A Guide to Customized Sampling Plans for National Transit Database Reporting

Xuehao Chu, University of South Florida
Ike Ubaka, Florida Department of Transportation

Abstract

For estimating the system total unlinked passenger trips and passenger miles of a fixed-route bus system for the National Transit Database (NTD), the sampling plans approved by the Federal Transit Administration (FTA) may either oversample or do not yield FTA's required confidence and precision levels for the specific conditions of a transit agency. This guide helps transit agencies avoid these problems by developing sampling plans customized to their fixed-route bus services. Detailed steps are provided to calculate the statistical variation in passenger miles and unlinked passenger trips and the correlation between them. The sampling plans that transit agencies develop using this guide maintain the main features of the original FTA plans. More important, transit agencies can easily use this guide to develop customized sampling plans in an Excel environment using three existing pieces of information: an NTD sample that they have collected with their current FTA approved sampling plan; annual days of service; and daily total number of one-way bus trips.

This guide complements FTA Circular C2710.1A in that transit agencies first select one of their own sampling plans that best meets their staffing needs, and then follow the procedures in the circular on sampling and collecting field data. Transit agencies should be able to use this guide and the related calculations as an approval of a qualified statistician as required by the current NTD reporting manuals.

Introduction

As part of its National Transit Database (NTD) reporting guidelines for fixed-route bus services, the Federal Transit Administration (FTA) provides a set of sampling plans for estimating unlinked passenger trips (UPT) and passenger miles (PM). These sampling plans are published in FTA Circular C2710.1A (UMTA 1988). The plans vary in the daily number of one-way bus trips sampled and the frequency of sampling. For example, the most popular plan requires sampling three one-way bus trips every second day with an annual sample of 549 trips. FTA requires that any sampling technique used for NTD reporting meet a confidence level of 95 percent and a precision level of ±10 percent. However, even FTA's own sampling plans do not always yield these levels of confidence and precision (Chu 2004). Furthermore, FTA's sampling plans may not minimize the sample size for the specific conditions of an individual transit agency (Chu 2004).

FTA also gives transit agencies the option of using alternative sampling techniques as long as they are developed and certified by a qualified statistician (FTA 2003). Many transit agencies, however, do not exercise this option and still use FTA-approved plans to estimate either their UPTs, PMs, or both, either because they are not aware of the problems in using the FTA plans or because they are reluctant to take the necessary steps to developing their own alternative sampling plans.

This guide is designed for transit agencies to customize sampling plans for their fixed-route bus services. By minimizing the sample size while meeting FTA's confidence and precision levels, the customized sampling plans represent more reliable and cost-effective alternatives to the FTA sampling plans. This guide is designed for transit agencies to develop these customized sampling plans with ease. Only three items of existing information are required: an NTD sample already collected using an FTA-approved sampling plan, annual days of service, and daily total number of one-way bus trips. Transit agencies only need to enter these data items in a prespecified format in Excel templates provided with this guide. This guide is also designed for transit agencies to transition from using the FTA sampling plans to using the customized sampling plans with ease. These customized sampling plans retain the primary features of the original FTA sampling plans.

This guide complements "Sampling Procedures for Obtaining Fixed Route Bus Operating Data under the Section 15 Reporting System," FTA Circular C2710.1A. Essentially a transit agency replaces the table of sampling plans in Table II-1 in the circular with what it develops using this guide, selects of one of the new plans that

best meets its staffing needs, and then follows the procedures in the circular on sampling and collecting field data.

The second section provides background information, including the FTA sampling plans, how these plans are used in practice, problems with current practices, the objective of this guide, features of customized sampling plans, and benefits of customized sampling plans. The third section provides guidance on using an existing NTD sample to calculate its statistical characteristics. The fourth section describes the process of developing a new sample size table using an Excel template. The Appendix presents statistical formulas that have been built into the template. The Appendix is only for those who are interested in the related sampling theory. Reading the second section is not necessary, but would be helpful for following guidance given in the remainder of the guide.

Background

FTA Sampling Plans

Table 1 shows FTA sampling plans from Table II-1, FTA Circular C2710.1A (UMTA 1988), which are widely used because of two features:

1. They follow a simple framework. For any plan, the same number of one-way bus trips is sampled on each sample day, and the interval between the sample days is constant.

2. They are assumed to have universal applicability. Following any plan by any transit agency is said to guarantee FTA's required confidence and precision levels.

Table 1. FTA-Approved Sampling Plans

Plan Number	1	2	3	4	5	6
Frequency of sampling	Every day	Every 2nd day	Every 3rd day	Every 4th day	Every 5th day	Every 6th day
Daily sample	2	3	5	7	10	15
Annual sample	730	549	610	644	730	915

For later reference, a table like this one is referred to as a sample size table. Each sample size table contains six sampling plans. Each plan consists of a particular frequency of sampling, along with a corresponding daily sample size and annual sample size.

How Transit Agencies Currently Use FTA's Sampling Plans

Transit agencies' use of FTA sampling plans to estimate system total UPT or PM generally falls into the following categories:

- *Directly estimating unlinked passenger trips,* where there is no 100 percent count of system total UPT. In this case, the sample average unlinked passenger trips per one-way bus trip is multiplied by the total number of one-way bus trips operated to obtain the system total UPT.

- *Directly estimating passenger miles*, where there is no 100 percent count of system total UPT. In this case, system total PM is directly estimated by multiplying the sample average passenger miles per one-way bus trip by the system total one-way bus trips operated. This is the PM-based approach to estimating system total PM.

- *Indirectly estimating passenger miles*, where there is a 100 percent count of system total UPT. In this case, system total PM is indirectly estimated by multiplying the 100 percent count of system total UPT by the sample average passenger trip length (PTL). This is the PTL-based approach to estimating system total PM.

