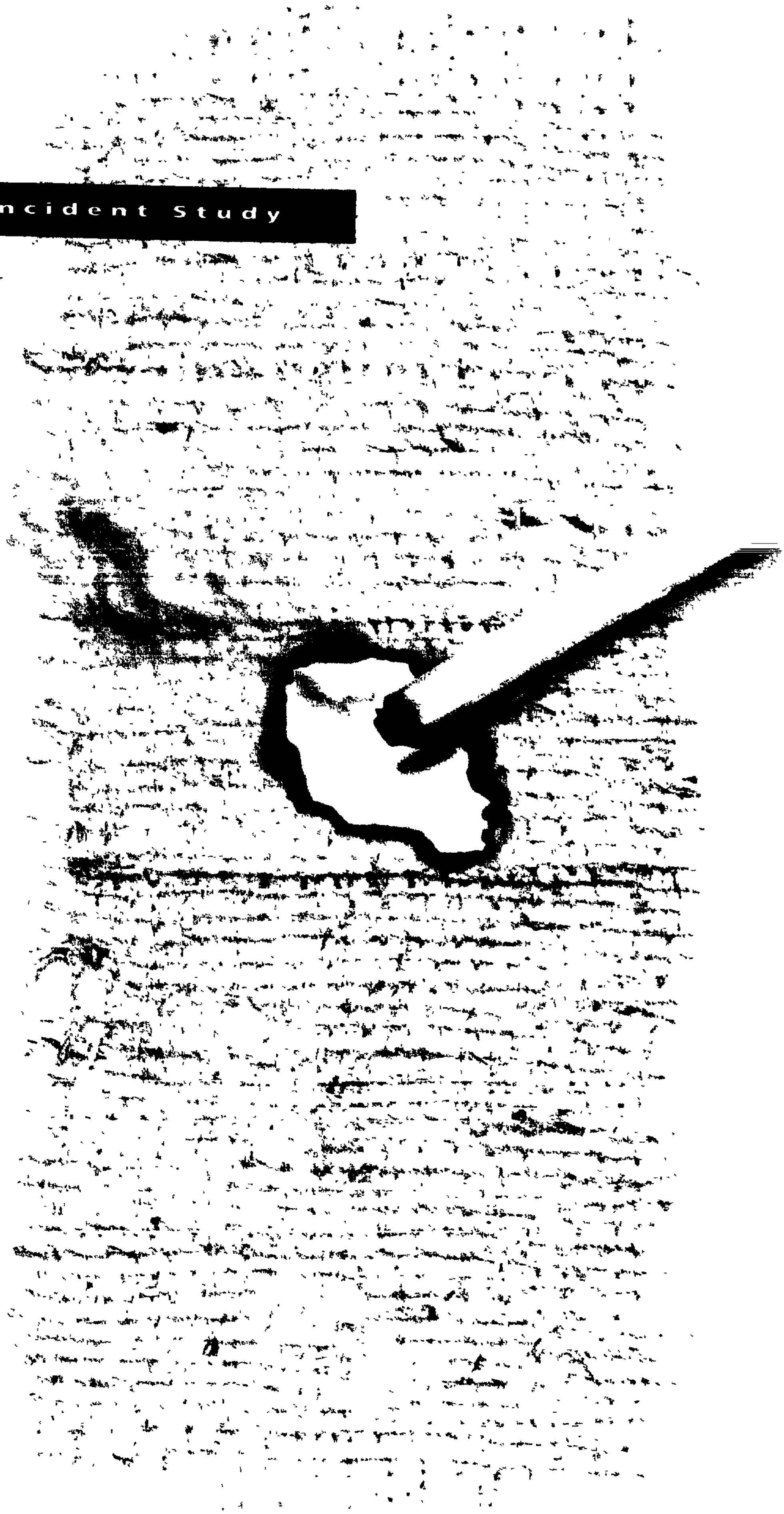
Cigarette Fire Incident Study



4

PREPARED FOR THE U.S. CONSUMER PRODUCT SAFETY COMMISSION BY THE NATIONAL FIRE PROTECTION ASSOCIATION AND MATHEMATICA POLICY RESEARCH, INC.

AUGUST 1993

Fire Safe Cigarette Act of 1990

Under the Cigarette Safety Act of 1984 (P.L. 98-567), the Technical Study Group on Cigarette and Little Cigar Fire Safety (TSG) found that it is technically feasible and may be commercially feasible to develop a cigarette that will have a significantly reduced propensity to ignite furniture and mattresses. Furthermore, they found that the overall impact of such a cigarette on other aspects of the United States society and economy may be minimal.

Recognizing that cigarette-ignited fires continue to be the leading cause of fire deaths in the United States, the Fire Safe Cigarette Act of 1990 (P.L. 101-352) was passed by the 101st Congress and signed into law on August 10, 1990. The Act deemed it appropriate for the U.S. Consumer Product Safety Commission to complete the research recommended by the TSG and provide, by August 10, 1993, an assessment of the practicality of a cigarette fire safety performance standard.

Three particular tasks were assigned to the National Institute of Standards and Technology's Building and Fire Research Laboratory:

- develop a standard test method to determine cigarette ignition propensity,
- compile performance data for cigarettes using the standard test method, and
- conduct laboratory studies on and computer modeling of ignition physics to develop valid, user-friendly predictive capability.

Three tasks were assigned to the Consumer Product Safety Commission:

- design and implement a study to collect baseline and follow-up data about the characteristics of cigarettes, products ignited, and smokers involved in fires,
- develop information on societal costs of cigarette-ignited fires, and
- in consultation with the Secretary of Health and Human Services, develop information on changes in the toxicity of smoke and resultant health effects from cigarette prototypes.

The Act also established a Technical Advisory Group to advise and work with the two agencies.

This report is one of six describing the research performed and the results obtained. Copies of these reports may be obtained from the **U.S. Consumer Product Safety Commission, Washington, DC 20207.**

4

Cigarette Fire Incident Study

Beatrice Harwood

Terry L. Kissinger

U.S. Consumer Product Safety Commission

Michael J. Karter, Jr., Alison L. Miller,

Rita F. Fahy, John R. Hall, Jr.

Donna Eisenhower, Peter Forbes, John Hall,

Mathematica Policy Research, Inc.

August 1993

Volume 4. Cigarette Fire Incident Study

Table of Contents

Cigarette Fire Incident Study (Overall Analysis)

Executive Summary.................................. i

General Overview................................... 1

Data Collection and Set-up for Analysis.......... 8

Preliminary Analysis............................... 11

General Logistic Regression Model................ 13

Complete Model of Cigarette Fire Risk............ 23

Concise Model of Fire Risk....................... 26

Conclusions.. 33

Appendix... 34

Cigarette Consumer Survey.............................. A1

Self-Proxy Comparisons for the Consumer Survey........ B1

Executive Summary

Data were collected from eight cities on a wide range of cigarette and smoker characteristics for a sample of smokers. Of these, 564 smokers had had fires and were identified through fire department response to those fires, while the other 1,611 smokers had not had fires and were identified through a telephone sample survey of the communities. The characteristics analyzed included those that had shown evidence of a relationship to the risk of a cigarette-initiated fire, either in laboratory studies or in previous statistical analyses of fire experience.

The smoker characteristics analyzed were (household) income, education, age, gender, and race. The cigarette characteristics analyzed were filter, tobacco column length, filter length, circumference, density, amount of tobacco, menthol, citrate, porosity and pack type. In addition, a variable was used to control for the smoker's city.

After controlling for all smoker characteristics and city, logistic regression modeling showed four cigarette characteristics to be significant - filter, filter length, porosity, and pack type. Filter, filter length and porosity all affect air intake, which therefore appears to be an important physical element in the combustion process associated with risk. Analysis limited to filtered cigarettes only showed the same characteristics to be significant, plus tobacco column length. Extension of the analysis to two-way interaction terms did not change any of the conclusions on which cigarette characteristics are important but did indicate that the role of pack type was different for men vs. women.

Sensitivity analyses, shown in the appendix, supported the main conclusions, which were that (1) cigarette characteristics are significant after controlling for smoker characteristics and (2) the four specific cigarette characteristics — filter, filter length, porosity, and pack — are the ones that are significant. These analyses checked the impact of cluster sampling, sensitivity to missing data on smoker characteristics, and sensitivity to non-fire smoker cases with responses by people other than the smokers themselves.

All this means there are already commercially available cigarettes that exhibit reduced ignition propensity when one controls for smoker characteristics.

Acknowledgements

This study would not have been possible without tremendous effort and cooperation from a great many people.

First and foremost, we wish to thank the fire departments in our eight cities. In particular, we wish to thank the following chiefs for agreeing to participate in this study and the project coordinators for all their hard work and determination: Fire Chiefs Peter O'Connor and Herman Williams and project coordinators Lieutenant Earl DeVincentz and Captain Kenneth Morris of Baltimore, Maryland; Fire Chief Walter Zimmerer and project coordinators Lieutenant Richard Klein and Captain Ray Masarik of Cleveland, Ohio; Fire Chief Harmon Dutko and project coordinators Captain Tom Huston and Captain Tim Carty of Columbus, Ohio; Fire Chief Dodd Miller and project coordinator Deputy Chief T.W. Oney of Dallas, Texas; Fire Chief Richard Gonzales and project coordinators Lieutenant Joe Sarconi and Captain Gregory Taft of Denver, Colorado; Fire Chiefs Robert Clayton and E.A. Corral and project coordinators Captain Michael Shrum, Assistant Chief Carl Hooker and Assistant Chief Dennis Holder of Houston, Texas; Fire Commissioners Roger Ulshafer and Harold Hairston and project coordinators Fire Marshal John Skarbek, Deputy Chief Robert Wauhop and Battalion Chief Michael Mulderrig of Philadelphia, Pennsylvania; and Fire Chief George Monogue, Fire Marshal Lynn Davis, project coordinator Assistant Fire Marshal Jim Crawford, and Captain Paul Barrett of Portland, Oregon.

Our thanks and appreciation extend to every member of the eight participating fire departments for gathering the information so vital to this project.

Thanks, too, to our colleagues at Mathematica - Donna Eisenhower, Randy Brown, and John Hall - who helped not only with the coordination of two complex sub-tasks but also with good ideas for every stage of the project.

We also wish to thank the members of the Technical Advisory Group, whose helpful comments on our research plans and draft reports - and extensive cooperation on data sources - varied from very useful to absolutely essential.

Thanks to our secretarial team of Helen Columbo, Nancy Schwartz and Norma Candeloro, whose tireless support throughout the project was a major factor in our success.

And finally, special thanks to Beatrice Harwood and Dr. Terry Kissinger of CPSC. Their contributions to our work and our analysis went far beyond the requirements of project officers.

General Overview

Introduction

The Fire Safe Cigarette Act of 1990, which led to the research reported here, had as its goal the completion of research begun under the Cigarette Safety Act of 1984, which had demonstrated that "it is technically feasible ... to develop cigarettes that will have a significantly reduced propensity to ignite upholstered furniture or mattresses."[1] The tasks identified included development and validation of a standard test method, additional studies of the societal costs of injuries from cigarette-initiated fires and of the effects on cigarette smoke toxicity of modifications to current cigarette designs, and the study described here on the characteristics of cigarettes and smokers involved in cigarette-initiated fires.

