counties as the reporting units.

2.2.3 Weighted Least Squares

It is shown in Appendix I that if, for example, census tract is the appropriate size of the reporting unit for modeling fire rates in terms of satisfying the assumptions of least squares estimation, then models based on other reporting units must involve violations of the homogeneity of variance assumption. This is a possible ex-

planation for the previous result that fire rates within a city could be predicted well using census data and census tracts as the reporting level, but that fire rates between cities could not be predicted well. One method to adjust or improve predictions would be to use weighted least squares to account for the differences in variances. The details are mathematical and are deferred to Appendix I. Census tracts were thought to be the appropriate reporting level for Michigan data, but could not be used for practical reasons. Weighted least squares was used on the Michigan data but did not improve the prediction. The use of counties as the reporting level and ordinary least squares seemed to work as well as weighted least squares.

2.3 Geographical and Seasonal Fluctuations

Data from both the Michigan Fire Incident Reporting System and State Farm Insurance Claims show variations from month to month. The two patterns are not identical, but both generally show a higher incidence of reported fires during warm weather months than during the cold weather months. The monthly variation is not smooth, so that it is not evident that a smooth seasonal pattern in fire incidence exists. Table 2-3 gives the monthly fire incident indices for both the State Farm Insurance data and from the MFIRS data. To some extent the differences may be due to the fact that summer months have a high proportion of outdoor fires--grass or trash, which would appear in the MFIRS data and are probably less likely to appear in the fire insurance claim data. That is, MFIRS reports fires, while State Farm data report fire insurance claims, which would be restricted to insured property.

There were also differences among county fire rates which appeared to be primarily geographical in nature. In the MFIRS data, these differences could be mostly explained by a population density or urban/rural split, dividing the counties with over 50 persons/square mile from these with under 50 persons per square mile. In insurance claim data, the differences could be related to the degree of market penetration by State Farm. That is, reported fire rates and prediction varied according to whether a high (over 10%) or a low (7%) proportion of the dwellings were insured by State Farm.

It is possible that geographical differences in fire rates are present across different regions of the U.S. If the season of year, or weather, is a factor in determining fire rate, then climates representative of different regions will influence the fire rates and imply regional differences. This presents a potential difficulty in interpreting fire data. The extent of such regional differences cannot be determined based on data from one state, or even from a few states, with similar climates and characteristics. The fire rates which are modeled here are yearly fire rates and do not provide seasonal

TABLE 2-3

SEAONAL FLUCTUATION*

Month	State Farm** Fire, Lightning, and Removal	MFIRS
January	.78	.66
February	.69	.69
March	.78	.91
April	.90	1.29
May	1.03	.99
June	1.50	1.23
July	1.48	1.13
August	1.49	1.24
September	1.04	1.07
October	.78	.86
November	.66	1.05
December	.87	.88

*A month with an index of "1" is an average; larger numbers indicate higher numbers of fires.
**Obtained from State Farm Fire and Casualty Company, Michigan Division.

estimates, although the factors in Table 2-3 could be used to determine seasonal effect. Similarly, the fire rates are modeled with no provision for geographic variability. To incorporate geographic factors would require additional data not currently available.

2.4 Models and Projections

Two different data sets have been used to derive predictive models relating fire rates to socio-economic and demographic variables. These are the insurance data provided by State Farm Insurance and the fire department incidence data for the State of Michigan, provided from the MFIRS. In both cases counties were the reporting units and the predictor variables were those available from the census county book. Either of the models could be used to obtain national projections by applying the models to the complete set of county data for the U.S. It should be noted that the two data sets are not the same--in one case fire rates are predicted from insurance claims while in the other case, fire rates are modeled from fire department records.

2.4.1 Models Based on Insurance Data

Fire claim insurance data for the State of Michigan provided by the State Farm Insurance Company were used to model two different dependent variables. These were the commercial property fire rate, and residential property fire rate. In Michigan, more than 10% of housing units were insured by the State Farm Insurance Company during the year 1975. However, the percent insured varied considerably from county to county. (The percentage of the market by county is not presented to preserve confidentiality of market information.) Over the entire state, models based on State Farm's claims data did not fit well. To a large degree this lack of fit appeared to be caused by the counties in which State Farm insured only a small proportion of the residences. Accordingly, counties were stratified according to high or low proportion of insured properties, and models fit only over the 42 counties with a high proportion of the eligible properties insured by State Farm. Basically, it seems that in the counties with a very small proportion of residences insured by State Farm, a rather special subset of the population is insured, while in counties with a relatively (10% or so) large proportion of the residences insured by State Farm,

the insured population seems to be a reasonable approximation of a random subset. As a result, the model was fit to the data from 42 counties and used to project for the rest of the state.

In the insurance data, the best predicting variable was the proportion of the housing units built prior to 1950. This single variable explained 53.7% of the variation in the fire rates among the counties. The model based on this one variable is:

$$\hat{R}_{FC} = -4.4169 + 0.6691X,$$

where $\hat{R}_{FC}$ is the predicted fire rate as fire claims per 1000 residential insurance policies in force, and X denotes the percent of housing units in the county which were built prior to 1950. This, the higher the percentage of older (prior to 1950) dwelling units, the higher the fire rate. One is tempted to speculate about possible causality in addition to mere association, since the older dwellings were constructed under less stringent fire codes, and heating equipment, wiring, and other sources of ignition may be in poor repair. However, much more information would be necessary to attempt to establish causative factors. So far, only association has been found.

There was also some association between fire rate and population density, as well as between the fire rate and the net migration for the county project in the insurance data. Thus it was found that the fire rate could be modeled more successfully in counties with high population density (150 or more persons per square mile) and also more successfully in counties with negative net migration. However, a stratified model with two strata defined by population density resulted in a very marginal increase in predicting power--only to a R^2 of 55.3%. A stratified model based on migration resulted in an even smaller increase in R^2 (to 54.2%). As a result, the increase is not significant and the more complex models are not justified.

It should be noted, however, that if one were interested in counties with a high population density (and a high percentage of the

dwellings insured by State Farm), the model incorporating one variable has an R^2 of 67.0%. Similarly, if attention is restricted to counties with negative net migration (and a high percentage of dwellings insured by State Farm), an R^2 of 95.5% can be obtained. Thus improved models are possible for special subclasses of counties, but there is some question about whether this is really a gain in predictive accuracy or an artifact of the reduced number of counties.

Models were less successful at predicting the commercial fire rate. The best model was

$$CFC = -0.1116 + 0.0214X,$$

where X is the percent of rental units which are vacant. Here CFC denotes the commercial fire claims rate. This model had an R^2 of 25.9%. No additional variables had coefficients which were significantly different from zero.

Applied statewide, the model based on insurance data predicts a total fire rate of 10.8 fires per 1000 persons. (This is obtained by determining the estimated number of fires for each county, summing over the counties to get a statewide total then dividing the statewide total by the state's population in thousands to get the rate per thousand persons.)

2.4.2 Models Based on MFIRS Data

Different combinations of independent variables were used to predict fire rates from the State of Michigan Fire Marshal's data. The per capita fire rate was best predicted by a combination of the percent of the population older than 65, the percent of families with income less than $3000 per year, the percent of Aid to Dependent Children recipients in the population, and the percent of housing changes from 1960 to 1970. The model was stratified into counties with low (50 or fewer persons per square mile) or high density. The predictive power as measured by R^2, the percent of variation explained

by the model, was 55.0% within the low-density counties and 49.4% within the high-density counties, for an overall R^2 of 53.4%. The final model* was:

$$F/C = -7.3824 + 1.445X_1 - 0.400X_2 + 115.5X_3 + 0.0216X_4, \quad \text{low-density}$$

$$F/C = 3.165 + 1.076X_1 - 0.588X_2 + 56.16X_3 + 0.548X_4, \quad \text{high-density}$$

where X_1 is the percent of the population older than 65,
X_2 is the percent of families with income less than \$3000/year,
X_3 is the percent of ADC recipients,
and X_4 is the percent of housing unit changes from 1960 to 1970.

The model for the number of fires per 1000 housing units used a different set of variables: The percent of the population older than 65, the percent unemployed, and the percent of families with a female as the head of the household. For this per residence fire rate stratification by density did not improve the model significantly. The predictive model is:

$$F/R = 30.66 + 0.3992X_1 - 2.390X_5 + 1.760X_6,$$

*Note. After the modeling was completed, it was discovered that the date from the State of Michigan Fire Marshal was not just for 1975, as assumed and requested, but actually also included data for the first six months of 1976. As a result, predicted rates from the models based on the MFIRS data would be for a year and a half rather than a year. Multiplying the predicted rates by 2/3 would approximately correct them back to an annual rate. Actually, a factor of 0.6753 derived from the seasonality table would be more appropriate. This correction has been incorporated into estimated rates, but is not incorporated into the models, which technically estimate rates for a year and a half.

where X_1 is the percent of the population older than 65,
X_5 is the percent unemployed, and
X_6 is the percent of families with a female as head of the household.

The R^2 for this model was 38.5%. The fire rate per residences is denoted by F/R.

2.4.3 Comparisons Among the Models

The total number of reported fires in the State Fire Marshal's Report for Michigan for 1975 was 74,970, which is a fire rate of 7.4 fires per 1000 persons. This may be viewed as a standard against which the models may be compared. The model based on the MFIRS data results in a state total estimate of 79,330 or a fire rate of 7.83 fires per 1000 persons. This is an error of about 6%--a slight overestimate. On the other hand, the model based on insurance data (from the State Farm Insurance Company only) resulted in an estimate total of 89,400 fires, or an estimated fire rate of 10.8 fires per 1000 persons. In comparison to the reported total, this represents a 46% difference.

Thus, the model based on insurance data, while showing some promise, does not appear adequate at the present time. There may be several reasons for this. First of all, there may in fact be a number of fires which are reported as fire losses and claims on insurance policies which are not reported through MFIRS. To a slight extent this could be underreporting in MFIRS. More likely, there may be some fires which result in property damage, but which are extinguished by the residents without a call to the fire department. These would naturally be excluded from the MFIRS data since it is based on fire department records. An additional possibility is that property insured by State Farm is a biased subset of such property in Michigan. This may apply to all insured property; that is, insured property may hve a different fire rate than non-insured property. It is also

possible that the property insured by this insurer has a higher fire rate than insured property in general. Probably the largest component in this inconsistency is reporting of small fires to insurance companies but not to fire departments. To the extent that this occurs, the data from NFIRS or any fire-department-based reporting system will underreport the true number of actual fires.

2.4.4 Application of the Model to Ohio

The model based on the MFIRS data was used with the county census data for the State of Ohio, and predicted a fire rate of 7.96 fires per 1000 persons for the State of Ohio in 1975. This fire rate is estimated to be close to that for Michigan (which is somewhat to be expected, since the two states are adjacent and quite similar in characteristics). The State Fire Marshal's report for Ohio indicates a fire rate of 4.0 fires per 1000 population. If that is correct, then the model estimates nearly twice as many fires for the State of Ohio as were reported to the State Fire Marshal's Office in Ohio. The total number of fires predicted for the year 1975 for Ohio by the model was 84,980. If the State Fire Marshal's report is accurate, then this would tend to invalidate the modeling approach based on census data. In that case, probably the best that can be done for national estimates of the fire rates is to combine data from all the reporting states, using simple expansion to arrive at a national estimate.

3.0 STATE OF MICHIGAN FIRE DATA

The Fire Marshal Division of the Department of State Police in Michigan collects, computerizes, and publishes an annual report on data submitted from the 978 fire departments in the State of Michigan. HSRI was supplied with a magnetic tape of these data in coded form for use in this study. The results in this section are based in part on these data, obtained from the Fire Marshal, and in part on published summaries.

In principle, all fire departments submit hard copy reports on all incidents weekly (for large departments) or monthly (for small departments). Even if no incidents occurred during the period, a summary is to be sent in indicating this, so that reports of no incidents can be distinguished from failure to report. The first full year of operation of the system, denoted by MFIRS, was 1975. As a consequence, there were a number of reporting and accuracy problems with the initial years' data. Many of these have been corrected, or improved, in 1976 data.

3.1 Completeness and Accuracy of Reported Data

In 1975, 962 out of the state's 978 fire departments submitted reports to MFIRS. This represents 98.4% of Michigan fire departments. There were some delays in reporting, incomplete data, and erroneous data, so that the actual data received are less than 98% accurate. The Fire Marshal's report estimates that 14.5% of total fire information is lacking. On the other hand, if all incidents are included, it appears that somewhat more of the data are missing--approximately 18%. Further, if one considers the reports with missing data, the accuracy of the reporting is even lower. For example, there were 318 fire fatalities in Michigan in 1975. However, only 201 of these were

sufficiently reported in reasonable completeness in the MFIRS to be used in various two-way tables relating nature of injury to age, part of body injured to nature of injury, or prior condition to reason for failing to escape, etc. Thus, only slightly more than 63% of the fatality data were complete. Similarly, about 70% of the injuries listed as casualties on the fire incident report were accompanied by a casuality report.

