A COMPUTER MODEL TO IDENTIFY RISK MARKERS FOR FOOD SERVICE SANITATION VIOLATIONS ASSOCIATED WITH OUTBREAKS OF FOODBORNE ILLNESS

A Thesis

Presented in Partial Fulfillment of the Requirements for the degree Master of Science in the Graduate School of the Ohio State University

by

James Robert Hartman, B.S.

* * * * *

The Ohio State University

1992

Master's Examination Committee:

Richard R. Lanese

John V. Gaeuman

R. Thomas George

Richard R. Lanese

Adviser

Department of Preventive Medicine

FIELDS OF STUDY

Major Fields: Preventive Medicine
emphasis on epidemiology

Microbiology
emphasis on food bacteriology

Registered Sanitarian. Job duties and professional interests include food service inspection and neighborhood sanitation in the OSU area, and the investigation of foodborne illness outbreaks city-wide.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I did all the work on this research project, but the credit for a great deal of the creativity it represents has to go to others. I wish I could claim its core idea, but it was Dr. Richard Lanese who suggested the use of computerized records from the Columbus Health Department's food protection program to predict important code violations. I merely set the stage by discussing the work of Irwin et al. in connection with my own unworkable research ideas. (I had wanted to set up an active foodborne illness surveillance system using an ongoing telephone survey. The work of Dr. Frank Bryan, now a food safety consultant in Lithonia, Georgia, originally sparked much of my interest in this area.)

John "Mac" Crawford and Gil Nestel of the Department of Preventive Medicine discussed various analytical methods with me. Dr. Thomas Santner of the Department of Statistics provided valuable insights about CART, the computer program I used for the main analysis.

Joe Damico of the OSU Academic Computing Center installed the CART program at OSU, including a larger-memory version for the Statistics Department, and got me started with that Statistics Department version. Joe and the other ACS mainframe and statistical consultants, especially Fred Ruland, Don Gibb, and Claudia Jordan, were indispensable throughout the process of reading and manipulating the data.

My work would not have been possible without the efforts of many fellow City of Columbus employees. Renee Closs of the Data Center copied the data onto tapes for me. Mark McDonnell, our liaison with the Data Center, smoothed the way. The data was collected by all the food sanitarians in the Columbus Health Department's Division of Environmental Health. Sanitarian Dan Smith also suggested a variable for testing. The data was keypunched by several competent clerical workers, most recently Barbara Pugh and (earlier) Julie Fultz--a process overseen by Linda Norris. My supervisor, Nick Malagreca, our Section Chief, Ted Strouth, and the Director of our Division, Mike Pompili, all provided encouragement.

A significant fraction of the financial support for this project was provided by the City through a tuition refund program, a benefit negotiated by Local 2191 of the American Federation of State, County and Municipal Employees.

My longtime companion, Bob Ramsey, put our home computer system together; otherwise I would have had to camp out at the OSU Computer Center. Bob suggested a

variable for testing, and it turned out to be one of the most powerful. And he tolerated my frequent unavailability and a yard overgrown with weeds while I slaved over this thesis.

VITA

May 14, 1951............................Born - Barberton, Ohio

1975....................................B.S., The Ohio State University, Columbus, Ohio

1979-present............................Sanitarian, Columbus Health Department, Columbus, Ohio

PUBLICATIONS

Two outbreaks of Shigella sonnei dysentery. Ohio Journal of Environmental Health 1983; 38 (6): 20-25.

Suspected foodborne outbreak, Columbus, Ohio. Ohio Journal of Environmental Health 1986; 36 (4): 6-13.

HACCP in three food services. Ohio Journal of Environmental Health 1987; 37 (6): 10-13.

HACCP in three food services. Ohio Journal of Environmental Health 1988; 38 (1): 14-17.

Food safety reviews. Ohio Journal of Environmental Health 1989-91 [bimonthly column].

A foodborne outbreak investigation with negative findings. Ohio Journal of Environmental Health 1990; 40 (4): 6-8.

Variables evaluated

End note

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER I

Introduction

The goal of this project was to identify groups of food service operations (restaurants) and food establishments (food stores) with higher rates of certain kinds of code violations associated with foodborne illness. It used CART (Classification and Regression Trees[1]) software to analyse the computerized inspection records of the Columbus Health Department's Food Protection Program. These records not only indicate inspection outcomes, but also contain variables with possible predictive power. The analysis also included "sociological" data, such as ethnicity and income levels of residents in the operations' zip codes, to test whether such variables can predict future inspection results before an operation even opens for business.

Researchers often refer to variables bearing a positive association with a disease as a "risk factor," but some have criticized this term because it seems to imply

1 Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Pacific Grove, CA: Wadsworth, 1984.

knowledge of causation when only an altered probability of disease is known. The term "risk marker"[2] is perhaps better.

The use of CART to identify risk markers for adverse inspection results may be more sophisticated than the focus of previous work in this field on inspection scores as a barometer of an operation's risk of causing outbreaks. A great body of information is available on the specific causes of foodborne illness, and because improper temperature control of potentially hazardous foods is the leader, the code violation representing this problem deserves special attention.

A place for environmental sanitation in public health

Walker[3] has discussed the impact of the National Academy of Sciences' 1988 report on the future of public health on environmental health programs. "Disarray, diffusion, confusion, and lack of support" characterize the present system. Within the field of environmental health there has been an emphasis on pollution control and on

---

2 McCormick J, Skrabanek P. Coronary Heart disease is not preventable by population interventions. The Lancet 1988: October 8; 839-41.

3 Walker B Jr. The future of public health. Journal of Environmental Health 1989; January/February:133-135.

participation by an increasingly knowledgable and environmentally conscious public. Environmental health programs' poor public image, arising partly from their enforcement orientation and partly from the overemphasis of personal health services as provided by nurses and physicians at the expense of environmental control programs, "interferes with the capacity of officials to mobilize support from the general public and from political leaders for the public health mission." Problems such as acid rain, toxic waste, and indoor air pollution have received attention, while many experts believe foodborne diseases are on the increase. "If local health departments take on additional responsibilities, however, the relative time spent on food protection will have to decline. Given the finite resources ever likely to be available for environmental health resources [sic], improved schemes for setting priorities and more efficient approaches to risk assessment will be necessary to ensure adequate services in all areas." "Environmental surveillance and biological monitoring have rightfully emerged as essential elements in the continuum of environmental health services ... but remain to be fully integrated into the total public health system."

The Columbus Health Department, consistent with the surgeon general's goals for the nation for the year 2000,[4] has targets for the reduction of illnesses caused by foodborne Campylobacter, Escherichia, Listeria, and Salmonella. This report describes a project that may serve not only to bring us closer to the target levels of these illnesses, but also, ideally, to improve the status of the food protection program by showing that careful research in this area, using modern methods, is possible.

---

4 Public Health Service. Healthy People 2000: national health promotion and disease prevention objectives. Washington, DC: U. S. Department of Health and Human Services, Public Health Service, 1990; DHHS publication no. (PHS)90-50212.

CHAPTER II

Background

The 1976 Food Service Sanitation Manual,[5] by the United States Food and Drug Administration (FDA), gives a brief history of this country's restaurant inspection program. The first proposed national "ordinance regulating eating and drinking establishments" was a mimeographed document promulgated in 1935. The Ohio Department of Health adapted the 1976 Model Code and approved the use of a 44-violation inspection report form reproduced here as Appendix A. (Some of the space for remarks was removed and the form was reduced.) The Columbus Health Department enforces this code locally. Its goals are to minimize foodborne illness, to ensure the "soundness" or purity of food, and to meet consumer expectations. Note that 13 items on the inspection report are marked with asterisks as "critical items requiring immediate attention."

The FDA Manual comments, "despite the progress made, foodborne illness continues to be a major public health

---

5 U. S. Department of Health, Education, and Welfare. Food service sanitation manual, including a model food service sanitation ordinance. Washington, D. C.: U. S. Government Printing Office, 1978.

problem." It may be the second most frequent cause of short-term illness in the United States (behind the common cold).[6] Archer and Kvenberg[7] used data from the National Ambulatory Medical Care Survey to estimate that the real incidence is between 18 and 61 million cases per year, and concluded that, including secondary cases, the U. S. has 24 to 81 million cases per year. In contrast, the official tally of (confirmed) cases of illness transmitted by food in 1983-1987 was 91,678 cases--an average of only 18,336 per year.[8] The CDC cautions that this data would be useless in trying to compare the relative incidence rates of these illnesses attributable to specific causes. Nevertheless, 41 to 58 percent of these reported cases were due to commercial food services; and, because restaurants are probably more likely to be reported than a home cookout, they probably contribute an even greater fraction of illness than the CDC has reported. This is not surprising, considering how common serious food service

---

6 Zaki MH, Miller GS, McLaughlin MC, Weinberg SB. A progressive approach to the problem of foodborne infections. American Journal of Public Health 1977;66:44-49.

7 Archer DL, Kvenberg JE. Incidence and cost of foodborne diarrheal disease in the United States. Journal of Food Protection 1985; 48:887-894.

8 Centers for Disease Control. Foodborne disease outbreaks, 5-year summary, 1983-1987. In: CDC Surveillance Summaries, March 1990. MMWR 1990; 39 (No. SS-1):15-57

health code violations are. In a quality-control survey conducted in Seattle and King County (Washington), 51 percent of the "complex-menu type operations" (restaurants with complex menus and food preparation procedures, and possibly large meal volumes) were in the "high to extreme hazard" category based on critical items violated.[9] In a recent federal survey of 15,000 nursing homes, 42.8 percent failed to meet food sanitation standards.[10] In Columbus for the year ending December 1, 1988 nine percent of all violations noted by inspectors were critical items. Clearly even better control of this already pervasively regulated industry is in order.

For many years studies of foodborne illness outbreaks have shown that certain food handling errors cause most of the problems.[8,11] They show up consistently from country to country and from year to year. The leaders (in order of importance) are improper holding temperatures of potentially hazardous foods, poor personal hygiene by infected workers, inadequate cooking, contaminated

---

9 Bernhardt RR. Seattle-King County Department of Public Health food protection program program review. Olympia, WA: Division of Health, Department of Social and Health Services, 1986:6.

10 43% of nursing homes flunk food sanitation. Columbus Dispatch; December 2, 1988: 1A.

11 Bryan FL. Factors that contribute to outbreaks of food- borne disease. Journal of Food Protection 1978;41:816.

equipment, and food from unsafe sources. See Table 1.

Ohio's Food Service Rules (Ohio Administrative Code Chapter 3701-21-W) give this definition:

> "Potentially hazardous food" means any food that consists in whole or in part of milk or milk products, eggs, meat, poultry, fish, shellfish, edible crustacea, tofu, baked or boiled potatoes, cooked rice, cooked beans, or other ingredients[,] including synthetic ingredients, in a form capable of supporting rapid and progressive growth of infectious or toxigenic microorganisms. The term does not include foods which have a pH level of 4.6 or below, or a water activity (A_w) value of 0.85 or less.

(Water activity is a measure of the amount of moisture available to bacteria; pure water has a value of 1.00.)

Section 5A of the Food Service Rules specifies time-temperature requirements for cooking, reheating, cooling, or storing these potentially hazardous foods. (See

Table 1. Factors contributing to foodborne disease outbreaks

1. Failure to refrigerate foods properly
2. Failure to heat-process or cook foods thoroughly
3. Infected workers practicing poor personal hygiene
4. Preparing foods a day or more before serving
5. Incorporating contaminated raw ingredients into foods that receive little or no cooking
6. Allowing foods to remain at warm temperatures at which bacteria can incubate
7. Failure to reheat cooked foods to temperatures that kill vegetative bacteria
8. Cross-contamination
9. Failure to clean and disinfect kitchen or processing-plant equipment

Appendix B, "Where to mark violations on food service operation inspection form," for lists of temperatures applicable to various foods and recommended time limits for processing.) Because violation of this section is the most important cause of foodborne illness, finding risk markers for "5A" violations will be the most important part of this study.

Ohio's Rules regulate restaurants, delicatessens, caterers, fast food operations, and similar facilities as "food service operations." Locations with food or beverage vending machines are also licensed as food service operations, although individual machines are not. Even vending locations serving only cold drinks or coffee are licensed, although similar operations staffed by people would not be, because of the possible absence of monitoring at the machine locations otherwise.

Unlike many other health departments, the Columbus Health Department also requires licenses and performs inspections for supermarkets, fish markets, carry-outs, ice cream parlors, and similar establishments, to which Ohio's Food Service Rules do not apply. Chapter 221 of the Columbus City Health Code designates these facilities as "food establishments." The program applies only to those operations having potentially hazardous food. The inspection form for food establishments is similar to the one used for food service operations (see Appendix C).

The Columbus Health Department has a contract to provide food protection and other public health services for the City of Worthington. City inspectors also inspect mobile operations licensed by other health departments if they operate in Columbus (at the Ohio State Fair, for example), as well as a few food vending machine locations licensed to operators outside of Columbus. This study examined all the food service, food establishment, and food vending records for Columbus and Worthington.

CHAPTER III

Studies of establishments that cause outbreaks

Relative risk ratios for restaurants compared to markets.

Kaplan and El-Ahraf[12] were apparently the first to tabulate data on reported outbreaks of foodborne illness according to the type of establishment involved. In a short, widely quoted article they presented work done with data from a large county in Southern California in 1979.

Table 2 summarizes the data presented by Kaplan and El-Ahraf on foodborne outbreaks by type of establishment. They refer to the ratios of the percentages (89%/70% = 1.27; 11%/25% = 0.44) as "relative risks"; and they, and later reviewers, refer to the ratio of these ratios (1.27/0.44 = 2.9) as the "relative risk ratio." The meaning they ascribe to this is that "...the average establishment in the category 'fast food and restaurant' is three times more likely to generate a reported outbreak than a food market [is]." (It is not totally clear from the authors' presentation whether an establishment could

---

12 Kaplan OB, El-Ahraf A. Relative risk ratios of foodborne illness in food service establishments: an aid in deployment of environmental health manpower. Journal of Food Protection 1979;42:446-447.

TABLE 2. Percentages of foodborne illness outbreaks reported by establishment type, Southern California, 1979.

establishment type	number	as % of total	out-breaks	as % of total	ratio of %'s
Fast food & restaurants	2,500	70	227	89	1.27
markets	900	25	28	11	0.44
liquor stores	185	5	1	-	-*
totals	3,585**	100	256	100	-

* The authors comment that this figure is "negligible" and that its standard deviation is large compared to its mean.
** The authors give this sum as 3,600.

TABLE 3. Factors associated with foodborne illness in restaurants, Seattle-King County, Washington, January 1, 1986 to March 31, 1987 (after Irwin et al.)

	Odds Ratio
Any improper food protection practice	15.8
Improper storage or handling of equipment and utensils	14.9
Potentially hazardous foods at unsafe temperature	10.1
Any "critical" violation	6.3
Inspection lasting 37 minutes or longer	5.6
Score of 86 points or below	5.4
Corporate owner	5.3
"Unsatisfactory" or "suspend permit" result*	3.9
Restaurant size 150 or more seats	3.4
Potentially hazardous food not cooked to proper temp.	**
American cuisine	0.2

* "Unsatisfactory" means score 70-85 or a critical violation; "suspend permit" means score below 70.
** odds ratio was indeterminate for this factor.

contribute more than one outbreak. If it could, the authors' choice of effect measures, relative risk, would not have been appropriate.)

Kaplan and El-Ahraf conclude that the differences in risk suggest "there is no logical basis for the traditional rule that all types of establishments [sic] must be inspected a given number of times." (The origins of this "traditional rule" are obscure, but a feeling that it is unfair to do extra inspections in some operations may have motivated it.) Perhaps administrators should "increase the surveillance of high-risk establishments and decrease that of low-risk ones. This would result in a more effective deployment of sanitarian manpower and related resources."

Vessel sanitation scores

The CDC began a passenger cruise ship food service and water quality control inspection program in 1975 on ships using U. S. ports because two percent of the cruises had five or more times the rate of enteric illness than the other 98 percent did.[13] It provided a rare opportunity to study the effects of sanitation on the health of a defined population, free of the influences of home meals and other

---

13 Dannenberg AL, Yashuk JC, Feldman RA. Gastrointestinal illness on passenger cruise ships, 1975-1978. American Journal of Public Health 1982;72:484-8.

factors normally present.

A unique surveillance system required the captain to report by radio, 24 hours before arrival in port, the number of diarrhea cases seen by the ship's physician. If necessary, epidemiologists could organize an investigation before dispersal of passengers on arrival.[11] An outbreak was defined as three percent or more of the passengers or crew seeking medical attention for diarrhea by the ship's physician. Illnesses that could be linked to meals on shore were excluded. There were 45 shipboard outbreaks during the ten years of the study.

By 1985 the CDC had completed almost 1,800 inspections on 172 vessels.[14] They classified inspections as "semiannual" (regular), "follow-up," and "other." The latter category included outbreak investigations. The CDC ranked ships according to their average scores from 905 semiannual inspections into three groups: the upper 20%, the middle 60%, and the lower 20%. When analyzed by average score, ships in the upper 20% had 1.8 outbreaks for every 10 million passenger-days, ships in the middle 60% had 3.5, and ships in the lower 20% had 8.1 outbreaks per 10 million passenger-days. Analysis by the percentage of "satisfactory" ratings (a score of 86 or above out of 100

---

14 CDC. Vessel sanitation scores. MMWR 1988;37:114-117.

points) and by shipping line showed a similar trend.

