used for county, regional, or statewide land-use capability analysis.

The ability to geographically relate linear systems to area conditions can provide data for both research and applied planning of many functions that concern federal, state, and local governments.

ACKNOWLEDGMENTS

The accident reporting system described in this report was prepared for the Indiana Department of Traffic Safety and Vehicle Inspection, in cooperation with the Indiana State Highway Commission, the Indiana State Police, the Indiana Department of Administration, and the U.S. Department of Transportation. We wish to acknowledge the contributions of the many staff members of Vogt, Sage, and Pflum Consultants for their collective thinking that has resulted in the detailed design and programming of INTRACS and their persistence and diligence in coding, editing, and processing the data files that are the basis of INTRACS.

The INTRACS working committee, comprised of state personnel representing the four agencies that will maintain and use the system, has offered constructive reviews, comments, and suggestions that have contributed to a better system.

Publication of this paper sponsored by Committee on Traffic Records.

Sampling Procedure Using Multistate Traffic Records to Select Accident and Exposure Data-Collection Sites

Chang S. Yoo and Martin L. Reiss, BioTechnology, Inc., Falls Church, Virginia

This paper describes a sampling plan developed to select 80 field data-collection sites. At each of these locations, large-truck accident rates were to be measured and vehicle exposure was to be sampled simultaneously for 1 year. The problems to be addressed were (a) to stratify the sites in each state so that the accident experience developed would be representative of the state (to preclude selecting only high-accident locations) and (b) to devise a roadway typology whereby sites were consistent across states selected. Accidents are now being investigated at sites selected by this procedure in California, Maryland, Michigan, Nevada, Pennsylvania, and Texas.

This paper describes a sampling plan developed to obtain the accident rates (based on exposure) of large trucks. Truck exposure is the term used to describe the number of trucks currently in use and the annual number of kilometers these vehicles are driven. Fairly accurate information is available at the state and national levels as to the number of trucks registered through state licensing agencies each year. But, as one attempts to classify trucks into categories on the basis of such considerations as body type, number of axles, mass, and length, the available information becomes vague, especially in terms of truck combinations (where different trailer units can be combined with a particular tractor for different trips).

A great deal of information is available about truck accidents in general. The National Safety Council each year provides a broad estimate of trucks involved in motor-vehicle accidents in the United States and the distribution between single-unit and combination vehicles. Little or no information is available that identifies the relationship between truck accident frequency and truck size and mass.

Although exposure data are available, it is impossible to discuss accident rates by truck classifications, except in the most general terms. The data that are currently available do not indicate whether longer trucks or heavier trucks are over involved or under involved in accidents, based on their representation in the traffic population at the accident locations.

Although the methodology described here was developed to address truck accident rates (per million vehicle kilometers of exposure), the same technique could be used to obtain details about other types of vehicles such as automobiles, motorcycles, or buses.

A sampling technique has been developed for selecting 80 roadway segments at which large-truck accident and exposure data will be collected. These segments comprise approximately 1609 km (1000 miles) of highway throughout six participating states, e.g., California, Maryland, Michigan, Nevada, Pennsylvania, and Texas.

A typology was created to partition all roadways into six exclusive types. Two classification variables were used: road location (two levels, urban and rural) and roadway type (three levels, primary, secondary, and Interstate). Roadway type was nested within road location.

In each state, a multistage stratified random sampling of roadway segments was drawn within each roadway type. The distribution of large-truck accidents experienced was then plotted for those segments sampled. Potential data-collection sites were identified by using a two-way stratification method based on historic truck-accident distribution curves. The final sites were selected by a team of trained field crews after on-the-scene evaluation of the potential sites. These crews based their decisions on previously specified selection criteria (e.g., weight data and ability to collect exposure data). The logic of the site selection process is shown in Figure 1. Table 1 illustrates the final distribution of selected sites in the framework of the six states and two roadway classification variables.

SELECTION OF COOPERATING STATES

A literature review of the existing truck-related research and accident data was conducted at the beginning of the project (1). From this information, states in candidate areas of the country were selected on bases of annual truck exposure (distance), truck accidents, or state laws permitting extremes in truck size, mass, or length. The six states selected and the reasons why are given in Table 2. The first three states (California, Texas, and Pennsylvania) were selected because they represent more truck kilometers of exposure per year than any other state. Michigan was selected because it permits the heaviest truck payloads in the United States. Nevada was selected because it permits triples (tractor plus three articulated trailer units) to operate over a larger variety of roads than any other state. Maryland was selected to permit convenient pilot testing of the field data-collection techniques and to represent a state allowing approximately average masses and sizes of vehicles.

The decision to choose 80 sites was made at the beginning of the project and is based on the number of national accident-exposure descriptors. It was reasoned that this number of locations would provide adequate quantitative data for statistical analysis of the variables of interest and their numerous subcategories (i.e., truck classifications, sizes, and masses).