Purpose

There are several purposes served by and potential benefits from the proposed study: (a) identification of specific cigarette characteristics correlated with differences in the risk of fire, which could be used in the development of secondary test methods, also called non-ignition tests; and (b) quantification of the relative importance of various cigarette characteristics in fire risk.

To understand how these questions rose to prominence, it is important to review what we already know. National fire incident data bases have shown consistently that cigarettes are the leading heat source in fatal U.S. fires. Even with recent declines, the 1,245 civilian fire deaths in 1990 that were estimated to have begun with lighted tobacco products (nearly all of them cigarettes) represented roughly one of every four civilian fire deaths in the U.S.[2] The frequency of fire death from cigarette fires, however, is known to vary among different sub-groups of the population, such as age groups and gender groups. At the same time, laboratory studies have shown that under standardized test conditions, the propensity of cigarettes to ignite fires is different for cigarettes

[1]Technical Study Group of the Cigarette Safety Act of 1984, *Toward a Less Fire-Prone Cigarette*, Final Report of the Technical Study Group on Cigarette and Little Cigar Fire Safety, October 1987, p. 1.

[2]Alison L. Miller, *The U.S. Smoking-Material Fire Problem Through 1990*, Quincy, MA: National Fire Protection Association, Fire Analysis and Research Division, March 1993, p. 1.

with different physical characteristics. Furthermore, risk-related cigarette characteristics are often correlated with each other, risk-related smoker characteristics are often correlated with each other, and it is likely that some risk-related cigarette characteristics are correlated with risk-related smoker characteristics through differences in the brand preferences of different smoker groups.

A cigarette fire is the result of three elements, each of which is necessary for fire to occur and each of which may be characterized by greater or lesser degrees of susceptibility. First is the contact between heat source (cigarette) and potential fuel source. Second is the susceptibility of the fuel source to ignition, when so exposed. Third is the propensity of the cigarette to ignite fires when put in this position.

Contact between cigarette and potential fuel source usually is a result of carelessness, and people differ in the degree of their carelessness. Therefore, smokers will have different probabilities of discarding cigarettes in place where ignition is possible, and cigarettes may have different probabilities of being discarded.

Potential fuel sources (e.g., upholstered furniture) may be more or less susceptible to ignition by a given intensity and time of heat exposure, for example, as a function of design or material composition. This translates into different probabilities of ignition, given exposure, for different substitutable items (e.g., two different couches), and this in turn translates into different probabilities of ignition, given exposure for different smokers, based on their relative likelihood of owning one or the other type of item.

Finally, cigarettes may differ in their propensities to ignite fires, given exposure to a specified potential fuel source. The probability of ignition, given exposure, may then be a function of cigarette characteristics that make it more or less likely that the cigarette will burn with a more or less intense heat or for longer or shorter periods of time after discard.

Smoker characteristics such as age, gender, poverty, and education are used to analyze this complex interaction of three types of elements. It is not asserted that poverty directly causes fires, but poverty may be associated, for example, with a higher likelihood of excess use of alcohol, which could lead to a higher degree of carelessness in handling cigarettes. Less education does not directly cause fires but may be correlated with a reduced awareness of or consistent practice of the rules of fire-safe behavior.

The principal hypothesis to be addressed in this study, then, is as follows: Are there cigarette characteristics that are statistically related to higher risk in cigarette-initiated fires, after adjusting for the effects of smoker characteristics?

If the answer is yes, then that would support the view that fire risk can be reduced through changes to cigarette characteristics. If the answer is no, however, then that could mean that cigarette modifications would not be an effective way to reduce fire risk, that risk only appears to be related to some cigarette characteristics because of their correlation with smoker characteristics. At the least, it would suggest that fire risk is not sensitive to changes in cigarette characteristics within the ranges of characteristics found in current commercial cigarettes.

If fire risk is related to cigarette characteristics and smoker characteristics, the relative strength of each relationship may not be the only, or even the primary, consideration in choosing which characteristics to emphasize in a program to reduce cigarette fire risk. Also important are the degree of sensitivity of fire risk to the characteristic and, as important, the feasibility and cost of making changes in the characteristic. If age and gender turn out to be strongly related to fire risk, for example, this has no direct practical value, because age and gender cannot be altered. It may have indirect value if one can develop an awareness program to change some other risk-related characteristic (e.g., carelessness) and tailor it especially to high-risk age or gender groups.

More subtly, if poverty or education proved to be highly correlated with fire risk, they can be changed, but only slowly and at very great cost. By comparison, changes to the physical characteristics of cigarettes can be made more quickly and less expensively. All this needs to be considered in interpreting the implications of this analysis.

To help answer these questions, this report will examine the effects of numerous cigarette and smoker characteristics all at the same time in order to adjust for associations and differences in the distributions among these characteristics. These results will indicate what characteristics are significant after adjusting for the effects of the other characteristics.

Smoker Characteristics Analyzed

Six smoker characteristics were proposed for analysis:

- *Income* (of the household)
- *Education*
- *Age*
- *Gender*
- *Race*
- *Number of cigarettes smoked per day*

The first five of these characteristics have been shown in numerous statistical analyses to be highly correlated with either the risk of fire or the risk of death in fire. It is also generally recognized that age, gender, and race are significantly related, through brand choice, with many cigarette characteristics, and it is likely that this relationship applies to the other smoker characteristics as well.

Cigarette Characteristics Analyzed

Eleven cigarette characteristics were proposed for analysis:

- *Filter* (present or not present)
- *Tobacco column length*
- *Total cigarette length*
- *Filter length*
- *Circumference*
- *Density* (of tobacco)
- *Amount of tobacco*
- *Menthol* (present or not present)
- *Citrate* (in wrapping paper)
- *Porosity* (of wrapping paper)
- *Pack type* (soft or hard)

Filter length was introduced as a separate variable late in the analysis as a response to a discussion of the implications of some of the preliminary analysis.

Amount of tobacco is a variable calculated from *circumference, density* and *column length.* In earlier studies, it has also been called tobacco mass.

Citrate refers to chemicals - sodium or potassium citrates - that are added to some cigarette paper to help the paper and the tobacco burn at the same rate.

Porosity is also a characteristic of the cigarette paper and measures the rate of air flow, which may affect the burning properties of the cigarettes. Porosity is measured in cubic centimeters of air per minute per square centimeter of paper, at a specified ambient air pressure. The measurement is expressed in "CORESTA" units, with higher values corresponding to higher air flow.

Pack type is a difficult characteristic to interpret since it is not inherently a characteristic of the cigarette. Two hypotheses have been identified for the influence of pack type on the risk of cigarette fire ignition: (a) It may be that cigarettes in soft packs are more likely to be physically altered, due to impact on the pack prior to being ignited, than are cigarettes in hard packs, which would be expected to shield more effectively against such incidental impacts. The bending or crushing or other modifications resulting from impact could affect ignition propensity. (b) It may be that cigarettes are designed differently as a function of pack type, either to cater to differences in customer preferences between choosers of hard vs. soft packs or to help preserve cigarette freshness under two sets of storage conditions.

Laboratory tests in the earlier study, under the Cigarette Safety Act of 1984, indicated ignition propensity differences for density, porosity, and circumference, in that order.[3] Citrate showed only slight differences, except for certain classes of cigarettes. Tobacco type (flue-cured vs. burley) showed negligible differences and has not been analyzed in this report.

In a feasibility study of the type of analysis in this report , the U.S. Consumer Product Safety Commission (CPSC) found potentially significant differences in density, tobacco column length, circumference, amount of tobacco, porosity, and filter.[4]

Note that both the cigarette characteristics and the smoker characteristics are taken from a much larger universe of definable characteristics. Smoker characteristics like frequency or degree of alcohol or drug use might be correlated

[3]Technical Study Group of the Cigarette Safety Act of 1984, *Toward a Less Fire-Prone Cigarette*, October 1987, Table 3, p. 8.

[4]Beatrice Howard and Linda Fansler, *Feasibility Study of Obtaining Field Data on Cigarette-Ignited Fires*, Technical Study Group of the Cigarette Safety Act of 1984, pp. 5, 7.

to the probability of careless disposal of a cigarette, but these are examples of characteristics that could not be reliably measured through field interviews. Similarly, there may be unidentified cigarette characteristics that, if known, would prove to be strongly linked to fire risk.

The possibility of important hidden smoker characteristics is a concern only if these characteristics have a strong link to fire risk that is also very different in form from the correlation to fire risk shown by the smoker characteristics in the model. To take the example cited, drug and alcohol use are known to be correlated with such characteristics as age, gender, and race - all variables included in the model.[5] The included smoker variables therefore may gain some of their apparent statistical strength by acting as proxies for variables like drug or alcohol use that are not explicitly included. This effect is not a problem because the analysis is concerned only with the strength of cigarette characteristics after controlling for smoker characteristics, not with the relative importance of individual smoker characteristics.

However, it is also possible that the hidden smoker characteristics could add to the apparent statistical strength of some of the cigarette characteristics. This theoretical possibility could occur only if a cigarette characteristic were strongly correlated with a hidden smoker characteristic and was not strongly correlated with any of the included smoker characteristics. This possibility is remote, and no patterns in the analysis or data from other sources have been put forward to indicate this is the case.

A more likely possibility is that one or more hidden cigarette characteristics could be adding to the apparent statistical strength of the included cigarette characteristics. For example, the measured citrates are probably not the only additives used. The possibility that hidden cigarette characteristics may be significant bears on the interpretation of the results of the analysis. If the analysis shows that cigarette characteristics are significant factors in fire risk after controlling for smoker characteristics, then it follows that cigarette redesign can lead to reduced fire risk, but the best approach to that redesign may not involve exclusive, or even primary, concentration on the particular cigarette characteristics found significant. Instead, these results may be more useful as

[5]See, for example, Table 198 and the general population tables in *Statistical Abstract of the United States: 1992,* 112th edition, Washington: Bureau of the Census, 1992.

part of a calibration for a standardized test method, which could then assess the impact of all types of changes in cigarette design.

Overview of Data Collection and Set-Up for Analysis

This section describes the steps leading up to the analysis, emphasizing key points needed to put the analysis that follows in proper context.

CPSC authorized two organizations to undertake data collection projects. NFPA conducted a cigarette fire incident data collection effort with the cooperation of eight fire departments. Mathematica Policy Research, Inc., under subcontract with Market Facts, Inc., conducted a study to collect data on a comparison group of smokers who did not have fires and lived in the same communities.

Selecting the Cities

With the approval of CPSC, NFPA selected eight fire departments to participate in the cigarette fire incident data collection effort. Among the criteria used to select the eight cities were willingness to participate and reported number of cigarette fires. (This meant the samples were not simple random samples of all U.S. smokers and cigarette fires.) The target total number of cigarette fires to be collected, based on sample size recommendations given by Dr. Chan Dayton of the University of Maryland, was between 400 and 600 fires for the eight communities. The eight cities selected to participate were Baltimore MD, Cleveland OH, Columbus OH, Dallas TX, Denver CO, Houston, TX, Philadelphia PA, and Portland, OR.

The data collection period was 13 months, from December of 1991 through December 1992. In all 647 forms were collected during that time period from the eight communities. In terms of useable forms for the study, defined as those with adequate information for precise cigarette brand identification, there were 564 useable forms.

Mathematica conducted their survey for the same eight communities so that there would be a comparison group of smokers who did not have fires. Mathematica designed and implemented a self-weighted sample of 1,532 households, obtaining information from all smokers in a household. Because the resulting sample was not a simple random sample - it included all smokers in a household - the possible effect of clustering on the analysis is of concern and will be examined later in the report.