The CDC program appears to have been effective despite limited data. Whereas the number of cruises and passenger-days increased continuously through the study period, the number of outbreaks per 10 million passenger-days decreased continuously.

Routine inspections can predict outbreaks

A detailed study was done in Seattle more recently by Irwin et al.[15] to examine the violations reported on the last routine inspection report before each of 28 outbreaks the restaurants had experienced. An agent was implicated in only 6 of the 28 outbreaks, but a food vehicle was identified in all but 4 of them. Improper temperature control of potentially hazardous foods was a contributory cause in 25 of the 28 outbreaks.

Irwin et al. set up a case-control study comparing case restaurants (ones causing an outbreak) to control restaurants (matched to cases by health district and routine inspection date). According to their results, the best predictor of which food services would later cause illness was "any improper food protection practice" ($\widehat{OR}$ =

---

15 Irwin K et al. Results of routine restaurant inspections can predict outbreaks of foodborne illness: the Seattle-King County experience. American Journal of Public Health 1989;79:586-590.

15.8). (Presumably this would be equivalent to any violation of 05F in Ohio--see Appendix B.) Improper temperature control of potentially hazardous foods was, surprisingly, third ($\widehat{OR}$ = 10.1) behind "improper storage or handling of equipment or utensils" ($\widehat{OR}$ = 14.9)--an outcome explained by the investigators as possibly being a statistical fluke. (There is also a chance unknown factors could be causing problems because of utensil-handling procedures, although present knowledge would not suggest this.) Specialization in American cuisine was protective, with an $\widehat{OR}$ of 0.2. (See "Variables not used by CART" in Chapter VI for further comments on restaurant ethnicity.) Table 3 (page 11) summarizes their results.

Food vending machines are safe

Available epidemiologic evidence suggests that food and beverage vending machines are unlikely to cause foodborne illness. They may be more likely to cause injuries by tipping over onto people who are trying to rob or vandalize them.[16] Their relative safety may be due partly to the self-regulating nature of the vending industry: it may be oversensitive to consumers' expectations of cleanliness, the absence of vermin, and the

16 McSwain David. Vending Program (Inservice sponsored by the Columbus Health Department), July 17, 1990.

palatability of food. The rapid turnover of product necessary for profitability probably also contributes to safety. Voluntary certification of machines by the National Automatic Merchandising Association (NAMA) probably helps, too, because the NAMA requires cleanability and safety features, such as a switch to prevent the vending of perishable food if the product has ever warmed to a temperature above 45° F. for any reason, such as a temporary power failure.

CHAPTER IV

Inspection strategies

Floors, walls, and ceilings

For many years the standard in food service inspection has been to check each operation a fixed number of times per year and to concentrate on structural problems. Until a few years ago the Ohio Department of Health inspection form listed "floors, walls, and ceilings" as the first item. Originally the rationale for the "floors, walls, and ceilings" inspection may have been to concentrate on fundamentals, often a necessary consideration in previous decades. Rats and filth were once overwhelming problems. This type of inspection lives on in what may be an application of the debunked Miasma Theory--the minds of many people equate filth with disease. Just as the Germ Theory replaced the Miasma Theory, more scientific techniques are replacing the "floors, walls, and ceilings" inspection.

Attempts over the last few years to improve inspections to control the risk of foodborne illness from licensed facilities have concentrated on three major areas: the application of hazard analysis to inspections, the manipulation of the frequency or scheduling of inspections, and the use of microbiological examinations of foods. The old style of inspections and the three new ones are not mutually exclusive: variable frequencies or microbiological examinations may be used with "floors, walls, and ceilings" inspections, for example.

HACCP

The most important development has probably been the introduction of hazard analysis and critical control point monitoring, or HACCP.[17] The contrast between the HACCP approach and the traditional "floors, walls, and ceilings" inspection technique exactly parallels the contrast between the Germ Theory of disease and the Miasma theory. HACCP was a spinoff of the U. S. space program--it resulted from food processors' adaptation of the NASA "zero defects" program to the production of food for astronauts.[18] The basic format of the HACCP approach is to follow food handling through time, paying attention to processes and procedures that may result in contamination by, or growth of, pathogens capable of causing foodborne illness. "[The

17 Bryan FL. Hazard analysis of food service operations. Food Technology, February 1981:78-87.

18 Bauman H. HACCP: concept, development, and application. Food Technology, May 1990:156-158.

HACCP] concept is really nothing more than what many good sanitarians and conscientious restaurant operators have been doing for generations. It is just more structured and formalized."[19] The distinction between HACCP and regular inspections can be thought of as the difference between a well-focused movie and a fuzzy still photograph.

The HACCP idea has been slow to catch on in health departments or the restaurants they regulate: a survey by the FDA in 1986 found that out of 2,700 state and local health departments, only 23 state and 8 local agencies expressed an interest in HACCP.[20] However, the FDA has been training state and local health departments in the use of HACCP in its Current Concepts in Food Protection program. Frank Bryan has discussed HACCP training and compiled a bibliography of training materials.[21]

Guzewich[20] and Bryan[21] have identified several impediments to widespread use of HACCP. The most important

---

19 Harrington RE. How to protect your restaurant against foodborne illness. NRA News, April 1986:33-34.

20 Guzewich JJ. Practical procedures for using the hazard analysis critical control point (HACCP) approach in food service establishments by industry and regulatory agencies. In: Food Protection Technology (papers presented at the Third Conference for Food Protection). Chelsea, Michigan: Lewis Publishers, Inc.:91-100.

21 Bryan FL. Teaching HACCP techniques to food processors and regulatory officials. Dairy, Food and Environmental Sanitation 1991;11:562-568.

may be resistance to change. Regulators and industry personnel are reluctant to invest time and money in training their people in rudimentary food microbiology and related subjects. Rapid turnover of food workers may require continuous training programs. Untrained managers may expect newly-trained employees to do things the same way they did before. Some of the resistance to HACCP is because the initial evaluation of an operation is time-consuming compared to the "floors, walls, and ceilings" inspection. The best way to compensate for this extra investment in time is probably to adjust inspection frequencies according to hazard category: time for HACCP evaluations comes from inspecting operations classified as low-risk less often. But many jurisdictions' codes require a fixed number of inspections annually, and many health department managers judge sanitarians' performance on the number rather than the quality of inspections. Some authorities have recommended laws requiring operators to monitor control points and maintain records of the monitoring. Subsequent visits by the health department sanitarian could entail a record review, rather than another complete HACCP evaluation, as another way to help make up the time spent on the initial evaluation. The food service industry seems to feel this savings would be at its expense--industry employees would waste time keeping records. (In fact, many fast food chains routinely keep food time-temperature logs. The OSU Hospitals kitchens have recording thermometers in their dishwashers. The OSU Residence and Dining Halls kitchens have incorporated HACCP concepts into recipes. None of this seems to be much of a burden.) There is a misconception that concentration on critical control points allows establishments to be filthy and vermin-infested.

One problem with instituting HACCP has not been discussed much in the literature: the impression the literature itself seems to convey that instituting a HACCP program has to involve extensive (and expensive) microbiological analysis of foods.

Variable inspection frequencies

Another major thrust of the improvements has been variable inspection frequencies, with the adjustment of the intervals between inspections according to various criteria. Frank Bryan has commented[21] that the food service industry has grown faster than most health department budgets, and that many food program budgets have shrunk in proportion to the rest of the health department budget due to de-emphasis of foodborne illness control programs. This implies that inspections must be shortened or reduced in number, or (as he advocates) a variable inspection frequency based on risk to the community must be implemented.

The FDA's 1976 model code[1] recommended semiannual inspections; Ohio's rules require at least an annual inspection. (Nevertheless, the Columbus Health Department reduced the inspection frequencies of food vending machine locations to once every two years in January of 1990 because of their good safety record.) The Seattle-King County Health Department experimented with a departure from their standard four inspections per year in 1970-1972 but discovered that one visit per year resulted in increased "food poisoning" complaints and decreased scores.[22] An experiment in the Ottawa, Ontario, region[23] in 1981 and 1982 found that decreasing the annual number of inspections from 12 to 7.5 did not influence the "proportion of establishments showing defects"; however, as the authors comment, there may be a threshold or saturation frequency beyond which more inspections do no more good. Both intervals seem like "overkill" by Columbus standards.

---

22 Bader M et al. A study of food service establishment sanitation inspection frequency. American Journal of Public Health 1978;68:408-410.

23 Corber S et al. Evaluation of the effect of frequency of inspection on the sanitary conditions of eating establishments. Canadian Journal of Public Health 1984; 75:434-438.

There have been few attempts documented in the literature to use an operation's inspection history as a basis for adjusting the frequency. Kaplan and El Ahraf's idea of custom inspection frequencies was discussed above.[12] They were among the first to advocate this idea. Zaki et al.[6] also suggested it in 1977.

Frank Bryan[24] suggested the use of food-property, food-operations and average-daily-patronage risk coefficients to customize inspection frequencies. In his system, foods that have most often been vehicles of foodborne illness, such as roast beef, ham, and turkey, receive a value of 5; foods unlikely to support microbial growth because of a water activity below 0.85 or a pH below 4.6 get a food-property risk coefficient of 1; and other foods with an intermediate risk get intermediate values. Similarly, risky food processing steps, such as room-temperature storage of potentially hazardous foods, receive a food-operations risk coefficient of 5; safe practices like normal storage of canned foods get a coefficient of 1. The average-daily-patronage risk coefficient ranges from 1, for an operation with 100 or fewer customers, to 2.5, for one with more than 500 per day. It is calculated for each

---

24 Bryan FL. Foodborne disease risk assessment of food- service establishments in a community. Journal of Food Protection 1982;45:93-100.

menu item from the average number of units of the item sold in a day. A composite risk index for an operation is formed from the products of the first two coefficients, summed over all the foods served by an operation, multiplied by the average-daily-patronage risk coefficient. Bryan recommended placing each operation into one of three categories based on the composite risk index. Those in Category 1 "deserve a thorough hazard analysis from which critical control points should be determined and monitored." Category 3 includes taverns, for which only an annual permit-renewal inspection is due.

A strategy developed in Texas by Briley and Klaus[25] accepted Kaplan and El Ahraf's idea of custom inspection frequencies, and used Bryan's food-property and average-daily-patronage risk coefficients. They decided Bryan's food-operations risk coefficients were too difficult and time-consuming to calculate, so they replaced them with an ordinal scale based on the average score from the previous 5 inspections. For example, the highest-risk establishments, with mean scores below 76.49, received the value 5; operations with mean scores above 94.49 received the value 1. These averages were calculated from scores

25 Briley RT, Klaus EF. Using risk assessment as a method of determining inspection frequencies. Dairy and Food Sanitation 1985;5:468-474.

from a recent period when all operations were inspected with the same frequency. They considered the food-property risks to be additive; but the overall risk potential for an operation was the multiplicative effect of the three coefficients. Inspection intervals ranged form monthly to semiannually, based on this product. This Texas study used inspection report scores (100 minus violations, weighted by risk) as the outcome measure.

Briley and Klaus manipulated the inspection intervals (the period between visits) using SPIF, the same computer system the Columbus Health Department's Food Protection Program uses (see the section on Data below). For each food service they compared the average score from the 5 inspections before the study period, as a baseline, to the average score from all inspections during the study period, and to the average score from the last 5 inspections in the study period. Among the "high-risk" establishments, whose frequency was increased from once every 3 months to once every month or every other month, both sets of mean scores increased over baseline. There was no change in the other operations during the study period. The system was self-regulating: if an operation's score would fall, it would receive more frequent inspections, causing its score to rise again.

Also in 1985, Wodi and Mill[26] recommended the use of a "complex combination mathematical approach" formula to calculate an inspection priority score based on a predicted risk score derived from the last two inspection scores (and especially the last score), critical items violated during those inspections, and the populations at risk at the time of the last two inspections. The priority scores were decimal fractions. Each sanitarian was to inspect the operation with the largest priority score first. A subsequent letter to the editor[27] about their article made the unusual comment that "standards, based on site inspections by sanitarians, are fraught with major weaknesses because they assume the food service manager is capable of maintaining the establishment in compliance with the health department regulations between inspections," and goes on to recommend manager certification, the certified manager being "an extension of the health department." The authors responded that if certification improves inspection results, this consideration is already included in their model.

---

26 Wodi BE, Mill RA. A priority system model for sanitation management in food service establishments. American Journal of Public Health 1985;75:1398-1401.

27 LaBoccetta AC. (letter re:) A priority system model for sanitation management in food service establishments. American Journal of Public Health 1986;76:709-710.

Scott County, Iowa, has been using a hybrid approach[28] to schedule food service inspections for the last five years to avoid what might be called "management by crisis." Outbreaks of foodborne illness, frustration with "the same establishments repeating the same violations," and a general failure of some operators to acknowledge the seriousness of their problems motivated the department to go beyond informal hearings and increased inspection frequencies, measures they felt had been applied selectively, "with each situation being dealt with differently."

Scott County categorized operations by the average inspection score over the last four regular inspections as high (95+), mid-range (80-94), and low (79 or below). Of some 570 establishments, 14 low-scorers began receiving bimonthly inspections. The state code also requires that operations with two consecutive scores below 76 be posted with a designation of "poor." Finally, Scott County developed a protocol to apply progressive enforcement (letters and conferences) consistently. Fifty-eight high-scorers received blue ribbons, favorable newspaper and television publicity, and one inspection per year. The rest of the operations stayed on the state-mandated

---

28 Moore GA et al. Food sanitation enforcement. Journal of Environmental Health 1990; 53 (2): 17-18.

biannual schedule. The department's three food sanitarians saved 46 hours with the reduced frequency in the blue-ribbon restaurants, and spent 44 extra hours doing inspections, restaurant employee training, and "in-depth" explanations of violations in the low-scorers.

Microbiological approaches

Microbiological approaches to food risk control make up the fourth group of inspection strategies. A group of researchers at the Suffolk County (New York) Department of Health Services attempted to integrate a microbiological sampling plan for potentially hazardous foods with the 44-item FDA scoring system (based on a 1974 version of what presumably became the 1976 Model Code). Zaki and co-workers[6] found that the bacterial counts of 100 samples of perishable foods at the time of sampling were not significantly related to their storage or display temperatures (above or below 45° F.), the presence or absence of critical violations, or the lag between production and sampling. (These results are perhaps not surprising, considering that counts are artifacts of the quality of raw ingredients and of handling, especially time-temperature control, throughout the product's history, rather than just at the end.) Nevertheless, the researchers urged the development of microbiological criteria for foods and the use of bacteriological monitoring of potentially hazardous foods, because they used high counts to encourage compliance. They did not mention costs.

Anderson et al.[29] found, in contrast, that the logarithms of the aerobic plate counts of mesophilic bacteria in 366 cold food samples from 175 food service establishments increased linearly with the temperature (26° F. to 80° F.) at the time of sampling ($r = 0.79$, $p < 0.05$). Only 64% of the foods met minimum temperature requirements at the time of sampling (< 45° F.). Fifty-eight percent of the samples had counts exceeding a million colony-forming units per gram, an often-cited arbitrary standard. They comment that their results are consistent with those of other similar surveys. Anderson et al. describe some of the problems with microbiological standards for foods, and suggest that repeated sampling at specific control points during food preparation might be useful in setting limits. (Their study also considered the pH and the method of preparation and storage of the foods, but they did not discuss these parameters further.)

---

29 Anderson PS, Rutenberg GW, Bowen NL. Assessing food quality: the difficulty in establishing microbiological standards. Journal of Environmental Health 1989;52:79-82.

It is interesting to note that Tebutt and Southwell[30] found no correlation between microbiological results and visual inspection ratings in food manufacturing plants in Britain. There was one exception: a relation between poor personal hygiene and the presence of Staphylococcus aureus in dairy products. Their inspection ratings included parameters reflecting overall appearance, personal hygiene, risk of contamination, temperature control, and training and education.

Conclusions

The discussion in Chapters III and IV leads to the conclusions that poor inspection results are associated with increased foodborne illness, and that several strategies may be effective in identifying higher-risk operations and improving their inspection outcomes.

For more than half a century food sanitation programs have been controlling foodborne illness. At first, the rules and the organization of inspection programs were based on the theoretical links between food sanitation and public health. Gradually the theories have borne up to testing. In spite of the difficulties inherent in outbreak

---

30 Tebutt GM, Southwell JM. Comparative study of visual inspections and microbiological sampling in premises manufacturing and selling high-risk foods. Epidemiology and Infection 1989; 103:403-475-486.

reporting and investigation, evidence has mounted that restaurants are more likely than markets to cause illness. The presence of potentially hazardous foods, error-prone processing steps, and size (or average daily patronage) can further stratify risk. The profile of the dangerous restaurant has been emerging more and more clearly as one with low inspection scores and a history of food protection (especially time-temperature) violations. Food program managers have mostly just assumed the validity of these risk markers; limited data has, however, supported these assumptions well.

Strategies for improving inspection outcomes have stressed increasing inspection frequency, but little work to establish optimum frequencies has been reported. Microbiological testing regimens are probably most useful as part of HACCP evaluations. HACCP, manager certification, and other concepts may be useful educational tools. The International Association of Milk, Food and Environmental Sanitarians' Committee on Communicable Diseases Affecting Man has also recommended the use of HACCP evaluations as an effective enforcement tool.[31] Letters, hearings, adverse publicity, food embargoes or seizures, voluntary closures, and permit suspensions are

---

31 IAMFES. Procedures to implement the hazard analysis critical control point system. Ames, Iowa: IAMFES, 1991, p. 34.

probably also very effective, but the relative effectiveness of various enforcement procedures has probably never been measured.