Research from this study also revealed the following practices:

- Some agencies estimate both UPT and PM and report the estimated figures to NTD, even though they have a 100 percent count of system total UPT.

- Some agencies have a 100 percent system total UPT and report it to NTD, but do not use it in estimating system total PM.

- Some agencies use an FTA sampling plan to estimate the sample average PTL because it was the plan used when a 100 percent system total UPT count was not available.

- Many agencies are not aware that customized sampling plans could be more cost effective.

Problems with Current Practices

Given the required confidence and precision levels, two related factors determine the sample sizes of a particular sample size table: a specific quantity (PM, PTL,

etc.) and its statistical variation across one-way bus trips. For estimating system total PM, for example, either PM or PTL may be used as the quantity, depending on whether a 100 percent count of system total UPT is available. There are two dimensions of statistical variation under the general framework of this guide: between-day variation and within-day variation. Within-day variation measures how much the chosen quantity varies across one-way bus trips on the same day, while between-day variation measures the variation in the chosen quantity across different days.

In principle, one should choose the quantity that varies the least when choices are available, and the statistical variation should reflect specific conditions of individual transit agencies. Development of the original FTA sampling plans violated both principles. They are based on PM rather than PTL, which typically varies much less than PM across one-way bus trips. Further, they are based on a single set of assumed variations in PM.

Violation of these principles by the original FTA sampling plans leads to three major problems in practice associated with the use of FTA's sampling plans.

1. Under many realistic conditions, FTA plans do not satisfy the required confidence and precision levels under the PM-based approach to estimating system total PM.

2. FTA sampling plans may result in oversampling under the PM-based approach to estimating system total PM. Agencies may attain FTA-required levels of confidence and precision using smaller samples. For example, in estimating system total PM, they may do so with fewer one-way bus trips under the same sampling frequency. If an agency sampled every second day, for example, it may only need to sample two, rather than three, one-way bus trips on each day.

3. FTA sampling plans are likely to result in oversampling under the PTL-based approach to estimating system total PM.

In addition to the sampling problems identified above, research for this guide also uncovered the following nonsampling problems.

•   When transit agencies have a 100 percent system total UPT count but choose to report an estimate to NTD, they are also most likely to report the count to entities such as their governing board, state transit assistance

programs, etc. As a result, ridership reported to these entities differs from NTD ridership reports.

- When a 100 percent count of system total UPT is available and is reported to the NTD, but is not used in estimating system total PM, inconsistencies exist among the 100 percent UPT count, the estimated PM, and the implied PTL. For example, the ratio of the estimated PM and a separately estimated PTL would not be necessarily close to the 100 percent count of system total UPT.

Objective of the Guide

This guide is designed for transit agencies to overcome these problems in their practice. Specifically, transit agencies can develop a sample size table customized to their own conditions in an Excel environment with an existing NTD sample and information on their annual days of service and daily total number of one-way bus trips.

Developing a new sample size table internally by a transit agency is one of several options for the agency to determine a customized sampling plan in any given year. The flow chart in Figure 1 outlines these options:

- Whether this is First Year (vs. Other Years) in which the agency is developing a customized sampling plan.

- If First Year, whether to determine the sampling plan using internal or external sources (e.g., consultant, CUTR, etc.).

- If external, whether a transit agency wants a consultant or CUTR to determine its new sample size table. (If CUTR, send the sample to *xchu@cutr.usf.edu* for a free service.)

- Either done internally or externally, the result is a new sample size table. Select one sampling plan from the new table, based on factors such as sample intervals, daily number of one-way bus trips on each sample day, and staff availability.

- If not First Year, whether to readjust the previously developed sample size table after a major change in services (Adjust vs. Don't Adjust). If the choice is to readjust the sample size table, the agency is in the same situation as if this is First Year.

- If no readjustment to the existing sample size table, whether to change the previously used sampling plan (New Plan vs. Old Plan).

- If a new plan is used, it would come from the previously developed sample size table.

This guide helps transit agencies when they choose the option where they want to develop a new sample size table internally, by themselves, as indicated by the bold arrow.

Features of Customized Sampling Plans

This guide is designed for transit agencies to develop customized sampling plans with ease. Information needed is: annual days of service; daily system total one-way bus trips; and an existing NTD sample. An agency that has not collected NTD data before would need to use one of the FTA-approved sampling plans for at least one year in order to use this guide for its customized sampling plans. Sampling theory is presented in the Appendix and has been built into an Excel environment. Transit

Figure 1. Flow Chart on Options for Determining a Sampling Plan
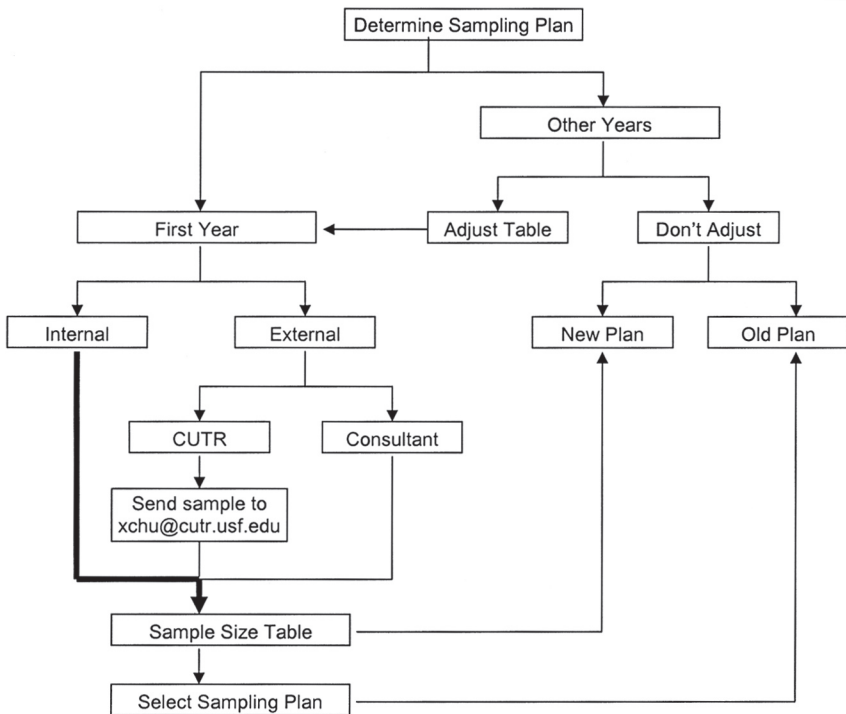
agencies only need to enter the required information in a prespecified format in Excel templates provided with this guide.