3.2 Descriptive Statistics from MFIRS Data

Table 3-1 gives distributions of the number of fires, the residential fires, civilian injuries, fire service injuries, and fire fatalities by the cause of the fire. The causes have been ordered by priority as in the NFPCA. In all categories, the causes noted as "unknown" and "other heat" are unfortunately high proportions of the totals. These are rather non-specific causes and probably represent cases where insufficient data were available to determine specific cause. Even with this non-specificity, there are some interesting patterns.

The largest cause (16%) of residential fires was heating. This cause was the second most frequent cause of fatalities (16%) and the third most frequent cause of civilian injuries (11%). Smoking was the leading cause of residential fire injuries (21%) and of fatalities (19%) and the second most frequent (after cooking) cause of all civilian fire injuries. However, smoking was relatively low (7%) as a cause of fire service injuries. Among residential fires, the second, third, and fourth most frequent causes were cooking (12%), flammable liquids (11%), and smoking (10%), respectively. These causes also contributed substantially to the civilian injuries, representing 21%, 4%, and 15% of these, respectively. Heating also was the cause associated with many of the fire service injuries (11%), and with many of the fatalities (16%), but cooking was relatively infrequently a cause of either fire service injuries (3%) or fatalities (8%).

TABLE 3-1

DISTRIBUTION OF FIRES AND INJURIES BY CAUSE
(in percent)

Cause	Fires	Residential Fires		Injuries		Fatalities
		Injury	Fires	Civilian	Fire Service	
Exposure	1.6	1.1	2.0	1.5	6.4	1.2
Natural	1.5	0.2	1.8	0.8	2.2	0.6
Incendiary/ Suspicious	2.2	4.0	2.7	3.5	3.3	4.9
Explosives/ Fireworks	1.0	0.3	1.1	0.4	1.4	0.9
Smoking	7.2	20.8	10.2	14.9	7.0	19.0
Children	5.0	5.7	5.3	3.7	4.0	1.8
Heating	6.6	14.0	16.2	11.3	11.2	16.2
Cooking	4.1	15.0	12.0	20.9	2.9	7.6
Air Condition/ Refrigeration	0.3	0.9	0.6	0.7	0.8	0.6
Electrical Distribution	5.4	4.8	7.1	5.1	7.8	7.3
Appliances	2.8	5.3	6.8	4.0	3.1	2.4
Other Equipment	6.4	2.5	0.8	6.3	4.6	2.4
Gas	1.6	0.9	2.0	1.9	1.7	1.5
Flammable Liquid	4.3	2.4	10.6	4.0	3.7	6.7
Open Flame, Spark	16.0	7.7	1.9	5.7	12.7	5.2
Other Heat	7.4	5.0	6.5	4.4	7.7	7.3
Unknown	26.0	9.2	12.5	11.5	19.6	14.1

TABLE 3-2

DISTRIBUTION OF FIRES AND FIRE INJURIES BY GROSS PROPERTY CLASS
(in percent)

Property Class	Fires	Injuries		Fatalities
		Civilian	Fire Service	
Public Assembly	2.1	1.1	3.3	0.3
Education	0.8	7.9	1.7	0.0
Institutions	0.7	1.2	0.7	2.1
Residential	28.2	54.3	53.3	74.0
Merchandising	2.8	2.3	9.1	0.9
Utilities	0.5	0.1	0.2	0.0
Industrial Manufacturing	1.9	2.4	4.1	1.2
Storage	7.0	7.3	8.7	4.9
Building Construction	2.1	0.2	1.9	0.3
Bridges, etc.	0,2	0.1	0.1	0.3
Special	27.1	2.6	5.0	9.5
Other	26.5	20.5	11.6	6.4

TABLE 3-3

GROSS STRUCTURE BY CAUSE

	Exposure	Natural	Incendiary Suspcious	Explosives Fireworks	Smoking	Children	Heating	Cooking	Air Condition-ing/Ferrigeration	Electrical Distribution	Appliances	Gas	Flammable Liquid	Open Flame Spark	Other Equipment	Other Heat	Unknown	Totals
Assembly																		
000-199	38	32	53	24	158	62	223	339	19	178	49	43	53	301	77	149	367	2165
Row	1.8	1.5	2.4	1.4	7.3	2.9	10.3	15.7	0.9	8.2	2.3	2.0	2.4	13.9	3.6	6.9	17.0	100.0
Col	2.4	2.0	2.3	2.4	2.1	1.2	3.3	8.0	5.4	3.2	1.7	2.6	1.2	1.7	1.2	1.9	1.4	2.1
Education																		
200-299	4	8	15	26	65	42	31	10	4	45	31	7	19	243	56	44	154	804
Row	0.5	1.0	1.9	3.2	8.1	5.2	3.9	1.2	0.5	5.6	3.9	0.9	2.4	30.2	7.0	5.5	19.2	100.0
Col	0.2	0.5	0.6	2.6	0.9	0.8	0.5	0.2	1.1	0.8	1.1	0.4	0.4	1.4	0.8	0.6	0.6	0.8
Instutition																		
300-399	2	4	93	3	151	1	27	40	7	32	41	3	4	66	40	28	143	685
Row	0.3	0.6	13.6	0.4	22.0	0.1	3.9	5.8	1.0	4.7	6.0	0.4	0.6	9.6	5.8	4.1	20.9	100.0
Col	0.1	0.3	4.0	0.3	2.0	0.0	0.4	0.9	2.0	0.6	1.4	0.2	0.1	0.4	0.6	0.4	0.5	0.7
Residential																		
400-499	599	516	779	308	2983	1551	4749	3504	175	2072	1994	233	588	3105	561	1898	3653	29268
Row	2.0	1.8	2.7	1.1	10.2	5.3	16.2	12.0	0.6	7.1	6.8	0.8	2.0	10.6	1.9	6.5	12.5	100.0
Col	37.3	33.0	33.5	30.3	40.2	29.7	69.3	83.0	49.9	37.3	68.1	13.9	13.3	18.0	8.5	24.8	13.6	28.2
Merchandising Office																		
500-699/888	45	51	101	55	251	53	261	46	37	355	147	48	135	424	289	210	439	2947
Row	1.5	1.7	3.4	1.9	8.5	1.8	8.9	1.6	1.3	12.0	5.0	1.6	4.6	14.4	9.8	7.1	14.9	100.0
Col	2.8	3.3	4.3	5.4	3.4	1.0	3.8	1.1	10.5	6.4	5.0	2.9	3.1	2.5	4.4	2.7	1.6	2.8
Utilities																		
600-654/656-659/																		
670-699	10	28	2	1	14	4	10	2	1	252	8	7	8	23	28	43	113	554
Row	1.8	5.1	0.4	0.2	2.5	0.7	1.8	0.4	0.2	45.5	1.4	1.3	1.4	4.2	5.1	7.8	20.4	100.0
Col	0.6	1.8	0.1	0.1	0.2	0.1	0.1	0.0	0.3	4.5	0.3	0.4	0.2	0.1	0.4	0.6	0.4	0.5
Industrial																		
700-799	22	117	38	10	80	14	157	28	10	11	107	20	73	158	613	151	227	1936
Row	1.1	6.0	2.0	0.5	4.1	0.7	8.1	1.4	0.5	5.7	5.5	1.0	3.8	8.2	31.7	7.8	11.7	100.0
Col	1.4	7.5	1.6	1.0	1.1	0.3	2.3	0.7	2.8	2.0	3.7	1.2	1.6	0.9	9.3	2.0	0.8	1.9
Storage																		
800-850/852-855/																		
857-879	373	253	134	117	323	784	533	58	12	387	128	107	208	1608	441	573	1204	7243
Row	5.1	3.5	1.9	1.6	4.5	10.8	7.4	0.8	0.2	5.3	1.8	1.5	2.9	22.2	6.1	7.9	16.6	100.0
Col	23.2	16.2	5.8	11.5	4.4	15.0	7.8	1.4	3.4	7.0	4.4	6.4	4.7	9.3	6.7	7.5	4.5	7.0
Construction																		
910-919	21	38	96	21	135	236	42	6	0	20	3	17	60	861	58	147	430	2191
Row	1.0	1.7	4.4	1.0	6.2	10.8	1.9	0.3	0.0	0.9	0.1	0.8	2.7	39.3	2.6	6.7	19.6	100.0
Col	1.3	2.4	4.1	2.1	1.8	4.5	0.6	0.1	0.0	0.4	0.1	1.0	1.4	5.0	0.9	1.9	1.6	2.1
Bridges, Highways																		
920-929	3	7	5	3	15	17	21	0	0	8	11	1	7	75	16	12	46	247
Row	1.2	2.8	2.0	1.2	6.1	6.9	8.5	0.0	0.0	3.2	4.5	0.4	2.8	30.4	6.5	4.9	18.6	100.0
Col	0.2	0.4	0.2	0.3	0.2	0.3	0.3	0.0	0.0	0.1	0.4	0.1	0.2	0.4	0.2	0.2	0.2	0.2
Special																		
900-909/930-999	306	323	594	265	2197	1872	333	60	46	943	182	432	1559	7461	2755	2565	6188	28081
Row	1.1	1.2	2.1	0.9	7.8	6.7	1.2	0.2	0.2	3.4	0.6	1.5	5.6	26.6	9.8	9.1	22.0	100.0
Col	19.0	20.7	25.5	26.1	29.6	35.8	4.9	1.4	13.1	17.0	6.2	25.7	35.2	43.3	41.6	33.5	23.0	27.1
Miscellaneous Other	185	187	417	183	1051	595	469	129	40	1157	228	760	1711	2907	1685	1829	13976	27509
Row	0.7	0.7	1.5	0.7	3.8	2.2	1.7	0.5	0.1	4.2	0.8	2.8	6.2	10.6	6.1	6.6	50.8	100.0
Col	11.5	12.0	17.9	18.0	14.2	11.4	6.8	3.1	11.4	20.8	7.8	45.3	38.7	16.9	25.5	23.9	51.9	26.5
Totals	1608	1564	2327	1016	7423	5231	6856	4222	351	5560	2929	1678	4425	17232	6619	7649	26940	103630
Row	1.6	1.5	2.2	1.0	7.2	5.0	6.6	4.1	0.3	5.4	2.8	1.6	4.3	16.6	6.4	7.4	26.0	100.0
Col	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 3-2 gives the distribution of fires, injuries, and fatalities by property class. The dominance of residential fires as a cause of injuries is noticeable. Although residential fires were only 28% of all fires, they accounted for over half of both the civilian and fire service injuries (53% and 53%, respectively), and for nearly three-fourths (74%) of the fatalities. The second most frequent property type was rather consistently storage facilities, being about 7% of the fires and 7% to 8% of the injuries. Again, "special properties" and "other" properties were quite frequent. This may indicate a need to revise the coding, or it may represent a tendency to report in general categories rather than to take the effort to determine the precise category.

Table 3-3 gives the distribution of fires by property type and cause, including row and column percentages. Row percentages are the most informative, since they indicate the distribution of causes by each property type. Note that the tables are based on 18 months of data, so that although the percentages may be fairly accurate, the frequencies represent more than annual estimates. From the table, one can see that (excluding unknown and miscellaneous categories) the most frequent scenario is a "special property" fire with an initial cause of open flames or sparks (7.2% of the cases). This is followed by several residential fires and associated causes: residential-heating (4.6%), residential-cooking (3.4%), residential-open flame or sparks (3.0%), and residential-smoking (2.9%), Other relatively frequent scenarios are special-property other-equipment (2.7%), special-property other-heat (2.5%), special-property smoking (2.1%), and residential-electrical distribution (2.0%). A more detailed table of property classifications by initial cause is presented in Table II-1 in Appendix II.

Attempts were made to relate the estimated property and contents losses in dollars to the property class and the initial cause of the fire. Unfortunately, in the 1975 MFIRS data tape that HSRI received from the Office of the State Fire Marshal, much of the needed information was missing. About 80% of fire incidents were missing

information on either the monetary losses from property loss, the property type, or the initial cause. Missing data on any of those three variables made the case unusable for estimating monetary cost by property type and cause. Similar missing data rates were encountered when dealing with the property loss of the contents of structures. In this category, about 63% of the cases were missing.

As a result of the large proportion of missing data, any totals or monetary amounts in losses would clearly be inaccurate and low. It is possible that the relative ordering of structure classes and/or causes by loss might be preserved, but even this seems suspect. The categories with the largest structure losses are: (1) Other and miscellaneous; (2) Codes 900-909 and 930-999, Special properties; and (3) Residential. Either the rather general gross property types are too large or they are being used as "catch-all" categories in the data reporting. Similarly, the "Remaining or unknown" cause category has the largest reported loss. This implies that the desired detail on losses by cause and structure type is not present.