CHAPTER V
Methods

A general description of CART will be followed by a brief overview of some of its special features. Many of these are described in Appendix L. Other procedures used in this study, particularly SAS procedures, will be discussed.

CART

The CART (Classification and Regression Trees) computer methodology[1,32] makes binary splits on data to form a prediction tree. Each node of the tree represents a question; data points for which the answer is "yes" are assigned to one branch, and the "no's" are assigned to the other. The leaves of the tree are called terminal nodes. CART determines the questions by identifying explanatory variables whose values can best discriminate between the values of the outcome variable. For a continuous explanatory variable, it then identifies the cut points. Nodes farther and farther from the root node get more and

---

32 California Statistical Software, Inc., Lafayette, CA., 1984. [software]

more concentrated, or pure.

As a recent review of newer computer-intensive statistical methods commented,[33]

> The 1980's produced a rising curve of new statistical theory and methods based on the power of electronic computation. Today's data analyst can afford to expend more computation on a single problem than the world's yearly total of statistical computation in the 1920's.

The article featured CART as an example of this emerging methodology, free from the mathematical tractability requirements of familiar statistics like means, standard deviations, hypothesis testing, analysis of variance, linear regression, etc.

A remarkable aspect of CART and its forerunner, the Automatic Interaction Detector program (AID),[34] is the small number of variables required to provide insight into a seemingly complex problem. In a multivariate study to identify variables useful in predicting teenage smoking,[35] for example, AID needed only three variables.

---

33 Efron B, Tibshirani R. Statistical data analysis in the computer age. Science 1991; 253:390-395.

34 Andersen R, Smedby B, Eklund G. Automatic Interaction detector program for analyzing health survey data. Health services research 1971; summer: 165-183.

35 Lanese RR, Banks FR, Keller MD. Smoking behavior in a teenage population: a multivariate conceptual approach. American Journal of Public Health 1972;6:807-813.

Johnson and Wichern's textbook on multivariate methods[37] contains a useful chapter on discrimination and classification. It discusses some features common to all systems, such as the use of a "learning sample" to develop the classification rules.

CART has several optional ways to estimate how misclassification is to be assessed. One is by means of a test sample, different from the learning sample. An even more accurate method is what CART calls "cross-validation:"

> It works by dividing the data into ten groups of equal size, building the tree on 90% of the data, and then assessing the tree's misclassification rate on the remaining 10% of the data. This is done for each of the ten groups in turn, and the total misclassification rate is computed over the ten runs. The best tree size is determined to be that tree size giving the lowest misclassification rate. This size is used in constructing the final tree from all the data. The crucial feature of cross-validation is the separation of data for building and assessing the trees: each one-tenth of the data acts as a test sample for the other nine-tenths.[32]

CART, as is standard in classification systems, allows the assignment of different penalties for different kinds of misclassification according to the real cost of the misclassification. In this study the cost of misclassifying a violator as a non-violator was assumed to be the risk of an outbreak of foodborne illness from the violation. The misclassification costs were kept as close as possible to the odds ratios given in Table 3, and were always within the confidence intervals given by Irwin et al.: 10 (95% CI, 2-46) for misclassifying a violator of the rule against time-temperature abuse of potentially hazardous foods, and 4 (95% CI, 1-11) for inspection failure. (The Columbus Health Department's criteria for "failure" resemble theirs for "unsatisfactory" or "suspend permit.")

Other procedures

The data used in this study was originally recorded using a Burroughs mainframe at City Hall. The Data Center copied it onto a tape with no further manipulation.

A program called JMLCOPY, developed at the Political Science Department at Ohio State,[36] copied the data from the tape to disks.

The analysis was begun using various procedures in SAS[37] to manipulate the data and transform it into a form CART could operate on. The work was done on an IBM 3081 mainframe computer at the OSU Instruction and Research Computer Center, which was renamed the Academic Computing Center in 1990. Because of limitations in sort space available, it was necessary to process the study's data in

---

36 Ludwig JM. JMLCOPY. Columbus: Instruction and Research Computer Center, 1983.

37 SAS Institute, Inc. SAS user's guide: basics. Version 5 Edition. Cary, NC: SAS Institute, Inc., 1985.

thirds, one for each administrative unit (District) subdividing the Environmental Health Division. Ironically, the 3081 was scheduled to be replaced early in 1992 to provide more disk space. The final data set was compiled from the three thirds in a fourth step. SAS routines similar to those used for preparing the data for one District and for making the data sets used by CART are included as Appendix H and Appendix I, respectively. (Most of the system messages at the beginning have been deleted from each job.)

An important feature of the SAS jobs used to make the district data shows up in Appendix H at the line

```
DATA WEEDED; SET WEEDWORK;
```

This feature allows the program to delete all observations after the last violation of interest for all operations that ever had the violation. The run in Appendix H deleted all inspection records after the last failure; similar runs deleted all observations after the last time-temperature violation. This feature can be turned off easily to use the last inspection for each establishment as the index observation. Figure 1 shows two identical series of inspection records, one using the last inspection and one using the last failure, to make this distinction clear.

Another important feature in these SAS runs is visible in Appendix H at the spot where these lines occur:

FIGURE 13. THE LAST INSPECTION VERSUS THE LAST FAILURE

ESTID	DATE	FAIL	ESTID	DATE	FAIL	ESTID	DATE	FAIL
6733	5/14	0	6734	5/14	0	6735	1/2	0
6733	9/14	0	6734	7/16	1	6735	5/2	0
6733	11/2	0	6734	9/10	1	6735	8/9	0
6733	12/1	0	6734	12/1	0	6735	12/1	1

A. THE LAST INSPECTION IS UNDERLINED.

ESTID	DATE	FAIL	ESTID	DATE	FAIL	ESTID	DATE	FAIL
6733	5/14	0	6734	5/14	0	6735	1/2	0
6733	9/14	0	6734	7/16	1	6735	5/2	0
6733	11/2	0	6734	9/10	1	6735	8/9	0
6733	12/1	0	6734	12/1	0	6735	12/1	1

B. THE LAST FAILURE IS UNDERLINED.

```
DATA OUTCOME PREDICTR; SET WEEDED; BY ESTID INSPDAT3;
IF LAST.ESTID THEN OUTPUT OUTCOME;
   ELSE OUTPUT PREDICTR;
```

This feature separates the last inspection (or the last inspection used in the analysis) for each operation from all that preceded it. Then many new variables, such as the number of regular inspections in the 5-year period covered by the study, could be calculated for each operation using just the "predictor" inspections (and licensing information). Finally, when all the data sets were combined, the record with the results of the last inspection (or the last one used) also contained the summary statistics calculated from its predecessors.

CHAPTER VI

Variables Evaluated

This chapter begins with a description of the sources of the data used. Its second section discusses the Columbus Health Department's most important variable for classifying food operations, the inspection interval, corresponding to the three broad categories of operation. Next comes an explanation of each of the other variables used in the analysis, first the outcome variables, then the predictor variables (in approximately their order of importance, as measured by CART). Variables are listed by an abbreviated name if an abbreviation was necessary to label a split; the names used for computer coding are given in parentheses. The last two sections cover variables examined but not used in a tree, and variables that were not even included in the analysis for various reasons.

Data sources

The Columbus Health Department issues about 950 food establishment and 3,527 food service licenses annually in Columbus and Worthington (including temporary food services and multiple licenses at some sites).[38] In February 1982 the Department began using the Sanitation Programs Information Formulator (SPIF) computer system.[39] Eventually about 35 state or local health departments adapted their own customized version of this system.[40] SPIF is a low-cost food inspection data processing system designed to provide violation frequency and manpower allocation data. It also provides sanitarians with reports intended to assist them in scheduling inspections at pre-set intervals. SPIF prints lists of establishments, mailing labels, form letters, and other data to help the clerical staff.

The dataset is incomplete in two ways. Data entered earlier than about five years ago was purged to save space. Also, the program the Data Center used to dump data to tapes seems to have had a problem of some sort, resulting in many inspection records without profile records to match, and vice-versa. There were 29,765 inspection

---

38 Moore RA. Columbus City Health Department food service program administrative & field sanitation survey. 1990: Ohio Department of Health [unpublished program review].

39 Guerin JP, Keeling H. System documentation. Vol. III of sanitation programs information formulator: user's guide. Washington, D. C.: U. S. Food and Drug Administration; 1975.

40 Hutchinson R, personal communication, May 14, 1990.

records, but 1,937 had no master record to match; and there were 3,795 master records, of which 929 were unusable because there were no matching inspections. (288 of the 929 were licensing information for pools, spas, and commercial sewage facilities, so they did not belong here anyway. 376, or 40%, of the rest were for food vending machine locations). However, there did not appear to be any relationship between missing records and outcome variables.

The inspection records analyzed here cover the period January 2, 1986 to December 4, 1990. Figure 2 shows the number of inspections in each year the data covers (not including two records with "year" miscoded). Likewise, Figure 3 breaks down the last inspection for each operation by year. (One had "year" miscoded.) It may reflect missing inspection records or business failures.

The SPIF system uses four data files, described in Unit 11 of the SPIF manual.[39] This study will use the Master File, in which each card represents a Profile Form (Appendix D); and the Inspection File, the most active file in the system, in which each card represents an Inspection Report (Appendices A and C). Appendix E is sample license applications. Note that the profile form contains all the information that the Ohio Department of Health requests on the applications. In this health department the operator

FIGURE 2

INSPECTIONS BY YEAR

YEAR	Freq	Cum. Freq	Percent	Cum. Percent
86	6830	6830	22.95	22.95
87	6265	13095	21.05	44.00
88	6588	19683	22.13	66.13
89	6226	25909	20.92	87.05
90	3854	29763	12.95	100.00

500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500

Inspections

FIGURE 3

LAST INSPECTIONS BY YEAR

YEAR	Freq	Cum. Freq	Percent	Cum. Percent
86	677	677	11.11	11.11
87	724	1401	11.88	22.99
88	918	2319	15.06	38.05
89	1550	3869	25.43	63.49
90	2224	6093	36.49	99.98
99	1	6094	0.02	100.00

200 400 600 800 1000 1200 1400 1600 1800 2000 2200

Inspections

normally does not complete the application itself except for the date and signature; instead, sanitarians or clerks collect the information on a profile form.

Inspection interval

The inspection interval ("inspin"), in the Master File, is the recommended interval based on potential risk. It is continuous, but is usually used as ordinal (ordered categorical), and is usually 120, 180, or 360 days. The Columbus Health Department simultaneously adapted HACCP and a variable inspection frequency scheme with these three frequencies, based informally on the potential risk each category represents because of the type of food and processing and the clientele served. Thus, it incorporates Frank Bryan's risk coefficients (see page 24). Appendix G, "Risk assessment of food service operations," gives the actual criteria.

In a nutshell, full-menu restaurants are to be inspected three times per year. Fast-food establishments and Food Establishments (that is, markets and carry-outs) are to be inspected twice annually. Bars and coffee shops are inspected annually, the minimum frequency permitted by state law. Food vending machine locations are classified with markets and carry-outs due to their excellent safety record, and are scheduled to receive annual inspections.

(See pages 9 and 10, pages 16-17, and page 50 for additional comments about food vending machines.) It makes sense to analyze these three groups separately, because they <u>are</u> different. But another, perhaps more compelling, reason to separate them is to avoid some confusion that would result otherwise. Variables such as the number of extra regular inspections received, the average or most recent interval between inspections, the average duration of inspections, and even the number of previous violations of a particular kind, would all have different meanings in these different groups.

Outcome variables

<u>Time-temperature violation</u> ("timetemp")--violation of section 3701-21-05A of the Ohio Administrative Code, signifying inadequate temperature control in potentially hazardous food. Violators may also have other critical or non-critical violations or a score below 90.

<u>Inspection failure</u> ("fail")--Officially, this is impossible, because neither the Ohio Department of Health nor the Columbus Health Department has an official definition of "fail." However, our policy requires sanitarians to schedule a follow-up inspection whenever there is at least one critical violation (including a time-temperature violation) or the score is below 90.

Critical violations are marked with an asterisk (*) on the inspection forms (see Appendices A and C). These items are widely believed to be associated with outbreaks of foodborne illness. They match lists of causes of outbreaks, such as Table 1, a relation between these practices and outbreaks is biologically plausible, and many of them were implicated by the Seattle study.

The item labeled "follow-up" on our inspection forms probably matches "fail" quite closely. But there are some differences. If a critical violation is corrected on the spot, the inspector may not want to schedule a follow-up inspection. When "follow-up" is "yes," SPIF lists the operation on a monthly delinquent list, and supervisors investigate. On the other hand, department policy also requires a follow-up if there are "chronic repeated violations," but this would not necessarily mean critical violations or a score below 90.

Predictor variables

<u>Ave. interval betw. insp's</u> ("avinint2")--the operation's average (arithmetic mean) interval, in days, between regular inspections over the calendar year before the inspection used in the analysis (index inspection).

<u>SD of scores</u> ("scorevar")--the standard deviation of the operation's scores in all inspections over the calendar

year before the inspection used in the analysis (that is, the index inspection).

<u>(No.) extra insp's in prev. year</u> ("extra")--the number of extra <u>regular</u> inspections the operation received in the calendar year before the inspection used in the analysis. A full-menu operation normally receives 3 inspections per year, so "extra" would be "1" if such an operation were inspected 4 times.

<u>Prev. insp. (no.) days ago</u> ("datedif")--the number of days since the last regular inspection. The most recent interval between inspections.

<u>(No.) prev. 4A, etc</u> ("sum4a,...,sum18d")--the number of times the operation violated each of the 44 items on the inspection forms over the 5-year study period.

<u>Ave. duration (no.) min.</u> ("avintim")--the average duration of inspections in the year before the index inspection.

<u>Ave. income in zip code</u> ("zipincom")--this was a very crude estimate of household income. For each zip code, the median household income (in thousands of dollars, from the 1990 Census) in each census tract[41] in the zip code was summed over all census tracts in the zip code and divided by the number of census tracts. Some census tracts cross

---

41 Donnelley Marketing Information Services. Market Profile Analysis--Columbus, Ohio SMSA. New York, NY: Donnelley Marketing Information Services, 1991.

zip code boundaries.

FSO--coded "yes" for food service operations, "no" for food establishments.

Vending--a food vending machine commissary or location. There are 1,729 licensed food vending machine locations;[38] individual machines at these locations are not licensed separately.

Freq of fail, criticals, 4A, etc--the frequency of the violation in regular inspections: the ratio of the number of failures, inspections citing critical violations, etc., divided by the number of regular inspections in the calendar year before the inspection used in the analysis. Variables handled in this way were each of 44 individual violations, the 15 categories of violation (food, food protection, personnel, etc., as listed on the inspection report), instances of failure, a score below 90, and inspections in which at least one critical violation is marked.

On one hand, this type of variable had a moderately high importance; on the other hand, CART rarely made splits on it. A slight error was just discovered in this variable: the denominator should have been the number of regular inspections in the entire follow-up period, not just the number in the 365 days before the inspection of interest. The meaning of this is that a more stable expression for the idea was available than was used. To the extent that the previous year's number of regular inspections was representative of every year's, the variable is just too big by a factor of 5, the number of years represented in its numerator's data. CART works the same on data transformed in this way. Figures 10 and 11 have corrected labels for the splits.

Variables not used by CART

Variables indicating a time-temperature violation, srore below 90, any critical item, or an inspection failure in the <u>last inspection</u> were never used by CART.

Many variables described below do not figure into any trees presented here, but showed up in preliminary trees or as surrogate splits:

<u>Food, food protection, personnel</u>, etc. ("food," "foodprot," etc)--"ever violated" one of the 15 categories of violation as listed on the inspection report over the 5-year study period. For example, if an operation ever violated 5A or 5H, "foodprot" would be "1."

<u>Commercial</u> ("commerc")--coded "0" ("no") if the type of operation listed on the profile was 01N, 015, etc. "Type" refers to the type of establishment. Refer to Appendix F and Appendix H, respectively, for the types and the exact manner of coding of these variables.

<u>Any violation before</u> ("anyviol")--coded "yes" if the operation ever had a violation before. Believe it or not, some operations never have violations.

<u>Purpose</u> ("purpos")--the purpose of the index inspection, coded "1" for "regular," "2" for "follow-up," and "3" for all other purposes, mostly inspections elicited by complaints. The purpose of a HACCP evaluation would be coded "3." (A hazard analysis uses the standard inspection form, with "purpose" marked "8 (other)" to report violations; it also uses a special form and instructions included in Appendix G).

<u>Size</u>--seating capacity categories for food service operations. Coded "1" for 0-74 seats, "2" for 75-99 seats, and "3" for over 100.

<u>Ethnic</u> ("ethni")--coded "0" for American cuisine (the default, 97.1% of all establishments), "1" for ethnic but not Asian (73 establishments, 1.2%), and "2" for Asian (102, 1.7%). The variable is based on data on restaurant or food market ethnicity from an article in a local newspaper[42]. Some operations were also assigned an ethnic status on the basis of their name. The original categories were American, Soul/Caribbean (5 establishments); French (4); German (16); Greek (34); Indian (6); Mexican/Spanish

---

42 Nolan T, Mallett K. Ethnic foods add spice to Columbus. Columbus alive! 1989: Nov.23-Dec. 7; 8-12.

(7); Middle Eastern (6); Chinese, Japanese, Korean, Thai (106); and Slavic (3). Due to missing records or some other reason, a few ethnic operations could not be included in the analyses.

