Estimation of County Primary Road System Needs by Sample Survey Methods

DONALD O. COVAULT, Associate Professor of Civil Engineering, Georgia Institute of Technology, and HAROLD L. MICHAEL, Assistant Director, Joint Highway Research Project, Purdue University

Highway needs studies are costly and time consuming and require large engineering staffs for their performance. This study is an investigation of whether sample survey methods can be used to reduce the time and work required to make these studies. Complete inventory and cost data were available for the primary county road systems in Michigan and Minnesota and this information was used for statistical analysis. The value of population characteristics and the variances of these characteristics were computed for the complete data of each state.

Four different sampling methods were investigated for required sample size using five different combinations of margin of error and α risk for each population. The methods investigated were:

1. Simple random sampling.
2. Stratified random sampling with optimum and proportional allocation.
3. Simple cluster sampling.
4. Stratified cluster sampling with optimum and proportional allocation.

Using the five combinations of margin of error and α risk for each population, sample sizes required for the estimate of total cost of statewide needs were computed. Sample sizes required for the estimation of other population totals were also computed. For a given margin of error and α risk for each population all forms of cluster sampling used required substantially larger sampling rates for the estimate of total cost than those required for simple random or stratified random sampling. For all forms of sampling used, stratified random sampling with optimum allocation required the smallest sample rates for a given order of accuracy and simple cluster sampling required the largest.

Estimated sample survey costs required for an estimate of the total cost of the needs for five orders of accuracy were also investigated for each sample method studied. Very little difference in sampling survey costs was found between the cost for simple random sampling and stratified random sampling with optimum and proportional allocation. All forms of cluster sampling, however, required much greater expenditures to obtain a needs estimate with a comparable margin of error and α risk to that used for simple and stratified random sampling.

It was concluded from this work that sample survey techniques for the estimation of total highway needs on county road systems in an entire state are practical.

●THE PERFORMANCE of a complete highway needs study on all highway systems in any state is costly and time consuming. A typical study requires from one to three years to complete and a large staff of engineers, technicians, and clerical personnel. Complete cooperation of all state, county, and municipal highway and street agencies is also necessary if the work is to be accomplished quickly and efficiently, and a total expenditure of $200,000 to $600,000 is not unusual.

The evaluation of the needed improvements on county and township highway systems has always been particularly difficult because of the large mileage of highways involved in these systems and the absence of basic data. Although the total cost of eliminating the inadequacies on these systems may be comparatively low, large expenditures are required to evaluate this small portion of the total cost of the required highway, bridge, and railroad crossing improvements.

Because highway needs studies influence the financial and administrative policy within the various state legislatures, recent and reliable estimates of highway needs are essential. Present methods used in making these studies, however, require such a large expenditure of time and money that needs studies are only performed infrequently, if they are made at all. To be most effective these studies must be made on a continuing basis. The application of sample survey techniques to the estimation of highway needs offers an ideal means whereby the time and expenditures required to make each study may be substantially reduced. Sample survey methods are especially adaptable to the county primary road systems because of their large mileage and the lack of necessity for detailed knowledge of the specific needs for every mile of highway. This system usually has comparatively low variability for needed improvements and thus requires small sample sizes for adequate accuracy. This research was intended to serve two specific purposes. The first of these was to determine if the use of sample survey methods for the estimation of the total cost of eliminating highway needs for county primary road systems was practical. A second purpose was to apply sample survey theory to this problem in order to compute the necessary sample sizes, population variances, and the relative survey costs for different sampling plans. This type of information is essential for any agency which plans to estimate highway needs by sample survey techniques. However, the scope of this paper does not cover the detailed methods by which sample surveys are made.

Sample surveys may not always provide data that are sometimes required or may be a by-product of a complete study. Sample survey techniques will provide a feasible means to obtain the total cost of highway needs and other general data on a statewide basis. However, complete data are not obtained for a particular county which can be used for planning and programming in that particular county. Also, some physical data on a statewide basis such as the number of miles of reconstruction, resurfacing, etc., may not be obtained with any reliable degree of accuracy unless large sample sizes are drawn.

This research was concerned with the development of methodology and techniques using sample survey theory for the estimation of highway needs on county highways in Michigan and Minnesota. Complete highway needs data were available in these two states for the county primary highway systems. The values of population characteristics (for example, the total cost of needs on the county primary highway system in Michigan) and variances were computed using the complete highway needs data and then used to compute sample sizes and sample variances for different sampling plans using various orders of accuracy. The sampling plans which were investigated were: (a) simple random sampling; (b) stratified random sampling with optimum and proportional allocation; (c) simple cluster sampling; and (d) stratified cluster sampling and optimum and proportional allocation.

Stratified random sampling with optimum or proportional allocation is referred to as optimum or proportional stratified random sampling, respectively, in this study. Stratified cluster sampling with optimum or proportional allocation is referred to as optimum or proportional cluster sampling, respectively.

Of basic importance to any sample survey is the choice of a suitable order of accuracy. This may be measured in terms of a margin of error and α risk (4). Stated in a slightly different manner, a margin of error d_T in the total and a small risk α that the actual error is larger than d_T should be chosen before a sample size is computed. Stated in terms of a probability formula, the sample size must be large enough so that

$$\Pr\left[|\hat{Y} - Y| \geq d_T\right] = \alpha \qquad (1)$$

where the symbols are defined as follows:

$\hat{Y}$ = Estimated population total
Y = Population total

Alternately the sample size is sometimes specified as large enough to provide a confidence interval of one-half width d_T with a confidence probability of $(1 - \alpha)$ (4).

For the several sampling methods considered for each population, five different combinations of margin of error and α risk were studied for required sample size and variance. No specific combination is recommended in this study for the estimation of highway needs as this decision must be made by the highway agency making and using the results of the sample survey. Before arriving at such a decision, the responsible personnel in the highway agency must consider such things as the answers to the following questions:

1. Who will use the information obtained for highway needs?
2. How serious will be the consequences of an estimate in error by 10 percent? By 20 percent? Etc.
3. How quickly (in terms of another scheduled survey) can a possible large error be corrected?

Generally, a high accuracy is not required for the estimate of the total cost of making all required highway improvements, especially if the highway agency makes continuing needs studies. If a substantial error occurs in the estimated needs, the next needs evaluation may correct the previous estimate before it has seriously influenced policy. Furthermore, the concepts, costs, and standards upon which highway needs are based are continually changing and the estimate of the total cost of these needs, therefore, will also change.

Although Michigan and Minnesota data were used as the basis for statistical analysis, these populations may be similar to the populations in other states. The county road systems in these two states are typical of most of the states in the midwest. Furthermore, land usage is similar to that in many other states and most of the states in the midwest have a well developed county primary road system which is extensively used for agricultural purposes. The results obtained from the analysis of Michigan and Minnesota data, therefore, are probably similar to values which would be obtained for other states, especially those which have a well developed county primary highway system.

In this study for many of the investigations, the county highway primary system in each state was considered as composed of six populations. These six were all the highway sections, all the bridges, and all the railroad crossings on the rural system, and a similar listing for the urban system. Each of these six populations (hereafter referred to as separate populations) was sampled separately and the results added to estimate the total highway needs for the entire county road system in the state.

