REVISING CLASSIFICATION STRUCTURE
USING SURVEY DATA

DAVID SKURNICK, N. ROBERT HEYER AND G. RAY FUNKHOUSER

DESCRIPTION OF THE PROBLEM

Revising a classification structure requires both underwriting judgment and an analysis of statistical data. The data can be used to estimate the loss experience of the various proposed classifications and to compare it with the loss experience of the existing classifications. When the merging of two classifications is contemplated, the data on hand can be used to compile experience on both separate and combined bases and to test the effect of the merger. However, when the splitting of established classifications is under consideration, the proposed classifications are often not identified in the data base, so a special effort is required to obtain the necessary data. This may involve obtaining additional information from the applications, imputing classification data from other data on hand or conducting on-site inspections of a number of risks. This paper illustrates a method for revising a classification system in which a mail survey was used to obtain statistical data.

The specific problem addressed was the classification structure of workmen's compensation insurance pertaining to California farms. At the time of the study there were five categories of farms, viz.: Orchards; Truck Farms and Vineyards; Poultry Raising, Egg Production and Hatcheries; Dairy Farms and Sheep Raising; and Field Crops and Stock Farms. These classifications were established some twenty years ago, but since that time California agriculture has changed considerably. The size and organization of the farming units have tended toward larger, more efficient farms. New methods of farming and new kinds of farm machinery have been introduced. The amount of irrigation in the state has increased, and various kinds of farm-related operations, for example, packing sheds, are more frequently included among the agricultural operations on California Farms. These changes in California agriculture, and concomitant discussions among the various constituencies involved in, or touched by, farm workmen's compensation insurance, led the California Inspection Rating Bureau (C.I.R.B.) to explore whether changes in the classification structure were desirable and, if so, what they might be.

The existing classifications were based upon the crops or livestock pro-

duced on the farm and it was anticipated that any revised classifications would have the same basis. It was therefore necessary to collect crop and livestock data for a sample of farms. The data would be used to relate insurance experience (exposure and loss) for each farm to the specific crops or livestock produced on that farm and enable the C.I.R.B. to make assessments of the present classification structure and various alternatives. One of the two possibilities for gathering the crop data, on-site inspections, would have been inefficient because a very large sample of widely dispersed farms was required. A large sample was required because there are tens of thousands of farms in California with none large enough to dominate any of the farm classifications and because the data had to be sufficient for the evaluation of a variety of potential classification systems. It was decided that a mail survey would be more efficient in gathering the needed data, and Field Research Corporation (FRC) was commissioned to conduct the survey.

<div align="center">SURVEY PROCEDURES</div>

Development

The first step in the survey process was to conduct a pilot study to determine the feasibility of using this research technique in this context. Consultations were held with experts representing the insurance industry, the California Department of Food and Agriculture, the Farm Bureau Federation and other researchers who had conducted surveys on farm safety. Next, a questionnaire was designed which it was believed would gather the necessary data from a sample of farms concerning their products as well as data on size, other operations, machinery and equipment, and certain agricultural practices. This questionnaire was sent (with a cover letter and a stamped return envelope) to a sample of about 500 farms.

From the pilot study we found that the use of a mailed questionnaire to collect crop data from farms was feasible, as well as economical. We also found that crop data from the survey and insurance data from the C.I.R.B.'s unit statistical reports could be satisfactorily linked. Moreover, the pilot study brought to light some problems for which adjustments were made in the main survey that substantially increased the volume and the quality of the data from the farms.

For example, it was found in the pilot study that large farms were inherently harder to reach than small farms were, due to more complex management structures. Corporation ownership, decentralized management and

multiple crop activities typical of large farms in California all combined to reduce the effective response rate from large farms. However, the larger farms have a larger impact on insurance loss, and it was imperative that data be obtained from as many of these as possible. As a result of the pilot study, some changes were made in the questionnaire, a sampling plan was designed whereby larger farms were sampled at a higher rate than were smaller farms, and strategies were devised to bring about a higher rate of returns among large farms.