This guide is also designed for transit agencies to easily transition themselves from using FTA sampling plans to using the customized sampling plans. First, the customized sampling plans retain primary features of the original FTA sampling plans. They form a single sample size table such as Table 1, and they follow a simple framework, where the sample days have constant intervals and the same number of one-way bus trips is sampled on all sample days.

Second, steps following the development of customized sampling plans are identical to those following the FTA plans (Figure 2).

- Select a sampling plan from the new table.
- Sample one-way bus trips.
- Collect and record the data.
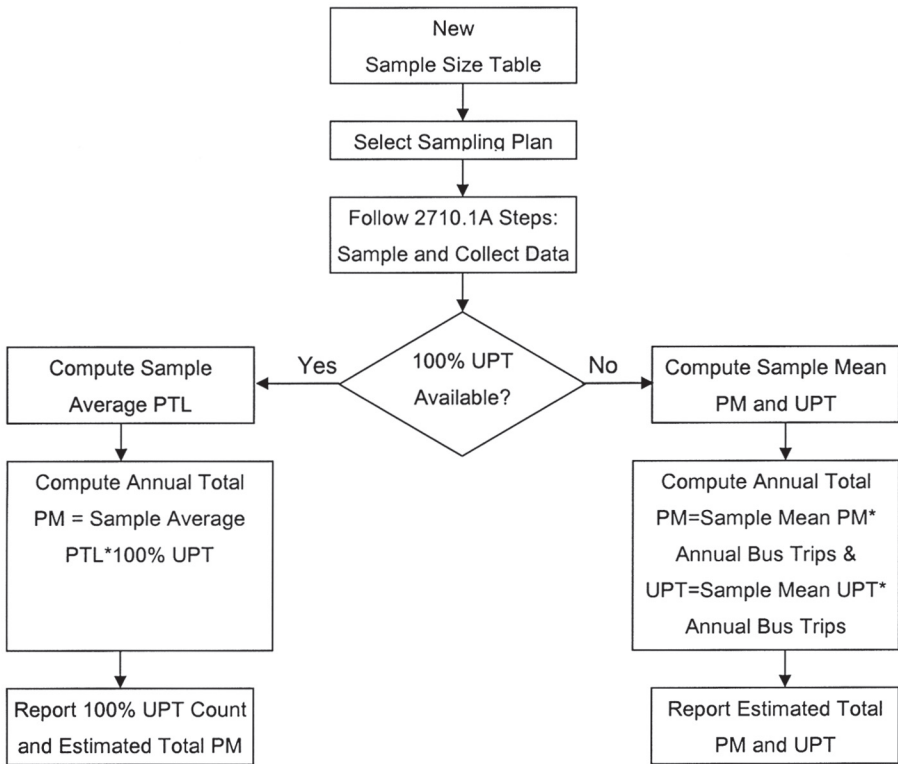- Estimate the annual system total ridership.

If the 100 percent system total UPT count is not available, FTA Circular C2701.1A provides guidance; the estimated system total PM and UPT can be reported to the NTD. If the 100 percent system total UPT count is available, FTA Circular C2701.1A provides guidance on sampling, collecting, and recording data. To estimate system total PM, the sample average PTL from the NTD sample and the 100 percent system total UPT count should be used. The 100 percent system total UPT count and estimated system total PM can be reported to the NTD.

Benefits of Customized Sampling Plans

Customized sampling plans, specific to agency conditions, may help avoid problems encountered with FTA plans, while still yielding FTA's required confidence and precision levels without oversampling. The exact benefit to a transit agency depends on whether the 100 percent count of system total UPT is available and how it is used.

If the 100 percent system total UPT count is available, the development of a customized sample size table will be based on the statistical variation in PTL across one-way bus trips. If the 100 percent system total UPT count is currently what is reported to the NTD, following any plan in the new sample size table should reduce sample size. If this count is not what is currently reported to the NTD, doing so is highly recommended, along with using the PTL-based approach to esti-

Figure 2. Flow Chart on Steps Following a New Sample Size Table

```
                    ┌─────────────────────┐
                    │        New          │
                    │  Sample Size Table  │
                    └─────────────────────┘
                              │
                    ┌─────────────────────┐
                    │ Select Sampling Plan│
                    └─────────────────────┘
                              │
                    ┌─────────────────────┐
                    │ Follow 2710.1A Steps:│
                    │ Sample and Collect Data│
                    └─────────────────────┘
                              │
                              ▼
      Yes          ◇ 100% UPT         No
  ┌─────────────┐◄─◇  Available? ◇──►┌──────────────────┐
  │Compute Sample│  ◇               │Compute Sample Mean│
  │ Average PTL  │                  │    PM and UPT     │
  └─────────────┘                   └──────────────────┘
        │                                   │
  ┌─────────────────┐            ┌────────────────────────┐
  │Compute Annual Total│         │  Compute Annual Total  │
  │PM = Sample Average│          │  PM=Sample Mean PM*    │
  │  PTL*100% UPT    │           │   Annual Bus Trips &   │
  │                 │            │UPT=Sample Mean UPT*    │
  │                 │            │   Annual Bus Trips     │
  └─────────────────┘           └────────────────────────┘
        │                                   │
  ┌──────────────────┐          ┌────────────────────┐
  │Report 100% UPT Count│       │Report Estimated Total│
  │and Estimated Total PM│      │   PM and UPT        │
  └──────────────────┘          └────────────────────┘
```

mating system total PM. Following any plan in a customized sample size table not only will reduce sample size, but also will lead to consistency in data reported.