Neither the Columbus Health Department nor the Ohio State University discriminate against anyone on account of race, religion, color, sex, handicap, age, national origin, or sexual orientation. Categorization of operators by ethnic category is not intended to result in differential services of any sort.[43]

Variables not included

Four variables listed in Table 4--city, district, subdistrict, and sanitarian assigned to the subdistrict--are not suitable for inclusion in the main part of the analysis for several reasons. For one thing, they have no meaning outside of Columbus. Also, the effect of sanitarian assigned to a subdistrict is nested in the effect of the subdistrict, if there is one (unless the sanitarian does inspections outside his or her assigned subdistrict), and similar nesting of the other variables precludes the assessment of the effect of each variable. Early CART runs did not split on these variables,

---

43 Myers WC. Statement of civil rights compliance for staff [memorandum]. Columbus Health Department 1990; June 29.

indicating that our food program is more or less uniformly administered among the districts. The average income in the subdistricts is probably a more useful dimension than any of these other four.

Certain variables bearing a close association with time-temperature violations or failures were deleted before the final analysis. These variables were "critical" and "ever had a critical," and "score below 90" and "ever had a score below 90" (in analyses for either outcome variable); "fail" and "ever failed before" (in analyses for time-temperature violations), and "time-temperature" and "ever had a time-temperature violation before" (in analyses for inspection failure). If these kinds of variables were presented to CART, CART would make splits on them, instead of splitting on variables with real explanatory power. Trees grown using this type of variable and data representing the last failure or time-temperature violation had poor predictive power when tested on data representing the last inspection.

Some variables were redundant. It was not obvious, but perhaps should have been, that it was a mistake to include them. "Number of regular inspections" ("purplyr") represented the number of regular inspections in the previous year. But "extra" represented the number of regular inspections beyond those required. Because the three different categories of inspection interval were analyzed separately, "purplyr" and "extra" differed only by an integer, so they are equivalent under CART; but "extra" is a bit clearer.

"Inspection ratio"--the ratio of the actual inspection interval to the scheduled interval. This variable was the ratio of the actual and recommended inspection intervals. CART found the actual interval more useful when operations were already classified by recommended interval.

CHAPTER VII

Results

A brief digression into how CART's performance can be measured by calculating sensitivity, specificity, and predictive value will be followed by the presentation of prediction trees for inspection failures and time-temperature violations for each of the three categories of operation (pre-classified by inspection interval). Only the first tree will be discussed in depth. The chapter concludes with a presentation of trees to predict outcomes using only information available at licensing, before the first inspection.

Performance measurement: sensitivity and specificity

The performance of a tree (or any screening test) can be measured in terms of sensitivity and specificity. A diagrammatic "confusion matrix"[44] showing actual versus predicted group membership is shown in Table 4. The sensitivity of a screening test for a disease is "the

---

44 Johnson RA, Wichern DW. Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1982, p. 488.

ability of a test to identify correctly those who have the disease,"[45] and is given by Equation 1. It is also the number that test positive and have the disease divided by the number with the disease, or the number of true positives divided by the sum of true positives and false negatives, and is often expressed as a percentage. Specificity is the ability to identify correctly those who do not have the disease, and is given by Equation 2. The positive and negative predictive value (PV+ and PV-), given by Equations 3 and 4, are the proportion of true positives and negatives, respectively, that are correctly identified by the test. These statistics are all influenced by the prevalence of the disease. The prevalence is the frequency, probability, or risk of having the disease. For a given test, the higher the prevalence, the higher the

Table 4. Disease states with results of a screening test

Test results (predicted class)	Disease state (true class)	
	no disease	disease
negative	true - (a)	false - (b)
positive	false + (c)	true + (d)
total	all without disease (a + c)	all with disease (b + d)

45 Mausner JS, Kramer S. Epidemiology--an introductory text. Philadelphia: W. B. Saunders Co., 1985.

sensitivity and predictive value.

$$\text{sensitivity} = \frac{d}{b + d} \qquad \text{(Eq. 1)}$$

$$\text{specificity} = \frac{a}{a + c} \qquad \text{(Eq. 2)}$$

$$\text{PV +} = \frac{d}{c + d} \qquad \text{(Eq. 3)}$$

$$\text{PV -} = \frac{a}{a + b} \qquad \text{(Eq. 4)}$$

$$\text{Prevalence} = \frac{b + d}{a + b + c + d} = p\,(D) \qquad \text{(Eq. 5)}$$

Risk markers for inspection failure

CART almost always did a much better job of predicting failures (or time-temperature violations) in the last inspection of each establishment when it used "enriched" data to form its tree, rather than using the last inspection results directly. The end of Chapter V (see Figure 1) explained how SAS generated this. The trees were formed using a 1,000-establishment learning sample from these "enriched" data sets. Larger learning samples would have been better, but the computer memory available to CART was insufficient to handle more observations. The trees were cross-validated, then tested with the population from which the learning sample was taken. Finally, the algorithm was evaluated using fresh data: all the records of the last inspection. Also, all trees presented were grown to maximize sensitivity, that is, to detect as many failures (or time-temperature violations) as possible. This was at the expense of a higher predictive power.

I. Full-menu restaurants. Figure 4 gives the algorithm CART generated for classifying full-service restaurants into categories with higher and lower "ever failed" rates. The test results are shown in Table 5. Appendix K is a copy of the output from the CART job used to create the algorithm.

In Figure 4, of the 1,621 full-menu establishment records available, the program has drawn a random sample of 1,000 as a learning sample. The sample contains 528 failures. Due to sampling error, the apparent failure rate was 53%. These last three statistics are shown in the first box at the left as N, n, and p. The first split was on question 1, "was the standard deviation of restaurants' scores in the previous year above 1.95?" Of the 573 establishments for whom the answer was "yes," 398 (69%) failed the index inspection. CART labeled these operations as failures; the "+" indicates, therefore, that the 398 failures can be counted as correctly identified in sensitivity calculations. The bar at the right end of this top box in the figure indicates that this was a terminal node: it was not split further. The 427 with a score SD of 1.95 or less were split further.

Not having had any extra inspections in the year before the index inspection (that is, having had only the three inspections full-menu establishments normally receive) split off a group of 159 with a 54% rate, which was not split further, and a group of 268 containing 44 that failed. One hundred seventeen restaurants, with only 3 failures, were removed in the third split: they had an average interval of more than 241 days between inspections. A fourth and final split on the remaining 151 restaurants resulted in separation of the 27% failure group into a 40% group and an 8% group. Restaurants receiving 2 or 3 extra inspections had the higher chance of failure.

Here is what all this means: higher failure rates are found in full-menu restaurants with a score SD above 1.95 (69%), OR ones that had no extra inspection (54%), OR ones with an average interval between inspections of no more than 241 days AND 2 or 3 extra inspections (40%). Any terminal node after the first can be thought of as representing an interaction. For example, the 159-restaurant node had relatively consistent scores AND no extra inspections.

This result generally makes sense, but is partly counter-intuitive. Restaurants with a variable score are not under effective control, and might be expected to have problems. If not receiving one or more extra inspections is a risk marker for failure, the inspections are generally doing what they are supposed to be doing. It is difficult to explain why having an interval exceeding 241 days would have a protective effect. Perhaps sanitarians know which restaurants will get along without an inspection for longer than the recommended 120-day interval, so they skip them. The precise way the variables interact is a little difficult to explain.

Sensitivity and specificity. Table 5 reformulates this tree's performance in terms of sensitivity and specificity, based on terminal groups. Results for the learning sample are given in the first of the four blocks of calculations because, although they are always overoptimistic, CART's classification tree diagram shows the learning sample results. In this study 10-fold cross-validation was always used for a more accurate assessment of the proportion of cases misclassified in the learning sample; the result is given in the second block. The third block gives the performance using all the "enriched" data, not just a sample of it. This may be the the most accurate assessment of the predictive power of the tree. The fourth block gives the results when the tree formed with the large number of last failures is tested on the series of last inspections. This is the most realistic situation, but the much smaller number of failures may have created instability. The next chapter comments on this problem.

Although all of the terminal nodes in Figure 4 contain both failing and passing establishments, CART identifies each terminal node as "pass" or "fail." Because three of the nodes were classified as "fail," indicated by the "+" signs in them, their failures can be added together to get the number of predicted failures: 398 + 86 + 36 = 520. This is shown in the first block in Table 5 at the intersection of the second column and second row. The algorithm missed 8 true failures (3 + 5). This is shown at the intersection of the second column and first row. Thus the tree's sensitivity (based on terminal groups in the learning sample) is 520/528 = 98%. The cost of such high sensitivity was a relatively low specificity, only 36%.

The sensitivity and specificity according to cross-validation in the 1,000-restaurant learning sample (the second block of Table 5) were 96% and 39%, respectively. These are more accurate statistics for the learning sample.

The fourth block in Figure 5 shows the result when the tree was retested on the original population of 1,621 restaurantsfrom which the learning sample was taken, with inspections after the last failure dropped. The sensitivity has deteriorated to 75% with no change in specificity from the cross-validation result using the test sample.

How does this scheme perform when applied to the last inspection (not "enriched" by using all the last failures)? The last of the four blocks in Table 5 shows that the sensitivity slips to 68%, and the specificity increases to 46%. Ultimately, this was the best CART could do with the real problem: predicting the next failures. This, then, shows the success of predictions on a "real-life" situation. Nine percent (139) of the 1,621 full-menu establishments for which records were available failed their last regular inspection (they had a score below 90, a critical violation, or both). This is shown in the prevalence, or probability of failure [P(fail)] line for the test sample, which was actually a population--all the data from the last inspection.

The CART output used to create this algorithm has been reproduced as Appendix K. The variable "purplyr" the printout shows was the number of regular inspections. Note that CART gives the variable "extra" as both a surrogate and a competitor for split 2. In this group, "extra LE 0.5" means the same as "purplyr LE 3.5"--no extra inspections. The observant reader who examines Appendix K may wonder about the meaning of the fact that competing splits are always present. An important aspect of these classification trees is instability. Changing one or more parameter settings can vastly change the topology of the resulting tree. Put another way, other variable combinations can have approximately the same predictive power. But be wary of full-menu restaurants with highly variable scores!

<u>II. Fast-food establishments, markets, and carry-outs</u> have potentially hazardous food, but generally do little processing. Therefore, they normally receive two inspections per year. The percentage that failed at least once in 5 years (42%) was less than for full-menu restaurants (49%).

Figure 5 gives the results when CART classifies fast food restaurants and licensed "food establishments" (FE's). As before, using the last failure, rather than the last inspection, gives the best results. Also, as before, the first split is on the standard deviation of inspection scores during the year before the index inspection, but now the cut point is 2.00. The next split (for the lower-risk group) is on the number of days since the last regular inspection; oddly, the higher-risk category in this group had their last inspection more recently. Of the 461 establishments in this group, 199, with a higher (43%) failure rate, form a terminal node based on their having had a shorter average interval between inspections. Of the remainder, the higher-risk set received no more than one extra inspection. Returning to split 2, operations with a longer lag since the last inspection were more likely to fail if they received no more than 1 extra inspection.

To summarize, fast-food establishments and markets were more likely to fail if they had variable scores (69% failure rate). If the scores were consistent AND the last inspection was within about a year BUT inspections were usually more frequent, the failure rate was 43%. If the scores were consistent AND the last inspection was within about a year AND inspections were infrequent AND one or fewer extra inspections took place, the rate was 51%. If the scores were consistent BUT the last inspection was a year ago AND one or fewer extra inspections took place, 33% failed.

Unfortunately, as Table 6 indicates, this classification scheme did no better than chance at predicting the result of the last inspection. A positive prediction was correct only 9% of the time. The failure rate was 7% in the last inspection series in fast-food operations and markets (it was 9% in full-menu food services). Yet the actual number of failures was higher here.

The results still seem somewhat credible, because some of the same risk markers show up consistently: variable scores and few extra inspections.

<u>III. Bars.</u> Figure 6 gives the results when CART classifies taverns, retail donut shops, coffee shops, and similar establishments according to risk markers it has identified. The variable that best differentiates operations with a failure in their history from those without one is, again, the standard deviation of the operations' scores. From the original "learning sample" of 1,000 operations with a 23% "eventual failure" rate, a score SD above 1.85 separates out a group of 140 with a 45% failure rate. From the 860 remaining, with a 19% failure rate, 689 (with the same failure rate) cannot be differentiated further. Establishments with consistent scores and an average interval of less than 274 days were at risk (28% failed) if their last inspection was less than about 2 years ago. This seemingly counter-intuitive result has a simple explanation: the 70-operation group (with one failure) that was removed probably contains largely food vending locations. CART lists vending as a competing split. Vending locations tend to be safer and less

frequently inspected.

This scheme did remarkably well when used to predict results of the last inspection: its sensitivity was 76%, as indicated by Table 7.

Risk markers for time-temperature violations

Figures 7 through 11 and Tables 8 through 12 show results for time-temperature violations, rarer events than inspection failure, but the tables still say "pass" and "fail."

<u>I. Full-menu restaurants.</u> Figure 7 shows the classification generated from the last violation. Table 8, block 3, shows that two hundred eighty-five full-menu restaurants (18%, almost a fifth) had the kind of violation most likely to cause illness at least once in 5 years, even though only 2% had it as of the last inspection (block 4).

Again, the first split was on variability of scores, but they had to be more variable than for mere failure. Operations with a score SD above 2.75 and a previous time-temperature violation had an extraordinarily high likelihood of having another--82%. Variable-score restaurants without that dangerous risk marker were still at risk (42% had the violation) if they received no extra inspections. Among consistent scorers, ones with no more than one extra inspection and an average interval of no more than 284 days had a 32% chance of a time-temperature violation.

The scheme correctly identified only 24% of violators in the last inspection; however, there were only 34 of them, in 2% of the full-menu restaurants.

Figure 8 and Table 9 show an alternate tree, formed using the last inspection directly, and associated testing. The first test sample here was the second test sample for Table 8 (note that the true class totals are identical). This tree did better for the last inspection (sensitivity = 62%); however, it only identified half the ever-violators (see Table 9). This one has the advantage of simplicity. Three percent of the operations whose average inspections last longer than 41 minutes, but none of the ones normally taking less time, were violators--a clean split. Size was a surrogate, consistent with the findings of Irwin et al. that larger operations are more likely to have outbreaks.

<u>II. Fast-food establishments, markets, and carry-outs.</u> Table 10 (block 3) shows that only 153 of the 2,051 operations in this inspection interval category (7%) ever had a time-temperature violation. The corresponding prevalence in the learning sample shown in Figure 9 was 81 of 1,000 (8%) (block 1), and only 14 (1%) had this violation in the last inspection (block 4). CART could not

predict any of the 14. It identified 91% of the establishments that ever violated this item, according to cross-validation in the learning sample (block 2), but its sensitivity when retested with that same enriched population (block 3) was only 4%. The predictive value of a positive test was only 8% on retest with that data.

Its first split was on average interval between inspections, with the higher-risk group having the shorter mean interval. Operations with an average interval less than 304 days and a score SD above 2.85 had twice the average risk of a time-temperature violation. Food services (but not markets) with an average interval less than 304 days and a score SD less than about 2.85 had a risk of violation double the average if they received more than 3 extra inspections, surprisingly. Yet operations with an average interval in excess of 303 days and no extra inspections had three times the average risk.

<u>III. Bars.</u> Figures 10 and 11 and Tables 11 and 12 show CART's classification of bars and coffee shops using the last violation and the last inspection, respectively. The complex tree in Figure 10 has its first split on whether the operation is a food vending machine location, which makes sense. The next split is on the frequency of dirty floors. Subsequent splits are on score SD (above 3.45) and extra inspections, in the usual directions.

Table 11 gages its performance. Its cross-validated sensitivity to the occurrence of the 15 past violations in the learning sample (block 12 of Table 11) was mediocre, and its ability to detect any of the 19 violations in the entire population of enriched data (block 3) or any of the 3 violations that occurred in the last inspection (block 4) was zero.

Figure 11 shows a scheme that seems to be sightly more effective, despite its extreme simplicity. This used the last inspection directly, and used the last violation as a test sample. According to the cross-validation analysis in Table 12, it detected 2 of the 3 cases found in the last inspection. It performed at the same zero-sensitivity level as the one in Figure 7 when tested on the last inspection or the last violation.

Risk markers for failure using only information available at licensing

Figure 12 and Table 13 explore the possibility of detecting establishments likely to fail without relying on any inspection history. CART could use a double-size learning sample because only 7 variables were used: inspection interval, vending, FSO, ave. income, commercial, size, and ethnicity. (The variable inspection interval was allowable here, because no confusion could arise.) The table indicates that it can be done--the cross-validated sensitivity is 96% using "enriched" data--but it works no better than chance at predicting the result of the next inspection. This may be satisfactory, however: more interest would center on this long-term outcome than on any particular inspection. No test sample was used for the last failure.

The first split classifies operations other than food vending machine locations as eventual failures, and classifies vending locations as failure candidates if they are in a neighborhood with a median household income between $15,600 and $18,400. But the tree could be accused of throwing everything into a high-hazard category to be sure to catch those really at risk. Its specificity is only 9%. It predicts that only 114 of the 2,000 will never fail, but 1,321 never do.

Risk markers for time-temperature violations using only information available at licensing

Figure 13 and Table 14 relate to a tree whose performance was unbelievably accurate, considering the meager amount of information available to it. Its cross-validated sensitivity, based on a 2,000-case learning sample from the 6,094 total cases and the last violation, was 94%, although, again, it was not specific. But it correctly predicted 58% of the problems when tested with the last inspection, with a specificity of 64%. As Table 14 also indicates, 1% of all inspections uncover time-temperature violations, and 8% of all licensed establishments eventually have at least one in 5 years.