The sampling units for the separate populations were different for each population when simple or stratified random sampling was used. The sampling unit for the highway population was a highway section; the sampling unit for the bridge population was a bridge; and the sampling unit for the railroad populations was a railroad crossing.

The sampling unit for simple and stratified cluster sampling was a county for all populations. All highway sections, bridges, and railroad crossings in each sample county were evaluated when the estimates of these populations were desired. Each highway section, bridge, and railroad crossing in this case was called an element.

For some of the evaluations, another population was used. This population was composed of all highway sections where each section included all the bridges and railroad crossings which occurred within it. This type of population is referred to as a composite population in this report. The total cost of improvement of a highway section in such a population, therefore, included the cost of improvement of the highway plus the cost of improving bridges and railroad crossings which were located within the section.

When the composite population was used, the sampling unit for simple or stratified random sampling was a highway section while for simple or stratified cluster sampling it was, as with the separate populations, the county. In the latter case, the highway sections were again referred to as elements.

ANALYSIS OF DATA

Because complete highway needs information was available in Michigan and Minnesota for the rural and urban county primary road systems, data from these two states were used for statistical analysis. Each state had obtained the total cost of their county primary highway needs and other pertinent information based on a 20-yr improvement program. All data had been placed on punch cards and thus could be readily used for analysis in business machines and digital computors.

Roads classified in the primary systems in Michigan and Minnesota had also been subdivided into homogeneous sections. Bridges and railroad crossings contained in each road section were also located so that they could be recorded and identified. Figure 1 indicates the rural primary road system in Missaukee County, Michigan, with appropriate identification for the highway sections, bridges, and railroad crossings.

Traffic volume information is essential to the evaluation of the cost of improvement of a highway system and is also necessary if stratified random sampling — with stratification on traffic volume — is to be used. Topographic information is also required for each road section in addition to the total road mileage in a road system for each county. All of this information was available for the county primary highway systems of Michigan and Minnesota.

The amount of prior information required for a sampling plan depends upon the type of plan that is to be used. The information required to draw simple or stratified cluster samples usually can be obtained from transportation maps, mileage records, economic, and fiscal data. Simple random sampling, for example, requires only a prior knowledge and appropriate identification of the various road sections, bridges and railroad crossings in each county. Stratified random sampling, on the other hand, requires the greatest

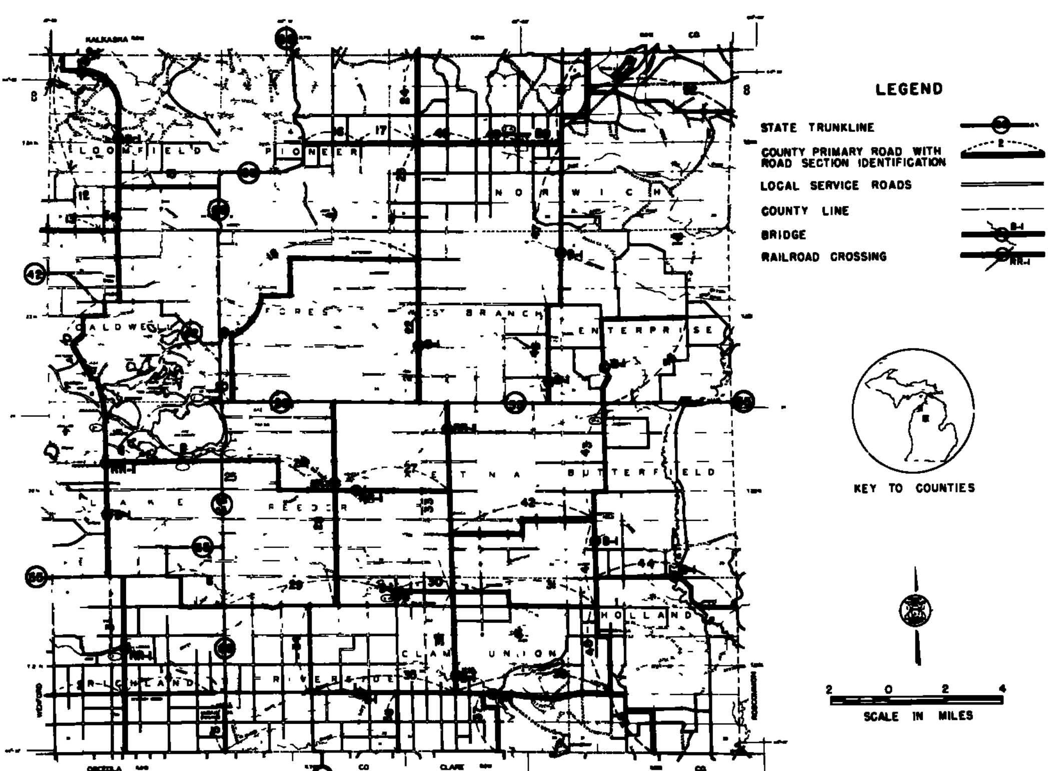

Figure 1. County primary roads, Missaukee County, Michigan (19).

amount of prior information about the various populations. Not only must the individual highway sections, bridges and railroad crossings be located and recorded, but additional information must also be obtained to permit stratification.

Because of the type of highway required for various traffic volumes and the different grading costs of various types of topography, traffic volume and/or topography were used as strata in stratified random sampling. Figures 2 and 3 illustrate how the rural and urban separate and composite populations were stratified into topographic and traffic strata for stratified random sampling. After considerable study of fiscal, economic, and mileage data counties were stratified for cluster sampling according to total primary road mileage in each county. The strata were defined as follows:

STRATUM	MILES OF PRIMARY ROAD
1	0 - 249.99
2	250.00 - 349.99
3	350.00 and above

Before population values and sample sizes for different margin of errors and α risks were computed, all existing and proposed multilane facilities and bridges and railroad crossings located in these highway sections on the rural system were removed from the original Michigan and Minnesota data. This was done because of the small number of such sections, the much greater cost of such improvement, and the effect these sections would have on the population variances. A final summary of the number of sampling units and elements used in the four methods of sampling studied are given in Tables 1 and 2 for the Michigan and Minnesota data. Table 3 presents the total costs and population variances for the various populations as determined from the total county primary highway needs data of each state. These data are for 21,120 miles on the rural and 402 miles on the urban primary county system in Michigan, and 27,000 miles on the rural system and 1,850 miles on the urban system in Minnesota.