The smoker survey of 1,532 households obtained information on 1,969 smokers. Discussions among NFPA, CPSC, and Mathematica analysts identified several categories that should be excluded from the analysis. These were (1) households definitely outside the fire service area, (2) households where the tract could not be determined or could only be imputed from zip codes, and (3) households that had cigarette fires (9 smokers). The first three were all groups that might not or did not fall in the boundaries of the fire service districts of interest. After excluding these four categories, the smoker survey file contained data from 1,130 households and 1,611 smokers.

Questionnaires Used and Resulting Data Elements

The questionnaire used to collect data from cigarette fires was designed in such a way that the data could be coded directly from the forms without the need for coding sheets. There were two versions of the questionnaire used in the course of the data collection phase of the study.

The education question originally had three choices -- completed high school, did not complete high school and undetermined. The choice added was "some college or more." The income question originally had four choices -- under $10,000/year, $10,001 to $20,000/year, over $20,000/year and undetermined. The revised form expanded on the third option. The choices for over $20,000 per year became $20,000 to $29,999/year, $30,000 to $39,999/year and $40,000 or more. The smoker survey used even more categories for income and education and more categories for race. In the end, the analysis needed all useable cases, so for analysis purposes, these added options were folded down to the choices on the original fire incident questionnaire.

The other continuous-variable smoker characteristic was the age of the smoker, which was measured in years and used in its continuous form.

Both NFPA and Mathematica collected data on census tract and used that data to assign proxy income and education variables to each smoker. In each case, the proxy variable was a characteristic of the tract (e.g., median household income, percent of population below poverty line). This was done as a backup in case the rate of non-reporting of income or education was so great that a sufficient sample size could only by achieved through the use of proxy variables.

Thanks to the excellent efforts of the eight participating fire departments, however, this backup proved unnecessary. There were complete data for 439 fire cases and 1,281 non-fire smoker cases. This was still within the target sample size for both groups. In addition, early analyses showed, as expected, that the proxy variables produced much weaker statistical models.

It should be pointed out that there was a problem on how the question on number of cigarettes smoked was asked on the two surveys. The cigarette fire incident survey used categories of up to a pack a day and more than a pack, while the smoker survey used categories of less than a pack a day and a pack a day or more. Analysis by Mathematica on part of their data confirmed that the response of exactly a pack a day was so common that the two classification schemes could not be treated as equivalent. Accordingly, it was necessary to omit this smoker characteristic from the analysis.

The cigarette identifiers collected on the survey forms were length, filter, pack type, and menthol, where length was a category variable of the type used in labeling the cigarette for sale (e.g., King). Taken with brand identification, these identifiers defined the unique packings, and three of them - filter, pack type, and menthol - were directly useable as cigarette characteristics in the analysis. In addition to these items, information on the physical characteristics of the cigarette, supplied by the manufacturers to CPSC, was added to each record by picking up the characteristics defined for the brand and type of cigarette. These cigarette characteristics included density, porosity, circumference, citrate, tobacco column length, and total cigarette length. Another characteristic - amount of tobacco - was calculated based on the circumference, density, and column length and added to the record. And a final characteristic - filter length - was calculated during the analysis.

The increasingly popular generic cigarettes were captured through the use of bar codes and label identification of company of manufacture. As noted earlier, no case was used unless a positive brand identification was possible.

Preliminary Analysis

Earlier sections indicated how results of simple comparisons of fire and non-fire data in past analyses have shaped the choices of cigarette and smoker characteristics to be analyzed in this project. The current effort provided a unique opportunity for more preliminary analysis of this type.

Tables 1-2 present simple comparisons, one variable at a time, of the fire and non-fire groups with respect to four smoker characteristics and three cigarette characteristics, respectively. The characteristics are the ones that were defined as categorical (as opposed to continuous) variables in the data collection. Statistical measures of the degree of difference between fire and non-fire groups on each of these characteristics are also presented. Note that these analyses are able to use more of the collected data, because only cases missing data on the one characteristic analyzed need be excluded.

Keep in mind that these are basic one variable tabulations and do not control for other smoker or cigarette characteristics.

Males accounted for a higher percentage of smokers who had cigarette fires (61.5%) than of smokers who did not have fires (52.6%). Whites accounted for a smaller percentage of smokers who had cigarette fires (48.0%) than of smokers who did not have fires (58.5%), while smokers who didn't complete high school comprised a much higher percentage of smokers who had cigarette fires (38.3%) than of smokers who did not have fires (18.4%).

Smokers with the lowest household incomes (under $10,000) accounted for a considerably higher share of smokers who had cigarette fires (45.6%) than of smokers who did not have fires (16.6%). Also, smokers with household incomes in the range of $10,000-19,999 comprised a larger percentage of smokers who had fires (30.3%) than of smokers who did not have fires (20.1%).

Unfiltered cigarettes were the choice of more smokers who had cigarette fires (10.5%) than of smokers who did not have fires (3.0%). Cigarettes from a soft pack were used by 73.8% of the smokers who did not have fires, but by 86.0% of the smokers who had fires. Menthol cigarettes were used about equally between the two groups. (This is the only one of the seven comparisons without a statistically significant difference.)

Table 3 displays a breakdown by city for the non-fire and fire groups. Some cities accounted for a considerably smaller percentage of the fire group than of the non-fire group. In particular, Columbus accounted for 11.0% of the non-fire group, but only 3.0% of the fire group.

City is neither a smoker nor a cigarette characteristic, but it may be correlated with any or all of those characteristics. It is not necessary for the fire and non-fire shares to match for each city, because the smokers in one city may, for a variety of reasons, be more likely to have fires than the smokers in another city. NFPA used reports to NFIRS and other information to estimate the completeness of reporting, which ranged from 80% or higher (Baltimore, Cleveland) to around 50% (Denver, Houston, Philadelphia, Portland) to around 25% (Columbus), with one undetermined (Dallas).

An appropriate means of addressing this is to insert dummy variables for the cities, and this was done. (A dummy variable is a variable whose only values are 1 and 0 and which can be used to indicate the presence or absence of a condition. In this context, there were dummy variables for all but one of the cities. For example, when the Baltimore variable was equal to 1, the fire or smoker was from Baltimore; otherwise, the fire or smoker was not from Baltimore. When all seven city variables were equal to 0, that indicated the city was Dallas.)

Table 4 compares the non-fire and fire groups with respect to the eight continuous-variable cigarette characteristics and the one continuous-variable smoker characteristic. The difference in means, relative to the standard errors, indicates the significance of the differences between fire and non-fire groups, through the Student's t statistic.

For filtered cigarettes only, filter length had a mean of 25.14 mm for the fire group and 23.52 mm for the non-fire group ($t = 7.69$, $p < .001$). Porosity had a mean of 32.44 for the fire group and 30.96 for the non-fire group ($t = 1.84$, $p < .1$). Amount of tobacco had a mean of 772.72 for the fire group and 748.50 for the non-fire group ($t = 5.37$, $p < .001$). Tobacco column length had a mean of 65.95 for the fire group and 64.98 for the non-fire group ($t = 2.91$, $p < .005$).

The last simple analyses prepared show the continuous cigarette characteristics in categorical form (see Table 5). These tables illustrate a general point cited earlier. Some characteristics that were significant risk factors may not be statistically significant in the field study because the range of values among commercial cigarettes is narrow, compared with the range examined in experimental cigarettes. Tobacco density and cigarette circumference are apt examples of such narrow ranges. The experimental cigarettes were made to show extremes: for tobacco density, the low range was from 130 to 170 mg/cc, the high range from 230 to 320 mg/cc. By comparison, among the commercial cigarettes identified in the field study, there were no cigarettes at the 130 to 170 low-density range, and very few with values under 200. More than 90 percent of the cigarettes were in the high-density range as defined for the experimental cigarettes.

Regarding circumference, the experimental cigarettes were either 21 mm or 25 mm in circumference. As can be seen from Table 5, only three percent of the non-fire cases and one percent of the fire cases involved cigarettes with a circumference within the range of the low circumference experimental cigarettes. About 90 percent were between 24.5 and 25.4 mm, the value for the large-circumference cigarettes.

These results for simple frequency counts and means for one variable at a time are one way to get an idea of how these cigarette and smoker characteristics are related to the risk of a cigarette fire. However, a better and more rigorous way is to examine a multivariate model that will adjust or control for the effects of other smoker and cigarette characteristics.

Dummy city variables were introduced so as to cover all the cities. Note that city was put in the model not to compare cities, but to control for possible differences in completeness of reporting of cigarette fires and other possible differences by city.

Table 6 shows how the categorical variables were coded.

One additional rule was followed in constructing the multivariate model. The model fits linear relationships of the variables, but three of the cigarette characteristics - total cigarette length, tobacco column length, and filter length - are linearly related. The first is the sum of the other two. This argues against using all three at one time. Of the three, the last two variables refer to more clear-cut, homogeneous physical properties - the tobacco and the filter - so they were the ones retained.

Table 1
Categorical Smoker Characteristics, by Group

A. Gender	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Male	846	52.6	347	61.5
Female	764	47.4	217	38.5
	1,610	100.0	564	100.0

Chi-square = 13.6 df = 1 p =.000 1 missing non-fire case
0 missing fire cases

B. Race	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
White	937	58.5	270	48.0
Nonwhite	664	41.5	293	52.0
	1,601	100.0	563	100.0

Chi-square = 18.9 df = 1 p = .000 10 missing non-fire cases
1 missing fire case

C. Education	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
High school graduate	1,260	81.6	300	61.7
Not high school graduate	285	18.4	186	38.3
	1,545	100.0	486	100.0

Chi-square = 81.6 df = 1 p = .000 66 missing non-fire cases
78 missing fire cases

D. Income	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under $10,000	223	16.6	223	45.6
$10,000-19,999	270	20.1	148	30.3
$20,000 and up	847	63.2	118	24.1
	1,340	100.0	489	100.0

Chi-square = 243.0 df = 2 p = .000 271 missing non-fire cases
75 missing fire cases

Chi-square is a test statistic of the hypothesis that both the non-fire and fire groups have a similar distribution with respect to the characteristic being analyzed. df refers to degrees of freedom. p is the probability that the chi-square statistic would be as large as or larger than the value shown if there were in fact no difference between the two distributions.

Non-fire refers to people who smoke and did not have a fire. Fire refers to people who smoke and did have a fire.

Table 2
Categorical Cigarette Characteristics, by Group

A. Filter	Non-fire		Fire	
	Number	Percent	Number	Percent
Filtered	1,562	97.0	505	89.5
Unfiltered	49	3.0	59	10.5
	1,611	100.0	564	100.0

Chi-square = 48.7 df = 1 p = .000 0 missing cases

B. Pack	Non-fire		Fire	
	Number	Percent	Number	Percent
Soft	1,189	73.8	485	86.0
Hard	422	26.2	79	14.0
	1,611	100.0	564	100.0

Chi-square = 35.0 df = 1 p = .000 0 missing cases

C. Menthol	Non-fire		Fire	
	Number	Percent	Number	Percent
Nonmenthol	978	60.7	331	58.6
Menthol	633	39.3	233	41.4
	1,611	100.0	564	100.0

Chi-square = 0.7 df = 1 p = .399 0 missing cases

Chi-square is a test statistic of the hypothesis that both the non-fire and fire groups have a similar distribution with respect to the characteristic being analyzed. df refers to degrees of freedom. p is the probability that the chi-square statistic would be as large as or larger than the value shown if there were in fact no difference between the two distributions.