If the 100-percent system total UPT count is not available, the development of the sample size table will be based on the statistical variation in PM across one-way bus trips. What this guide does in this case is to help transit agencies customize original FTA plans to their specific conditions. Any sampling plan in the new sample size table will yield FTA's required confidence and precision levels without oversampling.

In addition to avoiding the practical problems associated with the current use of FTA plans, using a customized sampling plan could also mean cost savings resulting from more efficient sampling (smaller sample size). These cost savings are likely to outweigh the minimal cost of developing customized sampling plans.

Relative Variances and Correlation

This section provides guidance on calculating five characteristics of an existing NTD sample in an Excel environment. These include: relative between-day and within-day variances of PM; relative between-day and within-day variances of UPT; and the correlation coefficient between PM and UPT. The variance of a quantity is a standard measure of its statistical variation. "Relative" indicates that the variances are normalized by the sample average of this quantity. Calculated values of these five characteristics become inputs to developing a new sample size table in the fourth section. This calculation is illustrated with a real NTD sample from the Hillsborough Area Regional Transit Authority (HARTline) in Tampa, Florida, for FY2002. This sample was collected with the second FTA approved plan, including a total of 549 one-way bus trips, with three one-way bus trips on each of 183 sample days. This sample and the related computation are shown in an Excel file, "FDOT_BC137_46_Relative Variances and Correlation.xls," available from the authors or at *http://www.dot.state.fl.us/research-center/Completed_PTO. htm* under project BC137-46.

Sample Data

Setting the sample data on PM and UPT in a spreadsheet format, as shown in Table 2, facilitates the computation. Some agencies keep their sample data at the stop level — boarding and alighting at individual stops and distances between consecutive stops for all sample trips identified by their sample dates. Some other agencies keep their sample data at the trip level — unlinked passenger trips and passenger miles for each sample trip identified by its sample date. In either case, one convenient way to convert sample data to the format in Table 2 is to use PivotTable in Excel. To do this, the sampled trips on each sample day need to be identified with the sample date and with the same set of identifiers, such as 1, 2, and 3, if three trips were sampled on every sample day.

Only the first 10 and last 10 sample days are shown in Table 2. The rows represent the individual sample days, and the columns represent the three one-way bus trips sampled on each sample day. The number 2.7 in cell B4, for example, indicates that two passengers in cell G4 boarded the first one-way bus trip sampled on the first sample day and collectively traveled a total of 2.7 miles. The two empty columns for either PM or UPT data are used later for computing the relative variances.

Table 2. Sample Data

	A	B	C	D	E	F	G	H	I	J	K
1	*Sample Days*			*PM*					*UPT*		
2		*Sample Trips*			*Daily Statistics*		*Sample Trips*			*Daily Statistics*	
3		*One*	*Two*	*Three*	*Mean*	*Variance*	*One*	*Two*	*Three*	*Mean*	*Variance*
4	1	2.7	13.1	84.1			2	4	31		
5	2	58.2	5.3	217.7			23	3	50		
6	3	48.1	36.4	95.9			14	10	18		
7	4	41.4	105.5	115.3			11	21	26		
8	5	160.2	154.2	179.0			50	23	42		
9	6	17.6	50.9	82.3			5	18	24		
10	7	70.2	58.5	201.6			29	13	40		
11	8	179.2	16.5	71.0			38	8	22		
12	9	95.6	55.5	7.7			26	15	9		
13	10	48.8	145.3	45.8			14	33	16		
177	174	94.4	6.7	4.6			20	2	2		
178	175	92.4	73.4	76.1			21	10	13		
179	176	100.2	29.6	208.2			19	4	49		
180	177	28.0	127.6	44.7			7	25	11		
181	178	79.9	49.0	131.3			13	12	8		
182	179	51.8	127.8	29.6			13	27	5		
183	180	168.4	68.3	90.3			41	12	27		
184	181	84.1	75.7	41.9			24	18	14		
185	182	153.6	62.1	39.0			25	18	9		
186	183	10.3	8.5	39.5			3	6	19		

Relative Variances

With the sample data ready, computing the relative variances is straightforward in a spreadsheet, taking advantage of existing Excel statistical functions (Table 3). The formulas involved are presented in the Appendix. Using PM data on the left of the table as an example, seven steps are involved.

1. Compute daily sample mean for individual sample days in column E, using function AVERAGE. For day 1, for example, AVERAGE(B4:D4) = 33.3 in cell E4.

2. Compute daily variances for individual sample days in column F, using function VARA. For day 1, for example, VARA(B4:D4) = 1962.5 in cell F4.

3. Compute the overall sample mean in cell E187, using function AVERAGE(E4: E186) = 79.9.

4. Compute the overall between-day variance in cell E188, using function VARA(E4:E186) = 1645.3.

5. Compute the overall within-day variance in cell F189, using function AVERAGE(F4:F186) = 4605.7.

6. Compute the relative between-day variance in cell E190 (=0.26) as the ratio of the overall between-day variance (E188) over the overall sample mean (E187) squared.