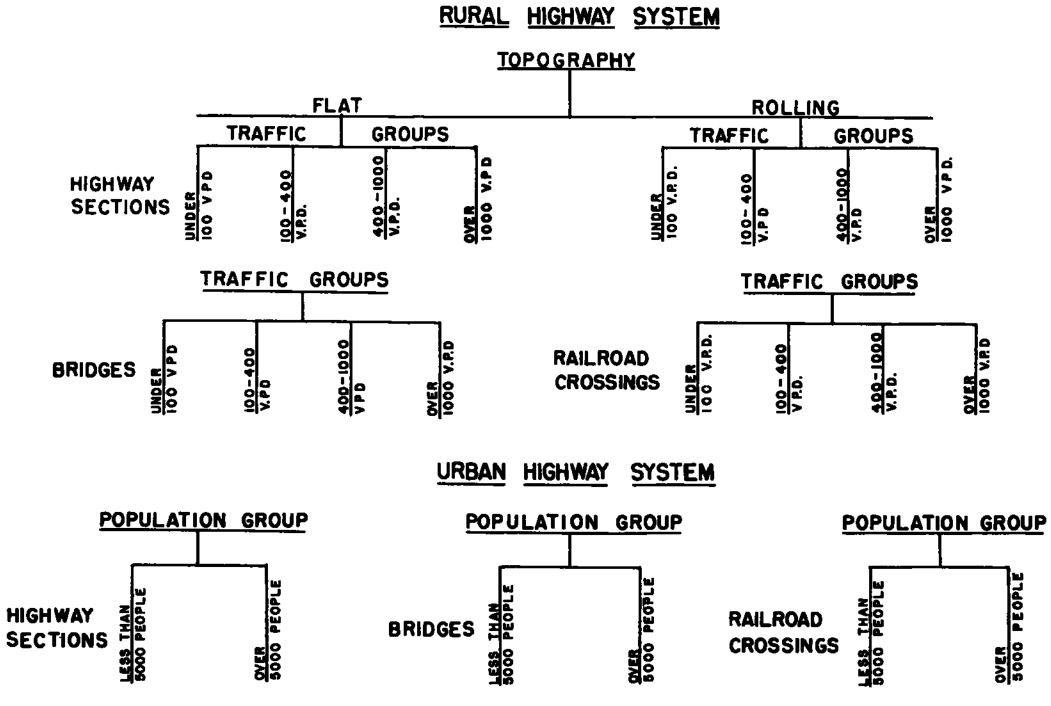

Figure 2. Diagram of stratification of separate populations for the rural and urban highway system.

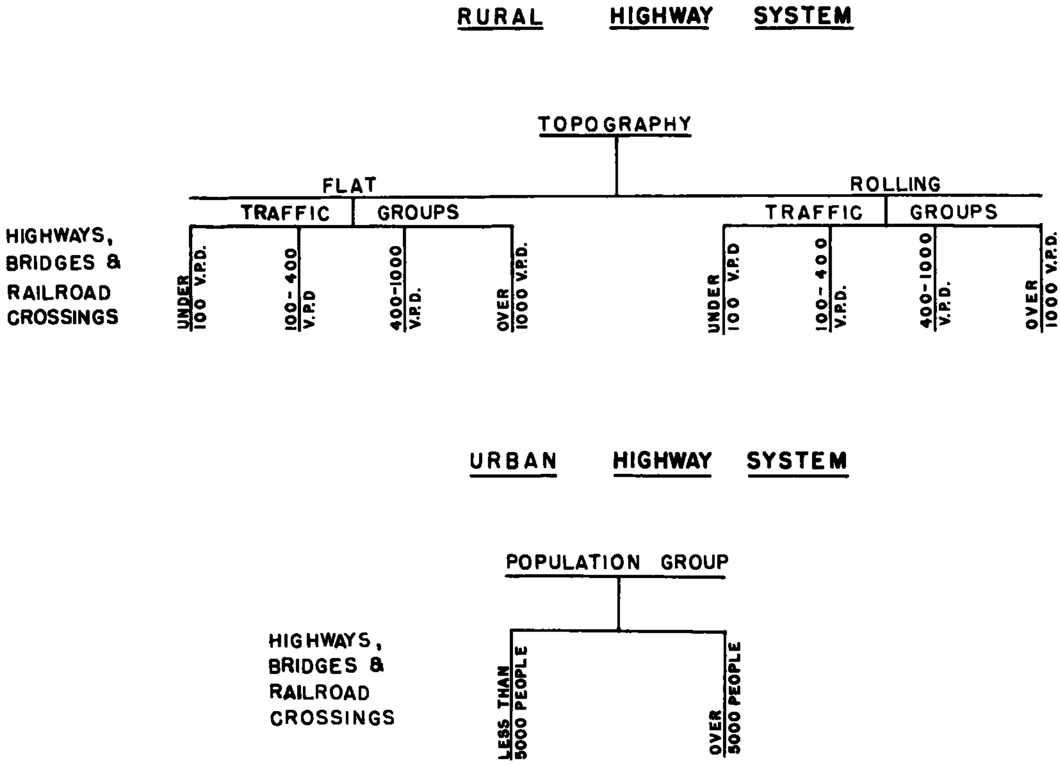

Figure 3. Diagram of the stratification of the composite populations for the rural and urban highway systems.

The choice of margin of error and α risk desired for the population characteristic being estimated, as stated before, is primarily a matter of policy which must be decided by the particular highway agency making and using the results of the survey. For this reason sample sizes for the estimate of various population characteristics were computed for five different combinations of margin of error and α risk for each population. These combinations define five different orders of accuracy for the estimate of each population characteristic. These five different combinations are listed below in ascending order of accuracy:

ORDER OF ACCURACY	MARGIN OF ERROR, d_T	α RISK
1	20 % of the total being estimated	0.10
2	10 % of the total being estimated	0.10
3	10 % of the total being estimated	0.05
4	5 % of the total being estimated	0.10
5	5 % of the total being estimated	0.05

For a given population and a specified order of accuracy, sample sizes required to estimate various population characteristics varied considerably. For each separate population and for the composite population the characteristic which required minimum sample size was the estimate of total cost of needs for that population. Other characteristics such as the total cost of needs for various time periods for the 20-yr program period, number of inadequate bridges, and the cost of various types of highway construction, etc., required much larger sample sizes than were required for the estimate of total cost. Tables 4-10 show for each population the sampling rates (sample size divided by the total number of sampling units in the population) required for the estimate of total cost. Only the highest and lowest order of accuracy used in this study are given in these tables but complete information for the other orders of accuracy and other population characteristics can be found in "Estimation of Highway Needs for County Primary Road Systems in Michigan and Minnesota by Sample Survey Methods" (7). Because the sizes of the respective populations varied, sampling rates of different populations cannot be directly compared. Sampling rates for the same population, using the various sampling methods, however, can be compared.

Table 4 indicates the sampling rates required for the separate rural populations for the Michigan data. Optimum stratified random sampling required minimum sampling rates for the respective populations. However, only a nominal difference occurred between this method and either simple random or proportional stratified random sampling. For example, when d_T = 20 percent of the total and α = 0.10, the sampling rates for the highway section population were 1.0, 0.9, and 1.0 percent, respectively, for simple random, optimum stratified random, and proportional stratified random sampling. When d_T = 5 percent of the total and α = 0.05, the rates for these respective forms of sampling increased to 19.2, 16.7 and 18.3 percent.

The bridge population required sampling rates of 8.0, 7.1, and 7.9 percent, respectively, for simple random, optimum stratified random, and proportional stratified random sampling when d_T = 20 percent of the total and α = 0.10. These rates increased to 66.3, 59.2 and 66.3, respectively, when d_T = 5 percent of the total and α = 0.05.