Non-fire refers to people who smoke and did not have a fire. Fire refers to people who smoke and did have a fire.

Table 3
City, by Group

City	Non-fire Number	Non-fire Percent	Fire Number	Fire Percent
Baltimore	203	12.6	110	19.5
Cleveland	139	8.6	78	13.8
Columbus	177	11.0	17	3.0
Dallas	204	12.6	75	13.3
Denver	99	6.2	26	4.6
Houston	296	18.4	68	12.1
Philadelphia	359	22.3	133	23.6
Portland	134	8.3	57	10.1
	1,611	100.0	564	100.0

Chi-square = 68.6 df = 7 p = .000

Chi-square is a test statistic of the hypothesis that both the non-fire and fire groups have the same distribution with respect to the characteristic being analyzed. df refers to degrees of freedom. p is the probability that the chi-square statistic would be as large as the value shown if there were in fact no difference between the two distributions.

Non-fire refers to people who smoke and did not have a fire. Fire refers to people who smoke and did have a fire.

Table 4
Continuous Cigarette and Smoker Characteristics, by Group

Non-fire

	Number of Cases	Mean	Standard Deviation	Minimum	Maximum
Density (mg/cc)	1,611	239.30	10.68	172	282
Porosity (CORESTA)	1,611	30.96	13.06	10	185
Circumference (mm)	1,611	24.61	.95	17	27
Citrate (%)	1,611	.78	.44	0	3.1
Column length (mm)	1,611	64.98	6.72	51	90
Filter length* (mm)	1,562	25.14	4.51	18	35
Total length (mm)	1,611	89.36	9.42	69	120
Amount of tobacco (mg)	1,611	748.50	82.40	429.44	1,060.70
Age of smoker (years)	1,577	39.11	14.48	12	88

Fire

	Number of Cases	Mean	Standard Deviation	Minimum	Maximum
Density (mg/cc)	564	240.02	10.79	172	266
Porosity (CORESTA)	564	32.44	17.52	14	185
Circumference (mm)	564	24.76	.44	21	27
Citrate (%)	564	.77	.35	0	3.1
Column length (mm)	564	65.95	6.76	56	88
Filter length* (mm)	505	23.52	3.98	18	35
Total length (mm)	564	87.00	8.32	69	120
Amount of tobacco (mg)	564	772.72	95.09	479.84	1,060.70
Age of smoker (years)	533	42.62	16.33	12	89

*Filter length includes only cases where filtered cigarettes were involved.

Non-fire refers to people who smoke and did not have a fire. Fire refers to people who smoke and did have a fire.

NOTE: Significance tests for the difference between two means are phrased in terms of the probability (p) that a difference in means would be at least as large as the measured difference if the true population means were equal. In this table, porosity was significant at the 0.1 level; column length was significant at the 0.005 level; and filter length, circumference and amount of tobacco were each significant at the 0.001 level. Density, citrate, and total length were not significant at the 0.1 level.

Table 5
Continuous Characteristics in Categorical Form

A. Density (mg/cc)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 170	0	0.0	0	0.0
170-199	22	1.4	9	1.6
200-229	84	5.2	26	4.6
230-239	977	60.6	357	63.3
240-259	479	29.7	158	28.0
260-282	49	3.0	14	2.5
Over 282	0	0.0	0	0.0
Total	1,611	100.0	564	100.0

B. Porosity (CORESTA)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 20	247	15.3	81	14.4
20-29	656	40.7	244	43.3
30-39	468	29.0	133	23.6
40-59	232	14.4	99	17.5
Over 59	8	0.5	7	1.2
Total	1,611	100.0	564	100.0

C. Circumference (mm)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 17.0	0	0.0	0	0.0
17.0-19.9	13	0.8	0	0.0
20.0-22.4	33	2.0	4	0.7
22.5-24.4	135	8.4	31	5.5
24.5-25.4	1,423	88.3	525	93.1
25.5-27.4	7	0.4	4	0.7
Over 27.4	0	0.0	0	0.0
Total	1,611	100.0	564	100.0

D. Citrate (%)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 0.6	430	26.7	117	20.7
0.6	564	35.0	195	34.6
0.7-0.9	369	22.9	181	32.1
Over 0.9	248	15.4	71	12.6
Total	1,611	100.0	564	100.0

Table 5
Continuous Characteristics in Categorical Form (Continued)

E. Column Length (mm)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 60	353	21.9	68	12.1
60-69	938	58.2	362	64.2
Over 69	320	19.9	134	23.8
Total	1,611	100.0	564	100.0

F. Filter Length (mm)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
No filter	49	3.0	59	10.5
18-25	698	43.3	320	56.7
Over 25	864	53.6	185	32.8
Total	1,611	100.0	564	100.0

G. Total Length (mm)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 80	148	9.2	60	10.6
80-84	869	53.9	356	63.1
85-99	383	23.8	97	17.2
Over 99	211	13.1	51	9.0
Total	1,611	100.0	564	100.0

H. Amount of Tobacco (mg)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 600	23	1.4	7	1.2
600-699	390	24.2	67	11.9
700-799	775	48.1	322	57.1
800-899	383	23.8	129	22.9
900-999	11	0.7	2	0.4
Over 999	29	1.8	37	6.6
Total	1,611	100.0	564	100.0

I. Smoker Age (years)	**Non-fire**		**Fire**	
	Number	**Percent**	**Number**	**Percent**
Under 35	736	45.7	222	39.4
35-64	766	47.5	269	47.7
Over 64	109	6.8	73	12.9
Total	1,611	100.0	564	100.0

Table 6
Codes for the Categorical Cigarette Characteristics and Smoker Characteristics Used in the Logistic Regression Analysis

Variable		Abbreviation
Filter	(1 = filter 0 = unfiltered)	Filter
Menthol	(1 = menthol 0 = non-menthol)	Menthol
Pack	(1 = soft 0 = hard)	Pack
Gender	(1 = male 0 = female)	Gender
Race	(1 = nonwhite 0 = white)	Race
Education	(1 = not high school graduate 0 = high school graduate)	Educ
Income1	(1 = income less than $10,000 0 = income not less than $10,000)	Inc1
Income2	(1 = income $10,000 - $19,999 0 = income not in the range of $10,000 - $19,999) (Note: When income1 and income2 are both 0, the reference group income is $20,000 or more.)	Inc2
City codes	(c1, c2, c3, c4, c5, c6, c7 are the codes representing dummy variables used for the cities c1 = 1 if city is Baltimore c2 = 1 if city is Cleveland c3 = 1 if city is Columbus c4 = 1 if city is Denver c5 = 1 if city is Houston c6 = 1 if city is Philadelphia c7 = 1 if city is Portland (Note: When c1 to c7 are all 0, the reference city is Dallas.)	Same as variable used

General Logistic Regression Model

The logistic regression model has been well-established for years, particularly for dose-response studies. Our situation is analogous to a dose-response problem. As a hypothetical example, if the "dose" of, say, citrate is zero, then the risk of a cigarette fire will be a function of the other variables. If the "dose" of citrate is then increased (and if the amount of citrate were positively correlated with risk), then the risk would rise, typically following an S-curve, with the effect on risk low at low doses, while at high doses, the risk of fire could approach certainty.

This technique requires a dichotomous outcome variable. In our situation, the possible outcomes are that the smoker had a cigarette fire or did not. Smoker and cigarette characteristics are the independent variables whose effects are to be examined. The model takes this form[6]:

$$\text{Prob (cigarette fire)} = \frac{\exp(Z)}{1 + \exp(Z)}$$

where exp (Z) is the base of the natural logarithm raised to the power of Z,

$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$,

β_0 is the intercept parameter,

X_i are independent variables,

β_i are regression parameters.

It should be noted that because sampling was disproportionate in the two groups (i.e., the fire group is overrepresented), the logistic equation cannot be used in this report to predict the probability of having a fire.

An important feature of the logistic regression method is that it provides a particularly simple form for calculating odds ratios. The odds is the quotient of the probability of fire to the probability of no fire. The odds ratio is the ratio of the odds corresponding to two different values of the independent variable. (It is a format for expressing probabilities most familiar in sports betting contexts, e.g., 10-to-1 odds that Team A will beat Team B.)

$$\text{Prob (no cigarette fire)} = 1 - \text{Prob (cigarette fire)}$$

[6]Marija J. Norusis/SPSS Inc., *SPSS Advanced Statistics User's Guide*, Chicago, IL: SPSS Inc., 1990.

Therefore

$$\frac{\text{Prob (cigarette fire)}}{\text{Prob (no cigarette fire)}} = \exp\,(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p)$$

The odds ratio can be used to determine the change in odds as a function of a change of so many units in a variable. For example, if column length is decreased by 10 millimeters, how much do the odds change of having a cigarette fire, after controlling for the effects of other cigarette and smoker characteristics?

Complete Model of Cigarette Fire Risk

The initial model included the following variables:

Cigarette characteristics — filter, filter length, porosity, pack, density, menthol, amount of tobacco, circumference, citrate, and tobacco column length.

Smoker characteristics — gender, age, race, education, and income (two dummy variables).

In addition, city was also included as a variable, using seven dummy variables.

The results of this logistic regression model can be seen as Model 1 in Table 7.

Filter and filter length, porosity, and pack were the cigarette characteristics found significant[7] after controlling for the effects of other cigarette characteristics, the smoker characteristics, and city. Density, menthol, amount of tobacco, circumference, citrate, and tobacco column length were all cigarette characteristics not found to be significant. Gender, education, and income were smoker characteristics found significant. Age and race were not found significant. The city variable was also found significant[8] and will be included in all other models.

Many models were run omitting certain variables. Particular attention was paid to models using different choices and combinations from the group of cigarette characteristics that could have posed multicollinearity problems, i.e., amount of tobacco and the three variables that make it up - density, tobacco column length, and circumference. In the end, no reason was found to do other than proceed to the most obvious concise model based on Model 1, as shown in the next section. As a further check, forward selection and stepwise regression techniques were used with significance criteria levels specified at the .05 level.[9]

[7]Significance is defined as $p < .05$.

[8]The city variable, as represented by the seven design variables, was found significant at $p < .005$.

[9]*SAS User's Guide: Statistics*, Cary, NC: SAS Institute, 1985.

Forward selection is a variable selection procedure by which variables are inserted one at a time into the model until a satisfactory regression equation is found. In stepwise regression, after each insertion of a variable, there is testing performed to see if any previously inserted variable can now be discarded due to relationships that exist among the variables currently in the model.

Results were similar, and the same four characteristics were found significant: filter, filter length, pack, and porosity.

Also numerous different combinations of characteristics were run, and results indicated that these four characteristics were the ones most often were found to be significant.

Table 7
Model 1 — The Most Complete
Logistic Regression Model

	Beta	p
Intercept	11.3542	.540
Filter[1]	1.8397	.003
Filter length[1]	-.0868	.000
Porosity	.0111	.015
Pack	.6081	.001
Density	-.0336	.171
Menthol	.0080	.960
Amount of tobacco	.0060	.420
Circumference	-.3189	.540
Citrate	-.1353	.440
Column length	-.0386	.660
Gender	.4348	.001
Age	.0044	.332
Race	-.2760	.066
Education	.4357	.003
Income1	1.8089	.000
Income2	1.2039	.000
c1	.1351	.560
c2	-.2034	.407
c3	-1.8605	.000
c4	-.9063	.007
c5	-.5665	.015
c6	-.1281	.550
c7	-.0837	.743

Model chi-square* 372.5 with 23 degrees of freedom
p = .000

Note that the model is based on 439 cases from the fire group and 1,281 from the non-fire group that had complete data (no missing data).