Table 3-4 presents the distribution of the civilian injuries by gross structure and by cause of the fire. Reference to Tables 3-1 and 3-2 will give the marginal distributions of injuries by causes and by structures. Table 3-4 can be used to identify the most frequent causes of fires which result in injuries within each property class. As an example of this use of the table, the 17 priority causes within residential fires, and their associated percentages in terms of the residential civilian injuries are listed in order of frequency of injuries as the second column of Table 3-1. A number of interesting differences can be noted. For example, smoking is the cause of only 10% of the residential fires, but is the cause of 21% of the residential fires which result in civilian injuries. On the other hand, flammable liquids are the cause of 11% of the residential fires, but only of 2.4% of the residential fires which result in injuries. Table II-2 in Appendix II presents the civilian injuries by more detailed property classifications and causes.

Table 3-5 presents fatalities distributed by property type and initial cause. Note that the total of 327 fatalities represents about a year and a half of fire fatalities reported in the fire incident reporting system. Fatalities reported on the incident reports are about two-thirds of the fatalities. Fatality reports are later updated to include persons who were injured in a fire, but died later in a hospital. Such cases would be reported as injuries on the incident reports rather than as fatalities. The number of fatalities is so small that many of the property-class and initial-cause combinations are empty. Except for residential fires, the cause distributions have little if any meaning due to the small numbers involved. Table II-3 in Appendix II presents these fatalities with additional detail in the property classification. This table approaches that of a case-by-case listing. Only scenarios in which a fatality occurred are listed.

Table 3-6 gives the distribution of service injuries by property type and cause of the fire. Again residential fires are associated with the largest number of injuries to fire service personnel. However, within the class of residential fires, there is more uniformity of the injuries by cause than for civilian injuries. Generally the number of service injuries seems to follow the number of fires. One notable exception appears. That is, cooking causes 12% of the residential fires, but results in only 4% of the service injuries. Electrical and open-flame-caused fires result in slightly higher injuries to firemen than would be expected based on the number of fires. A more detailed table of fire service injuries, by property class, is given in Table II-4 in Appendix II.

Table 3-4 "Civilian Injuries by Structure Type and Cause of Fire"

	Exposure	Natural	Inc'y/Susp.	Explosives, Fireworks	Smoking	Children	Heating	Cooking	Air Cond. Refrigeration	Elec'l Dist'n	Appliances	Gas	Flammable Liquid	Open Flame, Spark	Other Equipment	Other Heat	Remaining	Totals
Assembly 000-199	0	0	0	1	1	0	5	16	0	2	0	1	1	1	1	0	4	33
Education 200-299	0	0	0	0	0	0	4	214	0	10	0	0	3	0	8	0	1	240
Institution 300-399	0	0	15	0	11	0	0	1	0	1	6	0	1	7	0	1	9	38
Residential 400-499	14	3	50	4	257	71	173	185	11	59	65	11	30	95	31	62	114	1235
Merchandising Office 500-599/888	0	0	1	0	11	0	24	1	0	11	2	3	4	0	12	2	0	71
Utilities 600-654/656-659/ 670-699	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	3
Industrial 700-799	1	8	2	0	0	2	1	0	2	4	5	1	0	0	25	1	22	74
Storage 800-850/852-855/ 857-879	6	5	0	0	6	1	4	2	2	8	1	6	9	2	16	6	1	75
Construction 910-919	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	1	2	6
Bridges,Highways 920-929	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2
Special 900-909/930-999	6	0	2	1	9	0	4	3	0	2	2	5	16	7	15	4	2	78
Misc./ Other	5	0	2	2	11	3	18	10	0	6	2	13	19	4	20	13	82	210
TOTALS	32	17	73	8	307	77	233	432	15	105	83	40	83	117	130	90	237	2065

Table 3-5 "Deaths by Structure Type and Cause of Fire"

	Exposure	Natural	Inc'y/Susp.	Explosives, Fireworks	Smoking	Children	Heating	Cooking	Air Cond. Refrigeration	Elec'l Dist'n	Appliances	Gas	Flammable Liquid	Open Flame, Spakr	Other Equpt.	Other Heat	Remaining	Totals
Assembly 000-199	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Education 200-299	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Institution 300-399	0	0	1	0	5	0	0	0	0	0	1	0	0	0	0	0	0	7
Residential 400-499	3	0	10	0	55	6	50	25	2	20	7	2	3	13	0	15	31	242
Merchandising Office 500-599/888	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	3
Utilities 600-654/656-659 670-699	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Industrial 700-799	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	4
Storage 800-850/852-855/ 875-879	0	0	1	0	1	0	1	0	0	1	0	1	0	0	6	0	5	16
Construction 910-919	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Bridges,Highways 920-929	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Special 900-909.930-999	0	0	1	2	0	0	0	0	0	0	0	2	14	1	1	3	7	31
Misc./Other	1	0	3	0	0	0	1	0	0	2	0	0	5	2	1	6	0	21
TOTALS	4	2	16	3	62	6	53	25	2	24	8	5	27	17	8	24	46	327

Table 3-6 "Fire Service Injuries by Structure Type and Cause of Fire"

	Exposure	Natural	Inc'y/Susp.	Explosives, Fireworks	Smoking	Children	Heating	Cooking	Air Cond. Refrigeration	Elec'l Dist'n	Appliances	Gas	Flammable Liquid	Open Flame, Spark	Other Equpt.	Other Heat	Remaining	Totals
Assembly 00-199	0	2	6	0	4	1	16	15	1	13	5	0	5	4	0	4	15	91
Education 200-299	0	0	0	1	2	2	3	0	0	3	2	5	0	10	0	12	7	47
Institution 300-399	0	0	0	0	5	0	7	0	0	1	2	0	0	3	0	2	0	20
Residential 400-499	79	18	47	14	142	81	206	59	17	143	50	21	46	188	30	109	206	1456
Merchandising Office 500-599/888	61	3	8	4	12	0	18	0	0	26	5	1	14	23	8	18	47	249
Utilities 600-654/656-659/ 670-699	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	5
Industrial 700-799	3	14	4	1	7	0	7	2	2	2	6	0	0	17	33	6	8	112
Storage 800-850/852-855/ 857-879	18	8	5	6	7	15	42	0	0	8	9	7	9	31	21	18	35	239
Construction 910-919	1	0	16	2	0	3	0	2	0	0	0	0	4	12	0	0	13	53
Bridges,Highways 920-929	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	3
Special 900-909/930-999	2	1	2	4	4	4	0	0	0	4	6	10	5	27	25	19	26	139
Misc./Other	10	9	3	6	7	4	6	2	1	13	1	3	16	32	7	22	175	317
TOTALS	174	59	91	38	190	110	305	80	21	213	86	47	100	348	125	210	534	2731

4. CONCEPTUAL PROBLEMS

There are a number of conceptual problems in attempting to form national estimates of the numbers of fire incidents, fire injuries and fatalities, and property losses due to fires in the U.S. The most critical problem is what operational definition of "fire" is to be employed. If a general definition such as "any uncontrolled burning or smoldering of a substance" is used, then difficulties will arise in obtaining data indicating the total number of such fires. The vast majority of such "fires" would not be included in any reporting system or survey presently in existence, or envisioned. Restriction of the definition of "fire" to those fires reported by a specified agency may ignore a large number of fires which should be included. For example, restriction to fires to which fire departments were summoned would exclude a number of small residential fires which might be reported to insurance companies as well as a number of forest, grass, or agricultural fires which might be controlled by a different agency, such as natural resources or parks authorities.

The most useful approach to defining "fire" seems to be to specify some threshold in terms of injury causation, or property loss as a minimum criteria for fires of interest to be reported. Although the small unreportable fires pose the potential of growing into large fires if appropriate action were not taken, there seems no way currently of accurately measuring such fires.

4.1 Non-reported Fires

Fire department data obviously include only those fires (or incidents) to which a fire department was summoned. Consequently, small fires such as kitchen fires which are suppressed without professional aid will not be reported in any data system in which fire

departments remain as the sole reporting source. Similarly, fires within a factoryor a large commercial facility which have their own fire suppression personnel and equipment, and which do not require the assistance of professional fire fighters, would not be reported. In addition, for example, in Michigan, some forest and/or grass fires are fought by equipment and persons in the Department of Natural Resources rather than what is recognized as a conventional fire department. Such fires will not be reported in the MFIRS. In 1976, responses by the Department of Natural Resources amounted to 1,341, with an estimated monetary loss of $802,765. Similar figures for 1975 were unavailable, but are presumably of the same relative magnitude. It was not clear whether all of these "responses" were fires. If they were, then they amounted to approximately 1.7% of the fires reported by fire departments.

It should be feasible and relatively simple to incorporate fires responded to by other agencies into the total fire data set for a state. However, to include small fires for which no professional assistance was summoned would be an almost impossible task. If fire department reports were supplemented and matched with insurance claim reports for all insurers, some progress might be made. This would presumably include all fires for which either a fire department was summoned, or for which a fire insurance claim was filed, but there might still be a number of fires with insufficient loss to justify a claim and for which no fire department was called. Since these are "small" fires, one may be justified in excluding them. Presumably, however, they could have developed into potentially dangerous fires if they had not been detected and controlled sufficiently early. Thus they may represent a potential danger which should somehow be included in the estimation of the total fire hazard.

The only feasible method of including such fires currently would seem to be to use a household survey to estimate the number of such fires. This has serious data quality difficulties such as recall problems. It seems likely at this time that only some general

statement can be made to the effect that the current estimates do not include a large number of potentially dangerous fires which were extinguished without professional aid, and hence were not reported through conventional fire organization channels.

The fact that fire rates estimated from insurance claims data (for one insurer) were larger than the fire rates estimated from the state fire incident data for Michigan suggests that fire department data may not include all fires. That is, one explanation of the larger estimated rates from insurance data would be non-reporting of some fires to fire departments. Another--equally valid with the present state of knowledge--is that this insurance company or insured property as a whole had a larger fire rate than the state in general.

The point remains that fire department data alone will somewhat underestimate the number and amount of property loss due to fires. At a minimum some study to determine the amount of non-coverage of fires by fire department records seems valuable. Depending on the results of such an investigation, a parallel system of reporting from insurance companies might be attempted.

A similar conceptual problem occurs with fire injuries, and to a lesser extent, with fire fatalities. Fire department data gathered at the scene of the fire cannot be expected to include all fatalities, since typically, some persons injured in a fire will die later of their injuries. Thus, at a minimum some follow-up system is necessary to determine the actual number of fatalities resulting from fire. In addition, there may be a small number of fatalities and/or injuries which are caused by fire (e.g., by clothing ignition from a stove) and do not involve a fire department or fire reporting agency. The extent to which such injuries due to fires occur and are not included in fire department records should be explored further to assess the scope of this problem. If this number is not neglible, then some parallel system is indicated for reporting of fire injuries and fatalities keyed to health care providers (e.g., hosptials).

In summary, it seems that some threshold definition for "fire" is needed, either in terms of property damage or injury. This definition threshold should be high enough so that fires above that threshold will almost certainly be recorded in fire department records or in insurance claims. Even to include all fires exceeding this threshold will likely require supplementing the fire-department (NFIRS)-based data with either insurance data or medical-care-provider data, or both. Such a supplemented NFIRS could be expected to provide reasonably complete data on fires above the defined "fire" threshold. There would still remain a large number of small fires which would not be covered under any reasonable reporting system. These fires could be considered as being "trivial" in consequence. However, each presumably may have had the potential to become a larger, dangerous fire had they not been detected and extinguished while small. Thus they represent a potential danger whose magnitude could not be estimated through the data system. The only feasible method of estimating the number of such small, but potentially dangerous fires would seem to be through a survey. Such a survey should concentrate on the immediate past--say the last two weeks--and should contain questions designed to jog the memory of the persons being surveyed.

4.2 Advantages and Disadvantages of NFIRS for forming National Estimates

There are several advantages to the state fire incident reporting systems which would together combine into the NFIRS. One major advantage is that state data would be collected in computerized form at a central location. This would facilitate its use in research to aid in understanding both the magnitude and nature of the national fire problem. Consistent data forms for all the states would aid in achieving consistent data, which would permit more accurate comparisons between states aimed at determining which areas had unusually high or low fire rates. Identification of unusually high or low fire rates

would indicate the need for an attempt to determine the reasons for any disproportionate rates. Once causes for good, or bad, fire rates are known and understood, appropriate countermeasures can be implemented to try to reduce fire rates in the selected areas indicated.

An additional advantage of the NFIRS system is that it will enable comparisons to be made over time. That is, if a large federal program to reduce fire rates is implemented, NFIRS data would indicate if this program is effective. If the NFIRS data did not indicate the effectiveness of a program designed to reduce fire rates, then the program which is not achieving its desired goals should be revised or abandoned. That is, if the activities intended to reduce fires are not effective, new activities should be implemented.

The fact that all fire departments would be expected to complete the NFIRS forms means that some fire departments may become more aware of the causes of fires in their area. In addition, the potential of self-evaluation using data generated by the NFIRS may result in improved local efforts at fire prevention or more rapid response to fire emergencies.

