Field Data Collection and Sampling Procedures for Measuring Regional Vehicle Classification and Occupancy

Robert A. Ferlis, Peat, Marwick, Mitchell and Company, Washington, D.C.

Small-scale field survey procedures that can be used to measure vehicle classification and occupancy at predetermined levels of statistical precision are described. Potential uses of these data to support activities related to transportation system management, air quality, energy efficiency, validation of urban transportation planning models, and trend analysis are identified. Statistical sampling plans, including plans that use the results of parallel traffic-monitoring programs, are presented to enable local planning agencies to design cost-effective field surveys. Representative values for key sampling parameters based on previous surveys, which are needed to use these sampling techniques, are provided. These procedures are then applied to a problem of survey design that demonstrates that regional occupancy can be precisely measured through the use of a small field survey.

The growing need to get greater service and efficiency from the highway transportation system is leading urban transportation planning agencies to make a greater effort to monitor travel trends and to measure the impacts of policies and programs at the regional level. Vehicle classification and vehicle occupancy are increasingly being recognized as two key elements of an effective program to monitor urban travel (1).

Planning agencies and highway departments have typically conducted field surveys to measure vehicle classification and occupancy only on an occasional basis and to satisfy very specific needs. These studies have generally been extremely limited in scope because of the high personnel costs involved. Locations for data collection have usually been selected to address a specific need for information at one point or, more rarely, have been judgmentally selected as representative of particular geographic areas or highway types. These limited studies do not provide the kind of information that permits the development of valid regional estimates of vehicle classification and occupancy.

This paper presents a series of simple field survey procedures that can provide statistically valid estimates of vehicle classification and occupancy and derived travel measures at prespecified levels of precision. These survey designs are based on the random selection of locations and times at which data collection is performed. The specific objectives of this paper are to

1. Suggest possible uses for four regional travel-monitoring measures that can be estimated from field data: percentage of truck travel (TR), average occupancy of passenger vehicles (OCC), truck kilometers of travel (TRKT), and person kilometers of travel (PKT);
2. Present statistically sound sampling procedures to guide data collection in the field so that these measures can be economically estimated at preselected levels of precision; and
3. Illustrate how these procedures can be applied to provide useful travel-monitoring information at a relatively low cost.

DESIGN OF TRAFFIC-BASED MONITORING PROGRAM

Although a comprehensive program to monitor urban travel should encompass a wide range of measures at both the regional and local levels, this discussion is directed only toward regional measures that can be estimated on the basis of limited field surveys. These field surveys will therefore complement the relatively more extensive traffic-counting programs that agencies typically conduct with the aid of mechanical counters. Before a field survey program can be implemented, local agencies should identify their basic data needs and translate these needs into specific survey objectives.

Identifying Needs for Monitoring Data

Data on vehicle classification and occupancy and on truck and person travel can potentially be applied to five major categories of need: (a) evaluating the effectiveness of transportation system management (TSM) actions, (b) assessing changes in air quality indices in relation to travel, (c) monitoring the energy efficiency of travel, (d) validating urban transportation planning models, and (e) monitoring general trends in transportation characteristics. The following table gives a summary of the extent to which each monitoring measure is applicable to these five areas (XX = very useful and X = moderately useful):

Application	Percentage Trucks	Average Occupancy	Truck Travel	Person Travel
TSM actions		XX		XX
Air quality		X	XX	X
Energy efficiency		X	XX	X
Model validation		X	XX	XX
Trend analysis	X	XX	XX	XX

Although many TSM actions are directed toward alleviating extremely localized problems, the overall effectiveness of the TSM program can be assessed by measuring regional vehicle occupancy and person travel on an annual basis and by using this information to determine progress toward achieving more efficient patterns of vehicle use. These annual measurements must be very precise if small impacts are to be reliably discerned.