The p value is based on a Wald chi-square statistic for the Beta estimate. Any p value equal to or less than .05 is considered significant.

[1]Filter is reflected in two variables: filter (whether the cigarette is filtered) and filter length (if the cigarette is filtered, how long is the filter). Note in our sample the minimum value for filter length is 18 when a filter is present; this discontinuity explains why two model variables are needed to address filter.

*Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Primary Concise Model of Cigarette Fire Risk

The principal or primary model for the analysis was then run. Called Model 2, it is identical to Model 1, except that its cigarette characteristics were limited to those found significant in Model 1, i.e., filter, filter length, pack, and porosity. See Table 8, which also includes more of the relevant statistics for each characteristic than were shown in Table 7. Note that Model 2 also includes smoker characteristics not found significant, because the objective of the analysis is to control for smoker characteristic effects.

Table 8 includes Beta parameter estimates, standard errors, Wald chi-squares,[10] and p values for each variable. It also includes estimated adjusted odds ratios and 95% confidence intervals around the odds ratios.

The odds ratio is exp (Beta) for a categorical variable like pack. For example, the odds ratio for pack is exp (.6189) = 1.857. This means that after adjusting for the effects of other cigarette characteristics, smoker characteristics, and city in the model, the odds of a cigarette fire increase by 85.7% if the cigarette was from a soft pack rather than a hard pack.

For a continuous variable like porosity, the odds ratio is exp (c beta) where c is the unit of change.[11] For example, the odds ratio for porosity is exp (.0127) = 1.013. This means that for an increase of 1 CORESTA unit of porosity, the odds of a cigarette fire increase 1.3%. However, a unit of change will not have the same meaning for all variables. A more meaningful unit of change for a continuous variable like porosity might be the standard deviation, which is 13.8. If this is used as the unit of change, then the odds ratio for porosity is 1.192. This means that after adjusting for all the other variables in the model, the odds of having a cigarette fire increase 19.2% for every increase of 13.8 CORESTA units (one standard deviation) in porosity.

Two of the significant cigarette characteristics involve the filter - (presence of) filter and filter length. Therefore, it is necessary to consider them together. In

[10]The Wald chi-square is calculated by dividing the beta estimate by the standard error and squaring the result. Any p value equal to or less than .05 is considered significant.

[11]David W. Hosmer and Stanley Lemeshow, *Applied Logistic Regression*, New York, N.Y.: John Wiley and Sons, 1989.

going from an unfiltered cigarette to a filtered cigarette, one incurs the categorical-variable factor for filter presence (1.2204) and the factor for the length of the filter (-.0743 times filter length). Now the minimum value for filter length is 18(mm). So for a filtered cigarette with filter length of 18 the odds ratio is exp (1.2204 - .0743 x 18) = exp (-.1170) = .890. This means that after adjusting for all of the other variables in the model, a filtered cigarette with filter length 18 mm has an 11.0% (1.00-0.89) lower odds of being involved in a cigarette fire than an unfiltered one. For every additional unit of change (millimeter) in filter length, the odds ratio is exp (- .0743) = .928. This means for every increase of filter length of 1 unit, the odds of a cigarette fire decrease another 7.2% (1.000 - 0.928).

Now consider the odds ratios for the smoker characteristics.

The odds of a cigarette fire are increased by 507% for people with household incomes $10,000 or less compared to people with household incomes of $20,000 or more. The odds of having a cigarette fire are increased by 232% for people with incomes of $10,000 - $19,999 compared to people with household incomes of $20,000 or more. Income is clearly the strongest variable in the model.

The odds ratio for education is 1.540. This means that after adjusting for all of the cigarette characteristics and other smoker characteristics in the model, the odds of having a cigarette fire increase by 54.0% if a person didn't graduate from high school as compared to a person who did graduate. The odds ratio for gender is 1.528. This means that the odds of a cigarette fire for a male was 52.8% higher than for a female, after adjusting for all of the other smoker characteristics and cigarette characteristics.

In terms of the principal purposes of this project, Model 2 answers the question as to whether cigarette characteristics are significant factors in cigarette fire risk after controlling for smoker characteristics. They clearly are.

Of the four cigarette characteristics found to be significant in Model 1 and so used in Model 2 - filter, filter length, porosity, and pack - porosity had been identified as significant in earlier laboratory work and in the pilot study for this project. Filter had also been identified in the pilot study and had been only partially addressed in the earlier laboratory work. Pack type and filter length had not been considered in the earlier studies as characteristics. Pack type is an unusual variable, as noted earlier, and a possible explanation for its strength will emerge in the analyses on interactions among variables.

Tobacco column length had been a strong variable in the pilot study (and proved to be significant in the model of filtered cigarettes only, shown later).

Menthol and citrate had been found to have marginal significance in the earlier studies and were not found significant here. Density and circumference, which were found to be significant in laboratory studies of experimental cigarettes, were not found significant in this analysis. As noted earlier, density and circumference had so little variation in commercially available cigarettes that their potential impact on fire risk was not really measured by this study.

Note that filter, filter length, and porosity all bear on the air intake in some fashion. This suggests a common physical process of combustion bearing on ignition and risk. However, air intake is a complex phenomenon and occurs in a different form in a discarded cigarette than in a cigarette being smoked.

Filtered Cigarettes Only

With the apparent importance of filter, it was considered useful to see what model predicted cigarette fire risk best among filtered cigarettes.

An initial model was run and included nine cigarette characteristics (all except total cigarette length and filter), the smoker characteristics, and city. Four cigarette characteristics were found significant: filter length, pack, porosity, and column length. A model with just those four cigarette characteristics, the smoker characteristics, and city was run and is displayed in Table 9. The new variable added compared to the earlier model is tobacco column length. The odds ratio for column length is 1.034. This means that after adjusting for the effects of the other variables in the model, for each increase of 1 unit (mm) in column length, the odds of a cigarette fire increase by 3.4%.

Interaction Terms

The next model considered included interaction terms. Numerous two-way interactions among the four cigarette characteristics and the smoker characteristics terms in Model 2 were examined. All of the interaction terms were evaluated with all of the variables from Model 2 already in the model.[12] Three interaction terms were found significant: pack x gender ($p < .005$), education x income ($p < .005$), and race x income ($p < .05$). These interactions

[12]Significance was evaluated by comparing model chi-square for Model 2 (main effects) to Model 2 plus the interaction of interest.

were then combined in one model with all of the variables from Model 2 (see Table 10).

The strength of the two interaction terms involving pairs of smoker characteristics are of limited interest, once it is clear that they do not eliminate any cigarette characteristics from significance. Instead, the interaction term of greatest interest is pack x gender. Note that with this term present, both of the primary terms, pack and gender, cease to be significant.

Referring back to the general logistic regression model equations, it may be seen that, in calculating an odds ratios for two characteristics, all other characteristics are held constant. In the case of pack and gender, then, the odds ratios can be calculated based on exp (.0558 pack - .3722 gender + .9931 pack x gender), using the Beta parameters from Table 10. For females, this expression goes from 1.00 to 1.06 when switching from a hard pack to a soft pack, indicating a slight 6% increase in the odds of having a cigarette fire (which was found not to be statistically significant). For males, however, this expression goes from 0.69 to 1.97, for an increase of 185% the odds of having a cigarette fire if the cigarette is from a soft pack rather than a hard pack.

As noted earlier, pack type is an unusual variable. If the higher risk associated with a soft pack is a reflection of the added vulnerability of cigarettes in such packs to crushing, bending, or other impact-related modifications, then it may be plausible that these effects would be greater for males, carrying packs in their pockets, than for females, carrying packs in their purses. This remains speculation, however.

Table 8
Model 2 — Only Significant Cigarette Characteristics from Model 1

	Beta	Standard Error	Wald Chi-square	p	Odds Ratio	95% Confidence Interval
Intercept	- 2.3921	.4489	28.4	.000		
Filter[1]	1.2204	.4850	6.3	.011	*	
Filter length[1]	- .0743	.0163	20.8	.000	.928	(.899, .958)
Porosity	.0127	.0045	8.0	.004	1.013	(1.004, 1.022)
Pack	.6189	.1750	12.5	.000	1.857	(1.317, 2.617)
Gender	.4241	.1302	10.6	.001	1.528	(1.184, 1.972)
Age	.0052	.0045	1.3	.249	1.005	(.996,1.014)
Race	- .2752	.1411	3.8	.052	.759	(.575, 1.001)
Education	.4321	.1455	8.8	.003	1.540	(1.158, 2.048)
Income1	1.8034	.1624	123.3	.000	6.070	(4.415, 8.345)
Income2	1.1998	.1596	56.5	.000	3.319	(2.427, 4.539)
c1	.0880	.2245	0.2	.691		
c2	-.2499	.2419	1.1	.302		
c3	- 1.8740	.3800	24.3	.000		
c4	- .9326	.3332	7.8	.005		
c5	- .5816	.2318	6.3	.012		
c6	- .1910	.2102	0.8	.364		
c7	- .1167	.2533	0.2	.645		

Model chi-square** = 362.7 with 17 degrees of freedom p = .000

Note that the logistic model was based on 439 cases from the fire group and 1,281 from the non-fire group that had complete data (no missing data).

[1]Filter is reflected in these two variables: filter (whether the cigarette is filtered, and filter length (if the cigarette is filtered, how long is the filter); note in our sample the minimum value for filter length is 18 when a filter is present.

The Wald chi-square is calculated by dividing the beta estimate by the standard error and squaring the result. Any p value equal to or less than .05 is considered significant.

The odds ratio is exp (Beta) for a categorical variable like pack. For a continuous variable like porosity the odds ratio is exp (c Beta) where c is the unit of change and in the above table is assumed to be 1, though in the text other options are discussed.

*See discussion in text. Odds ratio for filter is only meaningful in combination with filter length.

**Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Table 9
Logistic Regression Results
Model for Filtered Cigarettes Only

	Beta	Standard Error	Wald chi-square	p	Odds Ratio	95% Confidence Interval
Intercept	- 2.8989	.8437	11.8	.000		
Filter length	- .0957	.0184	27.1	.000	.908	(.877, .942)
Porosity	.0145	.0046	9.9	.002	1.015	(1.005, 1.024)
Pack	.5779	.1767	10.7	.001	1.783	(1.261, 2.520)
Column length	.0339	.0140	5.9	.016	1.034	(1.006, 1.063)
Gender	.4317	.1345	10.3	.001	1.540	(1.183, 2.004)
Age	.0055	.0047	1.4	.237	1.006	(.996, 1.015)
Race	- .3117	.1473	4.5	.034	.732	(.549, .977)
Education	.3641	.1517	5.8	.016	1.439	(1.069, 1.938)
Income1	1.8569	.1676	122.8	.000	6.404	(4.611, 8.894)
Income2	1.2239	.1657	54.5	.000	3.400	(2.457, 4.705)
c1	.1732	.2299	0.6	.451		
c2	- .2298	.2476	0.9	.353		
c3	- 1.9719	.4134	22.8	.000		
c4	- .9263	.3433	7.3	.007		
c5	- .5747	.2388	5.8	.016		
c6	- .1337	.2179	0.4	.539		
c7	- .0664	.2619	0.1	.800		

Model chi-square* = 335.3 with 17 degrees of freedom p = 0.000

Note that the model is based on 397 cases for the fire group and 1,241 for the non-fire group that had complete data on filtered cigarettes (no missing data).