CART made its first split on inspection interval. Operations receiving one inspection per year (chiefly bars and coffee shops), other than food vending locations, had a 3% violation rate, but 354 vending locations (inspected no more than annually) had no time-temperature violation in 5 years. Operations inspected more than once annually were categorized as violators if they were full-menu and fast-food restaurants (12% rate). If they were markets or carry-outs, they were safe (1% had the violation) unless they were located in a lower-income neighborhood (8% had the violation, which also happened to be the overall average for all establishments).

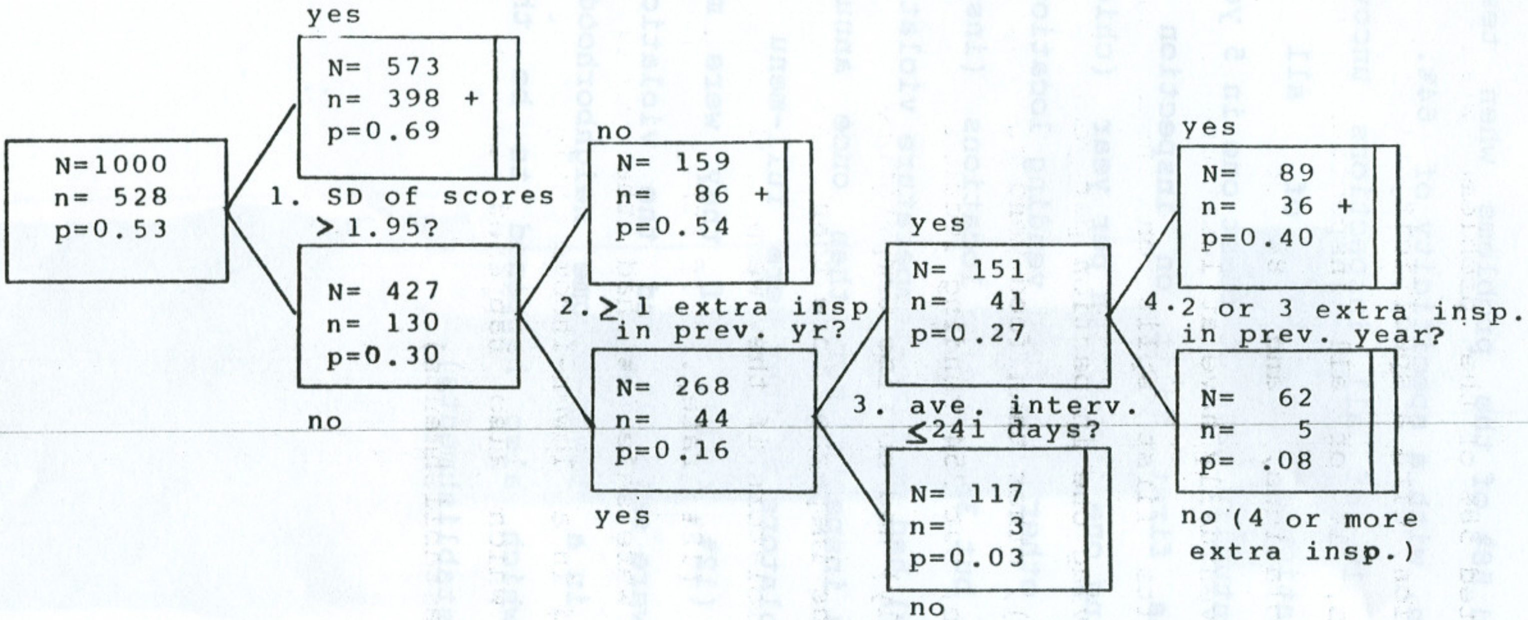

FIGURE 4. Inspection failure rates in a learning sample of 1,000 full-menu restaurants as a function of risk markers identified by CART.

Questions 1 through 4 about the risk markers classified 528 establishments that failed at least once, and others that never failed in 5 years, into groups with higher and lower failure rates (p).

TABLE 5. Sensitivity and specificity of the classification of full-menu restaurants according to risk markers identified by CART to explain or predict inspection failure.

		Learning sample (last failure)				Cross-validation (last failure)				Test sample (last failure)				Test sample (last inspection)		
		true class				true class				true class				true class		
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	171	8	179	pass	185	23	208	pass	324	202	526	pass	685	45	730
	fail	301	520	821	fail	287	505	792	fail	498	597	1095	fail	797	94	891
	total	472	528	1000	total	472	528	1000	total	822	799	1621	total	1482	139	1621

sensitivity	520 / 528 = .98	505 / 528 = .96	597 / 799 = .75	94 / 139 = .68
specificity	171 / 472 = .36	185 / 472 = .39	324 / 822 = .39	685 / 1482 = .46
PV +	520 / 821 = .63	505 / 792 = .64	597 / 1095 = .55	94 / 891 = .11
PV -	171 / 179 = .96	185 / 208 = .89	324 / 526 = .62	685 / 730 = .94
p (fail)	528 / 1000 = .53	528 / 1000 = .53	799 / 1621 = .49	139 / 1621 = .09

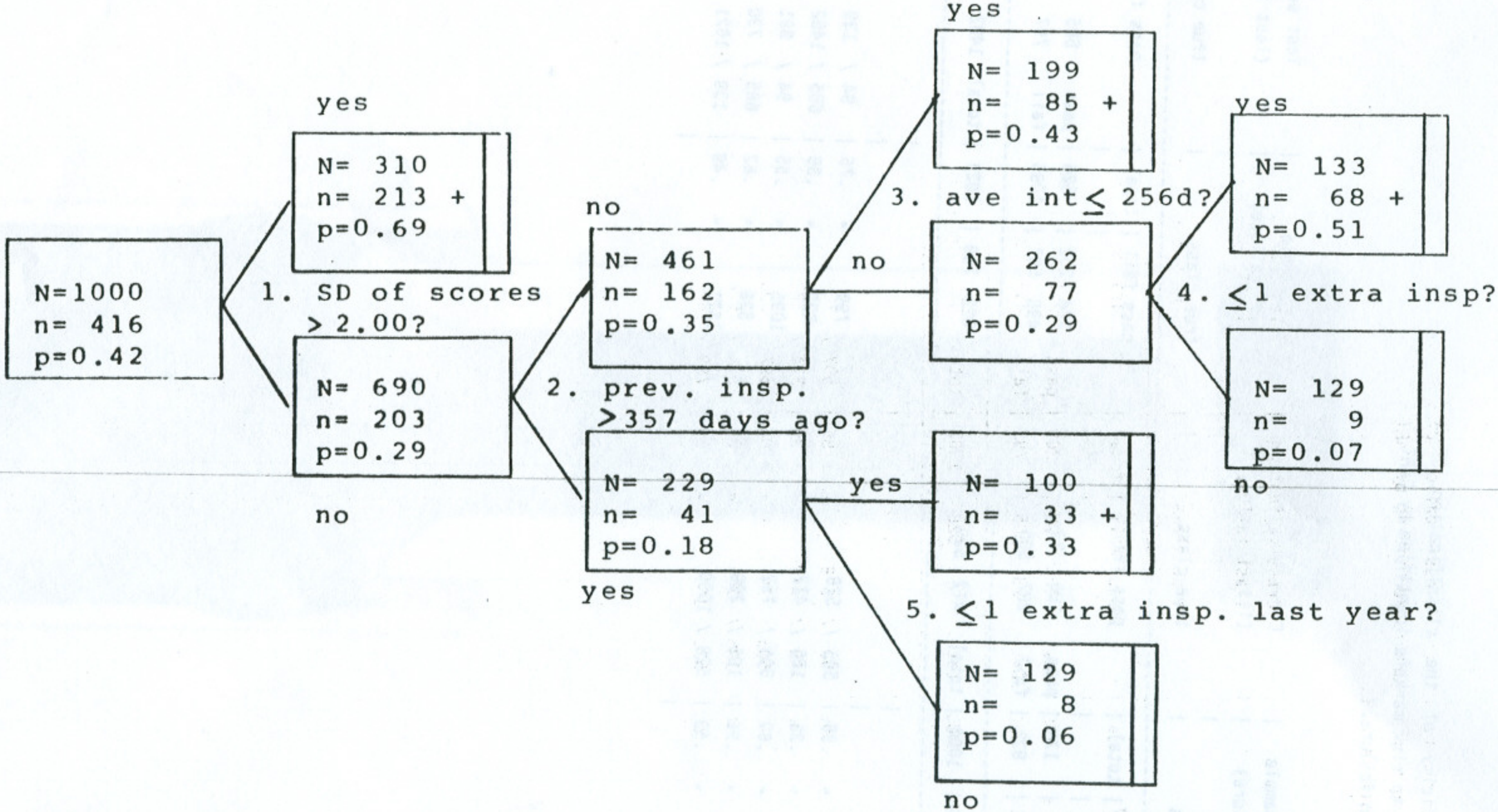

FIGURE 5. Inspection failure rates in a learning sample of 1,000 fast-food establishments, markets and carry-outs as a function of risk markers identified by CART.

Questions 1 through 5 about the risk markers classified 416 establishments that failed at least once, and others that never failed in 5 years, into groups with higher and lower failure rates (p).

TABLE 6. Sensitivity and specificity of the classification of fast-food establishments, markets and carry-outs according to risk markers identified by CART to explain or predict inspection failure.

	Learning sample (last failure) true class				Cross-validation (last failure) true class				Test sample (last failure) true class				Test sample (last inspection) true class			
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	241	17	258	pass	262	34	296	pass	660	328	988	pass	1190	76	1266
	fail	343	399	742	fail	322	382	704	fail	607	456	1063	fail	714	71	785
	total	584	416	1000	total	584	416	1000	total	1267	784	2051	total	1904	147	2051

	Learning sample (last failure)	Cross-validation (last failure)	Test sample (last failure)	Test sample (last inspection)
sensitivity	399 / 416 = .96	382 / 416 = .92	456 / 784 = .58	71 / 147 = .48
specificity	241 / 584 = .41	262 / 584 = .45	660 / 1267 = .52	1190 / 1904 = .63
PV +	399 / 742 = .54	382 / 704 = .54	456 / 1063 = .43	71 / 785 = .09
PV -	241 / 258 = .93	262 / 296 = .89	660 / 988 = .67	1190 / 1266 = .94
p (fail)	416 / 1000 = .42	416 / 1000 = .42	784 / 2051 = .38	147 / 2051 = .07

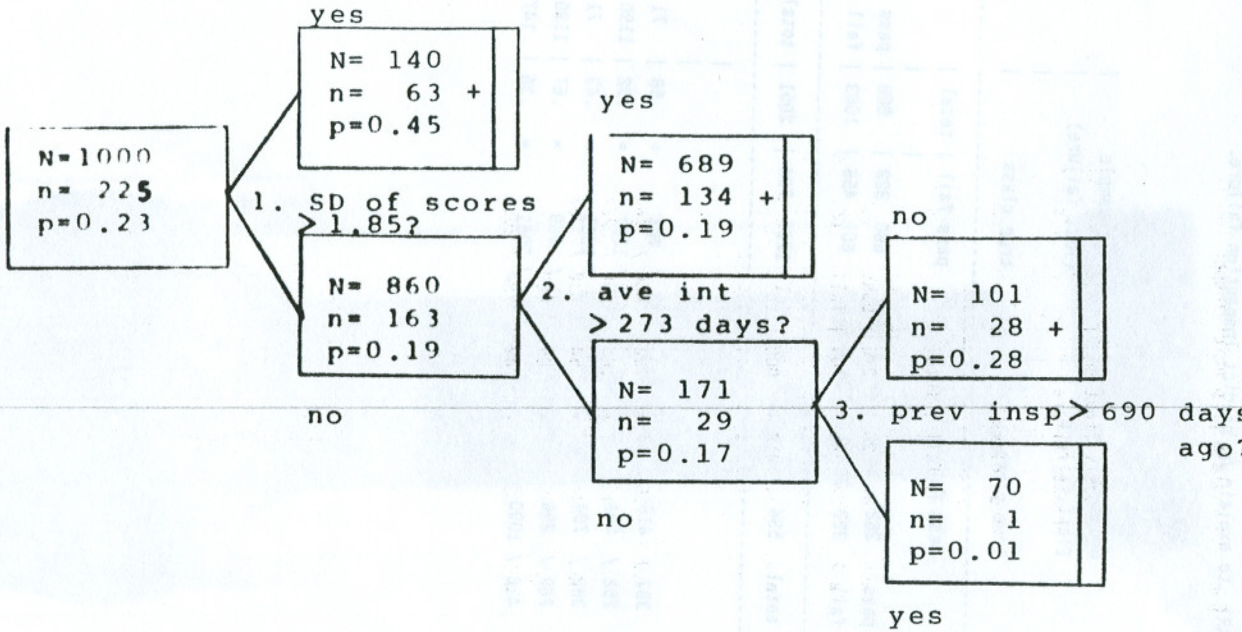

FIGURE 6. Inspection failure rates in a learning sample of 1,000 bars and coffee shops as a function of risk markers identified by CART.

Questions 1 through 3 about the risk markers classified 225 establishments that failed at least once, and others that never failed in 5 years, into groups with higher and lower failure rates (p).

TABLE 7. Sensitivity and specificity of the classification of bars and coffee shops according to risk markers identified by CART to explain or predict inspection failure.

	Learning sample (last failure)				Cross-validation (last failure)				Test sample (last failure)				Test sample (last inspection)			
	true class				true class				true class				true class			
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	69	1	70	pass	171	18	189	pass	253	58	311	pass	310	27	337
	fail	705	225	930	fail	603	208	811	fail	767	218	985	fail	874	85	959
	total	774	226	1000	total	774	226	1000	total	1020	276	1296	total	1184	112	1296

	Learning sample (last failure)		Cross-validation (last failure)		Test sample (last failure)		Test sample (last inspection)	
sensitivity	225 / 226	= 1.0	208 / 226	= .92	218 / 276	= .79	85 / 112	= .76
specificity	69 / 774	= .09	171 / 774	= .22	253 / 1020	= .25	310 / 1184	= .26
PV +	225 / 930	= .24	208 / 811	= .26	218 / 985	= .22	85 / 959	= .09
PV -	69 / 70	= .99	171 / 189	= .90	253 / 311	= .81	310 / 337	= .92
p (fail)	226 / 1000	= .23	226 / 1000	= .23	276 / 1296	= .21	112 / 1296	= .09

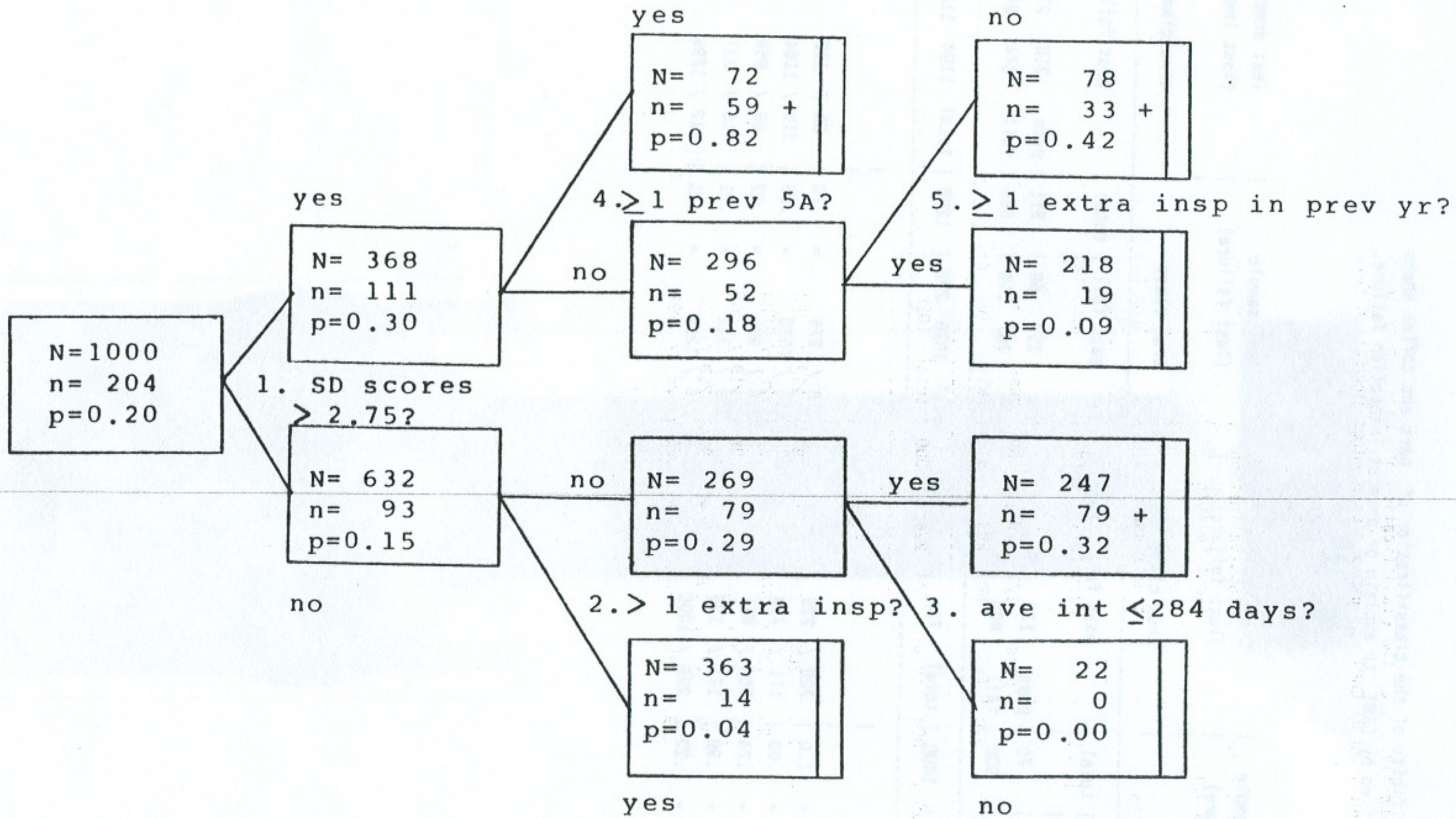

FIGURE 7. Time-temperature violation rates in a learning sample of 1,000 full-menu restaurants as a function of risk markers identified by CART.