Finally, the state design of the NFIRS organization implies the potential for feedback to local fire departments. This could take several forms. For example, if a local fire department instituted a large-scale prevention and information campaign, the state could use the data already computerized to aid that local department in measuring the effect of its campaign. Similarly, local departments might for example learn from NFIRS data that they had a predominance of their fires in one particular property type, or from some unusual cause. Such information might enable them to institute countermeasures to reduce the problem.

Although there are a number of strengths and advantages to the NFIRS system, there are some weaknesses and drawbacks. Conceptually, as had been noted, the NFIRS cannot include all fires. To obtain a complete picture of the fire situation would require supplementing the

NFIRS data with data from other agencies (e.g., Natural Resources, Federal Aviation Authority, insurance agencies, medical providers, etc.). Household surveys might also be necessary to determine the unrealized potential danger from fires.

There are also conceptual difficulties in making national estimates through only partial participation of the states in providing data. Although national projections can be made from various subsets of the states, these are not as accurate and reliable as samples of smaller units drawn across all states. Thus, from the national data information point of view, the NFIRS is not the most efficient source of data. Nor is it the most timely.

Currently, there are data quality problems in the NFIRS. This is most typical of a new data system. With concerned and concentrated effort, these problems can be reduced, but they must be expected as start-up problems with each state as it joins the system. In the Michigan fire data, a wide acceptance and participation of the individual fire departments has been obtained. Participation is about 98%, which is near complete state cooperation.

There are, however, some difficulties with the completeness and accuracy of the data. A number of cases have occurred where data was inadvertently placed on the wrong line of the form. This results, for example, in the number of firemen at the scene being incorrectly listed as the number of fire service casualties. In addition, all of the data that follows the incorrect item is likely to be coded in the wrong place in the record, resulting in erroneous or missing data. Careful editing of the forms before they are computerized would reduce much of this problem, and there is evidence that the 1976 Michigan data will be much improved over the 1975 in this respect. Nevertheless, some problems remain.

A more difficult problem with the data is in the area of its original recording. Some items are very often blank, such as census tract. Something in the order of 15% of the fires do not have cause determined, whether due to difficulty in determining the cause or

failure to properly record the cause on the form. In addition, there seems to be a tendency to report in general terms rather than specific terms--to use the "other" category rather than determining and entering the correct cause or property type. Some of this is due to the form being filled out after the fire and perhaps the data not being readily available. Another portion is likely due to failure to look up the specific code for designation to enter on the form.

These reporting problems are more difficult to correct than problems which can be corrected with careful editing. To correct basic data inaccuracies or incompleteness requires motivation and increased cooperation from the firefighters who complete the forms. It cannot be handled centrally, but needs to be improved in each department. Since most fire personnel do no particularly like to fill out forms and may regard than as an extra burden, it is a continual challenge to maintain data quality.

A uniform data form used by all states has been mentioned as a strength of the system. It should not, however, be thought of as unchangeable. It is likely that difficulties, or needs for different data, may make modifications of the form advisable. A change in the format of the form to reduce the possibility of entering the number of fire fighters used in the incident as the number of casualties would seem useful, for example. Other design changes might make the form easier to complete and reduce the change of error or incompleteness. The census tract variable is of questionable use, since it is only applicable in metropolitan areas and even there it is quite difficult for a fire department to determine.

There will likely be a need for more detailed information than can be reasonably be gotten from the NFIRS. For example, there has been a recent trend toward increased use of smoke detectors in residences. One reasonable question that might be asked of NFIRS data is whether this increased use of smoke detectors has resulted in any reduction in the incidence of fires, the amount of property damage from fires, or the injuries from fires. However, with the current data form, such a

question could not be answered because the data do not include information about the presence of smoke detectors. Similar questions of interest will arise in the future. Such detailed data cannot be collected from the entire NFIRS, but it would be useful to incorporate into the NFIRS a mechanism to answer such questions on a timely basis.

5.0 RECOMMENDATIONS--SAMPLE DESIGNS TO SUPPLEMENT NFIRS

5.1 Sampling to Parallel and Augment the NFIRS

As has been noted, coverage by the NFIRS is expected to be less than complete. In addition, the implementation of NFIRS is phased so that different numbers of states will be participating at different times. That is, participation is planned to increase gradually until all states participate. Further, current experience has shown that each state will experience start-up problems with the data collection and management. Thus, the data from any given state in the first year or two of its participation tend to be rather unreliable. As a result, the data from the NFIRS do not provide very precise or reliable estimates of the national fire experience. Nor will they provide a complete picture in the immediate future. Finally, even after the NFIRS is completely operational, careful monitoring will be necessary to ensure that the data quality from each state remains at an acceptably high level.

All of these considerations suggest that a supplementary or parallel sampling system would be useful to provide better national estimates sooner than they will be available from the NFIRS, to fill in some of the coverage gaps that the NFIRS will have, and eventually to serve as a validation or crosscheck on the data from the NFIRS.

Two general types of sampling seem indicated. The first type would be to supplement the NFIRS data to complete the coverage of fires and fire injuries. This might consist of a sample of emergency rooms and/or hospital records to identify fire injuries which were missed by the NFIRS. To get information on fires which are below the threshold which results in a fire department call, and are thus excluded from the NFIRS, the most promising type of sample would

seem to be a household survey. This could probably most efficiently be added to existing surveys. A second type of sampling would be aimed at providing better interim estimates until the NFIRS is fully operational and at providing a parallel validation or cross-check on the NFIRS data. Several types of samples could be utilized for this, depending on the level of effort and the degree of detail desired. These sample plans are discussed in the subsequent sections. An additional advantage of a parallel sampling system is that it could be used to obtain more detailed data than can be gotten from the NFIRS. It also could be used to investigate problems of special interest which may vary from time to time.

5.2 Types of Sampling

The nature of the sampling plan will depend largely on two items: (1) what data are desired, and (2) what resources are to be expended on sampling. Three possible approaches are detailed below. The three types are generally in increasing order of quality and detail of data that may be obtained, but also in increasing magnitude of cost. It is possible, of course, to implement a smaller scale of in-depth investigative sampling for about the same cost as a larger-scale sample of fire department records.

5.2.1 Approach One: Using Existing Fire Department Personnel

This approach would be similar in design to the Pedestrian/Bicyclist Accident Data Sampling and Analysis Program (PADSAP) developed and implemented by the National Highway Traffic Safety Administration (NHTSA). The PADSAP design involved assigning each of the 12,675 police agencies in the nation to one of 45 strata based on population size, region, and type. From these strata 348 police agencies were selected by a known probability method, and they were then asked to fill out supplemental accident report forms on pedestrian and bicycle accidents which took place within their

jurisdictions. For larger jurisdictions a system of subsampling only some of the pedestrian and bicycle accidents was planned.

A similar approach to creating a sample of fire departments could be used. There are about 39,000 units of local government in the United States, and most of these bodies have their own fire departments. Thus this approach would involve first listing these thousands of fire departments in defined regional and population size strata. The total number of fire departments to be selected would obviously depend on the amount of money available for the sampling program, but there should be at least 200 selected departments. The number of selections to be made in each stratum would be determined by a controlled selection computer program for allocating selections among strata on a probability basis in relation to the total strata populations. These selections within each stratum would be made by a random procedure utilizing the populations of all the fire departments within the given stratum. The selected fire departments would then be contacted and asked to complete a special FIRSS form on fires (and other fire department activities also, if desired) in their jurisdictions. Small jurisdictions would be asked to complete this form for all fires; in large jurisdictions some kind of subsampling system would be used, such as reporting fires from certain districts or which take place on certain days of the week or which have certain final digits on the record identification number.

In this approach it would probably be necessary to pay the fire departments for the extra work of completing the special FIRSS form, perhaps $10 or $20 per reported fire. Nevertheless, if the necessary cooperation could be obtained from the selected fire departments, this approach could be a relatively cheap way to obtain fairly detailed information on a reliable national sample of perhaps 30,000 fires.

5.2.2 Approach Two: Using Field Data Collectors

This approach would involve the employment of permanent, locally based field data collectors to visit or telephone the selected fire

departments in order to complete special fire incident reporting forms. To make this approach viable economically, it would be necessary for the selected fire departments to be clustered geographically into sample areas of at least 75,000 total population. Based on NFPA estimates, such an area would be expected to have about 1000 fire incidents annually, so a single field data collector located in this area would be completing forms on 4-5 fires per working day. Thus the employment of field data collectors in a sample of 30-40 sample areas would be expected to produce fairly detailed information on a reliable national sample of 30,000 to 40,000 fires.

Because of the necessity of geographic clustering in this approach, a different sample design would be used in order to select the participating fire departments in two stages. In the first stage a sample of geographic areas would be selected. This would involve dividing the nation up into geographic clusters of governmental units with a minimum population of 75,000. The majority of these clusters would be cities, parts of counties, or whole counties, but in rural areas it would often be necessary to cluster together many counties into one potential sample area in order to meet the minimum population criterion. These potential sample areas would be assigned to defined regional and population size strata; the total number of sample areas to be selected would be decided by the NFPCA staff; the number of sample areas to be selected from each stratum would be determined by the controlled selection computer program in relation to the total populations of the strata; and particular sample areas within a stratum would be chosen by a random procedure using the populations of all the potential sample areas in the stratum.

The second stage would involve selecting all of the fire departments in selected sample areas close to the minimum population size. In larger sample areas with many fire departments a further subsampling procedure would be necessary to select a number of fire departments likely to have about 1000 fires per year. Similarly in large cities a subsampling procedure using geographic districts, or days of

the week, or fire record identification numbers, or some other method, would be used in the second stage to reduce the number of investigations to be completed to a manageable one-person task.

In selected sample areas containing selected large professional fire departments the field data collector would probably travel to the department offices every day to transcribe the required fire incident information to the reporting forms from the existing records and to obtain any further information needed from fire department personnel. However, for small volunteer departments the data collector could probably do most of the work by telephone, perhaps just calling once a week to find out if the department had had any fires during the previous week and if so obtaining the necessary information for the data form over the telephone.

5.2.3 Approach Three: Using Field Fire Investigators

This approach would involve a multi-stage clustered sample design very similar to that of Approach Two. However, this approach would use field personnel not just to collect fire incident information from the records and personnel of the selected fire departments but also to conduct more detailed investigations of the sample fires by visiting the fire site, talking to victims, etc. These data investigators would need to be more highly trained than the data collectors in Approach Two, and it is obvious that a single data investigator could not cover nearly as many fires as could a data collector. To obtain this type of in-depth reports on 1000 fire incidents a sample area might require a team of four or five investigators, or some kind of combined team of a data collector and one or more investigators might be used to obtain general information on all of the fires, with in-depth information on a subsample of the fires in the area. Such subsample could be drawn in such a way as to give a greater probability of selection to the more serious injury or property damage fires which are of greates interest.

5.2.4 Discussion

Approach One would probably be the cheapest sample design to implement, and it also has the statistical advantage of dispersing the sample of fire departments more geographically than would Approaches Two and Three. The chief concern with Approach One is the problem of non-cooperation and incomplete data if the program were to rely entirely on existing fire department personnel. Even with the promise of payment for each completed form, one cannot expect all selected departments to be willing to participate or all participating departments to fill out the form completely on every eligible fire. Approach One would be particularly difficult to implement in fire departments requiring some kind of subsampling procedure for the selection of the fire incidents to be reported. Thus, while Approach Two would be expected to cost more than Approach One, it would also be expected to provide more complete and accurate data and thus more reliable estimates of the national fire experience. Of course, Approach Two would require some monitoring procedures to see that the field data collector is doing his job correctly, but supervision and control of data quality would be much easier in this approach than in Approach One.

Approach Three would obviously cost a great deal more than the other approaches, especially if detailed in-depth investigations were carried out on the entire national sample of fire incidents. However, by not just relying on information from fire department records and personnel, it would be expected to provide a uniformly higher quality of fire incident information. This higher quality of the resulting data might be considered sufficient to justify the higher costs of the Approach Three investigations, at least for a subsample of the national sample of fire incidents.

Table 5-1 gives estimates of the sampling errors and precision to be expected from the National Accident Sampling System (NASS) for various sample sizes, and different design effects. The estimates are

presented for estimates of proportions, such as the percent of fires which are residential or the percent of residential fires which are related to the heating system of the residence. Although the design effect cannot be calculated precisely before the sample is constructed and carried out, the design effect (DEFF)[1] is expected to be on the order of 4.0 or less for the NASS, and possibly about 1.5 for the PADSAP program. The PADSAP program costs something on the order of two million dollars per year, while the NASS, which conducts in-depth investigations, is projected to cost about six million dollars per year for about 32 primary sampling units, each conducting about 500 in-depth investigations per year. Presumably sample plans for fire investigations would be on the same order of costs.