The Wald chi-square is calculated by dividing the beta estimate by the standard error and squaring the result. Any p value equal to or less than .05 is considered significant.

The odds ratio is exp (Beta) for a categorical variable like pack. For a continuous variable like porosity the odds ratio is exp (c Beta), where c is the unit of change and in the above table is assumed to be 1, though in the text other options are discussed.

*Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Table 10
Logistic Regression Results
Model Including Interaction Terms

	Beta	Standard Error	Wald Chi-square	p
Intercept	- 1.8058	.4647	15.1	.000
Filter[1]	1.4473	.4900	8.7	.003
Filter length[1]	- .0819	.0165	24.6	.000
Porosity	.0122	.0044	7.7	.005
Pack	.0559	.2465	0.1	.821
Pack x Gender	.9931	.3303	9.0	.003
Gender	-.3722	.2987	1.6	.213
Age	.0060	.0046	1.7	.192
Race	- .7365	.2506	8.6	.003
Educ	- .2476	.3808	0.4	.516
Inc1	1.2695	.2329	29.7	.000
Inc2	.9396	.2168	18.8	.000
Educ x Inc1	1.1868	.4378	7.4	.007
Educ x Inc2	.3347	.4558	0.5	.463
Race x Inc1	.6227	.3327	3.5	.061
Race x Inc2	.7920	.3399	5.4	.020
c1	.1228	.2292	0.3	.592
c2	- 1.8923	.2460	0.6	.441
c3	- 1.8627	.3816	23.8	.000
c4	- .8879	.3352	7.0	.008
c5	- .5667	.2345	5.8	.016
c6	- .2116	.2126	1.0	.320
c7	- .1104	.2539	0.2	.664

Model chi-square* = 389.1 with 22 df p = .000

Note that the logistic model was based on 439 cases for the fire group and 1,281 for the nonfire group that had complete data (no missing data).

[1]Filter is a combination of these two variables: filter (whether the cigarette is filtered) and filter length (if the cigarette is filtered, how long is the filter); note that in our sample the minimum value for filter length is 18 when a filter is present.

The Wald chi-square is calculated by dividing the beta estimate by the standard error and squaring the result. Any p value equal to or less than .05 is considered significant.

*Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Conclusions

1. After controlling for all smoker characteristics and city, logistic regression modeling showed four cigarette characteristics to be significant - filter, filter length, porosity, and pack type. This means there are already commercially available cigarettes that exhibit reduced ignition propensity when one controls for smoker characteristics.

2. Filter, filter length and porosity all affect air intake, which therefore appears to be an important physical element in the combustion process associated with risk.

3. Analysis limited to only filtered cigarettes showed the same characteristics to be significant, plus tobacco column length.

4. Extension of the analysis to two-way interaction terms did not change any of the conclusions on which cigarette characteristics are important but did indicate that the role of pack type was different for men vs. women.

Appendix

Sensitivity Analysis

In addition to the analyses presented to this point, there were some methodological issues related to the project that needed to be addressed.

One concerned the apparent low response rate of Columbus with respect to the number of cigarette fires reported. A run was made of Model 1 (i.e., with all 10 cigarette characteristics) without Columbus. The same four cigarette characteristics found significant with all the cities were again significant: filter, filter length, porosity, and pack. It is also worth noting that the Beta coefficients for these four cigarette characteristics and the smoker characteristics earlier found significant changed very little. (See Table A-1.)

Another data quality question arose concerning the reliability of self-reported and other-reported data for the smoker survey, that is, self-reported answers vs. answers by others in the household. Mathematica did a separate report that addresses that topic and related issues. Mathematica concluded, "In the eventual analysis of the effects of smoker and cigarette characteristics on the likelihood of a smoking fire we believe that more credible results would be obtained if the proxy cases [i.e., reports by others] are included than if they are excluded. While excluding proxy cases would eliminate any potential biases due to misreporting by proxies, we [Mathematica] believe that these biases are likely to be relatively minor compared to the biases that would be created by deleting these cases."[13]

Mathematica also recommended that estimates be made with proxy cases removed, as a sensitivity test. Model 1 was then run with data from the smoker survey that included only self-report cases. It therefore excluded both smoker proxy and non-smoker proxy cases. This reduced the sample size for the non-fire smoker group from 1,281 to 709 cases. In examining the results from Model 1 with proxies excluded, three of the cigarette characteristics found significant earlier were again significant, filter, filter length, and pack, while porosity just missed ($p = .054$). The size of the Beta coefficient did change somewhat, but that is

[13]Donna Eisenhower, John Hall, and Randy Brown, *Self-Proxy Comparisons for the Cigarette Fire Safety Survey Final Report,* submitted by Mathematica Policy Research, Inc., Princeton, NJ to U.S. Consumer Product Safety Commission, Bethesda, MD, February, 1993.

to be expected with the change in sample size. (See Table A-2.)

Another issue is the sample design of the smoker survey. As noted earlier, the sample was not a simple random sample. Smokers were picked not independently but rather as part of a larger unit known as a cluster (in this case, household). If a household was selected, all smokers in the household were included in the sample. To determine the effect of clustering, two logistic regression runs were compared: (1) one where the sample selected was assumed to be chosen by a simple random sample, and (2) one where the sample selected was based on a clustered sample design, with household as the cluster. All of the logistic regression runs that have been made thus far were made under the simple random sample assumption.

A software statistical program, SUDAAN, developed by Research Triangle Institute, was used to perform logistic regression on a clustered sample design. The difference between the two sample design assumptions should be reflected in the standard errors of the Beta coefficients. The models examined were Model 2, with the four cigarette characteristics, both with and without the three interaction terms. Results for most of the variables seem to indicate little difference in the standard errors for the two sample designs (See Tables A-3 and A-4.) In particular, none of the cigarette or smoker characteristics that had been significant ceased to be significant.

Another issue is the missing data question. As was noted earlier, before the initial logistic regression run, the data base contained 564 cases for the fire group and 1,611 cases for the non-fire group. The data base of cases with complete data (no missing data) has 439 cases for the fire group and 1,281 for the non-fire group. This means that 79.1% of the original 2,175 cases were used in most of the logistic regression runs. This is more than adequate to do the analysis, but a question still arises about the 20.9% cases where there were one or more missing data items.

To help reduce that number and see whether results are affected, the following sensitivity analysis was done. The two data items where there was the largest number of missing-data cases were education and income. In this sensitivity analysis model, we decided to include cases where there was missing information only for either education or income (i.e., no missing data for any of the other variables). In the latter cases, two additional dummy variables were used to identify missing data for education and income, respectively.

This resulted in an expanded sample of 506 cases for the fire group, and 1,540 for the non-fire group, and it reduced the percentage of missing cases to just 5.9%. Model 1 with all ten cigarette characteristics was run with this data set. The same four cigarette characteristics again were found significant: filter, filter length, pack, and porosity. It is also worth noting that the Beta coefficients for the four cigarette characteristics changed very little from the original sample. (See Table A-5.)

Table A-6 displays a set of ratios directly from the data, showing the number of fires divided by the number of non-fire smokers, according to a three-way cross-tabulation by filter length, column length, and income. (Filter length was consistently one of the significant cigarette characteristics, column length was significant when only filtered cigarettes were studied, and income was the strongest smoker characteristic.) This display shows that even with the data divided into this many categories, the relationships between fire risk and certain characteristics are strong enough to show up in simple-to-read displays. Although these tabulations do not control for the effects of other smoker and cigarette characteristics, they agree with the logistic regression analysis that does control for other smoker and cigarette characteristics.

Table A-7 shows a similar display for the pack type and gender variables. As in the discussion of the interaction term on pack and gender, this simpler, less controlled analysis shows a high odds ratio for soft vs. hard packs for males smokers (2.53) and a smaller odds ratio for soft vs. hard packs for female smokers (1.79). Again, no other smoker or cigarette characteristics are controlled for in these calculations.

Other Patterns of Fire Cases

There were 23 civilian deaths and 57 civilian injuries in the 564 fire cases. Two of the deaths occurred in one incident, and 20 of the civilian injuries occurred in multiple-injury incidents - five incidents with two injuries, two incidents with three injuries, and one incident with four injuries.

Tables A-8 and A-9 display the patterns by form and type of material first ignited. Most fires began with the ignition of mattresses or bedding, upholstered furniture, or trash.

Table A-1
Logistic Regression
Model 1 Without Columbus

	Beta	p
Intercept	10.4679	.571
Filter	1.9849	.001
Filter length	-.0899	.000
Porosity	.0111	.015
Pack	.6039	.001
Density	-.0319	.193
Menthol	-.0211	.898
Amount of tobacco	.0055	.457
Circumference	-.2999	.564
Citrate	-.1715	.341
Column length	-.0333	.700
Gender	.4034	.003
Age	.0047	.313
Race	-.2592	.087
Education	.4331	.004
Income1	1.8016	.000
Income2	1.2308	.000
c1	.1419	.541
c2	-.2001	.415
c4	-.9154	.006
c5	-.5683	.015
c6	-.1251	.560
c7	-.0783	.759

Model chi-square* = 323.7 with 22 degrees of freedom p = .000

Note that the logistic model was based on 429 cases from the fire group and 1,142 from the non-fire group that had complete data (no missing data).

The p value is based on a Wald chi-square statistic for the Beta estimate. Any p value equal to or less than .05 is considered significant.

*Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Table A-2
Logistic Regression
Model 1 Excluding Proxy Cases*

	Beta	p
Intercept	11.1965	.545
Filter	2.6830	.001
Filter length	-.1147	.000
Porosity	.0094	.054
Pack	.4313	.037
Density	-.0326	.188
Menthol	.0237	.898
Amount of tobacco	.0055	.464
Circumference	-.3032	.559
Citrate	-.0794	.683
Column length	-.0323	.713
Gender	.8069	.000
Age	-.0002	.959
Race	-.2536	.135
Education	.6199	.000
Income1	1.6401	.000
Income2	1.1031	.000
c1	.3990	.140
c2	.0158	.900
c3	-1.7910	.000
c4	-.9253	.010
c5	-.5350	.039
c6	-.0960	.691
c7	.2026	.491

Model chi-square[1] = 328.2 with 23 degrees of freedom p = 0.000

*This model for the smoker survey (non-fire) includes only data from smoker self-report cases. It excludes smoker proxy and nonsmoker proxy cases.

Note that the logistic model was based on 439 cases from the fire group and 709 from the non-fire group that had complete data (no missing data).

The p value is based on a Wald chi-square statistic for the Beta estimate. Any p value equal to or less than .05 is considered significant.

[1]Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Table A-3
Logistic Regression
Comparison of Results
Based on Two Sample Designs

	Random Sample Standard Error	Cluster Sample Standard Error
Intercept	.4306	.44
Filter	.4856	.50
Filter length	.0164	.02
Porosity	.0046	.01
Pack	.1749	.18
Gender	.1304	.13
Age	.0045	.00
Race	.1413	.15
Education	.1458	.15
Income1	.1627	.17
Income2	.1599	.17
c1	.2246	.24
c2	.2423	.26
c3	.3802	.38
c4	.3333	.34
c5	.2321	.23
c6	.2106	.21
c7	.2535	.29

As noted in the text, the sample of smokers (non-fire) was based on a clustered sample. Smokers were not picked independently, but rather as part of a larger unit known as a cluster (in this case household). To determine the effect of clustering, two logistic regression runs were made: (1) where the sample selected was assumed to be chosen by a simple random sample and (2) where the sample selected was based on a clustered sample design and household is the cluster. The difference between the two sample design assumptions should be reflected in the standard errors.