TABLE 1

TOTAL NUMBER OF SAMPLING UNITS USED FOR SIMPLE AND STRATIFIED RANDOM SAMPLING MICHIGAN AND MINNESOTA DATA

Population	Total Sampling Units, N	
	Mich.	Minn.[2]
Rural separate:		
Highway sections	6,321	7,905
Bridges	1,985	2,761
Railroad crossings	521	662
Rural composite	6,321	7,905
Urban separate:		
Highway sections	276	3,307
Bridges	44	373
Railroad crossings	6[1]	467
Urban composite	276	3,307

[1] Incomplete data.
[2] Data from one county missing, containing approximately 90 road sections However, missing data did not seriously influence results obtained for the various sampling methods studied. Population assumed to consist of the number of sampling units shown.

TABLE 2

TOTAL NUMBER OF SAMPLING UNITS AND ELEMENTS USED FOR SIMPLE AND STRATIFIED CLUSTER SAMPLING FOR THE VARIOUS POPULATIONS MICHIGAN AND MINNESOTA DATA

Population	Michigan		Minnesota[2]	
	Total Sampling Units, N	Elements (M)	Total Sampling Units, N	Elements (M)
Rural separate:				
Highways	83	6,321	86	7,905
Bridges	83	1,985	86	2,761
Railroad crossings	83	521	86	662
Rural composite	83	6,321	86	7,905
Urban separate				
Highways	83	276	86	3,307
Bridges	83	44	86	373
Railroad crossings	83	6[1]	86	467
Urban composite	83	276	86	3,307

For footnotes see Table 1.

TABLE 3

POPULATION COST TOTALS AND VARIANCES FOR SEPARATE AND COMPOSITE POPULATIONS

Population	Michigan		Minnesota	
	Total Cost ($1,000)	S^2	Total Cost ($1,000)	S^2
Rural Highway System				
Separate populations				
Highways	$536,722	7,073	$479,080	3,133
Bridges[1]	63,957	2,634	55,724	1,122
Railroad crossings	2,757	26.03	1,762	34.12
Composite population (Highways, bridges, and railroad crossings)	603,475[2]	9,316	536,566	4,011
Urban Highway Systems				
Separate populations				
Highways	90,515	315,991	96,334	2,654
Bridges	14,261	328,000	23,307	35,003
Railroad crossings	Incomplete data		2,283	46.49
Composite population (Highways, bridges, and railroad crossings)	Incomplete data		121,924	7,584

[1] Includes approach costs for Michigan data only.
[2] Slight error in value because of minor coding errors in punch cards.

A large sampling rate was necessary for all forms of cluster sampling relative to the rate required for simple and stratified random sampling. Large sampling rates were also required for the bridge and railroad crossing populations for all methods of sampling.

TABLE 4

SAMPLING RATES FOR THE ESTIMATES OF TOTAL COST OF IMPROVEMENT FOR THE SEPARATE POPULATIONS ON THE RURAL HIGHWAY SYSTEM—MICHIGAN DATA

Method of Sampling	Population	Sampling Rate (%)[1] $d_T = 20\%$ $\alpha = 0.10$	Sampling Rate (%)[1] $d_T = 5\%$ $\alpha = 0.05$
Simple random sampling	Highways	1.0	19.2
	Bridges[2]	8.0	66.3
	Railroad crossings	10.7	73.3
Optimum stratified random sampling	Highways	0.9	16.7
	Bridges[2]	7.1	59.2
	Railroad crossings	9.8	69.1
Proportional stratified random sampling	Highways	1.0	18.3
	Bridges[2]	7.9	66.3
	Railroad crossings	10.2	71.2
Simple cluster sampling	Highways	34.9	91.5
	Bridges	45.7	95.1
	Railroad crossings	77.0	98.7
Optimum stratified cluster sampling	Highways	14.4	61.3
	Bridges	32.5	84.2
	Railroad crossings	47.0	75.8
Proportional stratified cluster sampling ($n_h \propto N_h$)	Highways	22.9	86.6
	Bridges	45.8	93.8
	Railroad crossings	73.5	100.0
Proportional stratified cluster sampling ($n_h \propto M_{Hh}$)	Highways	18.1	69.8
	Bridges	32.5	84.3
	Railroad crossings	57.8	75.8

[1] $\frac{n}{N}$ x 100 margin of error and α risk.

[2] Includes approach costs.

TABLE 5

SAMPLING RATES FOR THE ESTIMATES OF TOTAL COST OF IMPROVEMENT FOR THE SEPARATE POPULATIONS ON THE RURAL HIGHWAY SYSTEM—MINNESOTA DATA

Method of Sampling	Population	Sampling Rate (%)[1] $d_T = 20\%$ $\alpha = 0.10$	Sampling Rate (%)[1] $d_T = 5\%$ $\alpha = 0.05$
Simple random sampling	Highways	0.7	14.2
	Bridges	6.3	60.4
	Railroad crossings	32.9	91.6
Optimum stratified random sampling	Highways	0.6	12.2
	Bridges	5.1	49.3
	Railroad crossings	25.1	72.1
Proportional stratified random sampling	Highways	0.7	13.1
	Bridges	5.3	60.2
	Railroad crossings	31.9	91.2
Simple cluster sampling	Highways	19.5	84.0
	Bridges	40.2	93.1
	Railroad crossings	74.7	97.7
Optimum stratified cluster sampling	Highways	11.6	70.8
	Bridges	32.5	83.6
	Railroad crossings	63.8	96.2
Proportional stratified cluster sampling ($n_h \propto N_h$)	Highways	14.0	77.9
	Bridges	38.4	93.0
	Railroad crossings	75.5	100.0
Proportional stratified cluster sampling ($n_h \propto M_{Hh}$)	Highways	12.8	70.8
	Bridges	33.7	84.8
	Railroad crossings	72.1	96.5

[1] $\frac{n}{N}$ x 100 margin of error and α risk.

TABLE 6

SAMPLING RATES FOR THE ESTIMATES OF TOTAL COST OF IMPROVEMENT FOR THE SEPARATE POPULATIONS ON THE URBAN HIGHWAY SYSTEM—MICHIGAN DATA

Method of Sampling	Population	Sampling Rate (%)[1] $d_T = 20\%$ $\alpha = 0.10$	Sampling Rate (%)[1] $d_T = 5\%$ $\alpha = 0.05$
Simple random sampling	Highways	42.0	94.0
	Bridges[2]	81.8	97.8
	Railroad crossings	inadequate data	
Optimum stratified random sampling	Highways	30.1	75.7
	Bridges[2]	46.5	61.4
	Railroad crossings	inadequate data	
Proportional stratified random sampling	Highways	37.0	93.2
	Bridges[2]	79.5	97.7
	Railroad crossings	inadequate data	

[1] $\frac{n}{N}$ x 100 margin of error and α risk.

[2] Includes approach costs.