7. Compute the relative within-day variance in cell F191 (=0.72) as the ratio of the overall within-day variance (F189) over the overall sample mean (E187) squared.

The relative variances for UPT can be similarly computed in columns J and K. The relative between-day variance is 0.16 in cell J190, while the relative within-day variance is 0.46 in cell K191.

FTA-approved sampling plans are based on a relative between-day variance of 0.1 and a relative within-day variance of 1.0 in PM.

Correlations

Continue with this sample used in computing the relative variances, using the function CORREL to compute the correlation coefficient between PM and UPT within this sample. The result is that CORREL(B4:D186,G4:I186) = 0.77 in cell F192.

New Sample Size Table

This section provides guidance on developing the new sample size table in an Excel template, using the relative variances and correlation coefficient just computed.

100 Percent System Total UPT Count Unavailable

If the 100 percent system total UPT count is not available, use the Excel template called "FDOT_BC137_46_Without 100% UPT Count.xls" available from the authors or at *http://www.dot.state.fl.us/research-center/Completed_PTO.htm* under project BC137-46. The formulas involved are presented in the Appendix.

PM-Based Sample Size Table

To continue the current use of the PM-based approach to estimating system total PM, use the first sheet of the template called "Based on PM." This sheet is divided into INPUTS and RESULTS (see Table 4). This sheet is protected

Table 3. Computing Relative Variances in PM and UPT

	A	B	C	D	E	F	G	H	I	J	K
		PM						*UPT*			
	Sample Days	*Sample Trips*			*Daily Statistics*		*Sample Trips*			*Daily Statistics*	
		One	*Two*	*Three*	*Mean*	*Variance*	*One*	*Two*	*Three*	*Mean*	*Variance*
1											
2											
3											
4	1	2.7	13.1	84.1	33.3	1962.5	2	4	31	12.3	262.3
5	2	58.2	5.3	217.7	93.7	12225.4	23	3	50	25.3	556.3
6	3	48.1	36.4	95.9	60.1	993.7	14	10	18	14.0	16.0
7	4	41.4	105.5	115.3	87.4	1611.0	11	21	26	19.3	58.3
8	5	160.2	154.2	179.0	164.5	167.4	50	23	42	38.3	192.3
9	6	17.6	50.9	82.3	50.3	1046.8	5	18	24	15.7	94.3
10	7	70.2	58.5	201.6	110.1	6313.4	29	13	40	27.3	184.3
11	8	179.2	16.5	71.0	88.9	6858.1	38	8	22	22.7	225.3
12	9	95.6	55.5	7.7	52.9	1936.5	26	15	9	16.7	74.3
13	10	48.8	145.3	45.8	80.0	3203.6	14	33	16	21.0	109.0
177	174	94.4	6.7	4.6	35.2	2626.6	20	2	2	8.0	108.0
178	175	92.4	73.4	76.1	80.6	105.7	21	10	13	14.7	32.3
179	176	100.2	29.6	208.2	112.7	8091.1	19	4	49	24.0	525.0
180	177	28.0	127.6	44.7	66.8	2845.2	7	25	11	14.3	89.3
181	178	79.9	49.0	131.3	86.7	1728.3	13	12	8	11.0	7.0
182	179	51.8	127.8	29.6	69.7	2652.0	13	27	5	15.0	124.0
183	180	168.4	68.3	90.3	109.0	2767.3	41	12	27	26.7	210.3
184	181	84.1	75.7	41.9	67.2	499.0	24	18	14	18.7	25.3
185	182	153.6	62.1	39.0	84.9	3673.2	25	18	9	17.3	64.3
186	183	10.3	8.5	39.5	19.4	302.8	3	6	19	9.3	72.3
187	Overall sample mean				79.9					18.2	
188	Overall between-day variance				1645.3					53.0	
189	Overall within-day variance					4605.7					152.6
190	Relative between-day variance				0.26					0.16	
191	Relative within-day variance					0.72					0.46
192	Correlation of PM and UPT					0.77					

except for range G3:G6 in the INPUTS section where the values for the four parameters listed are entered. Once these values are entered, the template automatically shows the new sample size table in range A14:G15 in the RESULTS section. The FTA sample size table is also shown in range A12:G13 for comparison.

If the illustrative sample were the NTD sample for a particular transit agency and the computation in Table 3 were completed, 0.26 (from cell E190, Table 3) would have been entered as the relative between-day variance in cell G5

Table 4. Template without 100 Percent UPT Count—Based on PM

	A	B	C	D	E	F	G
1	Inputs						
2	**Parameters**						**Values**
3	Annual days of service						365
4	Total daily one-way bus trips						500
5	Relative between-day variance in PM						0.26
6	Relative within-day variance in PM						0.72
7							
8							
9	Results						
10	**Plan number**	1	2	3	4	5	6
11	**Sampling frequency**	Every Day	Every 2nd Day	Every 3rd Day	Every 4th Day	Every 5th Day	Every 6th Day
12	*FTA daily sample*	*2*	*3*	*5*	*7*	*10*	*15*
13	*FTA annual sample*	*730*	*549*	*610*	*644*	*730*	*915*
14	Customized daily sample—PM	1	3	6	20	#N/A	#N/A
15	Customized annual sample—PM	365	549	732	1840	#N/A	#N/A

Note: A particular plan is unavailable when "#N/A" is present.

and 0.72 (from cell F191, Table 3) as the relative within-day variance in cell G6. This example assumes that the agency's fixed-route bus system operates every day of the year with 500 one-way bus trips every day.