SAMPLING PROCEDURE

Roadway-Segment Sampling

Each state was divided into three to five nonoverlapping geographical regions. A sample (areas partitioned by

Figure 1. Site-selection procedure.

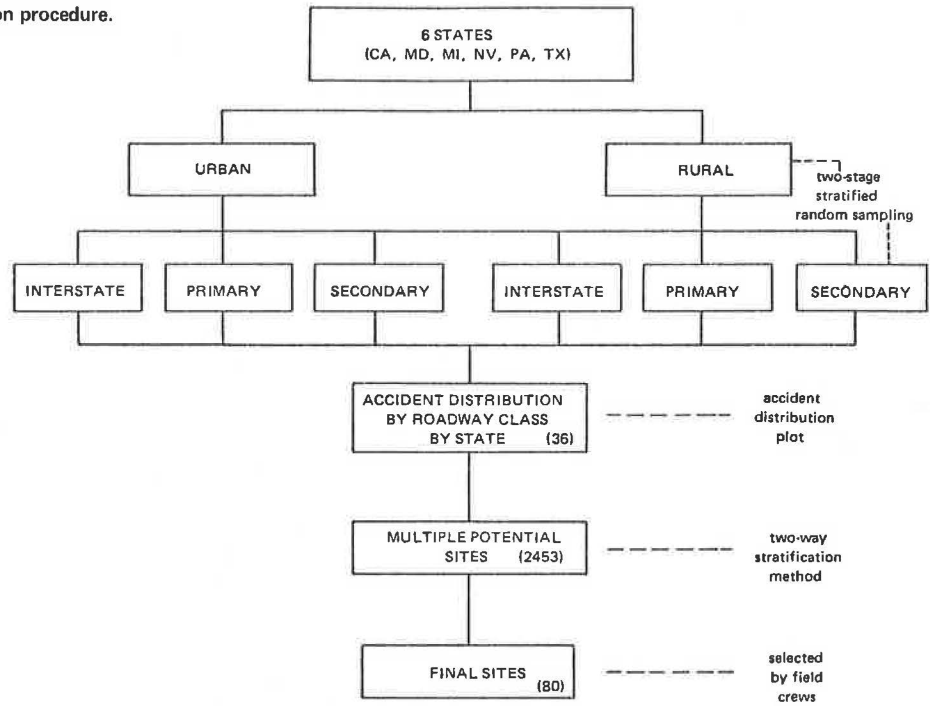


Table 1. Number of data-collection sites.

Roadway Class	No. of Sites: California	Maryland	Michigan	Nevada	Pennsylvania	Texas	Total
Rural							
Interstate	5	2	4	3	3	3	20
Primary	3	1	2	3	2	1	12
Secondary	3	1	2	2	2	2	12
Urban							
Interstate	5	2	4	2	3	4	20
Primary	2	1	2	1	1	1	8
Secondary	2	2	2	0	1	1	8
Total	20	9	16	11	12	12	80

Table 2. Vehicle characteristics in state participating in study.

State	Rank: Exposure[a]	No. of Accidents[b]	Max Truck Dimensions (m): Length	Height	Width	Comments
California	1	7	20	4.3	2.54	Allows doubles
Texas	2	4	20	4.1	2.44	–
Pennsylvania	3	1	16.8	4.1	2.44	–
Michigan	5	12	20	4.1	2.44	Heaviest axle loading allowed
Maryland	–	–	20	4.1	2.44	Pilot test procedures
Nevada	–	–	32	4.1	2.59	Allows triples

Note: 1 m = 3.3 ft.

[a]1974 Federal Highway Administration data for all trucks and all roads. [b]1973 Bureau of Motor Carrier Safety.

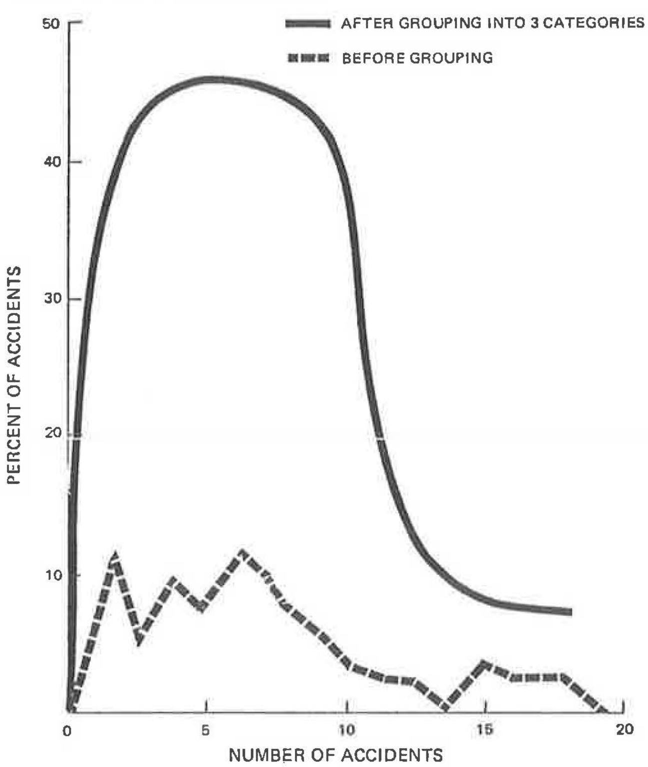
county or highway-district boundaries) was drawn from each. The drawing was made independently in each region (stratified random sampling). In California, for example, the sample was based on highway districts: Four districts—numbers 1, 3, 10, and 11—were sampled from among the 11.