TABLE 7

SAMPLING RATES FOR THE ESTIMATES OF TOTAL COST OF IMPROVEMENT FOR THE SEPARATE POPULATIONS ON THE URBAN HIGHWAY SYSTEM—MINNESOTA DATA

Method of Sampling	Population	Sampling Rate (%)[1] $d_T = 20\%$ $\alpha = 0.10$	Sampling Rate (%)[1] $d_T = 5\%$ $\alpha = 0.05$
Simple random sampling	Highways	6.0	59.2
	Bridges	61.8	97.2
	Railroad crossings	22.0	86.5
Optimum stratified random sampling	Highways	5.1	52.9
	Bridges	40.2	64.3
	Railroad crossings	21.8	85.9
Proportional stratified random sampling	Highways	5.5	57.3
	Bridges	60.8	97.2
	Railroad crossings	21.8	86.4

[1] $\frac{n}{N}$ x 100 margin of error and α risk.

Table 5 gives sampling rates computed for the separate populations in Minnesota. Optimum stratified random sampling gave minimum rates for each population. Again all forms of cluster sampling required much larger sampling rates than those required by simple or stratified random sampling.

Table 6 gives the rates required using the separate urban populations for the Michigan data. Large sampling rates were required for these populations for all combination of margins of error and α risks used in this study. Large sampling rates were required for these populations because of the comparatively few miles of highway on the urban system in Michigan.

Table 7 gives rates required using the Minnesota data for the separate urban populations. When d_T = 20 percent of the total highway cost and α = 0.10, sampling rates of 6.0, 5.1 and 5.5 percent, respectively, were required using simple random, optimum stratified random, and proportional stratified random sampling. When d_T = 5 percent of the total and α = 0.05, the rates increased to 59.2, 52.9 and 57.3 percent, respectively.

TABLE 8

SAMPLING RATES FOR THE ESTIMATES OF TOTAL COST HIGHWAY COST, BRIDGE COST, RAILROAD CROSSING COST FOR THE COMPOSITE POPULATIONS ON THE RURAL HIGHWAY SYSTEM—MICHIGAN DATA

Method of Sampling	Desired Estimated Value	Sampling Rate (%)[1] d_T = 20%, α = 0.10	Sampling Rate (%)[1] d_T = 5%, α = 0.05
Simple random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	1.1	19.9
	Highway cost	Same as Table 4	
	Bridge cost	11.3	74.3
	Railroad crossing cost	19.4	84.4
Optimum stratified random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	0.9	16.9
	Highway cost	Same as Table 4	
	Bridge cost	10.0	67.4
	Railroad crossing cost	14.9	69.8
Proportional stratified random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	1.0	19.0
	Highway cost	Same as Table 4	
	Bridge cost	11.2	74.0
	Railroad crossing cost	18.7	82.6
Simple cluster sampling	Total cost (All highway, bridge, and railroad crossing improvements)	32.5	91.5
Optimum stratified cluster sampling	Total cost (All highway, bridge, and railroad crossing improvements)	14.4	61.3
Proportional stratified cluster sampling (n_h oc N_h)	Total cost (All highway, bridge and railroad crossing improvements)	22.9	85.5
Proportional stratified cluster sampling (n_h oc M_{Hh})	Total cost (All highway, bridge, and railroad crossing improvements)	18.1	68.6

[1] $\frac{n}{N}$ x 100 margin of error and α risk

TABLE 9

SAMPLING RATES FOR THE ESTIMATES OF TOTAL COST HIGHWAY COST, BRIDGE COST, RAILROAD CROSSING COST FOR THE COMPOSITE POPULATIONS ON THE RURAL HIGHWAY SYSTEM—MINNESOTA DATA

Method of Sampling	Desired Estimated Value	Sampling Rate (%)[1] d_T = 20%, α = 0.10	Sampling Rate (%)[1] d_T = 5%, α = 0.05
Simple random sampling	Total cost (All highway, bridge and railroad crossing improvements)	0.7	14.5
	Highway cost	Same as Table 5	
	Bridge cost	8.7	68.3
	Railroad crossing cost	38.7	93.2
Optimum stratified random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	0.6	12.5
	Highway cost	Same as Table 5	
	Bridge cost	7.4	58.2
	Railroad crossing cost	27.9	82.9
Proportional stratified random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	0.7	13.5
	Highway cost	Same as Table 5	
	Bridge cost	8.7	68.4
	Railroad crossing cost	38.4	93.2
Simple cluster sampling	Total cost (All highway, bridge, and railroad crossing improvements)	19.5	84.0
Optimum stratified cluster sampling	Total cost (All highway, bridge, and railroad crossing improvements)	12.6	69.7
Proportional stratified cluster sampling (n_h oc N_h)	Total cost (All highway, bridge, and railroad crossing improvements)	14.0	79.0
Proportional stratified cluster sampling (n_h oc M_{Hh})	Total cost (All highway, bridge, and railroad crossing improvements	12.8	70.8

[1] $\frac{n}{N}$ x 100 margin of error and α risk

TABLE 10

SAMPLING RATES FOR THE ESTIMATE OF TOTAL COST HIGHWAY COST, BRIDGE COST, RAILROAD CROSSING COST FOR THE COMPOSITE POPULATIONS ON THE URBAN HIGHWAY SYSTEM—MINNESOTA DATA

Method of Sampling	Desired Estimated Value	Sampling Rates (%)[1] d_T = 20%, α = 0.10	Sampling Rates (%)[1] d_T = 5%, α = 0.05
Simple random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	4.3	32.8
	Highway cost	Same as Table 7	
	Bridge cost	15.2	38.8
	Railroad crossing cost	12.8	38.0
Optimum stratified random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	3.1	22.4
	Highway cost	Same as Table 7	
	Bridge cost	13.0	36.7
	Railroad crossing cost	12.7	38.0
Proportional stratified random sampling	Total cost (All highway, bridge, and railroad crossing improvements)	4.1	29.7
	Highway cost	Same as Table 7	
	Bridge cost	14.9	38.7
	Railroad crossing cost	13.3	38.0

[1] $\frac{n}{N}$ x 100 margin of error and α risk

Comparative large sampling rates were also required for the estimates of total costs of bridge and railroad crossing improvements in Minnesota. For example, when d_T = 20 percent of the total and α = 0.10 a sampling rate of 40.2 percent was required for the bridge population using optimum stratified random sampling. For the railroad crossing population when d_T = 20 percent of the total cost and α = 0.10, the sampling rate for optimum and stratified random sampling was 21.8 percent.

Tables 8, 9, and 10 contain sampling rates required for the rural composite populations in Michigan and Minnesota. For a given order of accuracy and method of sampling, the sampling rates required for the estimate of total cost of all highway, bridge, and railroad crossing improvements for the Michigan data were only slightly higher than or equal to the corresponding rates required for estimating the total cost of improvement of the separate highway populations. A similar statement can be made for the Minnesota data.

COST CONSIDERATIONS

The proper choice of a sampling method should include a consideration of the estimated total cost of making the study. To make the cost comparison it was assumed that the amount and accuracy of information which would be obtained by the use of the various sampling methods would be the same for the respective populations and that certain basic data such as maps, traffic information, soil types and so forth were available.

The total cost of a sample survey is composed of several component costs including those due to overhead and general supervision, sampling unit listing, travel, data collection, and data analysis. These costs will vary with sample size and with the type of sample used. For example, total travel cost costs between sampling units for cluster sampling may be less than for simple or stratified random sampling. On the other hand, cluster sampling will usually require a larger sample and the collection of more data than simple random or stratified random sampling.