Given the relative variances in PM, plans 5 and 6 would not be available. This means that no matter how many bus trips are sampled on each sample day, FTA's minimum confidence and precision levels would not be satisfied if every fifth or sixth day were sampled. The FTA plans allow sampling every fifth or sixth day because the assumed relative between-day variance of 0.1 does not reflect agency conditions. FTA's minimum confidence and precision levels would be violated if FTA's plan 5 or 6 were used.

Among the four plans available, sampling every day will reduce the sample by half; sampling every second day works just as well. Sampling every third

or fourth day, however, requires sampling 6 and 20 one-way bus trips on each sample day, respectively. While these two daily sample sizes appear to be larger than those of the corresponding FTA plans, they will meet FTA's minimum confidence and precision levels. Using FTA plans assumes statistical variation in PM (relative between-day variance of 0.1 and relative within-day variance of 1.0) and does not reflect the agency's actual conditions.

PTL-Based Sample Size Table

The agency may be interested in further reducing the sample size from using the PTL-based approach to estimating system total PM. Reduced sample size could be a factor, for example, in a decision on whether efforts should be made in obtaining the 100 percent system total UPT count. In this case, developing a PTL-based sample size table is necessary. The second Excel sheet, called "Based on PTL," would be used, entering the relative variances in UPT in cells G7 and G8, respectively, and the correlation coefficient between PM and UPT in cell G9 (from cell F192, Table 5). The PTL-based sample size table shows in range A19:G20. For comparison, both the FTA table and the PM-based table are also shown.

When based on PTL, plans 5 and 6 would become available. More important, sample size could be further reduced from the PM-based sample size table for plans 2 to 4. Furthermore, the agency could do much better with its own PTL-based sampling plan (1 trip a day) than the corresponding PM-based FTA plan (3 trips a day), if the agency samples every second day.

100 Percent System Total UPT Count Available

If the 100 percent system total UPT count is available, use the Excel template "FDOT_BC137_46_With 100% UPT Count.xls," available from the authors or at *http://www.dot.state.fl.us/research-center/Completed_PTO.htm* under project BC137-46. The formulas involved are presented in the Appendix.

PTL-Based Sample Size Table

With the 100 percent system total UPT count available, the PTL-based approach to estimating system total PM should be used. To develop a cus-tomized sample size table, the first sheet of the template, "Based on PTL,"

Table 5. Template without 100 Percent UPT Count—Based on PTL

	A	B	C	D	E	F	G
1				Inputs			
2	**Parameters**						**Values**
7	Relative between-day variance in UPT						0.16
8	Relative within-day variance in UPT						0.46
9	Correlation coefficient between PM and UPT						0.77
10							
11							
12				Results			
13	**Plan number**		1	2	3	4	5	6
14	**Sampling frequency**		Every Day	Every 2nd Day	Every 3rd Day	Every 4th Day	Every 5th Day	Every 6th Day
15	*FTA daily sample*		*2*	*3*	*5*	*7*	*10*	*15*
16	*FTA annual sample*		*730*	*549*	*610*	*644*	*730*	*915*
17	Customized daily sample—PM		1	3	6	20	#N/A	#N/A
18	Customized annual sample—PM		365	549	732	1840	#N/A	#N/A
19	Customized daily sample—PTL		1	1	2	2	3	5
20	Customized annual sample—PTL		365	183	244	184	219	305

Note: A particular plan is unavailable when "#N/A" is present.

should be used. This sheet is divided into INPUTS and RESULTS (Table 6) and is protected, except for range G3:G9 in the INPUTS section where the values for the seven parameters listed are entered. Once these values are entered, the template automatically shows the new sample size table in range A17:G18 in the RESULTS section. The FTA sample size table is also shown in range A15: G16 for comparison.

If the illustrative sample were the agency's NTD sample and the computations in Table 3 were completed, 0.26 (from cell E190, Table 3) would have been entered as the relative between-day variance in PM in cell G5, and 0.72 (from cell F191, Table 3) would have been entered as the relative within-day variance in PM in cell G6. Also entered would have been 0.16 (from cell J190, Table

Table 6. Template with 100 Percent UPT Count—Based on PTL

	A	*B*	*C*	*D*	*E*	*F*	*G*
1	Inputs						
2	**Parameters**						Values
3	Annual days of service						365
4	Total daily one-way bus trips						500
5	Relative between-day variance in PM						0.26
6	Relative within-day variance in PM						0.72
7	Relative between-day variance in UPT						0.16
8	Relative within-day variance in UPT						0.46
9	Correlation coefficient between PM and UPT						0.77
10							
11							
12	Results						
13	**Plan number**	1	2	3	4	5	6
14	**Sampling frequency**	Every Day	Every 2nd Day	Every 3rd Day	Every 4th Day	Every 5th Day	Every 6th Day
15	*FTA daily sample*	2	3	5	7	10	15
16	*FTA annual sample*	730	549	610	644	730	915
17	Customized daily sample—PTL	1	1	2	2	3	5
18	Customized annual sample—PTL	365	183	244	184	219	305

3) as the relative between-day variance in UPT in cell G7 and 0.46 (from cell K191, Table 3) as the relative within-day variance in UPT in cell G8. The value 0.77 would have been entered as the correlation coefficient between PM and UPT in cell G9 from cell F192 in Table 5. In addition, this illustrative example assumes that the agency's fixed-route bus system operates every day of the year with 500 one-way bus trips every day.

Given these relative variances in PM and UPT and the correlation coefficient between PM and UPT, all six plans are available. In this case, use of any customized PTL-based sampling plan is preferable to using the corresponding PM-based FTA plan.

Improvements Over the PM-Based Sample Size Table

If an agency is interested in exploring the improvements it made by using the PTL-based rather than the PM-based sample size table, the second sheet, "Based on PM," shown in Table 7, can be used. The PM-based sample size table is shown in range A19:G20. For comparison, both the FTA table and the PTL-based table are shown.