Final Questionnaire

The questionnaire used as the data-gathering instrument for the full-scale survey was a shortened version of that used in the pilot study. Recipients of the questionnaire, which was addressed to the principal individual named in the C.I.R.B. policy master file, were asked to indicate which crops or other products they had raised in the years 1970 through 1973. The format was such that the entire questionnaire was printed on both sides of a single sheet of 8-1/2" X 14" paper. Included with the questionnaire was a cover letter that explained the survey and mentioned that it was endorsed by the Farm Bureau Federation and the California Department of Food and Agriculture.

Sampling

The universe sampled for this survey was the list of farms covered by workmen's compensation insurance and in the C.I.R.B. files. These files are accessible by computer, and a program was written to draw this sample.

The pilot study had indicated that different types and sizes of farms would have different rates of response. Also, some of the present classification codes covered more crops and products than others. Therefore, a stratified sampling plan was devised by which certain types and sizes of farms were sampled at greater rates from the C.I.R.B.'s list than others. Farms with annual exposures of $5000 or less were selected by ratios ranging from 1 out of every 50 to 1 out of every 16, and in higher categories of exposure smaller ratios were used. All farms with annual exposures of $75,000 or greater were selected for inclusion in this sample.

This differential sampling was used so that the farms with the greatest impact on insurance experience would be more likely to be included in the survey data. This plan was adopted to guarantee obtaining enough loss data to provide high credibility for critical classification categories; but because of

the disproportionate sampling plan it did not yield directly a true cross-section of California farms. Because of the random selection process within each stratum, a "true cross-section" (in the sense of representing farms of all sizes in direct proportion to their actual frequencies) could be statistically constructed. Since this kind of analysis was not essential for comparing the relative hazard among classifications, it was not performed.

To provide for invalid addresses and duplicate entries due to multiple classification of farms, all strata sampling intervals were increased by 25% so that the final sample size would still be faithful to the projected sampling needs. A total of 5089 different farms comprised the final, mailed sample. Table I shows the final sample draw, broken down by existing farm classification and by amount of exposure in the primary classification. The sampling ratios (and the number of farms drawn) are given for each of the 20 different strata shown in the table.

Table I

Sampling Ratios (and Number of Farms Drawn)

Prior Farm Class	(Total Farms)	Exposure Category			
		$1 to $5,000	$5,000 to $20,000	$20,000 to $75,000	Over $75,000
Orchards	(1355)	1/50 (278)	1/17 (254)	1/3 (534)	1/1 (289)
Truck Farms & Vineyards	(2003)	1/22 (399)	1/12 (419)	1/5 (685)	1/1 (500)*
Poultry	(259)	1/16 (43)	1/8 (63)	1/2 (83)	1/1 (70)
Dairy and Sheep	(554)	1/2 (85)	1/13 (132)	1/4 (249)	1/1 (88)
Field Crops and Stock Farms	(918)	1/25 (325)	1/13 (224)	1/3 (275)	1/1 (94)
Total Sample	(5089)	(1130)	(1092)	(1826)	(1041)

*A total of 936 farms were supplied for this category, but 500 were judged to be an adequate number for the survey.

Data Acquisition

The questionnaires were mailed to the 5089 farms in the sample. Prior to mailing, a serial number for each farm in the sample was entered on the questionnaire sent to that farm, and also on that farm's unit report. This serialization enabled us to relate the returned questionnaires to the farms in the sample, an essential operation because of the need to link the questionnaire data to the insurance experience provided for the farms in the sample.

A second wave mailing was sent two weeks after the first wave to all farms that had not yet responded. Shortly after the second wave mailing went out, a third stage effort was begun by telephone. Calls were made to all farms in certain categories (indicated in Table II) from which it was deemed especially important to secure a high rate of return. In the telephone contact interview, crop data were obtained from the farms, but, to make the interview as brief as possible, other peripheral items on the questionnaire were not asked. Because the crop data were the main objective of the survey, this procedure did not appreciably impair the data base.