Included in the cost of overhead and administration in a sample survey is the cost of office space, utilities, paper, pencils, and other supplies. Also included is the cost for general planning and administration and for secretarial and clerical help. It was assumed, however, that the overhead cost would be the same for all sampling methods, and it, therefore, was not included in the total cost computations of the sample surveys in this study.

The listing of a sampling unit involves finding the location of the terminii or each highway section on a map; writing a description of its location; determining traffic volume

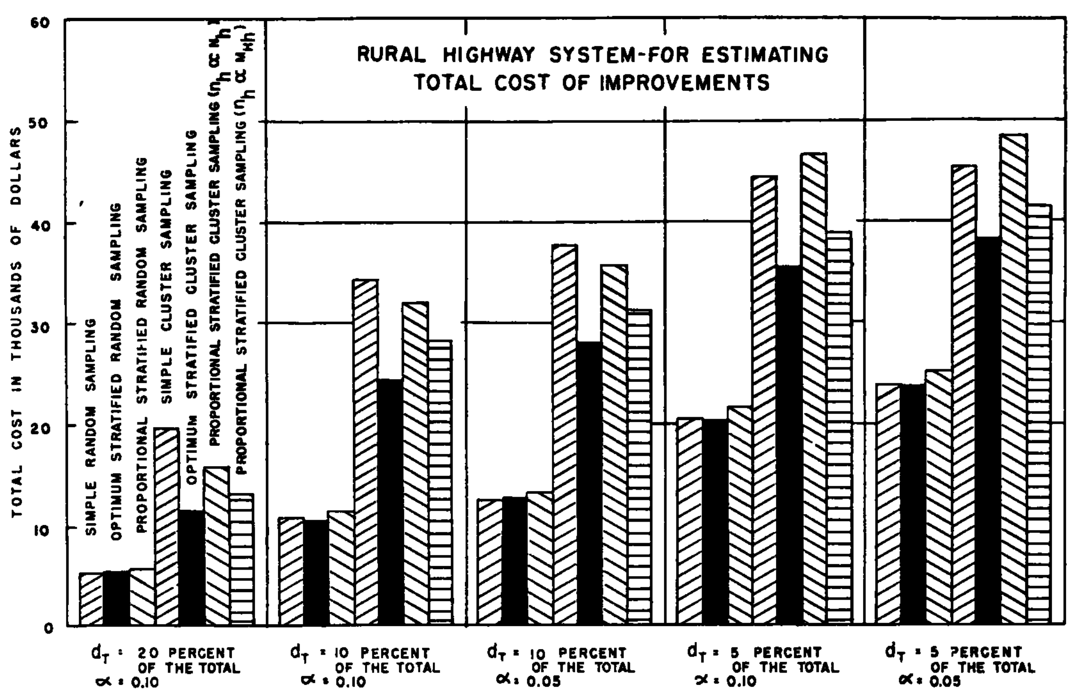

Figure 4. Estimated total cost of sample surveys for separate populations using simple and stratified random sampling and simple and stratified cluster sampling (Michigan data).

if required; punching the information on a card and checking the information for errors. For simple and stratified random sampling, this listing process must include every highway section, bridge, and railroad crossing in the road system being studied. On the other hand, complete listing need only occur in the units (counties) which are sampled when simple or stratified cluster sampling is used.

Travel cost is composed of several components. In addition to the actual expense of driving a vehicle from one sampling unit to another, the salaries of the survey crew while traveling must be included. A certain amount of travel expense is also caused by travel from and return to home or headquarters if the survey crew does not stay overnight away from home. If the survey crew is required to stay overnight away from headquarters, an additional subsistence expense occurs which is charged to travel cost. In this study all data were assumed to be collected by a survey party operating from a central headquarters or office within a given state and not by the local county engineer.

Travel cost does not vary directly with sample size. The larger the number of units sampled in a given area, the smaller will be the unit cost of travel. The total travel cost for a sample survey is (15)

$$C_T = C_1\sqrt{nA} \qquad (2)$$

when the symbols are defined as follows:

1. C_T - Total travel cost.
2. C_1 - Unit cost per mile of travel.
3. n - Number of units sampled.
4. A - Area of state in square miles.

Data collection cost includes salaries and wages paid to the survey crew for the gathering of information during the collection of inventory data on a highway system.

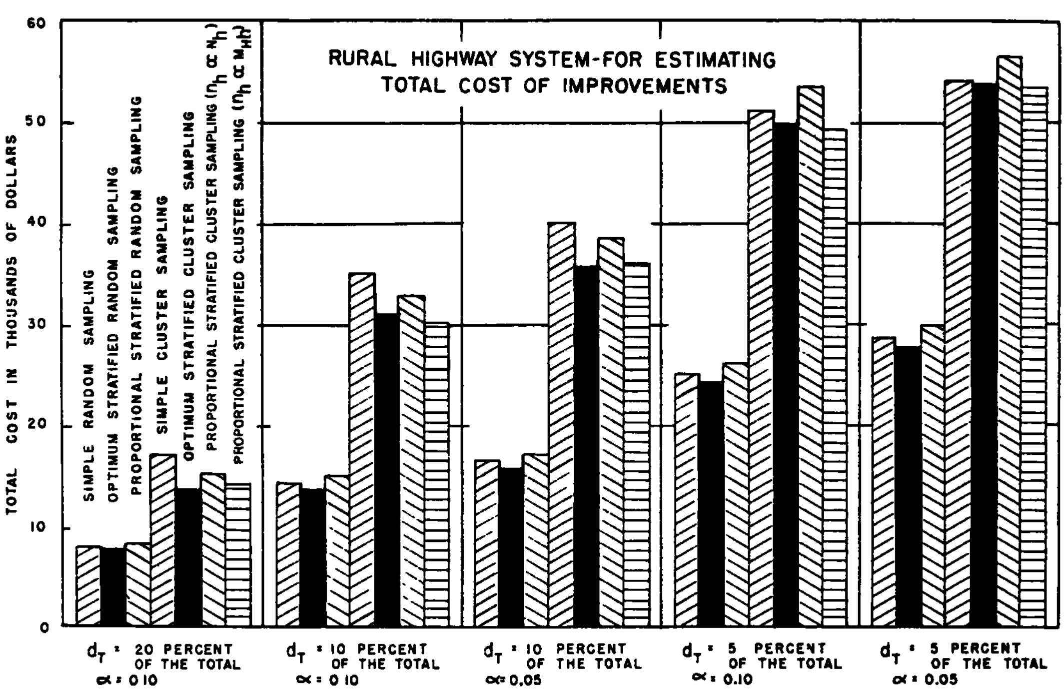

Figure 5. Estimated total cost of sample surveys for separate populations using simple and stratified random sampling and simple and stratified cluster sampling (Minnesota data).

Also included in the data collection costs are travel costs incurred while gathering information in the sampled highway section.

Data analysis is composed of several operations. Included in this cost, in addition to the actual statistical computations, is the cost of selecting the sample, evaluating inventory data, coding and punching information on punch cards, checking computations, and preparing the final reports.