Current patterns and annual trends in transportation-related air quality can also be assessed by using traffic-based measures. The magnitude and distribution of travel by vehicle class can be combined with emissions factors for representative vehicle classes to provide a quantitative estimate of regional transportation-related air quality on a periodic basis. An unusually detailed scheme of vehicle classification that includes breakdowns by size of vehicle and engine type (gasoline versus diesel) may be needed to provide accurate estimates of air quality. In addition, supplementary data, such as the number of cold starts, are required. By using these relatively simple procedures for estimating air quality, local agencies can anticipate future problems with air quality that may result from increasing travel and assess the efficacy of possible actions to reduce travel. Parallel measures of vehicle occupancy and person travel are mainly useful for computing air quality indices weighted by person travel rather than simply by vehicle travel.

Similarly, periodic estimates of regional travel by type of vehicle can support local studies of the energy efficiency of transportation and of changes over time. Again, the usefulness of travel data for this purpose will depend on the availability of factors that can convert travel by vehicle type into regional estimates of energy consumption. Occupancy and person-travel estimates will again be useful primarily for computing indicators of energy consumption based on person rather than vehicle movements.

Traffic-based monitoring data can also be used to validate the continuing adequacy of transportation planning models at the regional level. Major discrepancies between current-year model replications and monitoring estimates of total vehicle, truck, and person travel could indicate a need for updated input data or calibration parameters.

In addition to the specific uses cited above, traffic-based monitoring data can also serve as an important element of the urban transportation surveillance program. By periodically examining average vehicle occupancy, percentage of trucks, travel by vehicle class, and person travel, local agencies can identify emerging trends in transportation use that can be addressed through policy formation and planning at the regional, state, and federal levels. These might include the impacts of increases in fuel prices, restrictions on fuel consumption, weight restrictions on commercial vehicles, changes in speed limits, and the response of the motorist to measures designed to reduce unnecessary travel in urban areas.

Defining Survey Objectives and Approach

After carefully reviewing their needs for traffic-based monitoring data, agencies should translate these needs into a series of specific survey objectives. This process consists of three major steps:

1. Select the most important measure or measures. These measures will control the sampling plan.
2. Define the survey population. The population should be defined in terms of geographic scope, type of highway, time of day, day of week, and seasonal coverage.
3. Specify the desired level of precision for each of these measures. Level of precision can be defined here as the combination of two parameters: tolerance level D, which represents the acceptable difference between the estimated measure and the true value, and level of significance α, which represents the probability that the sample estimate will fall outside this range.

For example, an agency that desires to monitor average vehicle occupancy on a yearly basis could decide to estimate this measure for (a) all streets and highways within the standard metropolitan statistical area (SMSA), (b) the period from 7:00 a.m. to 7:00 p.m. on nonholiday weekdays only, and (c) the months of June through August. The agency could also specify a desired tolerance level of ±2 percent with an associated 5 percent level of significance. That is, the chance that the estimate of average occupancy will differ from the true value by more than 2 percent is only 1 in 20.

The technique for sampling and data collection can then be specified. The sampling procedures described in this paper were designed under four main assumptions:

1. All street and highway sections have an equal probability of being selected in the sample. This can be achieved by dividing the network into a set of "links" of uniform length for sampling purposes.
2. All days within the period of coverage have an equal probability of being selected in the sample. If the chance of selecting a holiday is felt to bias the sample, holidays should be explicitly excluded from the original population of days.
3. The sampling units of "link days" are randomly selected by using random numbers or similar techniques. The initial selection is not altered in an attempt to provide a more representative sample or to simplify the process of data collection.
4. If data collection is not conducted continuously throughout the selected link days, a systematic short-count technique is used to prevent bias and to maximize precision.

Since vehicle classification, vehicle occupancy, and traffic volume typically vary substantially during the day, a systematic short-count procedure in which observations are made for a fixed interval during each hour of the day offers the best potential for producing relatively accurate daily estimates while conserving personnel resources. In some urban situations, a single crew could conceivably collect data at several locations simultaneously by shifting from one site to the next on a fixed schedule so that each site is visited once an hour. A more likely approach, however, is for one person to systematically monitor one location by observing only a sample of lanes at any one time. For example, each lane on a three-lane, high-volume arterial could be observed for the same 15-min interval in each hour, and the fourth interval could be a rest period. This approach would produce a relatively accurate estimate of vehicle classification and occupancy while eliminating the need for a two-person crew.