The results for the cluster sample design were based on SUDAAN, a software statistical program developed by Research Triangle Institute, which performs logistic regression on complex sample designs.

Table A-4
Logistic Regression
Comparison of Results of Model
Including Interaction Terms
Based on Two Sample Designs

	Random Sample Standard Error	Cluster Sample Standard Error
Intercept	.4355	.44
Filter	.4906	.50
Filter length	.0166	.02
Porosity	.0045	.00
Pack	.2466	.25
Pack x Gender	.3306	.32
Gender	.1457	.14
Age	.0046	.00
Race	.2508	.26
Education	.3812	.38
Income1	.2333	.25
Income2	.2173	.23
Educ x Inc1	.4382	.45
Educ x Inc2	.4565	.46
Race x Inc1	.3329	.36
Race x Inc2	.3403	.36
c1	.2293	.24
c2	.2464	.27
c3	.3818	.38
c4	.3354	.35
c5	.2349	.23
c6	.2130	.21
c7	.2542	.29

As noted in the text, the sample of smokers (non-fire) was based on a clustered sample. Smokers were not picked independently, but rather as part of a larger unit known as a cluster (in this case household). To determine the effect of clustering, two logistic regression runs were made: (1) where the sample selected was assumed to be chosen by a simple random sample and (2) where the sample selected was based on a clustered sample design and household is the cluster. The difference between the two sample design assumptions should be reflected in the standard errors.

The results for the cluster sample design were based on SUDAAN, a software statistical program developed by Research Triangle Institute, which performs logistic regression on complex sample designs.

Table A-5
Logistic Regression Results
Model 1 (Expanded Sample Size)

	Beta	p
Intercept	15.5787	.370
Filter	1.8884	.001
Filter length	-.0946	.000
Porosity	.0107	.007
Pack	.5705	.001
Density	-.0388	.100
Menthol	.0472	.750
Amount of tobacco	.0075	.290
Circumference	-.4423	.373
Citrate	-.0793	.621
Column length	-.0548	.505
Gender	.3134	.009
Age	.0073	.079
Race	-.2159	.117
Education	.4454	.001
Education (Missing)	.7787	.008
Income1	1.7225	.000
Income2	1.1430	.000
Income (Missing)	-.0969	.661
c1	.0722	.740
c2	-.0204	.930
c3	-1.4995	.000
c4	-.6756	.028
c5	-.4144	.061
c6	.0402	.843
c7	.0032	.998

Model chi-square* = 427.8 with 25 degrees of freedom   p = 0.000

All of our logistic model runs had included only those cases from the fire and non-fire group where there was complete data (no missing data). In this model, we also included cases where there was missing information only for either education or income. Thus, this logistic regression model was based on 506 cases from the fire group and 1,540 from the non-fire group.

The p value is based on a Wald chi-square statistic for the beta estimate. Any p value equal to or less than .05 is considered significant.

*Model chi-square is the difference between -2 log likelihood for all the parameters in the model and -2 log likelihood with only the intercept in the model.

Table A-6
Ratio of Fires to No Fires
According to Column Length, Filter Length, and Income

Cigarette Characteristic	Income: Under $10,000	Income: $10,000 - 19,999	Income: At Least $20,000
Total	**1.00**	**.66**	**.14**
Filter length			
0 mm	1.91	1.50	.59
18 - 25 mm	1.18	.65	.20
26 - 35 mm	.70	.38	.06
Column length under 60 mm	**.46**	**.53**	**.08**
Filter length			
0 mm	--	--	--
18 - 25 mm	.45	1.00	.22
26 - 35 mm	.46	.45	.06
Column length 60 - 69 mm	**1.10**	**.57**	**.15**
Filter length			
0 mm	2.00	3.50	.40
18 - 25 mm	1.26	.62	.48
26 - 35 mm	.52	.30	.07
Column length over 69 mm	**1.17**	**.50**	**.17**
Filter length			
0 mm	1.88	1.10	.86
18 - 25 mm	--	--	--
26 - 35 mm	1.02	.39	.14

Table A-7
Ratio of Fires to No Fires
According to Pack and Gender

Pack Type	Gender: Male			Gender: Female		
	Fire	No Fire	Ratio	Fire	No Fire	Ratio
Hard	45	231	.19	34	191	.18
Soft	302	615	.49	183	573	.32

Table A-8
Fire Cases, by Form of Material First Ignited

Form of Material First Ignited	Number of Cases	
Mattress or bedding	256	(45.4%)
Upholstered furniture	143	(25.4%)
Trash	47	(8.3%)
Clothing	20	(3.5%)
Floor covering	14	(2.5%)
Papers	8	(1.4%)
Curtain or drapery	7	(1.2%)
Multiple items	7	(1.2%)
Structural element or framing	6	(1.1%)
Box or bag	6	(1.1%)
Other known form of material*	44	(7.8%)
Unknown form of material	6	(1.1%)
Total cases	564	(1.1%)

*Each specific form accounted for 4 or fewer cases.

Table A-9
Fire Cases, by Type of Material First Ignited

Type of Material First Ignited	Number of Cases	
Cotton, rayon, or cotton fabric or furnished goods	306	(54.3%)
Man-made fabric, or finished goods	77	(13.7%)
Untreated, uncoated paper	54	(9.6%)
Unknown-type fabric, textile, or fur	21	(3.7%)
Sawn wood	18	(3.2%)
Polyurethane	9	(1.6%)
Tobacco	8	(1.4%)
Multiple items	8	(1.4%)
Polyvinyl	7	(1.2%)
Unknown-type plastic	5	(0.9%)
Other known type of material*	38	(6.7%)
Unknown type of material	13	(2.3%)
Total cases	564	(100.0%)

*Each specific type accounted for 4 or fewer cases

Contract No.: CPSC-C-92-1001, Task Orders 003 and 005
MPR Reference No.: 8071

CIGARETTE FIRE SAFETY SURVEY

FINAL REPORT

OMB #3041-0110

January, 1993

Authors:

Donna Eisenhower
Peter Forbes
John Hall

Submitted to:

U.S. Consumer Product Safety Commission
5401 Westbard Avenue
Bethesda, Maryland 20816

Submitted by:

Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, N.J. 08543-2393
(609) 799-3535

Project Officer:
William Zamula

Project Director:
Donna Eisenhower

CONTENTS

Chapter		Page
I.	OVERVIEW OF THE STUDY	A-1
	1.1 Background	A-1
	1.2 Role of the Contractors	A-3
II.	INSTRUMENT DESIGN AND BRAND LIST FOR COMPARISON GROUP	A-5
III.	SAMPLING DESIGN AND PROCEDURES	A-7
	3.1 Target Population	A-7
	3.2 The Sample Frame and Its Coverage	A-7
	3.3 Sample Size	A-8
	3.4 Random Digit Dialing Methodology	A-9
	3.5 Stratification and Allocation of the Sample	A-10
IV.	DATA COLLECTION METHODOLOGY FOR THE COMPARISON GROUP	A-19
	4.1 Description and Schedule	A-19
	4.2 Response Rates Defined	A-22
V.	DATA PROCESSING AND ENTRY	A-27
VI.	ADDITION OF CENSUS AND OTHER DATA	A-29
	6.1 Census Tracts	A-29
	6.2 Census Data	A-30
	6.3 Other Merged Data	A-30
VII.	USING THE DATA	A-33
	7.1 Limitations in Using the Data	A-33
	7.2 Need for Weighting and Imputation	A-36
	7.3 Computing Sampling Errors	A-39
VIII.	FREQUENCY AND CROSS-TABULATIONS OF THE SURVEY DATA	A-45
APPENDIX	SIMPLIFIED VERSION OF THE QUESTIONNAIRE	

TABLES

Table		Page
III.1	HIGHEST MEDIAN INCOME EXCHANGE FOR LOW INCOME STRATUM	A-11
III.2	POPULATION, INTERVIEWS AND SCREENING BY SITE	A-13
III.3	POPULATION, INTERVIEWS AND SCREENING BY STRATUM	A-14
III.4	HOUSEHOLDS INSIDE/OUTSIDE FIRE SERVICE AREA (FSA)	A-16
III.5	TRACTS PARTIALLY INSIDE FIRE SERVICE AREAS	A-17
IV.1	DISTRIBUTION OF SMOKERS IN THE SAMPLE BY TYPE OF INTERVIEW	A-20
IV.2	FINAL STATUS OF HOUSEHOLDS ORIGINALLY SAMPLED--TOTALS	A-21
IV.3	RESPONSE RATES BY SITE	A-25
IV.4	RESPONSE RATES BY INCOME STRATUM	A-26
VI.1	NUMBERS AND PERCENTAGE OF SURVEY CASES MATCHING WITH MANUFACTURER'S DATA	A-32
VII.1	DISTRIBUTION OF HOUSEHOLDS IN SAMPLE AND POPULATION BY STRATUM	A-38

I. OVERVIEW OF THE STUDY

1.1 Background[1]

On August 10, 1990, Congress passed the Fire Safe Cigarette Act of 1990. The act authorized the U.S. Consumer Product Safety Commission (CPSC) to conduct research and assess the feasibility of developing a performance standard to reduce cigarette ignition propensity. As stated in the act, cigarette-ignited fires caused 1,492 deaths in 1987, 3,809 serious injuries, and $395 million in property damage.

The act requires the CPSC to conduct several research projects, including a study of fires started by cigarettes. A fire incident study was planned to identify the kinds of cigarettes, the characteristics of the smokers, and the other factors involved in such fires. The data will help the CPSC determine the relationship between various characteristics of cigarettes and smokers and the risk of fire.

In October 1987, CPSC staff, with the help of the International Association of Fire Chiefs (IAFC), conducted a pilot fire-incident study (Feasibility Study of Obtaining Field Data on Cigarette-Ignited Fires) to determine the relative risk of different cigarette types. Nine fire departments in cities throughout the United States participated. Data were collected over a period of two months and demonstrated the feasibility of a full-scale effort. This pilot study did not include collection for a comparison or control group.

The National Fire Protection Association (NFPA), under contract with the CPSC, has undertaken the current fire-incident study. It has been collecting data on actual incidents of cigarette-related fires at eight sites. The collection began in November 1991, and was completed in December 1992. Personnel of participating fire departments were trained to collect the information in person at the scene of a fire. The smoker or smokers thought to be responsible for igniting the fire were interviewed. If a person was incapacitated because of injuries, a family member was interviewed as a proxy respondent. The number of completed interviews for this part of the study was 564.

---

[1]Excerpts taken from the OMB package submitted in April 1992.

Mathematica Policy Research, Inc., under a subcontract with Market Facts, Inc., was selected to collect data for a comparison group. The potential respondent universe for the comparison survey consisted of cigarette smokers with telephones in the same eight sites across the U.S. The eight sites represented fire department service areas from which the fire-incident cases were sampled. The service areas were typically entire city areas, although some service areas also included census tracts outside the city. The boundaries of the survey areas are discussed in greater detail in section 6.1.