Questions 1 through 5 about the risk markers classified 204 establishments that had a time-temperature violation at least once, and others that never had one in 5 years, into groups with higher and lower violation rates (p).

TABLE 8. Sensitivity and specificity of the classification of full-menu restaurants according to risk markers identified by CART to explain or predict time-temperature violations.

	Learning sample (last violation)				Cross-validation (last violation)				Test sample (last violation)				Test sample (last inspection)			
	true class				true class				true class				true class			
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	570	33	603	pass	558	43	601	pass	1047	154	1201	pass	1226	26	1252
	fail	226	171	397	fail	238	161	399	fail	289	131	420	fail	361	8	369
	total	796	204	1000	total	796	204	1000	total	1336	285	1621	total	1587	34	1621

	Learning sample		Cross-validation		Test sample (last violation)		Test sample (last inspection)	
sensitivity	171 / 204	= .84	161 / 204	= .79	131 / 285	= .46	8 / 34	= .24
specificity	570 / 796	= .72	558 / 796	= .70	1047 / 1336	= .78	1226 / 1587	= .77
PV +	171 / 397	= .43	161 / 399	= .40	131 / 420	= .31	8 / 369	= .02
PV -	570 / 603	= .95	558 / 601	= .93	1047 / 1201	= .87	1226 / 1252	= .98
p (violate)	204 / 1000	= .20	204 / 1000	= .20	285 / 1621	= .18	34 / 1621	= .02

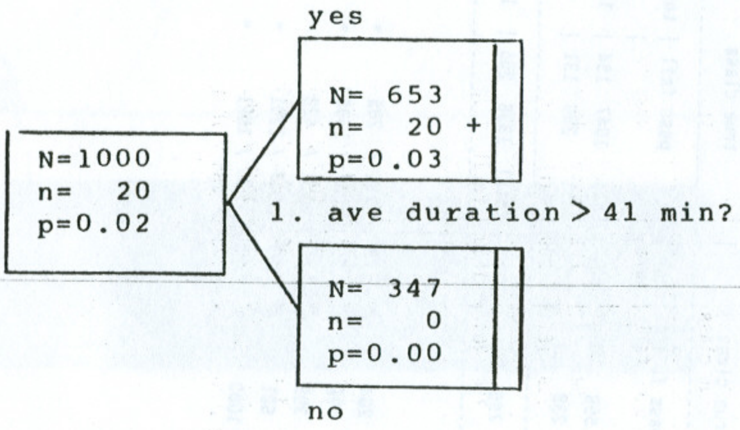

FIGURE 8. Time-temperature violation rates in a learning sample of 1,000 full-menu restaurants as a function of risk markers identified by CART using last inspection.

The risk marker question classified 20 establishments that had a time-temperature violation in their last inspection into groups with higher and lower violation rates (p).

TABLE 9. Sensitivity and specificity of the classification of full-menu restaurants according to risk markers identified by CART to explain or predict time-temperature violations using last inspection.

	Learning sample (last inspection)				Cross-validation (last inspection)				Test sample (last inspection)				Test sample (last violation)			
	true class				true class				true class				true class			
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted	pass	347	0	347	pass	416	5	421	pass	842	13	855	pass	663	140	803
class	fail	633	20	653	fail	564	15	579	fail	745	21	766	fail	673	145	818
	total	980	20	1000	total	980	20	1000	total	1587	34	1621	total	1336	285	1621
sensitivity	20 / 20 = 1.00				15 / 20 = .75				21 / 34 = .62				145 / 285 = .51			
specificity	347 / 980 = .35				416 / 980 = .42				842 / 1587 = .53				663 / 1336 = .50			
PV +	20 / 653 = .03				15 / 579 = .03				21 / 766 = .03				145 / 818 = .18			
PV -	347 / 347 = 1.00				416 / 421 = .99				842 / 855 = .98				663 / 803 = .83			
p (violate)	20 / 1000 = .02				20 / 1000 = .02				34 / 1621 = .02				285 / 1621 = .18			

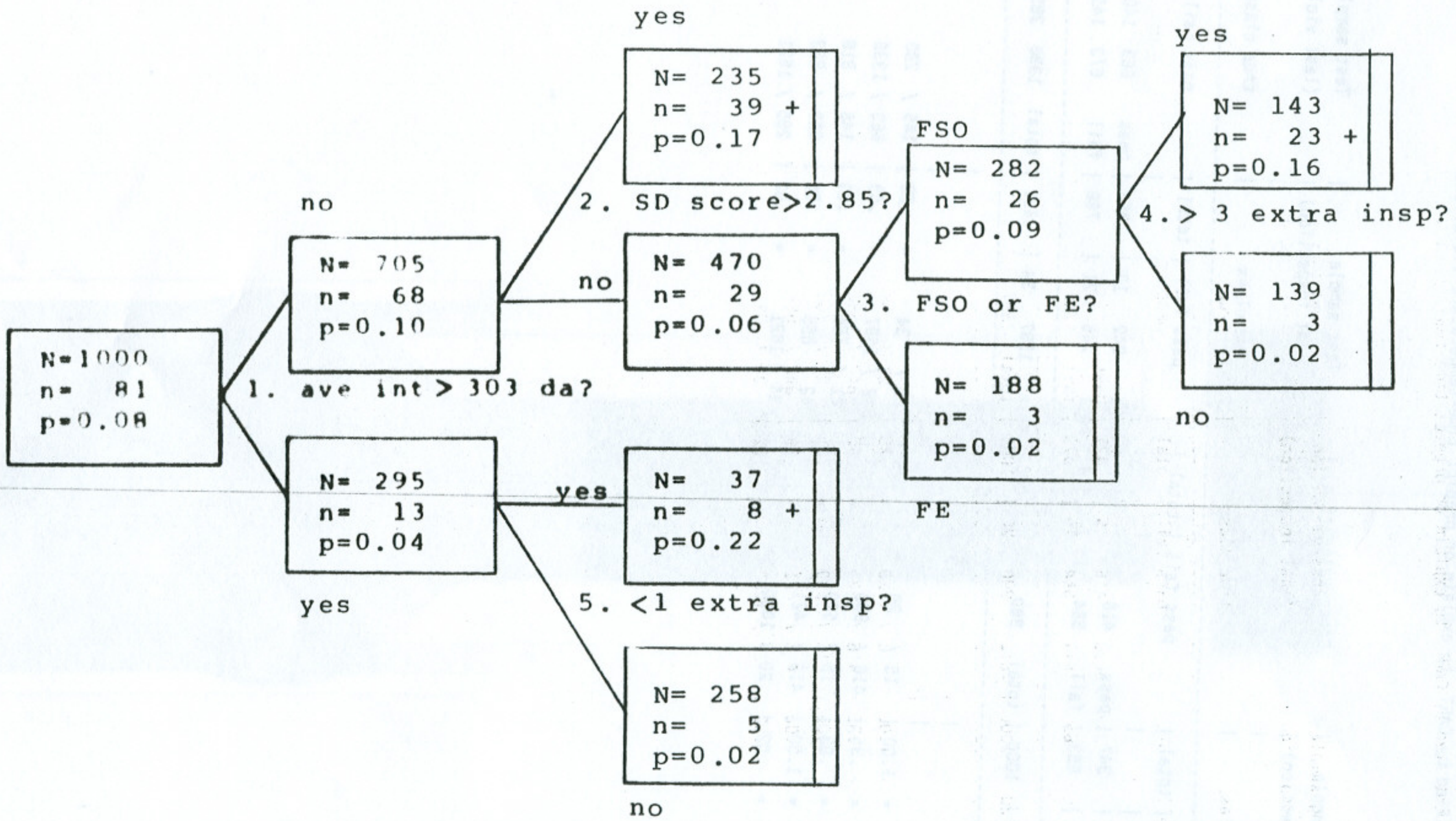

FIGURE 9. Time-temperature violation rates in a learning sample of 1,000 fast-food establishments, markets and carry-outs as a function of risk markers identified by CART.

Questions 1 through 5 about the risk markers classified 81 establishments that had a time-temperature violation at least once, and others that never had one in 5 years, into groups with higher and lower violation rates (p).

TABLE 10. Sensitivity and specificity of the classification of fast-food establishments, markets, and carry-outs according to risk markers identified by CART to explain or predict time-temperature violations.

	Learning sample (last violation) true class				Cross-validation (last violation) true class				Test sample (last violation) true class				Test sample (last inspection) true class			
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	574	11	585	pass	341	7	348	pass	1832	147	1979	pass	2031	14	2045
	fail	345	70	415	fail	578	74	652	fail	66	6	72	fail	6	0	6
	total	919	81	1000	total	919	81	1000	total	1898	153	2051	total	2037	14	2051

	Learning sample		Cross-validation		Test sample (last violation)		Test sample (last inspection)	
sensitivity	70 / 81	= .86	74 / 81	= .91	6 / 153	= .04	0 / 14	= 0.00
specificity	574 / 919	= .62	341 / 919	= .37	1832 / 1898	= .97	2031 / 2037	= 1.0
PV +	70 / 415	= .17	74 / 652	= .11	6 / 72	= .08	0 / 6	= 0.00
PV -	574 / 585	= .98	341 / 348	= .98	1832 / 1979	= .93	2031 / 2045	= .99
p (violate)	81 / 1000	= .08	81 / 1000	= .08	153 / 2051	= .07	14 / 2051	= .01

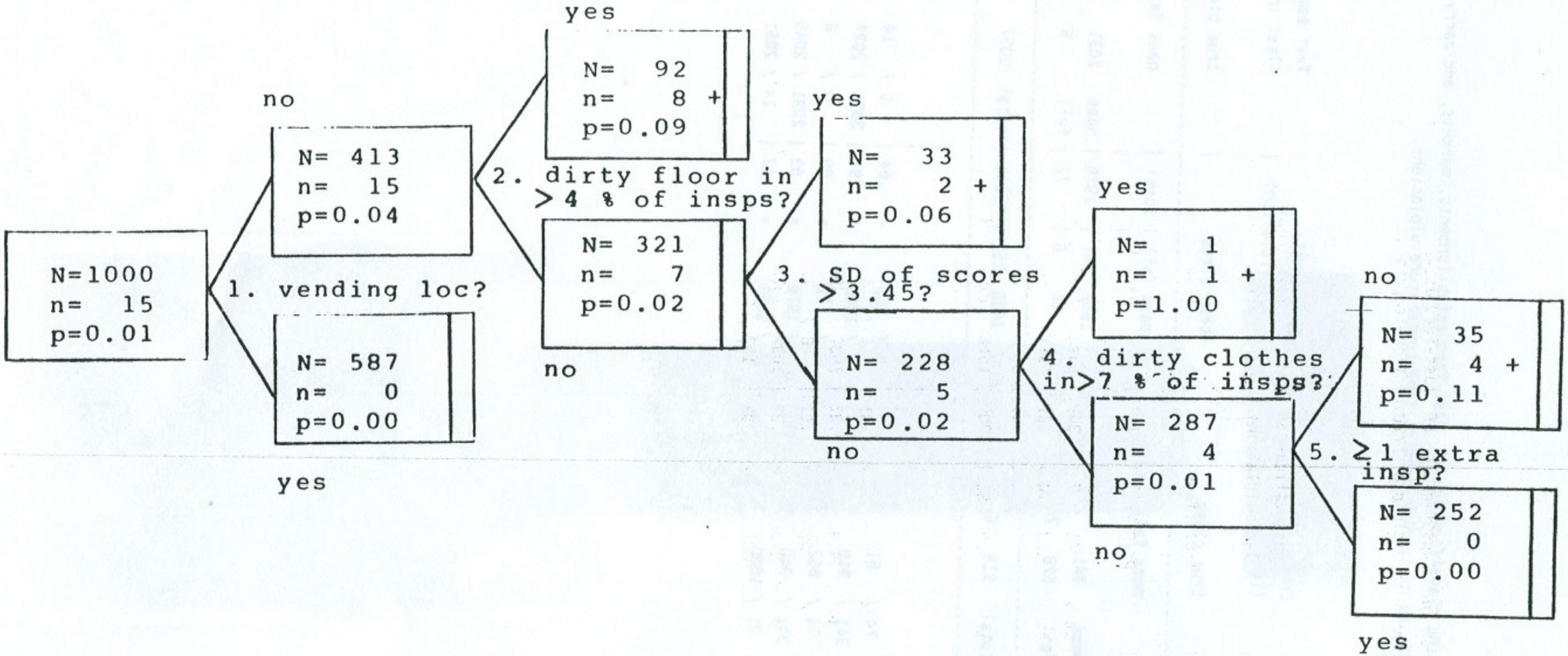

FIGURE 10. Time-temperature violation rates in a learning sample of 1,000 bars, coffee shops, and food vending machine locations as a function of risk markers identified by CART.

Questions 1 through 5 about the risk markers classified 15 establishments that had a time-temperature violation at least once, and others that never had one in 5 years, into groups with higher and lower violation rates (p).

TABLE 11. Sensitivity and specificity of the classification of bars and coffee shops according to risk markers identified by CART to explain or predict time-temperature violations.

	Learning sample (last violation)				Cross-validation (last violation)				Test sample (last violation)				Test sample (last inspection)			
		true class				true class				true class				true class		
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	839	0	839	pass	859	6	865	pass	1273	19	1292	pass	1293	3	1296
	fail	146	15	161	fail	126	9	135	fail	4	0	4	fail	0	0	0
	total	985	15	1000	total	985	15	1000	total	1277	19	1296	total	1293	3	1296

	Learning sample (last violation)		Cross-validation (last violation)		Test sample (last violation)		Test sample (last inspection)	
sensitivity	15 / 15	= 1.00	9 / 15	= .60	0 / 19	= 0.00	0 / 3	= 0.00
specificity	839 / 985	= .85	859 / 985	= .87	1273 / 1277	= 1.0	1293 / 1293	= 1.00
PV +	15 / 161	= .09	9 / 135	= .07	0 / 4	= 0.00	0 / 0	= ERR
PV -	839 / 839	= 1.00	859 / 865	= .99	1273 / 1292	= .99	1293 / 1296	= 1.0
p (violate)	15 / 1000	= .02	15 / 1000	= .02	19 / 1296	= .01	3 / 1296	= .00

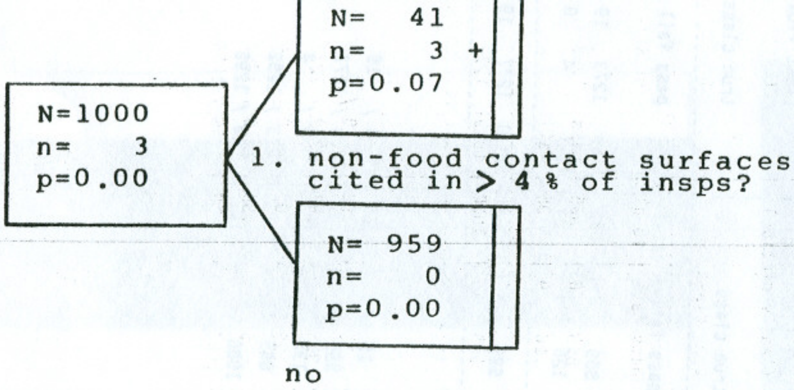

FIGURE 11. Time-temperature violation rates in a learning sample of 1,000 bars, coffee shops, and food vending machine locations as a function of risk markers identified by CART using the last inspection.

The question about a risk marker classified 3 establishments that had a time-temperature violation into groups with higher and lower violation rates (p).

TABLE 12. Sensitivity and specificity of the classification of bars and coffee shops according to risk markers identified by CART to explain or predict time-temperature violations using the last inspection.