SAMPLE SIZE FOR REGIONAL SURVEYS

The survey objectives and approach can now be translated into a sampling plan by computing the sample size of the link days needed to achieve the desired level of precision for each of the selected measures. As the formulas for sample size will show, the sample size is a function of level of precision and estimated sample variance. If an agency cannot accurately predict the sample variance that will be computed from the survey results, the estimate of the measure will still be valid but the level of precision may be either higher or lower than desired.

If more than one measure is felt to be important, the measure that requires the larger sample size controls the sampling plan. Although it is useful to specify concrete survey objectives before designing the sampling plan, an agency should also review the implications of the resulting plan in terms of sample size and survey cost. In some cases, an agency may choose to compromise its objectives to save costs.

As discussed earlier, these sampling formulas are designed for simple regional surveys that are appropriate for most traffic-monitoring needs. Agencies that wish to use relatively more complex stratified sampling plans, either to provide estimates for particular subsamples of the population or to reduce overall survey costs, will have to use more complicated sample-size formulas (2).

Percentage of Trucks

The first traffic-based monitoring measure that will be considered is percentage of trucks or, more conveniently, proportion of trucks, which represents a single regional ratio of truck travel to total vehicle travel. An agency should carefully define "truck" to meet the particular needs of the survey. The following procedure is directed toward a simple definition of the truck as a vehicle with double tires on at least one axle. The number of link days of data collection that would be needed to reliably estimate the proportion of trucks within a tolerance DTR at an α level of significance can be computed as

$$N = (Z^2 * ST^2)/DTR^2 \quad (1)$$

where

- Z = normal variate for the $(1 - \alpha)$ level of confidence (two-tailed test),
- ST = composite standard deviation of the proportion of trucks, and
- DTR = acceptable difference between the estimated proportion of trucks and the true value.

This formula assumes that the ratio of sampled link days to total possible link days is so small that the finite population correction factor need not be considered. The composite standard deviation depends on the design objectives and the approach of the survey. It can be estimated as

$$ST = (STL^2 + STS^2 + STW^2)^{1/2} \quad (2)$$

where

- STL = standard deviation of the proportion of trucks across link days within a season,
- STS = standard deviation of the proportion of trucks across seasons, and
- STW = standard deviation of the proportion of trucks across time periods during a day as a result of short counts.

The term for seasonal variation (STS) should only be included if the survey is intended to measure truck travel over an entire year and data collection will therefore extend throughout that period. If, on the other hand, data collection will be concentrated in a single season or in a period for which truck travel is not expected to vary significantly, this component should not be included in the computation of the composite standard deviation.

In the same way, the term for within-day variation (STW) should only be included if a short-count approach to data collection will be used. If all vehicles that pass a station during the selected link day will be observed by the data collectors, this term should not be included in the computation of the composite standard deviation.

Before local agencies can compute the minimum sample size, they must first estimate the composite standard deviation. The results of previous surveys should be used if they are available. Agencies that have not conducted such surveys should judgmentally estimate these variance terms. Table 1 gives representative values derived from surveys conducted in several urban areas. The seasonal standard deviation may be somewhat greater than what can be expected in most cities. The within-day standard deviation represents a systematic 25 percent sample of time periods within the day. These "default" values should be replaced by measured values after the first year. The key STL parameter should be replaced in all cases, but other terms should be replaced only if they can be estimated from either the survey results or special studies.

The relation among composite standard deviation, tolerance level, and resulting sample size at an assumed 95 percent confidence level is shown in Figure 1. If the composite standard deviation is estimated to be 0.045, a sample of approximately 20 link days will be required to estimate the proportion of trucks within ±0.02. If a considerably more precise estimate of the percentage of trucks is needed, a more complicated stratified sampling plan can be used instead of a simple survey to reduce survey costs (2).

Average Occupancy

A regional measure of average occupancy will probably be of considerably more interest to most agencies than the percentage (or proportion) of trucks. The sample size of link days needed to estimate average occupancy within a tolerance DOCC and a level of significance α can be computed as

$$N = (Z^2 * SO^2)/DOCC^2 \quad (3)$$

where DOCC = acceptable difference between the estimated average occupancy and the true value and SO = composite standard deviation of average occupancy.