Table 3. California rural Interstate-highway selection.

Highway District	Sample Area (counties)	Route[a]	Length (km)
1	5	None	–
3	11	A	145
		B	177
10	9	A	129
		B	48
		C	23
		D	26
11	3	A	32
		E	177
		F	32
All	28	6[b]	992[c]

Note: 1 km = 0.62 mile.
[a]Route numbers are not identified because the data collection in these areas is ongoing.
[b]Non-Interstate roadways.
[c]Sixty-two 16-km segments.

Figure 2. California large-truck accident distribution by 16-km (10-mile) segment of rural Interstate highways.



The next step was to identify the roadways to an initial limit of 10 to 20 in the sample area on the basis of their typology [i.e., urban versus rural and federal-aid Interstate (final) versus federal-aid primary versus federal-aid secondary (state)]. Although this task seems overwhelming, in no sample area was the number of roadways of a specific type (e.g., rural Interstate or urban primary) greater than nine. Table 3 illustrates the identification of rural Interstate highways in California.

Each roadway was divided into a number of segments. These were 16 km (10 miles) long for rural roadway categories, 4.8 km (3 miles) long for urban Interstates, and 3.2 km (2 miles) long for urban primary and secondary roadways. These lengths represent average distances between roadway exits where the average daily truck traffic remains constant.

Truck-Accident Distribution

Determining large (i.e., other than panel and pickup) truck-accident distributions by roadway classification was very time-consuming because it entailed counting the number of truck accidents in specific roadway segments. For each roadway classification, a 1974 truck-accident distribution was obtained from the proportion of those roadway segments in the sample having no truck accidents, one truck accident, two truck accidents, and so on. The site selection was primarily done on the basis of the accident distribution curves. These curves also permitted a direct comparison between the sites selected and the accident distribution for a specific class of roadway in a particular state. A large-truck accident distribution on the California rural Interstate highways sampled is shown in Figure 2. (Thirty-six of these distributions were plotted to display the six roadway types in each of the six states.) Such a distribution permits grouping all of the roadway segments into several accident categories, such as high, medium, and low. For California rural Interstate-highway segments, the following categories were derived:

1. Category 1 = low = 5 accidents/16-km (10-mile) highway segment,
2. Category 2 = medium = 6 to 10 accidents/16-km (10-mile) segment, and
3. Category 3 = high = more than 10 accidents/16-km (10-mile) segment.

Thus, in selecting the five rural Interstate-highway study sites allocated to California in Table 1, it was necessary to select from sixty-two 16-km (10-mile) highway segments within three accident categories and four geographical highway districts. A two-way stratification was used (2). (There were originally 2453 candidate sites from which 80 final study sites were selected.)

Table 4. Number and proportion of California rural Interstate segments in sample.

Accident Category	District 1 m_{ij}	District 1 P_{ij}	District 3 m_{ij}	District 3 P_{ij}	District 10 m_{ij}	District 10 P_{ij}	District 11 m_{ij}	District 11 P_{ij}	Total m_{ij}	Total P_{ij}
1	0	0.0	6	0.0968	3	0.0484	19	0.3065	28	0.4517
2	0	0.0	11	0.1774	7	0.1129	8	0.1290	26	0.4193
3	0	0.0	3	0.0484	4	0.0645	1	0.0161	8	0.1290
Total	0	0.0	20	0.3226	14	0.2258	28	0.4516	62	1.0000

Note: i = accident category (1, 2, or 3) and j = highway district (1, 3, 10, or 11).

Two-Way Stratification for Selecting Candidate Sites

In sampling the highway districts in California for rural Interstate study sites, sixty-two 16-km (10-mile) roadway segments were identified (Table 3). Table 4 represents the stratification of these 62 roadway segments into three accident categories and four geographic areas (highway districts). Thus, in district 3, there are 320 km (200 miles) of highway or 20 highway segments or potential sites. By using the 1974 truck-accident data, these 20 can be subdivided into 6 in category 1 (low), 11 in category 2 (medium), and 3 in category 3 (high). Similarly, the 14 segments in district 10 and the 28 segments in district 11 can be stratified by accident category. The number of segments is given by m_{ij} and the proportion of segments is given by P_{ij} ($P_{ij} = m_{ij}/62$).