All three waves of data gathering were conducted during a period of 25 days. The returns by classification and exposure are shown in Table II. Over-

Table II

Number of Completed Questionnaires (and return rates)

	Exposure Category			
Prior Farm Class	**$1 to $5,000**	**$5,000 to $20,000**	**$20,000 to $75,000**	**Over $75,000**
Orchard	141	131	278	122
	(51%)	(51%)	(52%)	(43%)
Truck Farms & Vineyards	203	197	328	219
	(47%)	(47%)	(48%)	(44%)
Poultry	27	53	62	61
	(63%)	(79%)	(75%)	(84%)
Dairy & Sheep	43	83	155	56
	(50%)	(63%)	(62%)	(64%)
Field Crop and Stock	148	146	177	70
	(45%)	(62%)	(63%)	(76%)

*Farms inside the dashed line received telephone follow-up calls.

all, questionnaires or telephone responses were obtained from 2476 farms during the period. This amounts to an effective return rate of 50.2% overall, after subtracting from the original sample 156 farms whose addresses proved to be invalid.

Data Linkage

The C.I.R.B. supplied FRC with the necessary insurance data for all of the farms drawn in the survey sample. These insurance data were compiled at the time the sample was drawn from the C.I.R.B.'s files, and were punched on cards in a format that enabled FRC to enter the data from the questionnaires directly onto the cards.

Two kinds of data cards were supplied by the C.I.R.B. First, one "farm card" was supplied for every farm in the sample containing the Bureau Number of the farm, its total exposure, its loss and subject premium for policy years 1970 and 1971 (the two most recent years for which complete insurance data were available), and its pure premium for those two years. The farm cards were used to compare the insurance characteristics of the responding farms and non-responding farms to test for response bias.

The second type of data card consisted of a variable number of "crop cards" for each farm, depending upon the number of class codes under which the farm was covered and the number of years the farm was so covered. A separate "crop card" for each class and for each year of coverage (1970 and/or 1971) was prepared for each farm. Each "crop card" was identified by class code, year, and Bureau Number, and contained the exposure, loss, pure premium and subject premium for that crop and that year.

Questionnaire Data

Crop data from the returned questionnaires were entered onto the "crop cards" by determining which crops reported on the questionnaire accounted for the coverage for each class code and year. The range of data available for coding was limited in some cases by the fact that certain individual crops reported on the questionnaires could not be separated when two or more of them fell within one category in the prior classification system. In such cases, where several crops could not be separated on a given "crop card", a combination crop code was used to designate what group of crops was represented. Thus, for example, a farm growing both oranges and lemons in one year was

coded as "combination of orchard crops", since both of these crops occur within the same prior Bureau classifications.

<center>DATA ANALYSIS</center>

Validity of the Survey

The insurance data and farm data were combined and statistically analyzed by computer. Table III shows a comparison between the insurance data representing the farms that responded in the survey and the farms that did not respond.

<center>Table III</center>

Comparison of Insurance Data between Responding and Non-Responding Farms

	Responding Farms	Non-Responding Farms
1. Number of Reports (crop cards)	6,475	6,993
2. Total Exposure	$368,694,656.	$466,475,904.
3. Average Exposure (per year) (2) ÷ (1)	$56,941.	$66,706.
4. Total Loss	$6,930,513.	$9,049,346.
5. Average Loss (per year) (4) ÷ (1)	$1,070.	$1,294.
6. Aggregate Pure Premium 100 × (4) ÷ (2)	1.880	1.940

This comparison shows that there is a slightly higher pure premium among the non-responding farms than among the responding farms. This may be a result of differential return rates among the different sizes and classes of farms, and it may have resulted from a slight tendency of farms with poorer insurance experience not to respond to the questionnaire. As the difference between the pure premiums of the two groups is in the range of 3%, it is doubtful that the discrepancy, no matter what its cause, would have any noticeable effect on the validity of the results of the survey. The conclusion was, therefore, that there is no serious bias in the insurance characteristics of the sample of responding farms, beyond the inherent bias that results from

the disproportionate stratified sampling plan discussed previously. On this basis, it was assumed that the questionnaire data on crops produced, linked with insurance experience on a farm-by-farm basis, could be interpreted as validly representing the originally drawn sample of California farms.

As a further check on the survey's validity, the prior classification system was also analyzed and compared to the actual, summary statistics in the C.I.R.B. records. This comparison is shown in Table IV for 1970 (the most recent complete data available at the time of the study).

Comparing the sample data with the summaries of the overall universes they represent, two things stand out:

1.   The pure premiums for the samples from each category are reasonably close to the actual universe values.

2.   A sample of about 7% of the reports, deliberately drawn to over-represent larger farms, accounted for almost one-quarter of the total insurance loss.