		Learning sample (last inspection)				Cross-validation (last inspection)				Test sample (last inspection)				Test sample (last violation)		
		true class				true class				true class				true class		
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	959	0	959	pass	971	1	972	pass	1293	3	1296	pass	1273	19	1292
	fail	38	3	41	fail	26	2	28	fail	0	0	0	fail	4	0	4
	total	997	3	1000	total	997	3	1000	total	1293	3	1296	total	1277	19	1296

	Learning sample (last inspection)	Cross-validation (last inspection)	Test sample (last inspection)	Test sample (last violation)
sensitivity	3 / 3 = 1.00	2 / 3 = .67	0 / 3 = 0.00	0 / 19 = 0.00
specificity	959 / 997 = .96	971 / 997 = .97	1293 / 1293 = 1.00	1273 / 1277 = 1.0
PV +	3 / 41 = .07	2 / 28 = .07	0 / 0 = ERR	0 / 4 = 0.00
PV -	959 / 959 = 1.00	971 / 972 = 1.0	1293 / 1296 = 1.0	1273 / 1292 = .99
p (violate)	3 / 1000 = .00	3 / 1000 = .00	3 / 1296 = .00	19 / 1296 = .01

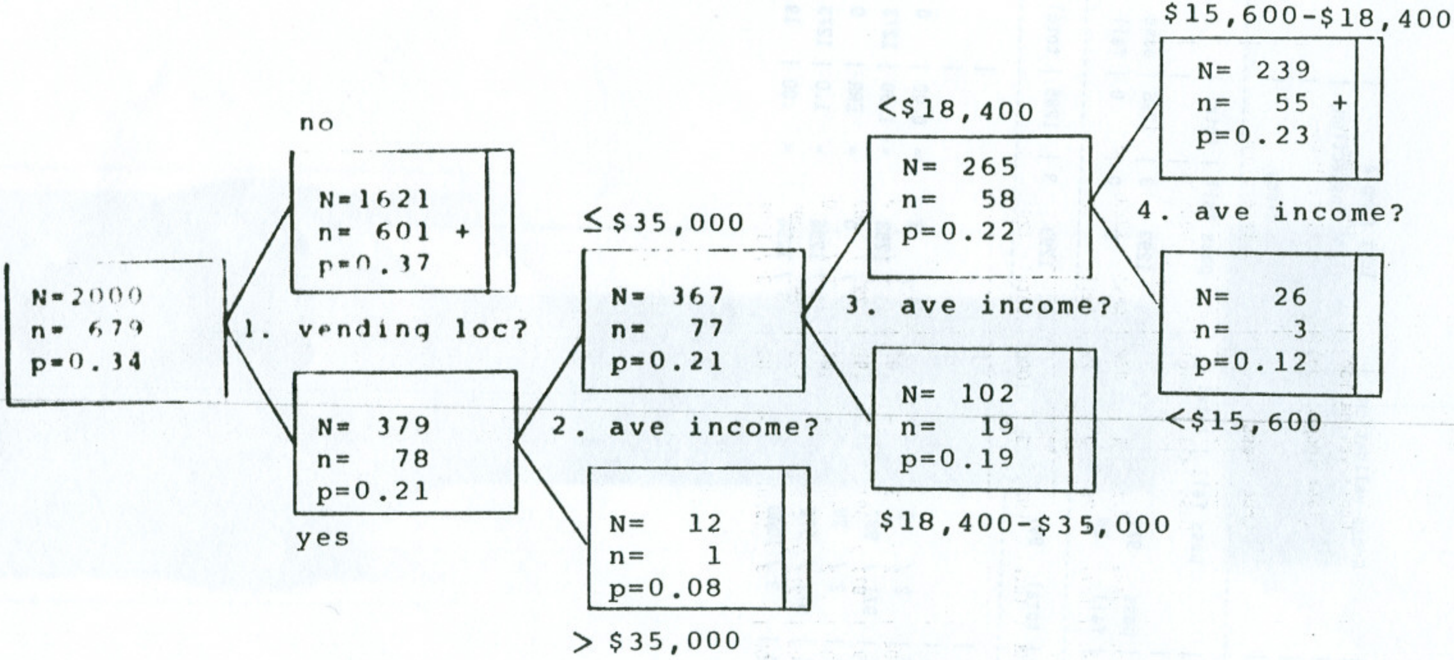

FIGURE 12. Inspection failure rates in a learning sample of 2,000 retail food operations as a function of risk markers identified by CART.

Questions 1 through 4 about the risk markers, involving only information available at licensing, classified 679 establishments that failed at least once, and others that never failed in 5 years, into groups with higher and lower failure rates (p).

TABLE 13. Sensitivity and specificity of the classification of all retail food operations according to risk markers identified by CART to explain or predict inspection failure, using only information available at licensing.

		Learning sample (last failure)				Cross-validation (last failure)				Test sample (last inspection)		
		true class				true class				true class		
		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	117	23	140	pass	114	24	138	pass	2871	257	3128
	fail	1204	656	1860	fail	1207	655	1862	fail	2725	241	2966
	total	1321	679	2000	total	1321	679	2000	total	5596	498	6094

	Learning sample		Cross-validation		Test sample	
sensitivity	656 / 679	= .97	655 / 679	= .96	241 / 498	= .48
specificity	117 / 1321	= .09	114 / 1321	= .09	2871 / 5596	= .51
PV +	656 / 1860	= .35	655 / 1862	= .35	241 / 2966	= .08
PV -	117 / 140	= .84	114 / 138	= .83	2871 / 3128	= .92
p (fail)	679 / 2000	= .34	679 / 2000	= .34	498 / 6094	= .08

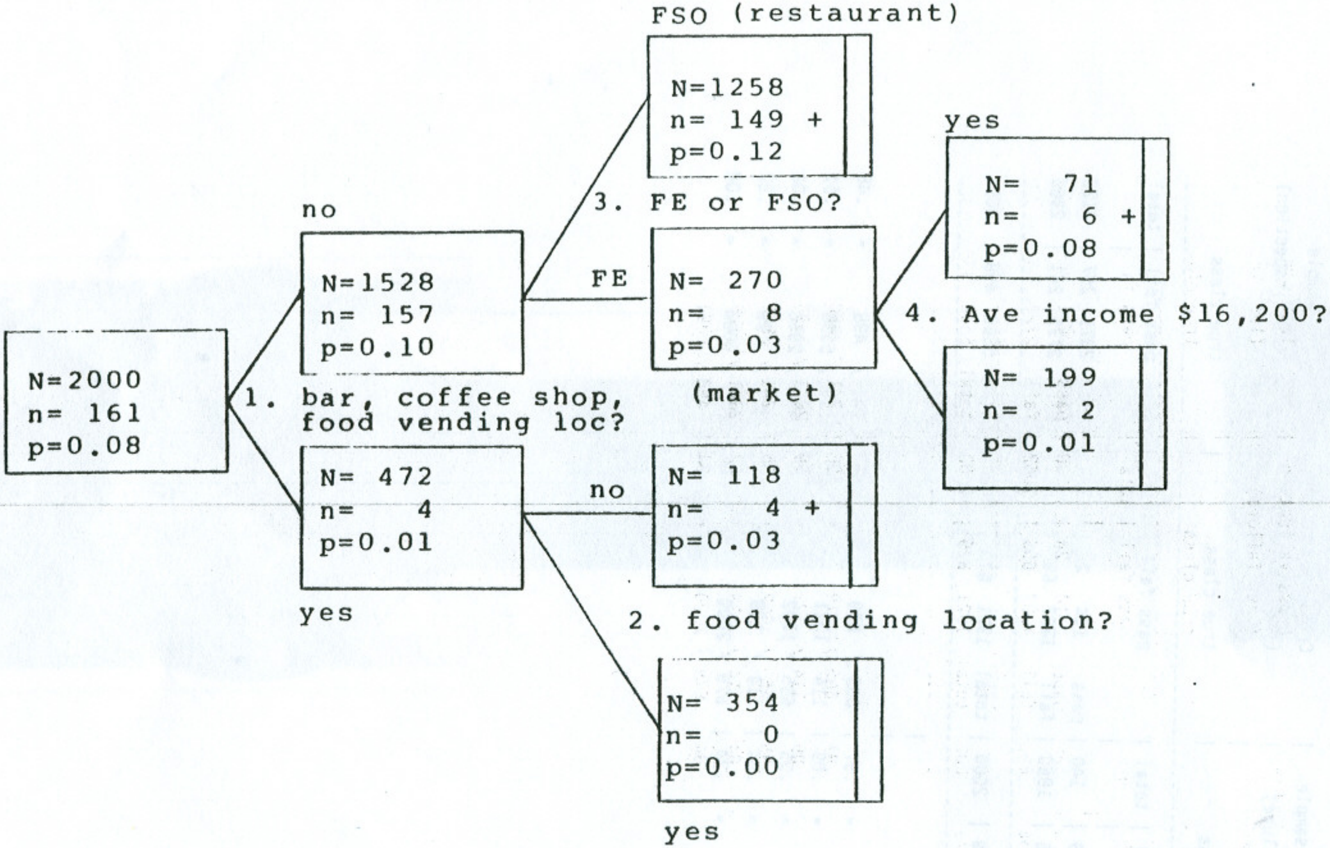

FIGURE 13. Time-temperature violation rates in a learning sample of 2,000 retail food operations as a function of risk markers identified by CART.

Questions 1 through 4 about the risk markers, involving only information available at licensing, classified 161 establishments that had a time-temperature violation at least once, and others that never had one in 5 years, into groups with higher and lower violation rates (p).

TABLE 14. Sensitivity and specificity of the classification of all retail food operations according to risk markers identified by CART to explain or predict time-temperature violations, using only information available at licensing.

	Learning sample (last violation)				Cross-validation (last violation)				Test sample (last violation)				Test sample (last inspection)			
	true class				true class				true class				true class			
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	551	2	553	pass	694	9	703	pass	3423	247	3670	pass	3847	24	3871
	fail	1288	159	1447	fail	1145	152	1297	fail	2180	244	2424	fail	2190	33	2223
	total	1839	161	2000	total	1839	161	2000	total	5603	491	6094	total	6037	57	6094

	Learning sample (last violation)	Cross-validation (last violation)	Test sample (last violation)	Test sample (last inspection)
sensitivity	159 / 161 = .99	152 / 161 = .94	244 / 491 = .50	33 / 57 = .58
specificity	551 / 1839 = .30	694 / 1839 = .38	3423 / 5603 = .61	3847 / 6037 = .64
PV +	159 / 1447 = .11	152 / 1297 = .12	244 / 2424 = .10	33 / 2223 = .01
PV -	551 / 553 = 1.0	694 / 703 = .99	3423 / 3670 = .93	3847 / 3871 = .99
p (violate)	161 / 2000 = .08	161 / 2000 = .08	491 / 6094 = .08	57 / 6094 = .01

CHAPTER VIII

Discussion and Conclusions

This chapter begins with a discussion of the limitations of the data and analysis. In spite of these limitations, some conclusions seem to emerge. The effects of most variables seemed in overall agreement with previously published findings. The Health Department could use failure predictions as an aid in targeting certain operations for HACCP evaluations or extra inspections. The predictive power of the algorithms could perhaps be improved by changing the way the Columbus Health Department maintains food complaint and enforcement records and by starting an active foodborne illness surveillance system. After a few other recommendations there is a summary.

Limitations of the data and analysis

The big problem with this study was that even the best prediction trees did not perform very well, especially when used to predict the results of the last inspection. As Table 5 shows, the predictive value of a positive test for the last failure, the proportion of true failures among those predicted to fail, was only 55% for a test sample of

TABLE 14. Sensitivity and specificity of the classification of all retail food operations according to risk markers identified by CART to explain or predict time-temperature violations, using only information available at licensing.

		Learning sample (last violation)				Cross-validation (last violation)				Test sample (last violation)				Test sample (last inspection)		
		true class				true class				true class				true class		
		pass	fail	total		pass	fail	total		pass	fail	total		pass	fail	total
predicted class	pass	551	2	553	pass	694	9	703	pass	3423	247	3670	pass	3847	24	3871
	fail	1288	159	1447	fail	1145	152	1297	fail	2180	244	2424	fail	2190	33	2223
	total	1839	161	2000	total	1839	161	2000	total	5603	491	6094	total	6037	57	6094

	Learning sample (last violation)	Cross-validation (last violation)	Test sample (last violation)	Test sample (last inspection)
sensitivity	159 / 161 = .99	152 / 161 = .94	244 / 491 = .50	33 / 57 = .58
specificity	551 / 1839 = .30	694 / 1839 = .38	3423 / 5603 = .61	3847 / 6037 = .64
PV +	159 / 1447 = .11	152 / 1297 = .12	244 / 2424 = .10	33 / 2223 = .01
PV -	551 / 553 = 1.0	694 / 703 = .99	3423 / 3670 = .93	3847 / 3871 = .99
p (violate)	161 / 2000 = .08	161 / 2000 = .08	491 / 6094 = .08	57 / 6094 = .01

full-menu restaurants; and this was for Figure 4, the most predictive algorithm. The overall proportion of failures was 49%--so CART did little better than chance alone would have. The positive predictive value when applied to the last inspection was no better, only 11% when the prevalence of failure was 9%.

Pruning trees by removing less effective questions can improve their performance. For example, the positive predictive value of the tree to predict failure in fast-food restaurants, markets and carry-outs (Figure 5) can be increased, at the expense of missing some failures, by cutting off the rest of the tree after the split on score SD:

$$\text{sensitivity} = \frac{213}{416} = 0.51 \qquad \text{(Eq. 6)}$$

$$\text{PV +} = \frac{213}{310} = 0.69 \qquad \text{(Eq. 7)}$$

The corresponding statistics in Table 6 were 96% and 54%. The performance of the tree shown in Figure 4 for full-service restaurants can also be changed by using only the split on score SD:

$$\text{sensitivity} = \frac{398}{528} = 0.75 \qquad \text{(Eq. 8)}$$

$$PV+ = \frac{398}{573} = 0.69 \qquad \text{(Eq. 9)}$$

The original sensitivity and predictive value were 98% and 63%. The predictive value of the tree in Figure 6 can be doubled in this fashion. However, this technique would not help some of the other trees.

The Columbus Health Department's classification of food operations into inspection interval groups based on potential risk (see Appendix G) was prophetic. Forty-nine percent of full-menu restaurants failed at least one inspection in 5 years; 38% of fast-food outlets and carry-outs, and 21% of bars and coffee shops, failed at least once. Nine percent of full-menu restaurants, 7% of fast-food outlets and carry-outs, and 9% of bars and coffee shops failed their last inspection. Corresponding time-temperature violation rates over five years were 18%, 7%, and 1%; and time-temperature violation rates in the last inspection were 2%, 1%, and 0%. These failure rates are consistent with the Health Department's prior expectations and inspection scheduling.

Because CART was applied only after operations were categorized into inspection interval groups, CART had to improve on an already fairly effective system. No highly correlated variables were available because the best predictor of the outcome variables--inspection interval--

was already taken.

A potentially important coding error was mentioned in Chapter VI (page 49). Better-formulated "frequency of violations" variables might have been more useful. Also, there was a potentially important omission: the study could have examined the effects, if any, of existing complaint records as predictor variables. The study looked at the effect of the purpose of the index inspection (see below and page 51), but omitted testing of a variable reflecting the number of previous complaints. Such a variable would have been similar to "extra," the number of extra inspections in the previous year. The annual numbers of food complaints from 1985 through 1989 were 645, 725, 704, 672, and 686. An average of about 12% of these allege illness. An average of 80% of the inspections scheduled in response to illness complaints find "no cause for action."[46]

Another problem in this study was that the records were not maintained for the purposes of the study. A more accurate measure of the income levels of neighborhoods, for example, would have been possible otherwise.

There is no guarantee that modifying risk markers will lead to decreased incidence of inspection failure or time-

---

46 Hartman J. 1985-1989 foodborne illness investigations. [Unpublished report to the Chief of District Operations.]

temperature violations.[2] For example, doing an extra inspection in the 159 full-menu operations with consistent scores (Figure 4) would not necessarily have changed failure rates in this subgroup to 16%. The term "risk marker" was suggested to avoid the implication in the more common term "risk factor" that an intervention is possible.

With this caveat in place, the next section will briefly review the important markers CART identified and examine some markers implicated in previous studies but not confirmed here.

Conclusions about specific variables

The findings in this study about specific variables seem to be in general agreement with the literature on the subject.

A shorter average interval between inspections was associated with higher failure rates in certain categories of full-menu restaurants (Fig. 4) and fast-food establishments and markets (Fig. 5); and with lower failure rates in bars and coffee shops (Fig. 6). A shorter average interval was also associated with higher time-temperature violation rates in certain categories of full-menu restaurants (Fig. 7) and fast-food stores and markets (Fig.9). A shorter <u>actual</u> interval between the last inspection and the index inspection was associated with higher failure rates in subcategories of fast food establishments and markets (Figure 5) and bars and coffee shops (Fig. 6). Much of this is inexplicable. Briley and Klaus[25] found that shortening the interval led to higher scores.

Briley and Klaus, Wodi and Mill,[26] and Moore et al.[28] used average scores to predict risk. The CDC[14] and Irwin et al.[15] found that low scores are associated with increased risk of causing outbreaks. The present study found that the standard deviation of scores was more informative than the average score, but that the higher rates of failure or time-temperature violations were, indeed, associated with the more variable scores.

The Ohio Department of Health food service inspection form has fewer categories than the one used in Columbus. ODH should consider use of a form more like the one used in Columbus, because score variability (as measured by the longer form) was an important risk marker for sanitation problems.

The role of the variable "extra" here seems to indicate that the optimum number of inspections is higher than the minimum requirements of the Columbus Health Department (see below).

This study identified a previous time-temperature violation in a full-menu restaurant (Figure 7) as a risk factor for another one. Irwin et al.[15] also found temperature violations to be associated with restaurants that cause outbreaks. Wodi and Mill[26] used critical items violated in the last two inspections (but not necessarily this particular violation) as a component of their measure of risk. However, Irwin et al. also found "any improper food protection practice," and also "food equipment violations" to be predictive, but this study did not.

Irwin et al.[15] found the average duration of inspections to be predictive of the risk of subsequent outbreaks. Here this effect showed up as a risk marker for time-temperature violations in full-menu restaurants (Figure 8). In contrast, an analysis (not included in this report) of "case-by-case" CART output from the failure tree for full-menu operations (Figure 4) showed that the average duration of inspections was about an hour regardless of predicted or actual outcome.