In summary, a sampling system is recommended for several reasons. First, it would provide reliable national estimates of fire incidents, injuries, and losses on a more timely basis than can the NFIRS because the sample could be implemented more rapidly. Second, a sample could be used to obtain more flexible data. That is, special questions could be posed and the appropriate data collected from a sample much more easily than from the entire NFIRS. Third, a parallel system could be used to supplement data from NFIRS and to add data which cannot be collected on the census scale. Examples of this are follow-up studies to ascertain the seriousness of injuries suffered in fires, to determine the completeness of injury reports, the actual amounts of property damage sustained in fires, etc. Fourth, the sampling system can be used to validate the data received from the NFIRS to ensure that it is

[1] The design effect (DEFF) summarizes the effects of various complexities of sample design, especially those of clustering and stratification. The DEFF is the ratio of the actual variance of a sample to the variance of a simple random sample of the same number of elements. The design effects in complex samples are almost always greater than one. However, it is possible in simple cases for a design effect of less than one to be achieved. The term is defined and its uses illustrated in Kish (Survey Sampling by Leslie Kish, John Wiley and Sons, 1965, p. 258ff.)

complete or to point out areas of difficulty. It should be mentioned, of course, that the concept of a sample implies that more effort at data quality control, management, and close liaison with the local departments in the sample is assumed than is possible with the NFIRS. That is, the advantage of a sample stems from the fact that it is a relatively small number of fire departments. As a result, it is possible for the sampling agency to expend a large amount of effort with each department. There is no advantage of a sample if only the same level of effort were to be expended on the sample units as on departments in the NFIRS. By concentrating limited resources for data collection on a few units in the sample, higher data quality and more detailed reporting may be obtained than from less effort on more units.

TABLE 5-1

SAMPLING ERROR AND PRECISION* OF NASS ESTIMATES FOR FIVE DIFFERENT PERCENTAGES WITH FIVE SAMPLE SIZES AND THREE DIFFERENT DESIGN EFFECTS

		Sampling Design Effect					
		DEFF = 1.0		DEFF = 4.0		DEFF = 9.0	
Sample Size	Sample Estimate	Sampling Error	Precision (in %)	Sampling Error	Precision (in %)	Sampling Error	Precision (in %)
1,000	1%	0.63	63.0	1.26	126.0	1.89	189.0
	5%	1.38	27.6	2.76	55.2	4.14	82.8
	10%	1.90	19.0	3.80	38.0	57.0	57.0
	25%	2.74	11.0	5.48	21.9	8.22	32.9
	50%	3.16	6.3	6.32	12.6	9.48	19.0
4,000	1%	0.31	31.0	0.63	63.0	0.94	94.0
	5%	0.69	13.8	1.38	27.6	2.07	41.4
	10%	0.95	9.5	1.94	19.4	2.89	28.9
	25%	1.37	5.5	2.74	11.0	4.11	16.4
	50%	1.58	3.2	3.16	6.3	4.74	9.5
8,000	1%	0.22	22.0	0.44	44.0	0.66	66.0
	5%	0.49	9.8	0.97	19.4	1.46	29.2
	10%	0.67	6.7	1.34	13.4	2.01	20.1
	25%	0.97	3.9	1.94	7.8	2.91	11.6
	50%	1.12	2.2	2.24	4.5	3.36	6.7
16,000	1%	0.16	16.0	0.31	31.0	0.47	47.0
	5%	0.34	6.8	0.68	13.6	1.02	20.4
	10%	0.47	4.7	0.95	9.5	1.42	14.2
	25%	0.68	2.7	1.37	5.5	2.05	8.2
	50%	0.79	1.6	1.58	3.2	2.37	4.7
24,000	1%	0.13	13.0	0.26	26.0	0.39	39.0
	5%	0.28	5.6	0.56	11.2	0.84	16.8
	10%	0.34	3.4	0.77	7.7	1.11	11.1
	25%	0.56	2.2	1.12	4.5	1.68	6.7
	50%	0.65	1.3	1.29	2.6	1.94	3.9

* The design effect is defined in the text.

APPENDIX I

STATISTICAL METHODOLOGY

APPENDIX I: STATISTICAL METHODOLOGY

Combining Data from Different Studies

If several sources report the same estimates, there is generally no problem. The common estimate is taken as acceptable. This is generally true so long as the several estimates are within plus or minus one or two standard deviations of each other. That is, they agree to within the sampling precision of the estimates. In this case, it is sufficient to check that the estimates have been made in a valid manner. It may be that although several estimates agree satisfactorily, that the reported errors are unacceptably large. In this case more data--a larger sample--must be collected. Typically this would be done in the same manner as used to form the previous estimates.

It is unfortunately often the case that several reported estimates of a phenomenon--number of deaths from fires, for example--differ by far more than could be due to sampling precision. In this situation it is difficult to determine the best estimate. Careful evaluation of each estimate is required before a consensus can be reached. This process is sometimes more art than science, but the following may serve as a guide or outline. The original estimates and their disparities should also be reported along with a warning that the consensus estimate may be unreliable.

Determine the exact definition of the population on which each estimate is based. It is frequently true that there are different thresholds defined. This is often the case if the phenomenon in question is an accident or an injury. Some sources may report all fires, some all fire department calls, some only fire in the property damage in excess of $10,000, etc. If differences in threshold level can be identified, then further comparisons would be within estimates

based on the same threshold. Also, it may be possible to state the estimates separately by level--e.g., so many fires involving fatalities, so many involving injury, etc.

Check the sampling on data collection procedures to ensure that the population actually sampled is the same as the target population and that these populations are the same in the different studies. Thus, samples to estimate the number of household fires using exactly the same data elements and forms could reach quite different populations and quite different conclusions if they were based on a telephone interview survey, a household interview survey, and a mail survey.

If the same populations have been reached and the same variables and definitions used, but results still differ by more than can be explained by sampling errors and missing data rates, then look for unsuspected variables which may be different in the different studies. These could be intervening variables such as time or a public safety program, or they might be inherent variables such as type of construction, different prevalences of types of heating fuel, or different weather conditions during the period during the sampling. If candidate variables which may explain the differences can be identified, hypothesis about the relation of these new variables to the phenomenon would be formulated and tested. From the "pure" point of view, these new hypotheses should be tested with new data. From a more practical point of view, the observed relationships would be investigated to the extent possible with existing data. It should be pointed out that this has been done and that any such post hoc relationships need to be verified in future work, but they may be advanced as tentative explanations. It may turn out, of course, that the data required to develop explanatory relationships with the new variables are not present in the existing studies. In this case, its explanatory power can only be conjecture.

Once the differences in results have been determined and explained to the extent possible, there still remains the desire to combine the

results into a common or consensus estimate. Some methods that have been used are:

(1) "Vote". Each of several experts who have reviewed the studies votes on the most appropriate estimate.

(2) "Count". The combined estimate is taken as the one most frequently reported. This is essentially taking each separate estimate as a data point and using the mode to represent the group.

(3) "Pick a Favorite". One estimate is selected as the best on the basis of data base quality, care of presentation, author's reputation, or other factors.

(4) "Pool". If the data on which the estimate are based are available, they may be pooled and re-analyzed to yield a pooled estimate.

(5) "Bayesian". The estimates themselves are each given a weight which reflects a judgment about their precision. The estimates are then combined using a weighted average.

Each of these methods can be appropriate under proper circumstances. Each also has potentially serious drawbacks. The first three represent selection of an estimate based on the judgments of several reviewers. The result will depend both on the quality of the original set of estimates and on the ability of the reviewers to select a good estimate. The fifth method depends on the ability of the reviewer to formulate appropriate weights based on the precision. To the extent possible the precision can be measured by the mean square error (variance plus bias^2). The subjectivity may come in in estimating the bias. The weight may also be adjusted to reflect recency of the data. That is to give less weight to studies done some time ago and more weight to more current studies. The fourth method-pooling the data and reanalyzing--is fraught with pitfalls and is generally best avoided. It requires not only the actual data from the several studies, but also assurance that sampling methods, data collection methods, and definition of variables were the same.

Further, the resulting combined sample must represent the target population appropriately. This is unlikely to be the case. In general stronger influences can be drawn from comparisons of results of separate studies including their discrepancies than from lumping all the data together and ignoring differences.

Generally the most widely applicable method is the fourth listed. That is, combining the individual results with each weighted according to its precision. In the case of categorical data, particularly for rates on dichotomies, this is known as the Mantel-Haenzel procedure and may be found in Fleiss.* In general, the weights attached to each individual result may be objective, subjective, or a combination. If the individual results are all from similar studies and the sampling variances are available, these would be used. The weights would be entirely proportional to the sampling variances. If the studies are of different types or if sampling variances cannot be determined, then expert judgment might have to be used in weighting the individual results. A combination may be the most appropriate. The method recognizes that there may be information about the precision and reliability of results which is not in the form of an estimated variance but allows this information to be incorporated

Estimation from Model Fitting

Data of sufficient detail and quality to address many of the relevant questions may be available only for relatively small and non-representative portions of the United States. This being the case, one may attempt to determine predictive models which predict the five rates (by type of structure, amount of loss, cause, etc.) as a function of other variables which are available for the nation as a whole. The variables which can be used for prediction are basically demographic, or population based. Such variables are available for the United States from the census, and are recorded for various levels

*Fleiss, J. Statistical Methods for Rates and Proportions. Wiley, 1974.

of aggregation such as census tract or enumeration district, county, etc. If a model can be determined which predicts fire rates accurately for those areas where adequate fire data are available, then one may form national estimates by applying the model to the national census data. In effect, some combination of demographic and geographical variables is used as a surrogate for the fire rate variables.

Selection of Variables

One of the key steps in this method of forming national estimates is the selection of variables. This task is often complicated by the fact that demographic variables may be highly intercorrelated. In addition to leading to non-unique sets of predictor variables, this multi-collinearity leads to difficulty in estimating the parameters of the model and to unacceptably large variances for the estimated parameters and consequently also for the predicted values.

Mason, Genst, and Webster* give five suggested techniques for determining whether multicollinearity is a serious problem in a given set of data for determining the degree of the problem. If multi-collinearity is a problem, there appear to be three essentially different methods of dealing with it. The first and most straightforward is to eliminate independent or predictor variables to remove the multicollinearity. The second is ridge regression which introduces a (hopefully) small amount of bias into the estimation to remove the inter-dependencies of the predictor variables and achieve a large reduction in variance. The third may be referred to as factor analysis, principle components or latent root regression analysis. It attempts to define a new set of orthogonal (i.e., independent) factors or predictor variables from the original set.

*Mason, R.L., Genst, R.F., and Webster, J.T. "Regression Analysis and Problems of Multicollinearity", Comm. in Statistics, 4(B): 277-292.

The first method--eliminating predictor variables--has both practical and theoretical difficulties. It is not always clear which variables should be eliminated. Models fit for different cities, for examples, might indicate that different variables should be retained. Yet if all the best variables from several different models are retained to try to fit a model to a combined set of data, the problem of multicollinearity may reappear. Another difficulty is that this method may result in eliminating all but a few variables, which do not provide a model with sufficient degrees of freedom to achieve a satisfactory fit. This approach does have the advantage of being intuitive and of using variables in the original set of predictor variables. Many times these variables are interpretable and may suggest methods to control or reduce fire rates if they have a causal connection as well as a simple association or predictive capability.

The second technique, Ridge Regression, was first proposed by A. E. Hoerl in 1962 and finally in its present form by Mr. Hoerl and R. Q. Kennard in 1970. Ridge Regression attempts to define a new point estimate of the coefficients that has smaller variance than the usual estimator (Ordinary Least Squares), and thus yields a more precise estimate. This is not accomplished, however, without some loss. This new estimator is now biased. Unbiasedness is usually a desirable property, but in the present situation where we wish to predict and extrapolate from our model, it is not necessary. Thus the properties of the Ridge estimator seem to coincide with the objectives of the study.

There are certain limitations in the application of Ridge Regression. The first is the relative uncertainty of the so-called Ridge parameter. A debate rages as to whether the parameter is a constant or a random variable. The general concensus is that it is a random variable and hence must be estimated from the data. There exist a number of techniques for choosing the parameter and the reader is referred to the Horking article for a complete listing of these techniques. Each has its advantages and disadvantages as compared to each other; there is no clear choice.

The second limitation is the problem of implementation. After the decision to use Ridge Regression has been made, the problem lies in getting a computer program to accomplish this end. Serious thought must be given to the greater amount of work and advantages of Ridge Regression versus the ease and lesser advantages of other techniques.

The third general method of eliminating the problem of multicollinearity is variously referred to as factor analysis, principle component regression, latent root regression, and others. Basically the technique is to define a new set of variables (factors or components) from the original set of predictor variables in such a way that the new set will consist of orthogonal (independent) variables. Typically - in situations where this technique proves useful--the number of variables will also be reduced by eliminating those in the new set which have the lowest association with the dependent variable(s).

In the principle components formulation, the covariance matrix of the predictor variables is used. The principle components are the eigen vectors of this covariance matrix, and the corresponding eigen root indicate how much of the variation each eigen vector accounts for. These eign roots can thus be used to eliminate those vectors which account for little of the variation. The new variables are linear combinations of the original ones. They are independent, so that the usual regression or linear model techniques may be used with ease. However, the new variables may have little meaning or interpretation. If the sole aim of the regression is prediction, this lack of interpretation is not a draw back.