The composite standard deviation will again depend on the design objectives and the approach of the survey. The composite standard deviation can be estimated as

$$SO = (SOL^2 + SOS^2 + SOW^2)^{1/2} \quad (4)$$

where

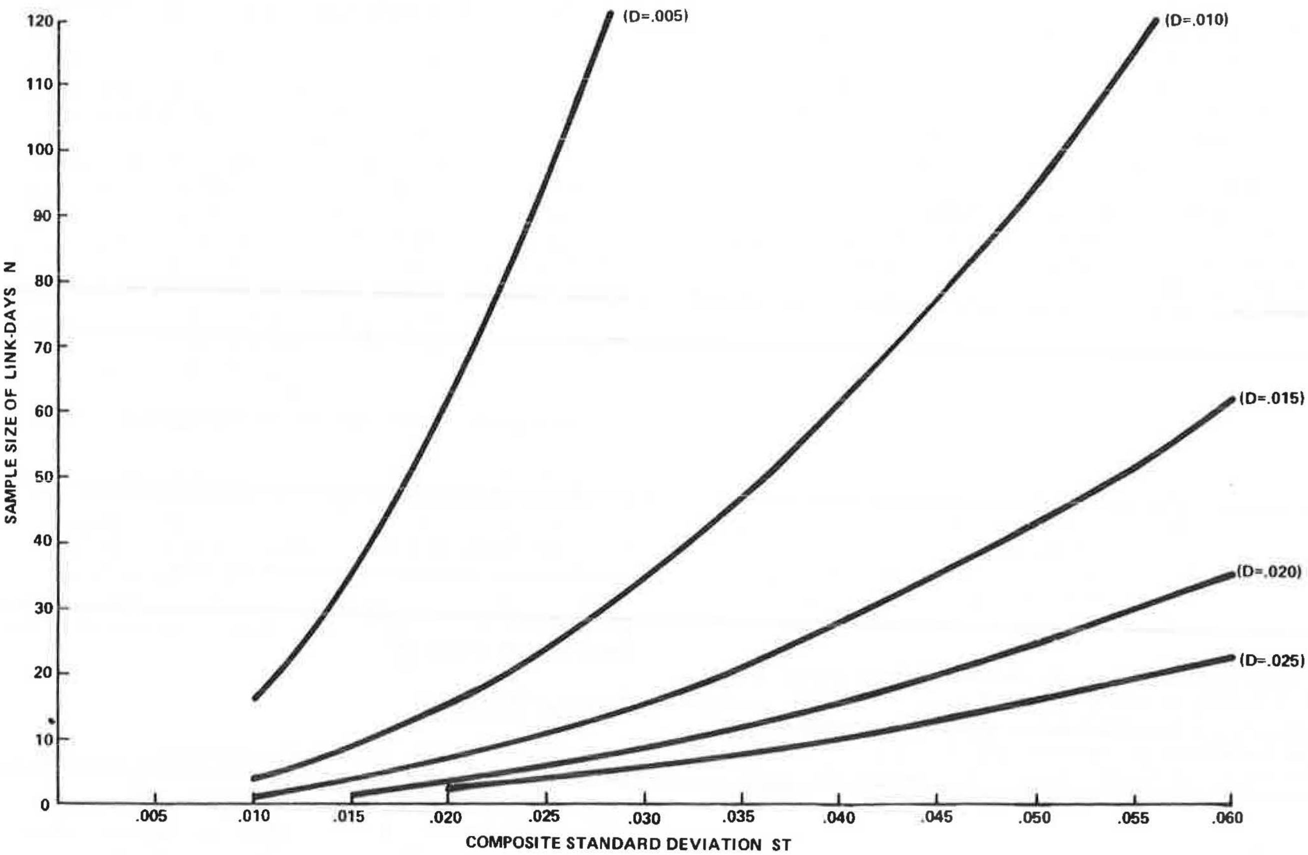
- SOL = standard deviation of average occupancy across link days within a season,
- SOS = standard deviation of average occupancy across seasons, and
- SOW = standard deviation of average occupancy across time periods during a day as a result of short counts.