Individual Crop Experience

A computer program summarized the total and average exposure and loss and the aggregate pure premium for each of the 112 crop categories measured in the study. The largest individual crop categories were "Dairy Cows and Calves" with 516 reports and $700,044 of incurred loss and "Chickens and/or Eggs" with 225 reports and $551,388 of incurred loss. However, with only $6,238,343 of incurred loss spread among 112 crops, most crops had very limited experience. Also, it was not possible to allocate exposure and loss for farms with two or more crops in a single prior classification. Experience of such farms had to be shown in appropriate combination codes.

Credibility considerations required that the design of new classifications be accomplished by the process of grouping the crops into various proposed classification schemes and comparing them on the basis of their insurance experience. Dealing with larger classifications also reduced the data lost on account of crop combinations, because, if a farm had two or more crops in one prior classification, but these crops were also in the same proposed classification, then its experience could be used.

Table IV

**Comparison of Loss and Pure Premiums
Between Survey Data and the C.I.R.B. Statistical Reports (1970)**

Class	Survey Results			C.I.R.B. Records		
	No. of Reports	Pure Premium	Total Loss ($)	No. of Reports	Pure Premium	Total Loss ($)
Orchards	849	2.62	970,211	16,958	2.26	4,118,045
Truck Farms and Vineyards	1,091	1.41	1,067,648	13,930	1.26	5,304,297
Poultry	191	1.30	286,631	1,294	1.74	647,530
Dairy and Sheep	334	2.62	374,898	3,867	2.64	1,507,456
Stock and Field Crops	710	3.21	727,802	9,991	3.80	3,501,322
Total	3,175	...	3,427,190	46,040	...	15,078,650

Selection of the Revised Classifications

A program was written to analyze the insurance characteristics of alternate classification systems proposed by the C.I.R.B. and others. Each proposed system of categorizing crops was entered into the computer, and an analysis was made of exposure, loss, and pure premium for each category in the system. The credibility of each of the proposed categories was calculated according to the formula:

Credibility $= (\text{Total Loss}/1,197,880)^{2/3}$, to a maximum of 1.0.

The credibility formula approximates the usual three part criterion used in workmen's compensation ratemaking. The amount of loss for full credibility was based upon 50 serious claims, 300 non-serious claims, and 80% of the non-serious amount for medical, but these three amounts were added together and compared to the total incurred loss, rather than applied separately to the three types of expected loss.

Since the insurance data now were linked with individual crops (or with specific combinations of crops), it was possible to form virtually any new set of classification categories desired. Because of the nature of agriculture in California, however, only a limited number of classification systems were considered reasonable.

The first step was to divide the crops into seventeen basic study groups and analyze their sample pure premiums. It was judged that the classifications ultimately chosen should be combinations of these seventeen groups. Four other classification schemes, ranging from four to eight categories, were then analyzed. One of these, Classification System 5, is summarized in Exhibit 1. Exhibit 2 shows how Classification System 5 was formed from the seventeen basic groups.

The commodity experts of the California Farm Bureau Federation had suggested a configuration which the C.I.R.B. modified in two respects to form Classification System 5. The Farm Bureau specialists had included potatoes and sugar beets with field crops, but the data indicated that the pure premium for potatoes and sugar beets was significantly lower than the pure premium for field crops; consequently, the C.I.R.B. decided to establish a separate classification for potatoes and sugar beets until a more comprehensive study could be made. Also, the Farm Bureau specialists had separated Beef Cattle (Farms) from Horses and Feed Yards, but the pure premiums were nearly the same; consequently, the C.I.R.B. decided to establish two new classification wordings placed under a single code number.

Classification System 5 would have been chosen to replace the prior system, were it not for the problem presented by cotton risks. Since they had been included in the Truck Farm classification enjoying a relatively low Manual rate, moving them in with the Field Crops would have resulted in a much higher rate for them. Furthermore, the data compiled in the seventeen study groups indicated that the pure premium for the Cotton group was slightly lower than the pure premium for Field Crops. It was therefore decided to establish a separate classification for Cotton until a more comprehensive study could be performed.

The final classification configuration as filed by the C.I.R.B. is shown in Table V.