The size of the target population in each site was estimated using national smoking rates by age group:

Site	Estimated Total Number of Cigarette Smokers
Denver, CO	110,200
Baltimore, MD	173,800
Cleveland, OH	119,200
Columbus, OH	151,300
Portland, OR	104,000
Philadelphia, PA	375,600
Dallas, TX	241,800
Houston, TX	388,900
TOTAL	1,664,800

The eight sites contain a total of 2,762,000 households.

The comparison group was interviewed by a random digit dialing survey conducted by a computer-assisted telephone interview (CATI). The survey was conducted with 1,532 households that had at least one smoker. One adult over 18 years of age in each household was interviewed about each smoker in the household. This method produced completed information on 2,206 smokers. Response rates for the comparison survey were 87% overall and 83% for households with smokers. A household where an adult reported that no one in the household smoked was considered a completed interview in the overall response rate. Response rates are discussed in greater detail in section 4.2.

The objective of the project is to plan and carry out the initial phases of a cigarette fire-exposure study that, when combined with a cigarette fire-incident study in the same areas, will enable the CPSC to assess the relative risk of a cigarette fire in relation to cigarette and smoker characteristics.

Analysis comparing the results of the two surveys will employ a logistic regression model or other statistical tests of differences, and the comparison survey data will allow valid use of logistic models. This analysis will be completed sometime in April 1993.

1.2 Role of the Contractors

Abt Associates, Inc., designed the study, including the sample and instrument, for the comparison group. NFPA was responsible for collecting the fire-incidence data and analyzing the entire data set for the study as a whole. Mathematica Policy Research, Inc., under a subcontract with Market Facts, Inc., was responsible for developing a CATI version of the questionnaire and collecting and processing the data for the comparison group. Mathematica with CPSC approval made some minor changes to the questionnaire and some changes to the sample design. Mathematica will be available as a statistical consultant to NFPA for the analysis.

The remainder of the report discusses the results of the survey conducted on the comparison group by Mathematica.

II. INSTRUMENT DESIGN AND BRAND LIST FOR COMPARISON GROUP

A copy of a simplified version of the questionnaire can be found in Appendix A. The questions appear in English and in Spanish. The draft instrument was pretested with nine respondents by Abt Associates before Office of Management and Budget (OMB) approval. Mathematica further pretested the CATI version of the instrument after receiving OMB approval. The latter pretest was used to develop the interviewer training materials for the study. The question-by-question specifications and other selected training materials are included as Appendix B of this report.

The questionnaire determined if the respondent was a household member over 18 years of age. Then the number of household members 12 years of age or older who smoke at least one cigarette a day was determined. Cigarette-related information was then reported by the respondent for every smoker in the household. This information included the following:

- UPC code if available from package
- brand of cigarette
- length
- filtered or not
- soft or hard pack
- mentholated or not

The following information was collected for each smoker:

- amount smoked
- gender
- age
- race
- education

Finally, the following information was collected for the household as a whole:

- number of people residing in the household
- household income
- home ownership
- incidence of fire
- address

The brand list used for this survey was developed under the direction of the CPSC's Division of Epidemiology. The list is part of the CATI instrument. The three-digit brand code includes the manufacturer code as the first digit. Three hundred brands were included in the list and "other, specify" answer choices were also utilized when an exact brand was not specified.

III. SAMPLING DESIGN AND PROCEDURES

In this section we describe the sample design, stratification, and selection procedures used in the CPSC cigarette fire survey. After defining the target population, we present the sampling frame and its coverage. Sample size calculations are then reviewed and the RDD methodology is briefly described. Finally, the stratification and allocation of the sample are explained.

3.1 Target Population

The target population for this study was smokers age 12 and over in households with telephones in the eight fire-service areas. The areas were defined in a variety of ways. Two (Baltimore and Philadelphia) were easily defined, because the city, the fire-service area, and the county were geographically identical. Three fire-service areas (Denver, Portland, and Dallas) were defined using a list of tracts provided by CPSC. Three other fire-service areas were defined in terms of city limits (Cleveland, Houston, and Columbus). Cleveland and Columbus each included three additional tracts outside the city.

3.2 The Sample Frame and Its Coverage

The sample frame for this survey was the set of working telephone exchanges associated with each of the eight service areas. Banks of 100 consecutive numbers with fewer than two listed household numbers were eliminated. In order to limit the sampling frame to these eight fire-service areas, a census tract/telephone exchange analysis was performed for each city. This analysis calculates the proportion of listed telephone numbers in a telephone exchange that were inside a specified geographic region (in this case, the fire-service area.) Using these reports, we were able to identify phone exchanges for households in the service areas.

For Cleveland, Houston, and Columbus, where fire-services area were defined in terms of city boundaries instead of tracts or counties, we used a two-step process to identify the telephone exchanges that served a fire-service area. Before performing the tract/exchange analysis, we identified the census tracts that comprise the city, because census tracts do not always match city boundaries. In Cleveland the match was not problematic. However, in Houston and Columbus many tracts were both inside and outside the city. We had to decide what minimum percentage of the tract had to be inside the city before we would classify it as being inside the city. In the case of Houston, we decided to use tracts that were 25% or more inside the city. For Columbus, we used tracts that were at least 50% inside the city. The tract inclusion criterion was stronger for Columbus than Houston because Columbus tracts matched city boundaries poorly, and we wanted to reduce calls to households outside the city.

Creating a list of telephone numbers for households inside a specified urban area is inherently difficult, because telephone exchanges do not perfectly overlap with city or census tract boundaries. The best possible frame would be a list of telephone numbers for all telephone households in the service areas and would exclude all others. Because that frame was not available, we had to decide which exchanges to include to achieve a high degree of coverage of households inside a service area while minimizing coverage of households outside it. In order to do that, some exchanges with listed numbers inside the service area were excluded from the frame. But few of the numbers in the excluded exchanges were within the service area, which was covered well by the remaining exchanges. In each service area, 95% or more of the listed households were on the frame of included exchanges.

3.3 Sample Size

The sampling design called for collecting data for all smokers age 12 years and older in 1,500 households. The survey design allowed for proxy reports by any adult about all smokers in the household. The assumptions Abt Associates used to determine the number of telephone numbers needed to produce interviews with 1,500 "smoker households" were:

- proportion of possible telephone numbers that were active, household numbers in sampled exchanges =.55
- smoking prevalence rate = 1/1.74 = .575

- response rate for smoker households = .60

On the basis of these assumptions, we estimated that the number of phone numbers needed to produce interviews with 1,500 smoker households would be $1{,}500/(.55 * .575 * .6) \approx 8000$. The working household telephone rate and response rate were conservative estimates based on experience with previous telephone surveys. The smoking rate was provided by Abt.

The observed working rate for RDD telephone numbers and the achieved response rate were higher than the rates used to determine the maximum number of sample telephone numbers needed. However, the rate for incidence of smoking was less than half the rate estimated by Abt. We found that an average of 3.9 households had to be screened in order to find one household in which anyone smoked. Consequently, the study required more sample telephone numbers than was originally expected. In all, 11,639 sample telephone numbers were used.[1]

3.4 Random Digit Dialing Methodology

The random digit dialing (RDD) sample for this study was created using a single-stage systematic selection procedure designed to produce a sample that has equal probability of selection within strata.[2] The selection procedure is "list-assisted" in the sense that it identifies "working 100-banks" of telephone numbers, from which the sample numbers are selected. (A 100-bank is a set of 100 telephone numbers associated with the eight digits beginning a phone number. A 100-bank is identified as "working" if it

---

[1]The observed rates were:

- working household telephone number rate =.603
- smoking prevalence rate = 1/3.89 = .257
- response rate for smokers = .833

[2]The sampling procedure was designed to ensure (within strata) an equal and known probability of selection for all residential telephone numbers. However, there was no guarantee that each and every telephone household had an equal probability of being selected, because roughly 7% of telephone households were served by two or more telephone numbers. As a practical matter it is difficult to measure the number of lines becasue respondents confuse the number of lines with the number of outlets.

contains two or more listed households.) Telephone numbers generated by the procedure are checked against a list of working 100-banks. Those in nonworking banks are not included in the sample. This procedure improves survey efficiency by reducing the number of calls made to nonworking telephone numbers.

3.5 Stratification and Allocation of the Sample

Stratifying telephone exchanges by income is not straightforward, because income estimates are available for geographically defined areas-counties, zip codes, and census tracts-that do not correspond to the areas covered by telephone exchanges. Further, the stratification cannot be consistent across sites because telephone companies use different criteria for assigning telephone exchanges. We defined 17 sampling strata for the survey. Within each service area, the sample was stratified by median household income, estimated for each telephone exchange. The initial plan was to define two strata per site, with the low-income stratum defined as those telephone exchanges with median incomes below the 20th percentile of median household income for all exchanges in the site. At some of the sites we found that the exchanges defined as low-income by this criterion contained substantially less than 20% of a site's total adult population. Using strata which contain too few households would not be effective in achieving our goal of increasing the representation of low-income households, so it was decided to increase the cut-off point to make the low income strata larger. Conversely, raising the cut-off too high would "dilute" the low income stratum, again making it more difficult to increase the representation of low income households. Thus, a maximum median income of $21,000 was set as a cut-off for inclusion in the low income stratum. The revised cut-off points are shown in Table III.1. The Dallas service area was divided into three income strata because after dividing into two strata, the distribution of median household income for the exchanges in the low income stratum suggested that this stratum contained two distinct subgroups, a very low income subgroup, and a low income subgroup.

TABLE III.1

HIGHEST MEDIAN INCOME EXCHANGE FOR LOW INCOME STRATUM

SITE	CUT-OFF ($)
Baltimore	18,890
Cleveland	20,776
Columbus	19,274
Dallas	16,973*
Denver	15,028
Houston	19,882
Philadelphia	20,776
Portland	19,905

*Cut-off for middle income stratum was $22,751.

Within each stratum, allocation of the sample was proportional to the estimated number of (total) households. In doing so, low income telephone households are oversampled, thus compensating for lower telephone coverage rate for low income households. Table III.2 summarizes the distribution of households, persons, screened households, and smoker households interviewed in each of the eight fire-service areas. Table III.3 provides additional information about the 17 site/income strata.

The Census/Estimated Population column of Table III.2 shows that the methods used to define the fire-service areas resulted in an overestimation of the populations in most cases. This overestimation resulted from a lack of fit between the fire-service areas and the areas serviced by the telephone exchanges included. The overestimation was an unavoidable consequence of including enough telephone exchanges to provide nearly complete coverage of the households inside the fire-service areas, because a sizeable number of households outside the fire-service areas were consequently included in the sampling frame.

In Philadelphia and Baltimore, the census population/estimated population ratio is very close to one. The closeness of fit between the fire-service area and the areas serviced by the included exchanges is due to the fact that Philadelphia County and Baltimore City County boundaries match the boundaries of the included telephone exchanges closely. The populations of the other fire-service areas were overestimated by anywhere from approximately 14% (Denver) to 37% (Columbus).

The implication of this lack of fit is that a number of interviews in some sites (specifically, those other than Philadelphia and Baltimore) were conducted with households outside the fire-service areas. To isolate the interviewed households known to be in the fire-service areas, we compared each household's census tract with lists of census tracts associated with each of the fire-service areas.

The lists for Denver, Portland, and Dallas were supplied with the initial sample documentation, written by Abt. The lists of tracts inside Baltimore, Houston, Philadelphia, Cleveland, and Columbus