Table 7. Template with 100 Percent UPT Count—Based on PM

	A	*B*	*C*	*D*	*E*	*F*	*G*
12	Results						
13	**Plan number**	1	2	3	4	5	6
14	**Sampling frequency**	Every Day	Every 2nd Day	Every 3rd Day	Every 4th Day	Every 5th Day	Every 6th Day
15	*FTA daily sample*	*2*	*3*	*5*	*7*	*10*	*15*
16	*FTA annual sample*	*730*	*549*	*610*	*644*	*730*	*915*
17	Customized daily sample—PTL	1	1	2	2	3	5
18	Customized annual sample—PTL	365	183	244	184	219	305
19	Customized daily sample—PM	1	3	6	20	#N/A	#N/A
20	Customized annual sample—PM	365	549	732	1840	#N/A	#N/A

Note: A particular plan is unavailable when "#N/A" is present.

When based on PM, plans 5 and 6 become unavailable, indicating that, no matter how many bus trips are sampled on each sample day, FTA's minimum confidence and precision levels would not be satisfied if sampling is every fifth or sixth day. The FTA plans allow sampling every fifth or sixth day because the assumed relative between-day variance of 0.1 does not reflect the agency's actual conditions.

If sampling is every third or fourth day, on the other hand, it would be necessary to sample 6 and 20 one-way bus trips, respectively, on each sample day. These two daily sample sizes appear to be larger than those of the corresponding FTA plans. The smaller sample sizes of these two FTA plans indicate that using these FTA plans would violate FTA's own minimum confidence

and precision levels. This occurs because the assumed statistical variation in PM (relative between-day variance of 0.1 and relative within-day variance of 1.0) by the FTA plans does not reflect the agency's actual conditions. More important, PTL-based sampling significantly reduces sample size from the customized PM-based sample size table for plans 2 to 4. In fact, the customized sample size would be reduced by two-thirds if sampling every second day or every third day. More significantly, sample size would be cut to one-tenth if sampling is every fourth day.

Summary

This article provides guidance for transit agencies to develop sampling plans customized to their fixed-route bus services for NTD reporting. With basic information on annual days of service and daily total number of one-way bus trips operated, this guide helps individual transit agencies develop sampling plans customized to their fixed-route bus services. The plans are in an Excel environment using an existing NTD sample collected with an FTA-approved sampling plan. The new sampling plans developed using this guide maintain the main characteristics of the original FTA plans. More important, by minimizing sample size while meeting FTA's confidence and precision levels, the customized sampling plans represent more reliable and cost-effective alternatives to the FTA sampling plans.

While sampling plans developed by using this guide do not preclude collecting data through methods other than ride check surveys, other sampling methods may be more efficient and/or more cost effective. An example is the use of automated passenger counters for data collection (Strathman and Hopper 1991; Kimpel and Strathman 2002).

This guide, along with several recommendations, has been discussed in person with the FTA in July 2004. One recommendation was to stop treating current FTA-approved sampling plans as "FTA-approved sampling techniques" in the NTD reporting manual. Another recommendation contains two options on the customized sampling plans developed using this guide. One option is for FTA to treat the customized plans as "FTA-approved sampling techniques." The other option is for FTA to add "FTA-approved alternative sampling techniques" to the NTD reporting manual as a third option, in addition to the current two options—"FTA-approved sampling techniques" and "alternative sampling techniques." Under the

second option, transit agencies would not need to provide an additional approval of a qualified statistician. Rather, they must provide the following in their files:

- This guide
- Estimation technique and sample averages used in estimation
- Spreadsheet used to calculate variances and correlation
- Spreadsheet used to determine the customized sample size table
- Chosen sampling plan.

The FTA has not formally accepted the recommendations; however, we do not see any reason for the FTA to reject the use of this guide and related data items listed above as the required approval of a qualified statistician.

Acknowledgments

Appendix

Two sets of formulas are presented. The first is used to compute the relative variances of a given quantity and contains three formulas: (1) relative between-day variance; (2) relative within-day variance; and (3) relative overall variance. These formulas apply to both PM and UPT. The second is used to determine sample sizes and contains two formulas. One where system total PM is directly estimated as the product of the sample mean PM per one-way bus trip and the annual total number of one-way bus trips (100 percent system total UPT count unavailable), and the other where system total PM is indirectly estimated as the product of the sample mean PTL and a 100 percent system total UPT count (100 percent system total UPT count available).

Basic Notations

Following Cochran (1977), these notations are used:

M = number of days in a year that fixed-route motor bus service is available

m = number of sample days (i.e., 365, 183, etc.)

N = number of one-way bus trips on each sample day

n = number of bus trips sampled on each sample day (to be determined)

y_{ij} = quantity obtained for the j-th sampled trip on the i-th sample day

$$\bar{y}_i \equiv \sum_{j=1}^{n} \frac{y_{ij}}{n} = \text{sample mean per one-way bus trip on } i\text{-th sample day}$$

$$\hat{y} \equiv \sum_{i=1}^{m} \frac{\bar{y}_i}{m} = \text{overall sample mean per one-way bus trip}$$

$$s_B^2 \equiv \frac{\sum_{i=1}^{m} \left(\bar{y}_i - \hat{y}\right)^2}{m-1} = \text{sample between-day variance}$$

$$s_{i,W}^2 \equiv \frac{\sum_{j=1}^{n} \left(y_{ij} - \bar{y}_i\right)^2}{n-1} = \text{sample within-day variance on } i\text{-th sample day}$$

$$s_W^2 \equiv \frac{\sum_{i=1}^{m} s_{i,W}^2}{m} = \text{sample within-day variance on all sample days}$$

Relative Variances

With these notations, the two formulas for computing relative variances are given by:

$$C_B^2 \equiv \frac{s_B^2}{\hat{y}^2} \qquad = \text{estimated relative between-day variance}$$

$$C_W^2 \equiv \frac{s_W^2}{\hat{y}^2} \qquad = \text{estimated relative within-day variance}$$

The relative between-day variance, in general, is smaller than the relative within-day variance. In the case of FTA-approved sampling plans, for example, the relative between-day variance is assumed to be 0.1, while the relative within-day variance is assumed to be 1.0. The quantity of interest in this case is PM.