Income or socioeconomic status is frequently implicated as a risk marker for disease, but apparently no previous reports have mentioned it in connection with food service code violations.

Kaplan and El-Ahraf[12] indicated that fast food operations and restaurants were more likely to cause outbreaks than were markets and liquor stores. This study found that restaurants had a higher risk of time-temperature violations among fast-food establishments and carry-outs (Figure 9) and in general (Figure 13).

McSwain[16] indicated that food vending machines are safe, and this study seemed to agree (Figures 12 and 13).

Irwin et al.[15] reported corporate ownership to be significantly associated with restaurants causing outbreaks; here, commercial status (which could mean a sole proprietorship or partnership as well as corporate ownership) was unimportant. They also reported size and ethnicity as risk markers for outbreaks; this study found neither to be associated with either outcome variable. In this study 3% of all establishments were ethnic; in theirs, 68% of restaurants causing outbreaks were ethnic. That could reflect a reporting bias if people were more likely to suspect a foodborne etiology when illness followed a meal at an ethnic restaurant. (These outbreaks were reported by the public, rather than uncovered by Seattle's active foodborne illness surveillance system, because "the pathogen was unknown for most outbreaks.")

Potential application of predictive models

There are at least two important uses for these models in spite of their weak predictive power. In some instances one could schedule a hazard analysis (HACCP) tailored to the likely causes of the next failure or time-temperature violation. In other instances an appropriate intervention would be to schedule additional inspections for operations predicted to fail. The original classification into inspection interval groups was for these same purposes, to aid in targeting food operations for HACCP evaluations and to provide more inspections where needed. The predictive models would merely provide additional prioritizing.

HACCP. Referring to Figure 2, one idea would be to perform HACCP evaluations in the 573 operations with high score variability. Figure 5 indicates that among operations with a variable score, a previous time-temperature violation is the best predictor of that particular violation. Perhaps other specific violations can be anticipated using the records of operations predicted to fail.

All critical violations would appear to be critical control points amenable to HACCP evaluations and monitoring. Suppose an operator has difficulty maintaining a dish machine's final rinse temperature, for example (perhaps due to an inadequate booster heater), and cannot afford to retrofit the machine with a chemical sanitizer injector. The health department could require maintenance of a temperature log for monitoring purposes. The sanitarian and manager would agree on procedures to implement if the temperature dips below the required level.

Disposable utensils or hand washing with a chemical sanitizer would be options. Planning for contingencies could avoid future violations.

The Environmental Health Division began instituting a HACCP program in 1987 in one work group, pronounced the experiment a success after a few years, purchased a pH meter and other equipment for HACCP evaluations, and has not emphasized it since. HACCP should receive continued emphasis, and all food sanitarians should use it.

IAMFES also recommends the passage of laws formally requiring HACCP evaluations.[31] Ohio has apparently taken the first steps toward agreement! The February 1992 revision of the Food Service Operation Law and Rules contains HACCP concepts in a new rule on heat treatment dispensing freezers (OAC 3701-21-071), complete with time-temperature record-keeping requirements. Such a freezer is

> .. a self-contained dispensing freezer with a product reservoir that processes previously pasteurized products, freezes the products, dispenses frozen dairy products, and maintains microbiological quality by elevating the temperature of the product using heating methods that are an integral part of the dispensing freezer.

The new rule requires that these freezers "shall be equipped with a critical control monitoring device" that maintains time-temperature logs.

The Columbus Board of Health (And the Ohio Department of Health) should enact rules authorizing the general use of HACCP and requiring the maintenance of needed records.

More inspections. The other strategy to head off problems would be to do more inspections. The 159 establishments in Figure 2 that did not receive an extra inspection in the year before the index inspection would get at least one, or perhaps 2 or 3, extra inspections.

Seattle-King County, Washington, found 4 inspections annually to be better than 1 at reducing foodborne illness complaints and increasing scores.[22] Corber[23] found that reducing the number from 12 to 7.3 made no difference. The implication of these reports and the present findings is that 3 inspections annually for full-menu restaurants may be inadequate.

More inspections would probably require hiring additional personnel. A possible alternative might be to cut back on inspections elsewhere, but this generally appears counterproductive. However, the Columbus Health Department already has fewer locations assigned to each sanitarian (or, rather, "full-time equivalent") than the maximum recommended by the Ohio Department of Health.[43,49] ODH recommends 380; the average in Columbus is 350. The Ohio Department of Health should consider decreasing its recommended number of locations in light of these results.

Recommendations to improve predictive power

Computerized complaint and enforcement logs. The District Operations Complaint System is computerized. (See Appendix L). It tabulates complaints, including sanitarian-initiated ones as well as those from the general public, about messy garbage storage, rats, weeds, etc., at apartment buildings, residences, restaurants, vacant lots, etc.--about every conceivable structure (sometimes including restaurants and markets). All complaints have a disposition entered. Records indicate when a sanitarian issues an order, when a consultation takes place, when an enforcement letter is sent, the date of a referred to the Night Prosecutor program, the date of an administrative or Board of Health hearing, when a trial date is set, and what the verdict, fine, or sentence are. The date of correction is shown.

No such detail exists for food program complaints. The SPIF system does show, indirectly, when an order is issued in response to a complaint and when the violation is corrected, but not whether a hearing is held. The nature of the complaint is recorded on paper logs (Appendix M). The electronic record in SPIF does not show the nature of the complaint unless the sanitarian cites a violation. If certain kinds of complaints are associated with violations, this information could help CART make predictions.

If enforcement works, it could be an "unseen hand" in the food program data. However, no enforcement actions are entered into SPIF.

These shortcomings would appear to be easy to rectify. The food complaint system and enforcement actions should be tallied by computer, preferably in the SPIF system. Future CART runs could test the predictive power of different kinds of complaints and the effectiveness of various steps in "progressive enforcement" in preventing further violations.

An active foodborne illness surveillance system. The IAMFES Committee on Communicable Diseases Affecting Man recommends using actual foodborne illness complaints wherever possible to guide in the selection of establishments to receive HACCP evaluations.[31]

The Seattle-King County Department of Health investigates isolations of enteric agents by hospital and other medical laboratories as potential foodborne illness complaints against any restaurants patients may have visited at the beginning of the likely incubation period of the illness.[47] In 1987 Seattle received 207 reports of *campylobacter*, 157 of *giardia*, 264 of *salmonella*, and 89 of

47 Grendon J. Report: Seattle-King County foodborne illness surveillance and outbreaks, 1986. Seattle, WA: Seattle-King County Department of Public Health, 1986.

shigella. Counting these 717 reports and 69 reports of other agents, their enteric illness reports totalled 786. Follow-up of these reports uncovered 30 confirmed or suspected outbreaks of foodborne illness. Nearly 60% included illness in one person.

The Ohio Department of Health received 608 reports of isolates of enteric disease germs from patients in Franklin County in 1991: 148 campylobacter, 186 giardia, 83 hepatitis A, 167 salmonella, and 24 shigella.

In addition to investigation of enteric isolates, other means are available to improve surveillance. Training the public (publicizing outbreaks, listing a phone number for foodborne illness reports), training physicians and emergency room personnel, and more thoroughly training sanitarians would be additional steps to take. The IAMFES Committee on Communicable Diseases Affecting Man has listed procedures for establishing foodborne illness surveillance systems.[48]

The Columbus Health Department should start an active foodborne illness surveillance system whether or not the use of CART and predictive models is anticipated, and perform HACCP evaluations wherever indicated.

48 IAMFES. Procedures to investigate foodborne illness, 4th ed. Ames, Iowa: IAMFES, 1988.

Other recommendations

This study has assumed that the effects of risk markers have been operating without biases introduced by differences in inspection techniques among sanitarians and supervisors. However, consistency in sanitarian performance requires verification. The Ohio Department of Health has staff available to standardize the inspection techniques of food program supervisors, who could then standardize their sanitarians' performance. Each District has a coordinator to manage the flow of paperwork, so supervisors ought to be free for field evaluations.

Predictions would probably have been better if more predictor inspections had been available, or if the outcomes had been failure or time-temperature violations over a period of time, rather than in just one index inspection. For example, perhaps using the first 2.5 years of data to predict failures in the second 2.5 years would have been more successful. Using just one outcome inspection may have been plagued by the same instability as using just the previous inspection, rather than 5 years of inspections, would have entailed. Similarly, the effective prevalence of violations could perhaps have been increased by using 7 years of records rather than 5, if more had been available. The Health Department should review and, if necessary, modify its policy on record retention periods.

Now that an in-house computer system is in use, electronic records should never be destroyed.

Recently a survey of convenience stores in Michigan found a correlation between manager knowledge of food sanitation, as revealed in an 8-question test, and the sanitary condition of the scores.[49] The researchers recommended mandatory food sanitation training for convenience store managers. A survey to assess the knowledge of food operation managers in Columbus could be done rather quickly, and would perhaps improve the power of these predictive models if the results were included in SPIF.

If education for food service workers is needed, the Division's Education Unit could schedule seminars. Also, the Columbus Health Department might consider requiring food service manager certification. Alternatively, a "food handler's permit" could be a requirement; to receive the permit, employees would have to pass a written test. This is a requirement in Seattle.

Summary

This project was to find risk markers for failure of food service operations, markets, and similar establishments to pass sanitary inspections. Risk markers

---

49 Burch NL, Sawyer CA. Food handling in convenience stores. Journal of Environmental Health 1991;54:23-27.

for time-temperature violations, that is, factors associated with higher probabilities of citations for mishandling of potentially hazardous (temperature-sensitive) foods, received special attention. Evidence shows this violation, as well as other "critical" violations identified during inspections, to be associated with outbreaks of foodborne illness in the community.

The Classification and Regression Trees (CART) program analyzed computerized inspection records of the Columbus Health Department. CART identified a high standard deviation of inspection scores as the best predictor of inspection failure or time-temperature violations. Among full-menu restaurants, receiving three or fewer regular inspections annually was also associated with these problems.

The predictive models can help identify food operations with higher probabilities of causing outbreaks. Proposed interventions for these operations include more frequent inspections, as well as Hazard Analysis Critical Control Point (HACCP) evaluations.

List of References

1 Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Pacific Grove, CA: Wadsworth, 1984.

2 McCormick J, Skrabanek P. Coronary Heart disease is not preventable by population interventions. The Lancet 1988: October 8; 839-41.

3 Walker B Jr. The future of public health. Journal of Environmental Health 1989; January/February:133-135.

4 Public Health Service. Healthy People 2000: national health promotion and disease prevention objectives. Washington, DC: U. S. Department of Health and Human Services, Public Health Service, 1990; DHHS publication no. (PHS)90-50212.

5 U. S. Department of Health, Education, and Welfare. Food service sanitation manual, including a model food service sanitation ordinance. Washington, D. C.: U. S. Government Printing Office, 1978.

6 Zaki MH, Miller GS, McLaughlin MC, Weinberg SB. A progressive approach to the problem of foodborne infections. American Journal of Public Health 1977;66:44-49.

7 Archer DL, Kvenberg JE. Incidence and cost of foodborne diarrheal disease in the United States. Journal of Food Protection 1985; 48:887-894.

8 Centers for Disease Control. Foodborne disease outbreaks, 5-year summary, 1983-1987. In: CDC Surveillance Summaries, March 1990. MMWR 1990; 39 (No. SS-1):15-57

9 Bernhardt RR. Seattle-King County Department of Public Health food protection program program review. Olympia, WA: Division of Health, Department of Social and Health Services, 1986:6.

10 43% of nursing homes flunk food sanitation. Columbus Dispatch; December 2, 1988: 1A.

11 Bryan FL. Factors that contribute to outbreaks of food- borne disease. Journal of Food Protection 1978;41:816.

12 Kaplan OB, El-Ahraf A. Relative risk ratios of foodborne illness in food service establishments: an aid in deployment of environmental health manpower. Journal of Food Protection 1979;42:446-447.

13 Dannenberg AL, Yashuk JC, Feldman RA. Gastrointestinal illness on passenger cruise ships, 1975-1978. American Journal of Public Health 1982;72:484-8.

14 CDC. Vessel sanitation scores. MMWR 1988;37:114-117.

15 Irwin K et al. Results of routine restaurant inspections can predict outbreaks of foodborne illness: the Seattle-King County experience. American Journal of Public Health 1989;79:586-590.

16 McSwain David. Vending Program (Inservice sponsored by the Columbus Health Department), July 17, 1990.

17 Bryan FL. Hazard analysis of food service operations. Food Technology, February 1981:78-87.

18 Bauman H. HACCP: concept, development, and application. Food Technology, May 1990:156-158.

19 Harrington RE. How to protect your restaurant against foodborne illness. NRA News, April 1986:33-34.

20 Guzewich JJ. Practical procedures for using the hazard analysis critical control point (HACCP) approach in food service establishments by industry and regulatory agencies. In: Food Protection Technology (papers presented at the Third Conference for Food Protection). Chelsea, Michigan: Lewis Publishers, Inc.:91-100.

21 Bryan FL. Teaching HACCP techniques to food processors and regulatory officials. Dairy, Food and Environmental Sanitation 1991;11:562-568.

22 Bader M et al. A study of food service establishment sanitation inspection frequency. American Journal of Public Health 1978;68:408-410.

23 Corber S et al. Evaluation of the effect of

frequency of inspection on the sanitary conditions of eating establishments. Canadian Journal of Public Health 1984; 75:434-438.

24 Bryan FL. Foodborne disease risk assessment of foodservice establishments in a community. Journal of Food Protection 1982;45:93-100.

25 Briley RT, Klaus EF. Using risk assessment as a method of determining inspection frequencies. Dairy and Food Sanitation 1985;5:468-474.

26 Wodi BE, Mill RA. A priority system model for sanitation management in food service establishments. American Journal of Public Health 1985;75:1398-1401.

27 LaBoccetta AC. (letter re:) A priority system model for sanitation management in food service establishments. American Journal of Public Health 1986;76:709-710.

28 Moore GA et al. Food sanitation enforcement. Journal of Environmental Health 1990; 53 (2): 17-18.

29 Anderson PS, Rutenberg GW, Bowen NL. Assessing food quality: the difficulty in establishing microbiological standards. Journal of Environmental Health 1989;52:79-82.

30 Tebutt GM, Southwell JM. Comparative study of visual inspections and microbiological sampling in premises manufacturing and selling high-risk foods. Epidemiology and Infection 1989; 103:403-475-486.

31 IAMFES. Procedures to implement the hazard analysis critical control point system. Ames, Iowa: IAMFES, 1991, p. 34.

32 California Statistical Software, Inc., Lafayette, CA., 1984. [software]

33 Efron B, Tibshirani R. Statistical data analysis in the computer age. Science 1991; 253:390-395.

34 Andersen R, Smedby B, Eklund G. Automatic Interaction detector program for analyzing health survey data. Health services research 1971; summer: 165-183.

35 Lanese RR, Banks FR, Keller MD. Smoking behavior in a teenage population: a multivariate conceptual approach. American Journal of Public Health 1972;6:807-813.

36 Ludwig JM. JMLCOPY. Columbus: Instruction and Research Computer Center, 1983.

37 SAS Institute, Inc. SAS user's guide: basics. Version 5 Edition. Cary, NC: SAS Institute, Inc., 1985.

38 Moore RA. Columbus City Health Department food service program administrative & field sanitation survey. 1990: Ohio Department of Health [unpublished program review].

39 Guerin JP, Keeling H. System documentation. Vol. III of sanitation programs information formulator: user's guide. Washington, D. C.: U. S. Food and Drug Administration; 1975.

40 Hutchinson R, personal communication, May 14, 1990.

41 Donnelley Marketing Information Services. Market Profile Analysis--Columbus, Ohio SMSA. New York, NY: Donnelley Marketing Information Services, 1991.

42 Nolan T, Mallett K. Ethnic foods add spice to Columbus. Columbus alive! 1989: Nov.23-Dec. 7; 8-12.

43 Myers WC. Statement of civil rights compliance for staff [memorandum]. Columbus Health Department 1990; June 29.

44 Johnson RA, Wichern DW. Applied multivariate statistical analysis. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1982, p. 488.

45 Mausner JS, Kramer S. Epidemiology--an introductory text. Philadelphia: W. B. Saunders Co., 1985.

46 Hartman J. 1985-1989 foodborne illness investigations. [Unpublished report to the Chief of District Operations.]

47 Grendon J. Report: Seattle-King County foodborne illness surveillance and outbreaks, 1986. Seattle, WA: Seattle-King County Department of Public Health, 1986.

48 IAMFES. Procedures to investigate foodborne illness, 4th ed. Ames, Iowa: IAMFES, 1988.

49 Burch NL, Sawyer CA. Food handling in convenience stores. Journal of Environmental Health 1991;54:23-27.

Food, Food Protection, ..., Miscellaneous

VARIABLES EVALUATED

Outcome variables

time-temperature violation
inspection failure
    time-temperature
    critical violation
    low score

Predictor variables

(Inspection interval)
Ave. interval between inspections
SD of scores
No. of extra inspections in previous year
previous inspection (no.) days ago
(No.) previous 4A, 4B, ..., 18D
Ave. duration (no.) minutes
Ave. income in zip code
FSO
Vending
Frequency of fail, criticals, 4A, 4B, ..., 18D

Predictor variables available to CART but not used

Food, Food Protection, ..., Miscellaneous
Commercial
Any violation before?
Purpose
Size of FSO
Ethnicity