When the highway sections, bridges and railroad crossings are treated as separate populations, the evaluation of the needs for each population consists of a separate sample survey. The total survey cost for a particular road system is, therefore, the sum of the costs for the sampling and analysis of each highway, bridge, and railroad crossing population. These populations are not totally independent, however, if the following procedure is used in collecting data. If bridges or railroad crossings are located in a highway section included in the highway section sample, these bridges and railroad crossings may be "forced" into the bridge or railroad crossing sample. If one assumes for simple random sampling that each highway section contains one bridge and one railroad crossing, or one bridge and no railroad crossing, or no bridge and one railroad crossing, or no bridge or railroad crossing at all, any bridge and railroad crossing which is included will also be chosen at random. Some road sections contain two or more bridges and/or railroad crossings, however, and this assumption of randomness is not exactly true. This error, however, should not be serious. "Forcing" bridges and railroad crossings in stratified random, and simple and stratified cluster sampling plans also can be done. This method of sampling has the definite advantage of reducing travel costs for the collection of data required for a sample survey.

A substantially smaller travel, data collection, and data analysis cost will occur for the composite population than that required for the separate populations. The data are somewhat easier to obtain because only those bridges and railroad crossings located in the sampled highway sections are sampled. However, the information which is obtained from the composite population may have limitations. If estimates of the bridge

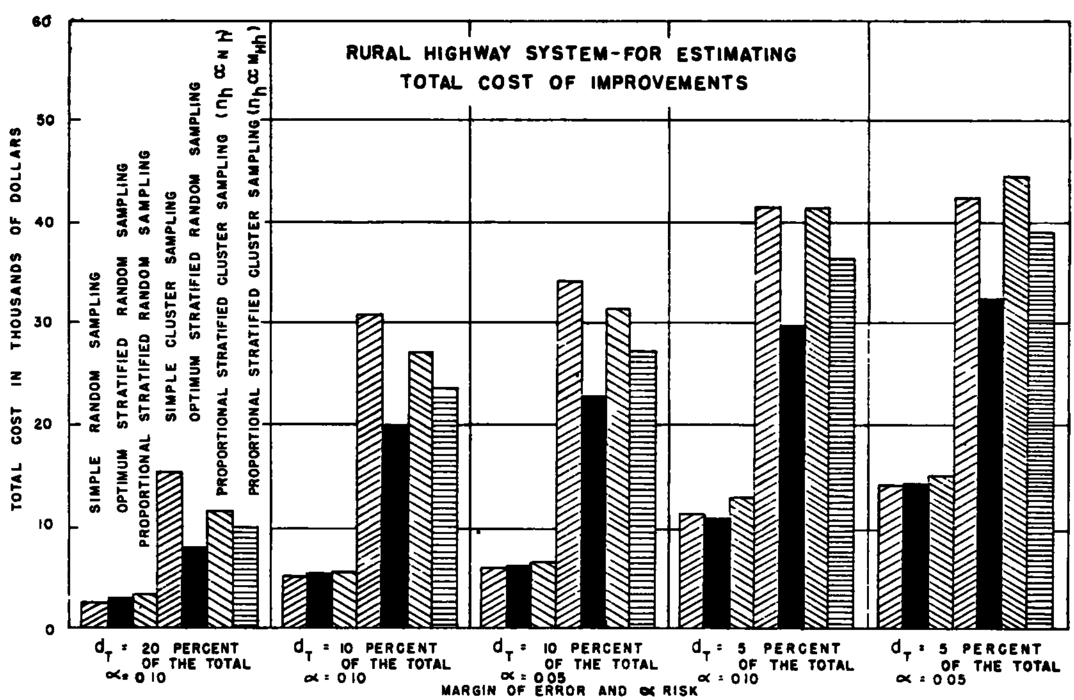

Figure 6. Estimated total cost of sample surveys for composite populations using simple and stratified random sampling and simple and stratified cluster sampling (Michigan data).

and railroad crossing improvement costs are desired, much larger sample sizes are required for comparable accuracy than are required for the separate populations. The decision to use this population should be made only after careful consideration has been given to the information desired from the sample and the relative costs of sampling from separate and composite populations.

Estimated total sample survey costs as found in this study for simple random, optimum stratified random, proportional stratified random, simple cluster, optimum stratified cluster, and proportional stratified cluster sampling are compared in Figures 4 and 5 for the separate populations. Optimum stratified random sampling gave minimum total survey costs for the Michigan and Minnesota data. The difference in survey costs for simple random and proportional stratified random sampling, however, was small and anyone of these methods could have been used without a great difference in cost. From the standpoint of ease of understanding the method and ease of selecting the sample, simple random sampling is, without doubt, the most practical. All forms of cluster sampling required much larger expenditures for comparable orders of accuracy.

The Minnesota data required higher expenditures than the Michigan data for comparable orders of accuracy. When d_T = 20 percent of the total and $\alpha = 0.10$, the total cost for simple random and stratified random sampling of the Michigan data was approximately $5,400 and of the Minnesota data was $7,800. These costs increased to $24,700 for the Michigan data and $28,700 for the Minnesota data when d_T = 5 percent of the total and $\alpha = 0.05$.

Figures 6 and 7 illustrate the total sample survey costs obtained for the various orders of accuracy and sampling methods using the composite populations for the Michigan and Minnesota data. For both groups of data simple random sampling generally gave minimum cost, although optimum stratified random sampling for the Michigan data gave a slightly smaller value when d_T = 5 percent of the total, $\alpha = 0.10$ and d_T = 5 percent of the total, $\alpha = 0.05$. However, only slight differences in total cost were

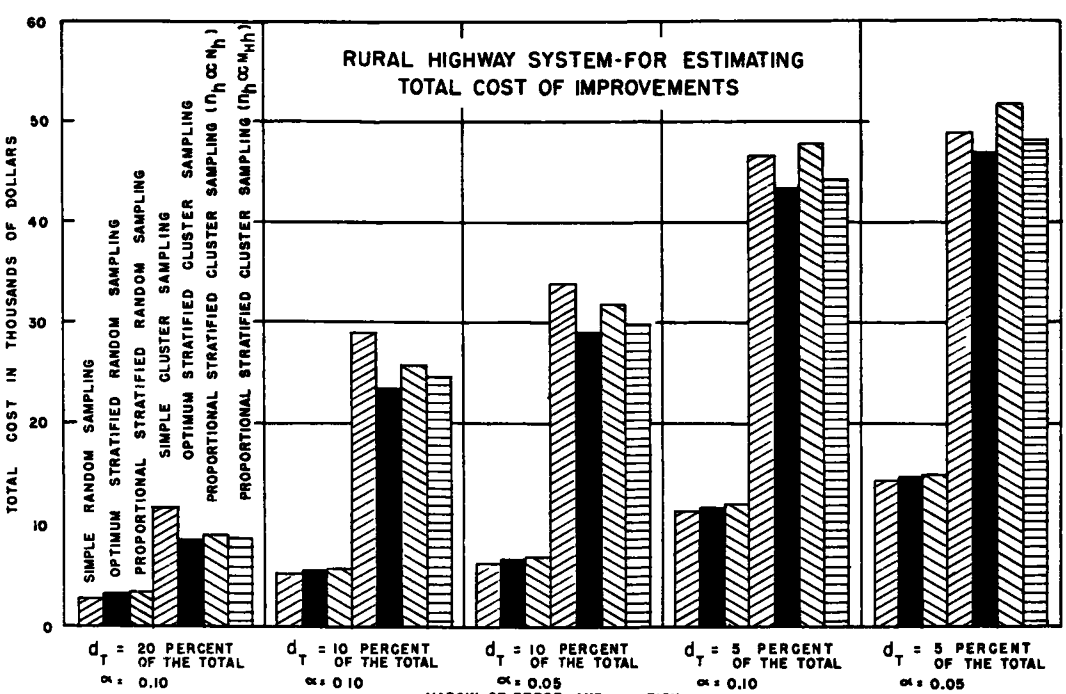

Figure 7. Estimated total cost of sample surveys for composite populations using simple and stratified random sampling and simple and stratified cluster sampling (Minnesota data).