As in the case of the proportion of trucks, these terms should only be included if they are appropriate.

Table 1. Standard deviations for percentage trucks and average occupancy.

Measure	Source of Variation	Symbol	Value	City
Percentage trucks	Location and day	STL	0.021-0.054	Washington, DC; Kileen-Temple, TX
	Season	STS	0.014	Various Ohio cities
	Within day	STW	0.009	Washington, DC
Average occupancy	Location and day	SOL	0.057-0.069	Kileen-Temple, TX; Seattle
	Season	SOS	0.011-0.019	Minneapolis, Albany
	Within day	SOW	0.012-0.022	Seattle; Washington, DC

Figure 1. Sample sizes for estimating regional percentage trucks within tolerance D at 95 percent confidence.



The results of previous local surveys should ideally be used to estimate these terms. Representative ranges of the standard deviation terms for average occupancy are also given in Table 1. Again, these values should be replaced by actual variances after the initial survey is completed, both to assess the level of precision actually achieved and to compute sample size for future surveys of occupancy.

The relation among composite standard deviation, tolerance level, and sample size at the 95 percent confidence level for average occupancy is shown in Figure 2. An assumed composite standard deviation of 0.067 based on the default recommendations would provide an estimate of average occupancy within ±0.03 with a sample of 20 link days. Thus, if the survey results show that the estimate of the composite standard deviation is accurate and the estimated average occupancy is 1.45, the probability that the true regional occupancy is between 1.42 and 1.48 is 95 percent.

Truck Travel

Truck kilometers of travel (TRKT) can be developed directly from a series of field counts. But a more economical approach for urbanized areas, which measures total vehicle kilometers of travel through extensive mechanical traffic-counting programs, is to combine the estimate of vehicle kilometers with an estimate of the proportion of trucks derived from a relatively more limited program of field data collection. The number of link days of field data collection needed to estimate TRKT within a relative tolerance ETRKT with an α level of significance can be computed as

$$N = (Z^2 * ST^2)/(ETRKT^2 - EVKT^2) * TR^2 \qquad (5)$$

where

ETRKT = acceptable relative error between the estimated TRKT and the true value,
EVKT = computed relative error between the estimated vehicle kilometers of travel and the true value based on the recent survey, and
TR = estimated regional proportion of truck travel.

The definitions of Z and ST are the same as before. In contrast with previous sample-size formulas, the tolerance levels are expressed in relative rather than in absolute terms—as proportions of total travel. The precision of the estimated TRKT can approach but never equal the precision of the estimated vehicle kilometers of travel from the recent study.

Person Travel

Person travel in person kilometers can also be estimated by combining the results of a limited field survey with a recent estimate of vehicle kilometers. Since the estimate of person kilometers of travel incorporates the effects of three separate variables (i.e., average occupancy, proportion of automobile travel, and total vehicle travel), a relatively more complex formula is required. The number of link days of data collection needed to estimate person kilometers of travel within a relative tolerance EPKT with an α level of significance can be computed as

$$N = [Z^2 * (PV^2 * SO^2 + OCC * ST^2)]/(EPKT^2 - EVKT^2) * (OCC * PV)^2 \qquad (6)$$

where

PV = estimated proportion of passenger-vehicle travel (assumed to be equal to 1 - TR),
OCC = estimated average vehicle occupancy for passenger vehicles, and
EPKT = acceptable relative error between the estimated person kilometers of travel and the true value.

This formula assumes that the proportion of vehicles that are neither passenger vehicles nor trucks is insignificant. Therefore, the composite standard deviation for the proportion of trucks is used to estimate the composite standard deviation of the proportion of passenger vehicles.