In the latent root formulation the correlation matrix of the original predictor variables is used rather than the covariance matrix. Thus this is a scaled version. Again the new variables are linear combinations of the previous ones and, while they are independent and may predict well with fewer variables, the same problems of interpretation arise.

Factor analysis may be thought of as a series of transformations applied to the new variables defined by principle components to

achieve interpretability. In factor analysis, one assumed that each of the original predictor variables is a linear combination of independent factors that are unknown. The technique provides a method of identifying these factors. The outcome of the analysis is a matrix with the variables as rows and factors as columns. The cell entries are the correlation coefficients of the variables with the factors. Since the factors are undefined, they usually are identified by the set of variables which have high correlation with that particular factor but very low correlations with other factors. In essence, factor analysis separates the variables into independent sets; within each set, the variables are highly correlated.

Other techniques can be used to identify the factors. These consist of defining a different measure of "distance" between groups of variables, such as the Minkowski distance. Variables close together in terms of this distance are grouped and identified with some factor. Thus the "distance" between variables thus grouped is small, while the "distance" between groups is relatively large. In this sense, factor analysis verges into cluster analysis, with the clusters identifying the factors.

Applications to Cities Residential Fire Rates

Previous studies have reported some degree of success in finding regression models which predicted fire rates among census tracts within cities. However the authors also reported somewhat discouraging results when the modeling technique was used to try to predict fire rates between cities (over different years). This section will suggest some possible reasons for this along with possible methods to improve prediction of the inter-city fire rates, as well as methods to validate the models suggested from inter-city comparisons.

The models previously developed should be tried on the fire data for the new cities. The authors understand that NFPA is planning to do this. This will give an estimate of how much predictive accuracy is lost in applying the model to new sets of data. This is crucial

to an understanding of what degree of accuracy may be expected of national estimates derived by using census data together with regression models to obtain predictions for the United States. The measure of accuracy obtained from a regression--such as R^2--are generally overly optimistic from predictions or explanations based on different data sets.

One possible reason for the disappointing predictions for the inter-city models is that the data did not satisfy the assumptions required for ordinary (least squares) regression. Indeed, if the usual regression assumptions are valid for the models predicting fire rates from census variables on the census tract level, the assumptions cannot hold for data aggregated to the city level. Since the regression model was fairly successful in that it explained about 60% of the variation in fire rates among census tracts for data from fire communities, there is some evidence that the regression assumptions are reasonably valid and that the communities shared common coefficients. (Methods for testing for common models and for possibly improving prediction are discussed below in the next section.)

Thus the model suggested can be represented as

$$f_{ij} = u + \sum_i b_t x_{ijt} + e_{ij} , \qquad (1)$$

where e_{ij} are indepedently normally distributed with mean 0 and variance σ^2. The f_{ij} notation represents the number of fires per 1000 persons (or households) for community i tract j; x_{ijt} denotes the demographic variable t of census tract j, community i; u and b_t are parameters.

Based on equation (1), one can see that the fire rate for the community i, f_i, is normally distributed with a variance which is highly related to the population variation between census tracts. The less the population variation among census tracts, the more stable the fire rate of the community. This can be seen as follows.

Let n_i denote the total population of community i and n_{ij} be the total population of tract j of community i. Then from equation (1), we have

$$f_i = \sum_j f_{ij}$$

$$= u + \Sigma_t b_t(\Sigma_j n_{ij}\, x_{ijt}/n_i) + \Sigma_j n_{ij}\, e_{ij}/n_i \ . \qquad (2)$$

Since the e_{ij} were independently normally distributed with mean 0 and variance σ^2, f_i, is normally distributed with mean

$$u + \Sigma_t b_t(\Sigma_j n_{ij}\, x_{ijt}/n_i) \ , \qquad (3)$$

and variance $\sigma^2 \Sigma_j n_{ij}^2/n_i^2$.

Note that the variance now depends on the variation of population among census tracts. Consequently, the error terms in the inter-city models do not have constant variance. This implies that ordinary least squares regression is inappropriate. Weighted least squares should be used instead, with the weights the inverse of the variances. Of course it might turn out that the quantities $\Sigma_j n_{ij}^2/n_i$ are very nearly equal, in which case the weighted least squares would not change the results much.

This result--that if the assumptions for least squares regression hold for one degree of aggregation of the data, they will be violated for other degrees of aggregation--has implications for applying this technique to get national estimates. Namely, it must be determined which level of aggregation most appropriately meets the assumptions. If this is not the level of aggregation at which the data are to be used, the appropriate modifications in the techniques (use of weights) should be made. At the present it seems likely that the census tract will be the most feasible level of aggregation to use. Census tract is a variable in the NFIRS, and the census data are readily available. There is some question about the appropriateness of the dependent

variable at that level, since it has relatively few values in most cases. Block data would have definite problems in the definition of the dependent variable, as well as posing a more formidable modeling problem in terms of data processing. It would also be a large task to obtain the block identifications from the NFIRS data. Further investigation is small settings of the use of block data is called for, however. Conceivably it might be necessary to aggregate to a city or county level to use the NFIRS data initially. That is, the census tract data may suffer from a large missing data rate in the early stages of the program. In this case it seems likely that weighted least quares may be more appropriate than ordinary least squares.

Models to Improve Inter-City Estimates

In addition to the possible need for using weighted least squares mentioned above, two other possibilities may contribute to low inter-city fire rate prediction. The different communities may not have the same parameters in the regressions or they may have the coefficients, but differ in a constant term. The following procedure may be used to test whether a set of cities have a common set of regression parameters. (This may be referred to as the hypothesis of hyperparallelism.)

Suppose we have a set of variables $x_{ij1},\ldots,x_{ijk}$ for aggregation unit j within city i. These variables are assumed to explain the dependent variable f_{ij}, with variance σ^2, as

$$Ef_{ij} = \sum_{k=0}^{K} a_{ik} x_{ijk}$$

where $x_{ij0} = 1$. This can be rewritten as

$$Ez_m = \sum_{n=1}^{N} b_n w_{mn}$$

where $z_1,\ldots,z_{M0} = y_{11},\ldots,y_{1J},\ldots,y_{I1},\ldots,y_{IJ}$,

$$[b_1,\ldots,b_N] = B' = [a_{10},\ldots,a_{I0},\ldots,a_{1K},\ldots,a_{IK}] \quad ,$$

and w_{mn} is either x_{ijk} for some i, j, and k, or else 0.

The hypothesis H_o: $a = a_{km}$ for a certain k, j, k, and m is equivalent to the hypothesis that the contrast CB = 0, there c is in the positions corresponding to a_{ij} and a_{km}. Several such hypotheses can be tested simultaneously by putting 1's and -1's in the rows of a matrix, being careful to maintain full row rank. For example H_o: $a_{ij} = a_{1k}$ for all i and k = 1, ..., K becomes H_o: CB = 0 where the (i,k)-th row of C has 1 in the a_{ik} position, -1 in the a_{1k} position, and zeros elsewhere. Of course C has no row corresponding to i = 1 for any k. Data can be analyzed with respect to hypotheses such as these using a general linear hypothesis program. Usually, hypotheses of hyperparallelism are restricted to the equality of coefficients of variables other than the X_{ijo}. The case where the coefficients of the X_{ijo} are different corresponds to different intercepts or levels of the different cities and is discussed later.

The hypothesis that the predictor variables $x_{ij1},\ldots,x_{ijK}$ have the same effects for every city i is the hypothesis given above, H_o: $a_{ij} = a_{1k}$ for all k = 0 and all i. If this hypothesis is rejected, we may want to test subhypotheses like H_o: $a_{ik} = a_{1k}$ for some specific k and all i. A stepwise elimination procedure can be formulated as follows:

1) For each $k \neq 0$, test H_o: $a_{ik} = a_{1k}$ for all i.
2) If the least significant F-ratio in step 1 is greater than a specified quantile of the corresponding F-distribution, step. Otherwise go on to step 3.
3) Restrict the linear model by requiring $a_{ik} = a_{1k}$ for all i, where k corresponds to the least significant F-ratio in the previous step. Remove this k from consideration in step 1 and go back to step 1.

The denominator sum of squares for the F-ratios is the residual sum of squares from the full model. It has d.f.

$$M - \text{rank}[w_{ij}] = \Sigma J_i - \Sigma \text{ rank } X_i$$

$$\text{where } X_i = \begin{matrix} x_{i10} & \cdots & x_{i1K} \\ \vdots & & \vdots \\ x_{iJ0} & \cdots & x_{iJK} \end{matrix}$$

All the X_is have full column rank if an only if $[w_{ij}]$ has full column rank. In the absence of multicollinearity, there are

$$M - N = \Sigma J_i - I(K+1)$$

denominator d.f. The numerator of each F-ratio is the reduction in sum of squares divided by its d.f., which is I-1 in the absence of multicollinearity.

Such a stepwise procedure can easily involve many hypothesis tests, even with a small value of K. This problem is present in the statistical analysis of all but the simplest of data sets. In effect we are using the significance levels of the F-ratios as measures of concordance of the data with the hypotheses, rather than as true significance levels.

If the value of K is very small, we may want to fit the model with $a_{ik} = a_{1k}$ for all i, for k in each subset (1,...,K). This procedure, of course, does not help with the problem of multiple tests of hypotheses.

In the event that cities appear to satisfy the hypothesis of hyparallelism (to have the same regression coefficients for a common set of predictor variables), there might still be a significant city effect. This would result in a model of the type

$$Ef_{ij} = \mu + C_i + \sum_K b_K X_{ijK} \quad ,$$

where the C_i represent different levels of fire rates among the cities. If the C_i are different from zero, this might be due to different reporting practices among the cities. For example, one city might not report fires

involving no loss (trash, grass, etc.) or fires with loss below a minimum figure. If reporting is related to severity, if that fact can be determined, and if the data include information on severity, thus a model incorporating level of severity could be developed. Such a model would use data from all cities reporting at a given threshold level of severity of fire.

In the case where city parameters are significantly different from 0, another variable which represents the severity of the reported fires could be added to the model if severity is also reported. The severity might explain much of between city differences. Suppose as an example, that three levels of severity are found. In the model, there will be three different dependent variables; severe fire rate, moderate fire rate, and minor fire rate. Cities having all three different severities of fires reported will have three different dependent variables for each of their census tracts. Cities with only two levels of fires reported--presumably severe and moderate, will have two dependent variables for each of their census tracts, while cities with only the severe fires reported will have only one dependent variable for each of their census tracts. Thus the models are:

$$Ef_{j(i)} = a + u_1 + C_i + \Sigma b_k \cdot X_{j(i)K}$$

$$Ef'_{j(i)} = a + u_2 + C_i + \Sigma b_K \cdot X_{j(i)K}$$

$$Ef''_{j(i)} = a + u_3 + C_i + \Sigma b_K \cdot X_{j(i)K} \qquad (2)$$

$$g_{j'(m)} = a + u_1 + C_m + b \cdot X_{j'(m)} + e$$

$$g'_{j'(m)} = a + u_2 + C_m + b \cdot X_{j'(m)} + e$$

$$R_{j''(n)} = a + u_1 + C_n + b \cdot X_{j'(n)} + e$$

$f_{j(i)}$, $f'_{j(i)}$ and $f''_{j(i)}$ represent the reported severe, moderate, and minor fire rates for census tract j at city i, which reports all three levels of fires. $q_{j'(m)}$, $q'_{j(m)}$ are reported severe and moderate fire rates of census tract j' at city m, which reports only severe and moderate fires, and $R_{j''(n)}$ is the reported severe fire rate of census tract j" at city n, which reports only severe fires. The parameters u_1, u_2, and u_3 represent the effects from the severe, moderate, and minor fire reporting systems. It should be mentioned that f, f', and f" are not independent nor are the q and q'.

One could addume that f, f', and f" are independent as well as q and q' and apply the regular regression procedure to estimate the parameters. These are approximate estimations. The estimated parameters should not be too different from the "legal" estimates.* Again, the significance of the city parameters c_i should be tested against the assumption that $c_i = 0$. If it is not significant, the model without the city parameter could then be used to estimate the reported fire rates of different severity, either nationally or locally.

*Since f, fi, and f" are not independent, the sum of the total reported fire rate is a better dependent variable to manipulate. Clearly $f_{j(i)} = f'_{j(i)} + f''_{j(i)}$ can be expressed as

$$\hat{f}_{j(i)} = \hat{a} + \hat{u}'_3 + 3\hat{c}_i + \Sigma\hat{b}_k \cdot 3x_{j(i)K} \quad ,$$

where

$$\hat{u}'_3 = 2\hat{a} + \hat{u}_1 + \hat{u}_2 + \hat{u}_3 \quad .$$

Similarly $Q_{j'(m)} = q_{j'(m)} + q'_{j'(m)}$ and

$$\hat{Q}_{j'(m)} = \hat{a} + \hat{u}_2' + 2\hat{c}_m + \Sigma\hat{b} \cdot 2x_{j'(m)K} \quad ,$$

with

$$\hat{u}_2' = \hat{a} + \hat{u}_1 + \hat{u}_2 \quad .$$

In order to incorporate the supplemental data available from some cities, it might be useful to fit two models--a rural and an urban one. The final urban model could then incorporate the data from the cities as well as from the urban portions of the NFIRS states. In any event, the cities can serve to provide an additional estimate of the validity of the national estimates model by estimating the fire rates for the cities and comparing with the observed rates.