apparent for simple random, optimum stratified random, and proportional stratified random sampling for the respective Michigan and Minnesota data. Any one of these three sampling methods could have been used without any substantial difference in expended funds.

All forms of cluster sampling exhibited much greater costs than simple and stratified random sampling. These large costs indicate the impracticality of the county as a sampling unit. A much smaller unit is needed to provide a large number of sampling units and moderate variance so that small sample sizes can give reliable estimates of needs.

Comparison of sample survey costs of the separate populations and the composite populations for similar orders of accuracy for total road system needs shows that the total survey cost using the separate populations required a 70 to 150 percent larger expenditure than that required when using the composite populations.

SUMMARY

The following are some of the results obtained in this study:

1. For a given population and specified order of accuracy, the sample size required to estimate various population characteristics varied considerably. Total cost of needs was the population characteristic investigated which required the smallest sample size.
2. Sample sizes required using porportional stratified random sampling were only slightly larger than sample sizes required for optimum stratified random sampling for the estimate of total cost of needs for a specified order of accuracy for the respective separate and composite highway populations. Sample sizes required for simple random sampling also were only moderately larger than those required for proportional stratified random sampling.
3. For a specified order of accuracy for all types of sampling studied, the composite populations required much larger sample sizes for an estimate of bridge and railroad crossing cost than those required for the respective separate populations.
4. The sample sizes required for the estimate of highway cost using the separate population for the various sampling methods studied were only slightly less or equal to the sample sizes required for estimating the total cost of all highway, bridge, and railroad crossing improvements for the composite population.
5. For a specified order of accuracy all forms of cluster sampling for the respective populations required substantially larger sampling rates for the estimate of total cost than those required for simple random or stratified random sampling.
6. For a specified order of accuracy using the separate populations little difference in sample survey costs was obtained between simple random, optimum stratified random, and proportional stratified random sampling for estimating total cost of highway needs. Similar observations were also noted for composite populations in regard to sample survey cost.
7. All forms of cluster sampling required substantially greater expenditures of funds than simple or stratified random sampling for comparable orders of accuracy. Maximum sample survey costs found in this study usually occurred for simple cluster sampling.

CONCLUSIONS

1. Sample survey techniques were found feasible for the estimation of the costs of improvements on rural two-lane and urban county primary highways in Michigan and Minnesota. For an estimate of similar information magnitudes of sample size similar to those found in this study may be expected for two-lane county primary highways in other states, especially those which have a well-developed county primary road system which is used for agricultural purposes.
2. For the estimate of total cost of all highway, bridge, and railroad crossing improvements for each order of accuracy, the sample survey cost of the composite sample was less than the sum of the sample survey costs of the separate highway, bridge, and railroad crossing population samples. The information which can be obtained from the composite sample is more limited, however, than the information which can be obtained from the separate samples.

3. Sample survey techniques were not feasible for the estimation of the total cost of needs for multi-lane highways such as rural and urban expressways and freeways on the county primary road system because of their relative rare occurrence.

4. The county is too large a sampling unit for the estimation of total cost of needs on the county primary road systems. Perhaps townships would be a more feasible sampling unit, especially for the estimation of needs on the local or township road system. The use of area sampling (1) may also be applicable to this problem.

REFERENCES

1. "Application of Probability Area Sampling to Farm Surveys." Agriculture Handbook No. 67, U.S. Dept. of Agriculture, Washington, D.C. (May 1954).
2. Branham, A.K., Covault, D.O., and Michael, H.L., "Progress Report on State Highway Needs in Indiana." Proc. of the 42nd Annual Road School, Purdue University (April 1956).
3. Burr, Irving W., "Engineering Statistics and Quality Control." New York, McGraw-Hill Book Co., Inc. (1953).
4. Cochran, William G., "Sampling Techniques." New York, John Wiley and Sons, Inc., (1953).
5. "Codes and Coding Instructions for Michigan Highway Needs Study." Michigan Legislative Highway Study Committee (1955).
6. "County-Municipal-State Aid Needs Study Procedure." City Engineers' Needs Study Manual, Minnesota Department of Highways State Aid Division (Unp.).
7. Covault, D.O., "Estimation of Highway Needs for County Primary Road Systems in Michigan and Minnesota by Sample Survey Methods." Ph.D. Thesis, Purdue University (Jan. 1959).
8. Covault, D.O., "Indiana's Highway Needs Study." HRB Proc., Vol. 37 (1958).
9. Covault, D.O., Michael, H.L., and Branham, A.K., "Summary Report on State Highway Needs in Indiana." Proc. of the 43rd Annual Road School, Purdue University (April 1957).
10. "Engineering Procedures and Instructions for Determining County and Township Road Needs." Minnesota Highway Study Commission (1953).
11. "Engineering Procedures and Instructions for Determining County Road Needs." Michigan Legislative Highway Study Committee (1955).
12. "Engineering Procedures and Instructions for Determining Municipal Street Needs." Michigan Legislative Highway Study Committee (1955).
13. "Engineering Procedures and Instructions for Determining State Highway Needs." Minnesota Highway Study Commission (1953).
14. "Engineering Procedures and Instructions for Determining Street Needs in Cities under 5,000 Population." Minnesota Highway Study Commission (1953).
15. Hansen, Morris H., Hurwitz, William N., and Madow, William G., "Sample Survey Methods and Theory." Vol. I and II, New York, John Wiley and Sons, Inc. (1953).
16. Jebe, Emil H., "Estimation of Sub-Sampling Designs Employing the County as a Primary Sampling Unit." American Statistical Association Journal, Vol. 47 (1952).
17. Madow, Lilian H., "On the Use of the County as the Primary Sampling Unit for State Estimates." American Statistical Association Journal, Vol. 45 (1950).
18. Medley, R.D., "Evaluation of Adequate Maintenance Cost in Kentucky from a Probability Sample." Kentucky State Highway Department (1955).
19. Michigan State Highway Department, County Highway Transportation Maps.
20. Minnesota State Highway Department, County Highway Transportation Maps.
21. Minnesota State Highway Department, County Traffic Flow Maps.
22. Ostle, Bernard, "Statistics in Research." The Iowa State College Press (1954).
23. Sukhatme, Pandurang V., "Sampling Theory of Surveys with Applications." The Iowa State College Press (1954).