APPLICATIONS OF THE SAMPLING FORMULAS

Consider a planning agency that wants to conduct a limited field survey to meet the following specific objectives:

1. Estimate regional average occupancy during the fall within a tolerance level of ±0.02 with a 95 percent level of confidence. This measurement will be repeated in subsequent years to assess the progress of the comprehensive TSM program.
2. Estimate regional person travel produced by private passenger vehicles within a relative tolerance level of ±5 percent with a 95 percent level of confidence. This measurement is desired for long-range trend analysis.
3. Estimate regional truck travel to validate certain aspects of the results of the urban transportation planning model, which will be used to assess air quality. No specific level of precision is required.

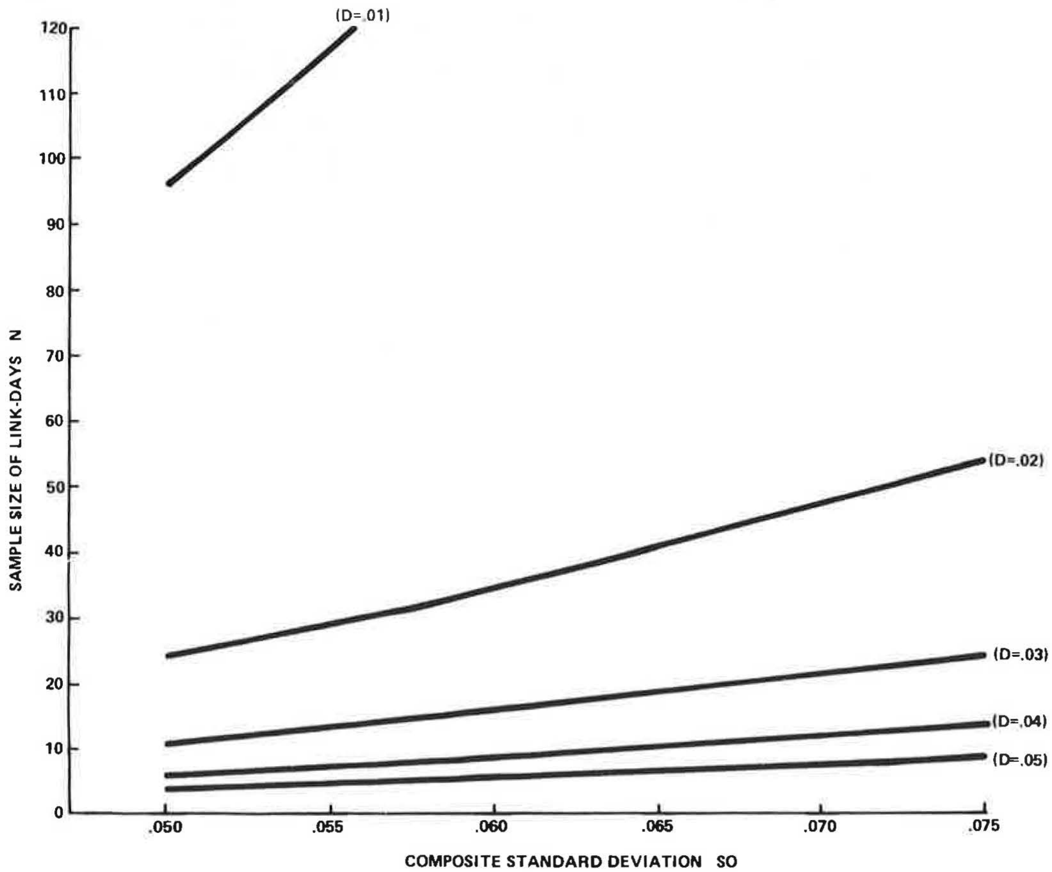
Assume that the survey will encompass all streets and highways in the region. The period of concern is a three-month period in the fall for which only travel between 7:00 a.m. and 7:00 p.m. on nonholiday weekdays is considered. A short-count procedure in which the field crew systematically monitors each lane for precisely 20 min/h is used for most data collection stations.

Before the sample sizes required for each of the measures can be computed, the composite standard deviations must be estimated. Assume that the agency previously conducted a survey of occupancy and estimates SOL as 0.062. Since the data collection effort will be performed in only one season, SOS can be assumed to be zero. Finally, since the short-count data collection plan involves a 33 percent coverage rather than the 25 percent coverage used to compute the default values summarized in Table 1, the agency selects an SOW value of 0.010. The composite standard deviation can then be computed as $SO = (0.062^2 + 0.010^2)^{1/2} = 0.063$.