The next procedure is to give each segment an approximately equal chance of selection and each accident category and highway district its proportional representation. In this instance, five study sites ($n = 5$) have been specified, and the numbers $n_{i.} = n\sum P_{ij}$ and $n_{.j} = n\sum P_{ij}$ are computed. These products are rounded to the nearest integers (with a further minor adjustment if required), so that both $n_{.j}$ (the total number of sites in the highway districts) and $n_{i.}$ (the total number of sites in the accident categories) add to n. The next step is to draw $n = 5$ cells with the probability $(n_{i.} \times n_{.j})/n^2$ for the ij th cell by constructing an n by n matrix. In row 1 of this matrix, one column is drawn at random. In row 2, one of the remaining columns is drawn at random, and so on. At the end, each row and column contains one unit. The results of one draw are indicated by X's in the table below.

Row (accident category)	Column (highway district) 1 (3)	2 (3)	3 (10)	4 (11)	5 (11)
1 (1)	–	–	–	X	–
2 (1)	–	–	–	–	X
3 (2)	–	X	–	–	–
4 (2)	X	–	–	–	–
5 (3)	–	–	X	–	–

Columns 1 and 2 were assigned to district 3 because $n_{.2} = 2$. Similarly, rows 1 and 2 were assigned to accident category 1 because $n_{1.} = 2$, and so on. This completes the allocation of the sample to the 12 cells. The three cells in district 1 were out of the draw because $n_{.1} = 0$. The allocation is given in more compact form below. (An alternative approach would be to convert P into ranges of integers and then select from a table of random numbers.)

Accident Category	District 1	3	10	11	Total
1	0	0	0	2	2
2	0	2	0	0	2
3	0	0	1	0	1
Total	0	2	1	2	5

The 62 potential sites have now been reduced to 34 potential sites. Specifically, the two sites representing the low accident category can now be selected from 19 potential sites in highway district 11, the two sites in the medium accident category can now be selected from 11 potential sites in highway district 3, and the one site in the high accident category can now be selected from 4 potential sites in highway district 3. The final sites were selected by the same crew that traveled to the six states and used a defined set of criteria.

Final Site Selection

There were nine primary criteria used in selecting the 80 sites. These were

1. Well-defined points of egress and entrance (control of exposure and accident data),
2. Vehicle volumes (confidence that vehicles traversed the entire segment),
3. Vehicle and truck mix (confidence that specific types of vehicles were consistent through the entire segment),
4. Truck-accident rates (high accident rates at selected sites within an accident category should be offset by low accident rates at the remaining sites to be selected within that category),
5. As long a segment as possible (given that the first three criteria are met),
6. Possible collection of exposure and speed data (field data collection by observers or cameras must be covert and not affect vehicle performance),
7. Possible collection of truck-mass exposure data (rest area, truck stop, or weighing station close enough to the site that there is high probability that the data collected will be directly applicable to the site under consideration),
8. Clear field of view and light source (to permit sampling of vehicle exposure both day and night), and
9. Cooperation of the local jurisdiction (it was necessary to have police and highway departments to assist in gathering the exposure data and to identify all large-truck accidents occurring within a 1-year period of time at each of the sites).

SUMMARY

This rather elaborate sampling procedure made possible the first large-scale, multistate, simultaneous field data collection of both accident and exposure data. The multistage stratified random sampling of roadway types will permit an analysis of large-truck accidents that are representative of the experiences of each of the six states that are cooperating in the study.

ACKNOWLEDGMENTS

This paper relates to one area of a study by BioTechnology, Inc., for the Federal Highway Administration. We wish to thank the individuals responsible for the traffic records of the states comprising the field study sites. We wish to acknowledge the efforts of Gerald R. Vallette of BioTechnology, who was in charge of the state coordination and site-selection activities and thank Wallace G. Berger of the U.S. Senate Staff (formerly of BioTechnology), and Julie Anna Fee, Martha Schwendeman, and Michael Freitas of the Federal Highway Administration for their suggestions. The views expressed in the paper are those of the authors and do not necessarily reflect the views or the policies of the Federal Highway Administration or the U.S. Department of Transportation.

REFERENCES

1. M. Reiss, G. Vallette, and C. S. Yoo. Accident and Traffic Operations Implications of Large Trucks: State-of-the-Art Review of Truck-Related Literature. BioTechnology, Interim Rept., Sept. 1975.
2. W. G. Cochran. Sampling Techniques. Wiley, New York, 2nd Ed., 1963.

Publication of this paper sponsored by Committee on Traffic Records.