Throughout the discussion, the term "estimates of fire rates" has been used. What is actually desired, of course, is a series of estimates by type of structure, cause, and loss of injury category. The previous discussion applies with the understanding that it is multivariate regressions or multivariate linear models that are being referred to. That is, each component of the "national estimates" is part of the vector of dependent variables used in the modeling process. This could be computationally quite cumbersome. It might be necessary to do the modeling only for the total, and rely on the relative distribution of the components in the NFIRS states to obtain national estimates for each component.

Finally, the simple expansion method ought to be tried as a preliminary step. That is, each state's data would be expanded to national size. Comparisons of these among the states as well as with the more sophisticated models should also be made.

We finally have three equations:

$$f_{j(i)} = a + u'_3 + c'_i + \sum_K b_K \cdot 3x_{j(i)K} + e \quad ,$$

$$Q_{j'(m)} = a + u'_2 + c_m' + \sum_K b_K \cdot 2x_{j'(m)K} + e \quad , \text{ and}$$

$$R_{j''(n)} = a + u_1 + c_n + \sum_K b_K \cdot X_{j''(n)K} + e.$$

Regular regression techniques can then be used to estimate the parameters m, u', c, and b.

It may be that the city parameters c_i are significantly different from zero, but cannot be identified as due to severity or other reporting policies. In this case an attempt should be made to identify cities with like c_i. A variable such as geography, population size, or some other city characteristic should be sought which will correspond to the c_i. That is, since it is desired to apply the model to areas beyond where the parameters have been estimated, the model cannot include terms unique to one area and still be useful. That is, a useful model must use available variables to predict.

Models for the NFPA Fire Department Survey Data. The data from the survey of a national sample of fire departments may represent the data set most nearly national in scope. However, these may be lacking in detail. Regression models should be estimated from these data to the extent possible. It may not be practical to obtain census tract data, or even aggregated census data, but these should be used in the regression model if feasible. In any event it should be possible to include geographical location and/or climate as a variable, as well as size of city and perhaps type of fire department. These regression models might be useful directly for obtaining national estimates. They might also be used to cross-check of validate the regression models developed from the cities or from the NFIRS data. That is, parameters of the various models could be compared, or the NFIRS model could be used to predict fire incidence for these communities and the results checked against the reported figures to obtain an independent estimate of the error in national estimates based on NFIRS regression models and census data.

APPENDIX II

DETAILED DATA TABLES

APPENDIX II
DETAILED DATA TABLES

This appendix includes detailed data tables from the NFIRS data. As mentioned earlier, they are based on fire incidents by building types from January, 1975 until June, 1976, so totals do not reflect annual figures. Table II-1 gives the causes by the codes used in the tables. Table II-2 gives the frequencies with row and column percentages of fires by cause and structure. The row percentages can be interpreted as the percent of fires with the given cause which involve each structure type.

Table II-3 gives the civilian injuries from fires classified by structure type and cause of fire. Table II-4 gives the civilian deaths in the same fashion. Combinations of structure type and cuase which resulted in no deaths or no injuries are omitted. Table II-5 gives the fire service injuries by structure type and cause of fire.

TABLE II-1

FIRE CAUSE CODE NUMBERS

1	Exposure
2	Natural
3	Incendiary Suspicious
4	Explosives/Fireworks
5	Smoking
6	Children
7	Heating
8	Cooking
9	Air Conditioning/Refrigeration
10	Electrical Distribution
11	Appliances
12	Gas
13	Flammable Liquid
14	Open Flame/Spark
15	Other Equipment
16	Other Heat
17	Unknown

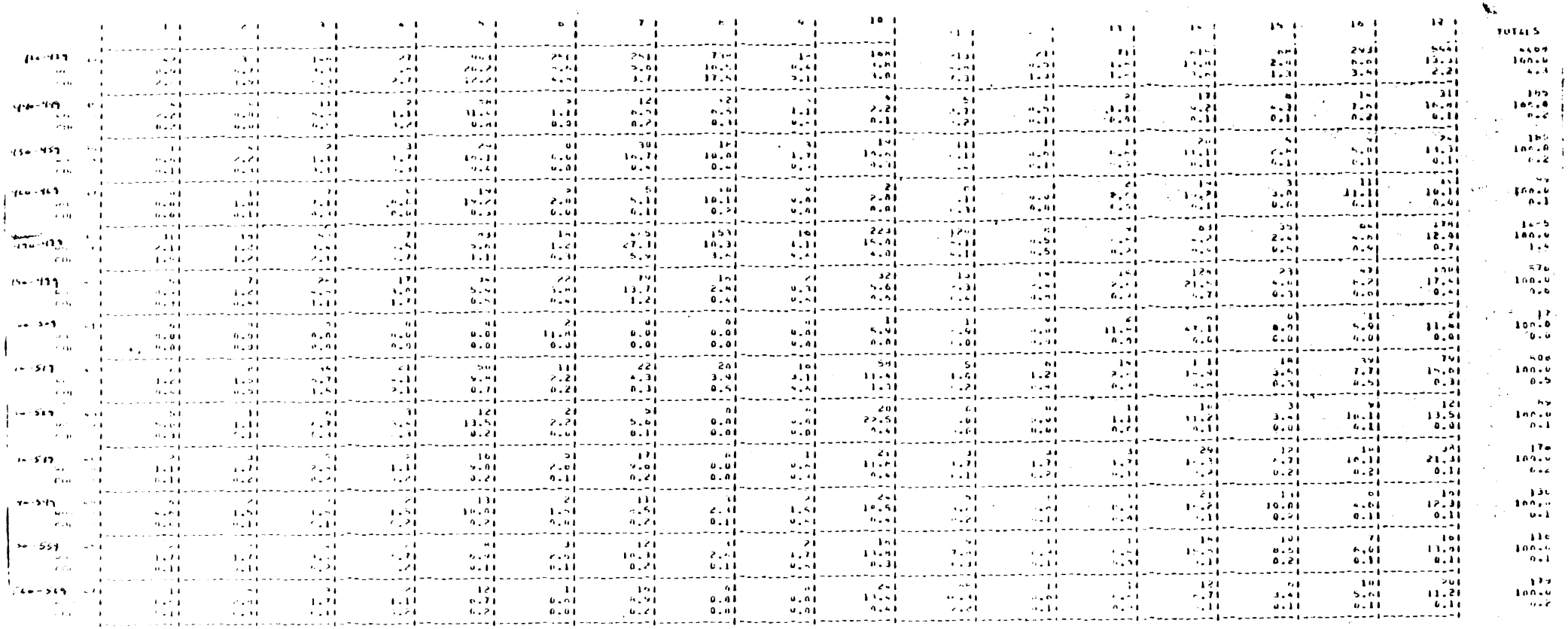
Table II-2

Fire Incidents by Cause and Property Type
Michigan 1975 - June 1976

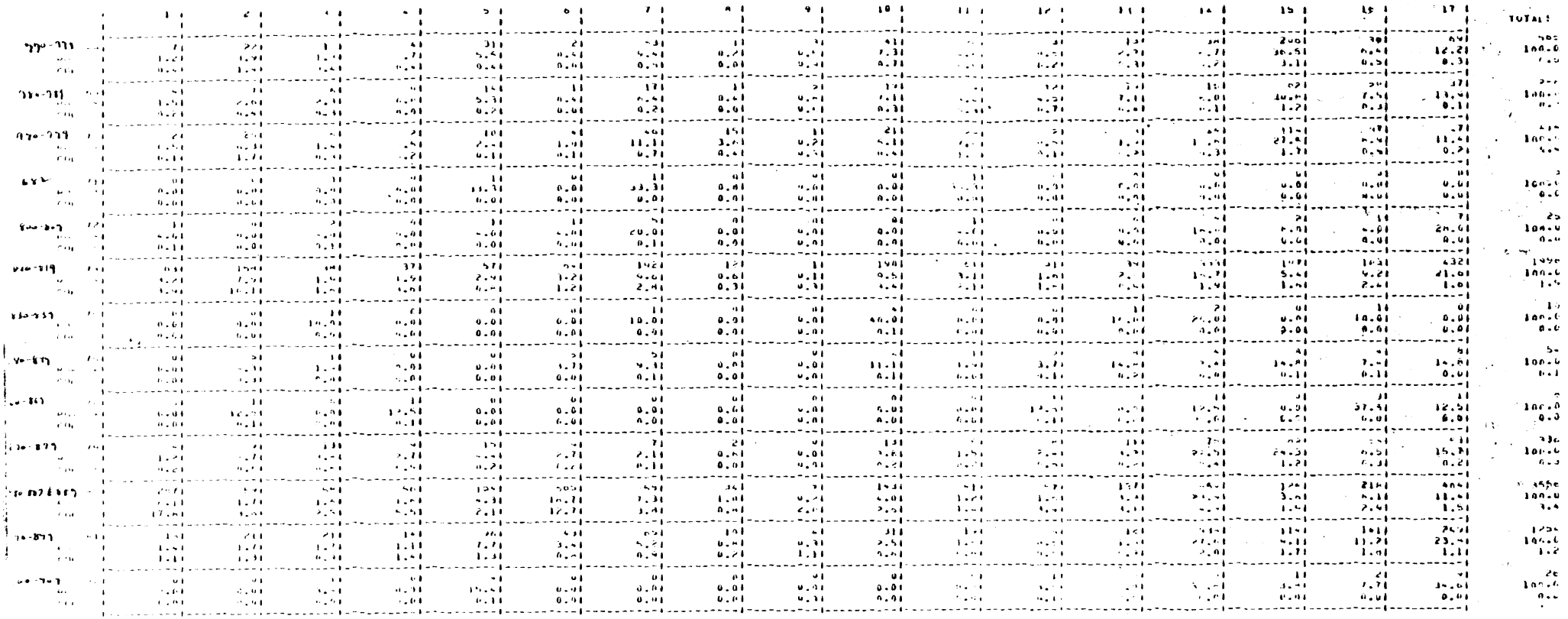
[illegible]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	TOTALS
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1 1.9 [illegible]	2 3.5 0.0	[illegible] [illegible] 0.0	1 [illegible] [illegible]	[illegible] 7.5 0.1	1 [illegible] [illegible]	3 [illegible] 0.2	[illegible] 17.[illegible] [illegible]	6 11.3 [illegible]	3 5.7 0.0	3 5.7 0.5	17 32.1 0.1	5[illegible] 100.[illegible] 0.[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5 4.9 0.1	[illegible] [illegible] 0.2	[illegible]	12 11.9 0.2	[illegible]	1 1.[illegible] 0.1	2 2.[illegible] [illegible]	16 1[illegible] [illegible]	[illegible] 3.6 0.5	5 4.9 0.1	25 23.[illegible] 0.1	10[illegible] 100.[illegible] [illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] 11.[illegible] [illegible]	[illegible]	[illegible]	1 1.5 0.0	[illegible]	2 2.9 0.0	[illegible]	[illegible]	[illegible]	19 27.5 [illegible]	5 7.[illegible] 0.1	4 5.9 0.1	18 26.0 0.1	[illegible] 100.[illegible] [illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	9 3.5 0.2	18 7.2 0.3	[illegible] 3.[illegible] 0.2	[illegible]	17 [illegible] 0.3	[illegible]	[illegible]	[illegible]	75 2[illegible] [illegible]	9 3.6 0.1	13 5.2 0.2	41 16.3 [illegible].2	25[illegible] 100.[illegible] [illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	24 20.3 [illegible]	13 11.0 0.1	2 1.7 [illegible]	15 1[illegible].7 0.3	[illegible]	[illegible]	[illegible]	[illegible]	3 2.5 0.0	15 12.7 0.2	14 11.[illegible] 0.1	11[illegible] 100.[illegible] [illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1 9.1 0.0	[illegible] 0.0 0.0	0 0.0 0.0	0 0.0 0.0	0 0.0 0.0	3 27.3 [illegible]	[illegible]	[illegible]	[illegible]	2 1[illegible].2 [illegible]	1 9.1 0.0	[illegible] 1[illegible].2 0.0	1 9.1 0.0	11 100.[illegible] 0.[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	63 7.1 [illegible]	17 1.[illegible] 0.3	[illegible] 7.3 [illegible].9	28[illegible] 32.4 [illegible]	13 1.5 3.7	75 [illegible].5 1.3	[illegible]	[illegible]	13 1.5 [illegible].3	[illegible]	22 2.5 0.3	57 6.5 0.7	109 12.4 0.4	[illegible] 100.[illegible] [illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] 13.3 [illegible]	[illegible]	1 [illegible] 0.0	0 0.0 0.0	0 0.0 0.0	0 0.0 0.0	[illegible]	[illegible]	[illegible]	[illegible]	1 [illegible].3 [illegible]	1 [illegible].3 0.0	2 16.7 0.0	12 100.[illegible] 0.[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] 4.5 0.0	[illegible] 9.[illegible] 0.1	[illegible]	12 [illegible] [illegible]	[illegible]	[illegible]	[illegible]	[illegible] 1[illegible].7 [illegible]	1 2.[illegible] 0.0	1 2.[illegible] 0.9	4 [illegible].5 0.0	[illegible] 100.[illegible] 0.[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible] [illegible] 0.6	103 16.[illegible] 1.5	14 2.2 0.1	2 0.3 [illegible]	3[illegible] [illegible] 0.7	[illegible]	31 [illegible] 1.[illegible]	[illegible]	[illegible] 15.2 [illegible]	2[illegible] [illegible] 0.4	[illegible] 7.0 0.6	1[illegible] 22.5 0.5	[illegible]27 100.[illegible] 0.[illegible]