Assuming that no previous vehicle classification studies were conducted, the composite standard deviation for the proportion of trucks can be estimated judgmentally from the default values as $ST = (0.045^2 + 0.008^2)^{1/2} = 0.046$.

Finally, the agency estimates the remaining parameters for sampling purposes. These estimates are derived from the survey objectives or the results of

Figure 2. Sample sizes for estimating regional average occupancy within tolerance D at 95 percent confidence.



previous surveys or judgmentally, as appropriate:

Parameter	Estimate	Process or Source
Z	1.96	For 5 percent level of significance
OCC	1.40	Judgment
TR	0.06	Judgment
PV	0.94	1 - TR
EVKT	0.04	Recent study
DOCC	0.02	Survey objectives
EPKT	0.05	Survey objectives

By substituting into Equation 3, the sample size of link days required to measure average occupancy can be computed as

$$N = (1.96^2 * 0.063^2)/0.02^2 = 39 \quad (7)$$

Similarly, Equation 6 can be used to compute the sample size of link days of field data collection needed to factor the estimate of vehicle kilometers of travel from the traffic-counting program:

$$N = [1.96^2 * (0.94^2 * 0.063^2 + 1.40^2 * 0.046^2)]/(0.05^2 - 0.04^2) * (1.40 * 0.94)^2 = 19 \quad (8)$$

The sample size required to estimate average occupancy is greater and is therefore used for the survey. The level of precision that can be expected for the estimate of truck travel can now be computed by rearranging the terms in Equation 5 as

$$ETRKT = \{[(Z^2 * ST^2)/(N * TR^2)] + EVKT^2\}^{1/2} = \{[(1.96^2 * 0.046^2)/(39 * 0.06^2)] + 0.04^2\}^{1/2} = 0.25 \quad (9)$$

The estimate of TRKT can be expected to have a relative error of almost 25 percent of the true value. If the objectives of the survey design had included accurately estimating this measure, either the sample size would have to be substantially increased or a relatively more efficient stratified sampling plan would have to be used.

After the survey is completed, the agency can compute many of the variance and other parameters by using the survey results. The sample-size formulas should then be rearranged and applied to compute the actual level of precision achieved. Depending on the accuracy of the various parameter estimates, the actual precision may be higher or lower than anticipated. Parameter values should therefore be conservatively estimated in cases where failure to achieve minimum levels of precision can seriously compromise the utility of the survey.

CONCLUSIONS

This paper identifies potential uses for traffic-based monitoring measures, presents cost-effective sampling and data collection procedures to estimate these measures, and demonstrates how these procedures may be applied. The following conclusions can be made:

1. Traffic-based regional measures of vehicle classification and occupancy, truck travel, and person travel can support a variety of possible uses that relate to TSM actions, air quality, energy efficiency, model validation, and trend analysis.
2. Simple regional field surveys can be designed to estimate these measures at predetermined levels of statistical precision.
3. These field surveys can complement existing or anticipated programs of vehicle counting and travel estimation.
4. Systematic short-count data collection techniques can be used to substantially reduce crew size without seriously reducing the accuracy of daily summary data.
5. Average vehicle occupancy and person travel can be estimated at a high level of precision with a small number of field surveys.
6. If percentage of trucks and truck travel are used, a relatively larger sample of field surveys is required to produce highly precise estimates.

ACKNOWLEDGMENT

This paper is based on the results of research sponsored by the Federal Highway Administration, U.S. Department of Transportation.

REFERENCES

1. Transportation Research Board. Proposed Urban Transportation Data Reporting Requirements for States and Metropolitan Planning Organizations. Federal Highway Administration, U.S. Department of Transportation, 1976.
2. Peat, Marwick, Mitchell and Company. Guide for Estimating Urban Vehicle Classification and Occupancy. Federal Highway Administration, U.S. Department of Transportation, Oct. 1978.

Publication of this paper sponsored by Committee on Transportation Information Systems and Data Requirements.