Table V
Revised Farm Classification System

Prior Code	Crop	Proposed Code
0016	Orchards	0016
0017	Vegetables	0172
0017	Potatoes	0041
0017	Grapes	0040
0017	Cotton	0044
0034	Poultry Raising	0034
0036	Dairy Farms	0036
0037	Stock Farms	0038
0037	Field Crops	0171

The premium rates and expected loss rates for the proposed codes were calculated from the pure premiums selected. Classifications 0016, 0034, and 0036 were essentially unchanged, so their pure premiums were based on past experience. Pure premiums for classes 0038 and 0171 were selected to be proportional to their sample pure premiums, so balanced that the premium generated by the two new classes would equal the premium that would have been generated at the rate indicated by the past experience of class 0037. Similarly, the pure premiums for Codes 0040, 0041, 0044, and 0172 were based on their sample pure premiums and balanced to the past experience of Code 0017. Selected pure premiums for the new classes were limited to a 25% increase over the pure premiums indicated by past experience, and a balancing adjustment was made so that the total premium income for the farm classes would be unaffected by the change in classification structure.

Exhibit 1

Classification System 5

Category		Number of Reports	Total Exposure	Total Loss	Aggregate Pure Premium	Percent Credibility
Orchards &	Total	1739	78542120.	1657874.	2.111	100%
Nuts	1970	845	36584144.	960895.	2.627	86%
0016	1971	892	41957976.	696979.	1.661	70%
Vegetables	Total	371	36988796.	413803.	1.119	49%
0017	1970	181	17234596.	207809.	1.206	31%
	1971	190	19754200.	205994.	1.043	31%
Grapes	Total	636	19019076.	224278.	1.179	33%
XXXX	1970	311	7834982.	44939.	0.574	11%
	1971	325	11184094.	179339.	1.604	28%
Cotton &	Total	842	19430658.	559748.	2.881	60%
Field Crops	1970	413	8940592.	282692.	3.162	38%
YYYY	1971	429	10490066.	277056.	2.641	38%
Potatoes &	Total	78	6390912.	63267.	0.990	14%
Sugar Beets	1970	41	2977909.	32239.	1.083	9%
ZZZZ	1971	37	3413003.	31028.	0.909	9%
Poultry &	Total	439	46806156.	686618.	1.467	69%
Sheep	1970	218	22547196.	288928.	1.281	39%
0034	1971	221	24258960.	397690.	1.639	48%
Dairy	Total	516	25447544.	700044.	2.751	70%
0036	1970	252	11933190.	345814.	2.898	44%
	1971	264	13514354.	354230.	2.621	44%
Stock Farms	Total	268	13419345.	451933.	3.368	52%
& Feed Lots	1970	129	6241234.	180790.	2.897	28%
0037	1971	139	7178111.	271143.	3.777	37%
Overall Summary						
	Total	4887	246044607.	4757565.	1.934	100%
	1970	2390	114293843.	2344106.	2.051	100%
	1971	2497	131750764.	2413459.	1.832	100%
Reports Not Classified Under This System						
	Total	1602	122825072.	2173041.	1.769	100%
	1970	785	57509992.	1083084.	1.883	94%
	1971	817	65315080.	1089957.	1.669	94%

Exhibit 2

Formation of Classification System 5
From the Seventeen Study Groups[*]

Seventeen Study Groups

	Pure Premium	Percent Credibility	Prior Class Code	Classification System 5 Code
Beef Cattle Farms	3.554	23%	0037	0037
Feed Yards	3.425	27%	0037	0037
Horses (Hogs)	3.362	15%	0037	0037
Field Crops	2.997	55%	0037	YYYY
Dairies	2.750	70%	0036	0036
Orchards	2.341	94%	0016	0016
Cotton	2.220	14%	0017	YYYY
Poultry	1.470	65%	0034	0034
Sheep (Goats)	1.223	5%	0036	0034
Grapes	1.202	33%	0017	XXXX
Vegetables	1.174	46%	0017	0017
Potatoes & Sugar Beets	0.989	14%	0017	ZZZZ
Nuts	0.943	11%	0016	0016
Melons & Strawberries	0.819	10%	0017	0017
Dry Beans & Peas	0.244	1%	0017	0017
Bush Berries	0.128		0017	XXXX
Hops	0.116		0017	XXXX

116