		1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	14	15	[illegible]	17
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	0	2	0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	0.0	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	100.0	0.0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	123
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3	0.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.5
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	0.0	[illegible]	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	100.0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	2	[illegible]	[illegible]	[illegible]	[illegible]	1	1	4
	[illegible]	[illegible]	0.0	[illegible]	[illegible]	22.2	0.0	0.0	4.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.7	3.7	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1	0.0	[illegible]	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	0.0
[illegible]	[illegible]	1	[illegible]	2	[illegible]	6	2	[illegible]	1	1	5	[illegible]	[illegible]	[illegible]	7	11	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4.2	[illegible]	[illegible]	[illegible]	11.5	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1	0.0	0.0	0.0	[illegible]	0.1	[illegible]	[illegible]	[illegible]	[illegible]	0.2	[illegible]	0.0
[illegible]	[illegible]	1	2	[illegible]	1	5	1	6	2	[illegible]	5	1	[illegible]	[illegible]	[illegible]	2	6	16
	[illegible]	1.2	2.5	[illegible]	1.2	[illegible]	1.2	7.5	2.5	0.0	6.1	1.2	[illegible]	[illegible]	[illegible]	2.5	7.5	20.0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1	0.0	0.1	0.0	0.0	0.1	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.1	0.1
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	11	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10	3	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	22.3	0.0	12.5	8.7	[illegible]	5.8	20.[illegible]	[illegible]	[illegible]	[illegible]	9.7	2.4	7.4
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.3	0.0	0.2	0.2	[illegible]	0.1	[illegible]	[illegible]	[illegible]	[illegible]	0.2	0.0	0.0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	[illegible]	[illegible]	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	50.0	0.0	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	0.0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	0.0	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	0.0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7	[illegible]	[illegible]	20	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	20	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	25.9	[illegible]	1.5	[illegible]	1.4	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	5.7	[illegible]	25.9
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.1	0.7	1.7	[illegible]	[illegible]	[illegible]	[illegible]	.2	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	1	[illegible]	1	3	[illegible]	[illegible]	[illegible]	[illegible]	3	[illegible]	[illegible]	[illegible]	15	1	2	7
	[illegible]	[illegible]	1.2	[illegible]	1.2	3.7	0.0	5.0	0.0	0.0	3.7	[illegible]	[illegible]	[illegible]	[illegible]	1.2	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	0.1	0.6	[illegible]	0.1	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	[illegible]
[illegible]	[illegible]	1	1	[illegible]	[illegible]	7	0	3	[illegible]	[illegible]	3	[illegible]	[illegible]	[illegible]	[illegible]	3	1	19
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6.7	[illegible]	2.2	[illegible]	[illegible]	[illegible]	[illegible]	11.1	6.7	[illegible]	22.2
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	0.0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1	[illegible]	1	1	[illegible]	1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2	1
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.0	0.0	10.0	10.0	0.0	10.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	20.0	10.0
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	0.0	0.0	[illegible]	0.0	[illegible]	[illegible]	[illegible]	[illegible]	0.0	0.0	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	2715
	[illegible]	[illegible]	[illegible]	[illegible]	1.1	[illegible]	[illegible]	17.[illegible]	11.[illegible]	[illegible]	7.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	12.2
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	57.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	10.1



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	TOTALS
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]



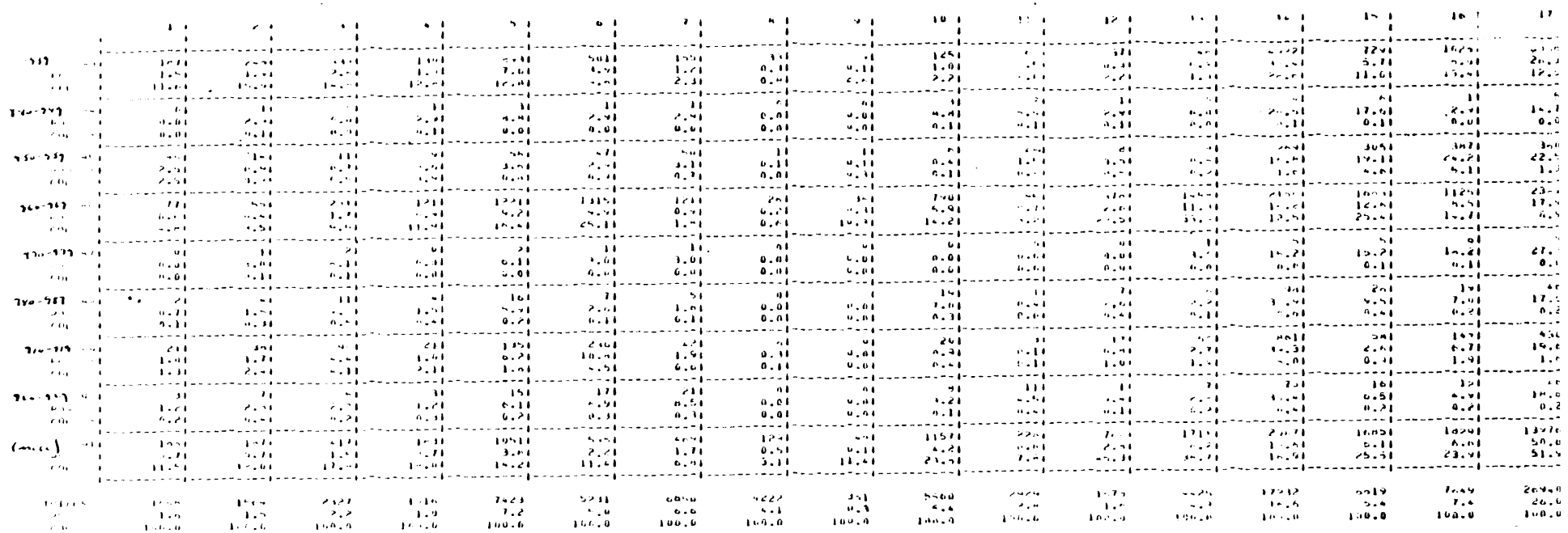


Table II-3. Civilian Injuries by Property Type and Cause

Property Class	Initial Cause Category																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
001-009 & 101-109													1					1
010-019 & 110-119					1													1
030-039 & 130-139				1			3	2										6
040-049 & 140-149								2										2
060-069 & 160-169							2	11		1				1	1		1	17
090-099 & 190-199								1		1		1					3	6
210-219							1	214					3		7		1	226
240-249										10					1			11
290-299							3											3
310-319			2					1			2			1				6
330-339			4		11						1			6		1	1	24
340-349			9							1	3		1				6	20
390-399																	2	2
410-419	11	3	31	4	153	41	141*	139	7	45	49	6	21	56	26	45	67	845*
420-429	1		13		89	26	17	30	2	8	11	4	9	28	1	14	32	285
440-449			4		3			1		1				4		1	7	21
450-459					2				1					1				4
460-469					1									2			1	4
470-479			2		7	3	11	15	1	5	4			3	1	2	6	60
490-499	2				2	1	4				1	1		1	3		1	16
510-519							1	1		2								4
520-529			1		1													2
530-539											1				1			2
540-549							1											1
550-559							5								1			6
560-569											1							1
570-579							7			5		1	4		6			23
580-589							10			4		2						16
590-599					10										4	2		16
640-649										2								2
670-679		1																1
710-719									2									2
720-729			2															2
750-759	1	1				2				1					1		20	26
760-769		4													9			13
770-779		2								1	2	1			12	1	2	21

Table II-3. (Cont.)

Property Class	Initial Cause Category																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
780-789											1							1
790-799		1					1			2	2				3			9
800-809					1													1
810-819	2	4					1		1	5	1	1	1		8*	5		29*
840-849															2		1	3
870-879										2		1	1					4
880-887 & 889	4	1			3	1	3	1	1	1		4	7	2	6	1		35
890-899					2			1										3
930-939			1		3					1			3	5	2	2		17
950-959	5						1											6
960-969	1		1	1	6		3	3		1	2	4	13	2	9	2	1	49
970-979															1			1
980-989												1			3			4
910-919			1		1									1		1	2	6
920-929															2			2
(Misc.)	5		2*	4	11	3	18	10		6	2	13	19	4	20	13	82	210*
Total	32	17	73*	8	307	77	233*	432	15	105	83	40	83	117	130*	90	237	2079*

*Indicates an entry where an obvious coding error has been corrected.

Table II-4. Deaths by Property Type and Cause

Property Class	Initial Cause Category																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
030-039 & 130-139				1														1
310-319			1		1													2
330-339					4						1							5
410-419	2		4		41	5	32	21	2	16	6	1	3	6		13	23	175
420-429	1		4		10	1	6	2		4	1			5		1	7	42
440-449					2													2
470-479			2		2		7	2						2		1	1	17
490-499							5					1						6
550-559							1											1
570-579										1								1
590-599					1													1
750-759																	3	3
760-769		1																1
810-819										1		1					5	7
880-887 & 889			1		1										6			8
890-899							1											1
930-939													2	1*	1	1		5*
960-969			1	2								2				2	2	9
970-979													12				5	17
910-919														1				1
920-929		1																1
(Misc.)	1		3				1			2			5	2	1	6	*	21*
Total	4	2	16	3	62	6	53	25	2	24	8	5	22	17*	8	24	46*	327*

*Indicates entries which have been corrected for obvious coding errors.

Table II-5. Fire Service Injuries by Property Type and Cause

Property Class	Initial Cause Category																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
010-019 & 110-119			1							2								3
030-039 & 130-139		1					1							1				3
040-049 & 140-149							2										1	3
060-069 & 160-169			4		3	1	12	15	1	8	5		4	3		4	9	69
090-099 & 190-199		1	1		1		1			3			1				5	13
210-219					1	1	3				2	1		9		1	7	25
230-239					1													1
240-249						1				3						11		15
290-299				1								4		1				6
310-319							4				1							5
330-339					5					1				2		2		10
340-349											1							1
390-399							3							1				4
410-419	73	17	28	12	99	61	159	41	15	109	37	18	34	139	26	85	144	1097
420-429	5		9	2	38	20	23	14	2	18	9	1	11	36	3	20	45	256
440-449			7				1							2			15	25
450-459					1					1				1		1	1	5
460-469														1				1
470-479	1	1	2		4		9	4		12	4		1			1		39
490-499			1				14			3		2		9	1	2	1	33
500-509														1				1
510-519		1	3	2	1								10	2	1	3	12	35
520-529	36		3				1			3			1				4	48
530-539	8				2		4			5	1			3			1	24
540-549										1				9				10
550-559	8						3			2				2			3	18
560-569				1	1		2				3			1				8
570-579					2		5			3			3	1	2	2	2	20
580-589		1	2		5		1			5		1		1		6	8	30
590-599	9	1		1	1		2			7	1		1	3	5	6	17	54
888																1		1

Table II-5. (Cont.)

Property Class	Initial Cause Category																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Total
640-649																	2	2
670-679		3																3
710-719					4				1									5
720-729			2								2							4
740-749							1		1						1			3
750-759		13	1				2							12	2		3	33
760-769		1					1				2			1	11	1		17
770-779			1	1	3		1			2					9	4	1	22
780-789														2	1			3
790-799	3						2	2			2			2	9	1	4	25
810-819		5		3		1	14			5	8	1	1	7	6	1	11	63
830-839										1								1
840-849																	1	1
870-879	1													2	4	1		8
880-887 & 889	12	3	4	3	2	13	11			2	1	6	6	14	8	10	7	102
890-899	5		1		5	1	17						3	8	3	6	16	65
930-939		1	2	2	2							2		13	8	7	18	55
940-949											1					2	2	5
950-959	1				1									1	1			4
960-969	1			2	1	4				3	5	8	5	13	15	9	6	72
970-979																1		1
980-989										1					1			2
910-919	1		16	2		3		2					4	12			13	53
920-929		1												1	1			3
(Misc.)	10	9	3	6	7	4	6	2	1	13	1	3	16	32	7	22	175	317
Total	174	59	91	38	190	110	305	80	21	213	86	47	100	348	125	210	534	2730