According to Cochran (1977), the relative overall variance can be related to these two relative variances as follows:

$$C^2 = \frac{M-m}{M}\frac{C_B^2}{m} + \frac{N-n}{N}\frac{C_W^2}{mn} \qquad\qquad \text{(A-1)}$$

Sample Size

Sample size depends on a set of common factors:

- Agency conditions, such as number of service days (M) and average number of one-way bus trips operated on sample days (N).

- The agency's choice of sampling frequency, which determines the annual total number of sample days (m).

- FTA's minimum confidence and precision levels. The minimum confidence level is set at 95 percent, indicating a z value of 1.95. Following UMTA C2701.1A (UMTA 1988), this z value is rounded up to 2. The minimum precision level is at ±10 percent ($r = 0.1$).

In addition to these common factors, sample size also depends on the quantity to be directly estimated from the sample. Different quantities have different degrees of statistical variation, leading to different sample size requirements. If the PM-based approach to estimating system total PM is used, the quantity to be estimated from the sample is the sample mean PM per one-way bus trip. If the

PTL-based approach to estimating system total PM is used, the quantity to be estimated from the sample is the sample mean PTL. The formulas are presented separately for these two cases.

PM-Based Approach

In this case, sample design is based on PM. To determine sample size, relative overall variance in PM must be linked to FTA's minimum confidence and precision requirement, $C_M^2 = (r/z)^2$, where C_M^2 is given by equation (A-1) with the two relative variances measured for PM. More explicitly, the following relationship exists:

$$\frac{M-m}{M}\frac{C_{MB}^2}{m} + \frac{N-n}{N}\frac{C_{MW}^2}{mn} = \left(\frac{r}{z}\right)^2 \qquad \text{(A-2)}$$

Where:

M has been added to the subscript of the two relative variances to indicate that they are measured for PM.

Having equation (A-2), sample size can be determined by explicitly solving equation (A-2) for n. The exact formula depends on whether sampling is done every day or less frequently. If sampling every day, that is, $m = 365$, the minimum number of one-way bus trips is given by the integer no less than n as given by a simple formula:

$$n = \frac{NC_{MW}^2}{mN\left(\dfrac{r}{z}\right)^2 + C_{MW}^2} \qquad \text{(A-3)}$$

The minimum sample size would be two one-way bus trips every day, for example, if n is either 1.1 or 1.9. If sampling at a lesser frequency, i.e., $m < 365$, the minimum number of one-way bus trips necessary to sample is given by the integer no less than n as given by a more complex formula:

$$n = \frac{MNC_W^2}{M\left(C_W^2 - NC_B^2\right) + mN\left[M\left(\frac{r}{z}\right)^2 + C_B^2\right]} \qquad \text{(A-4)}$$

These last two formulas have been built into the Excel template called "FDOT_BC137_46_WITHOUT 100% UPT COUNT.xls" available from the authors or at *http://www.dot.state.fl.us/research-center/Completed_PTO.htm* under project BC137-46.

PTL-Based Approach

In this case, sample design is based on PTL. Since the sample average PTL is the ratio of the sample total PM over the sample total UPT, the statistical basis of the following formula is the theory of ratio estimates (Smith 1983). The relative overall variance in PTL (C_L^2) is related to the relative overall variances for both PM (C_M^2) and UPT (C_T^2) as well as their correlation coefficient (ρ). Specifically,

$$C_L^2 = C_M^2 + C_T^2 - 2\rho \, C_M \, C_T \qquad \text{(A-5)}$$

Where:

C_M^2 and C_T^2 are both given by equation (A-1) for PM and UPT, respectively.

Similar to the first case, this relative overall variance must be linked to FTA's minimum confidence and precision requirement, $C_L^2 = (r/z)^2$. This equation does not have a closed solution for n. Equation (A-5) has been built into the Excel template called "FDOT_BC137_46_WITH 100% UPT COUNT.xls" available from the authors or at *http://www.dot.state.fl.us/research-center/Completed_PTO.htm* under project BC137-46.

Equation (A-5) can also be adapted for computing the between-day and within-day relative variances for PTL, using the relative variances for PM and UPT as well as the correlation coefficient between PM and UPT. For example, the between-day relative variance for PTL, C_{LB}^2, is given by

$$C_{LB}^2 = C_{MB}^2 + C_{TB}^2 - 2\rho\, C_{MB}\, C_{TB} \qquad\qquad (A\text{-}6)$$

Where:

C_{MB}^2 and C_{TB}^2 are the between-day relative variances for PM and UPT, respectively.

References

Chu, Xuehao. 2004. Another look at FTA-approved sampling plans for fixed-route bus services. Submitted to the *Journal of Transportation and Statistics*.

Cochran, W. G. 1977. *Sampling techniques*, 3rd ed., John Wiley and Sons, Inc., New York.

FTA. 2003. *National Transit Database 2003 reporting manual*. U.S. Department of Transportation, *http://www.ntdprogram.com/NTD/ntdhome.nsf/?Open*.

Kimpel, Thomas, and James G. Strathman. 2002. *Automatic passenger counter evaluation: Implications for National Transit Database reporting*. Portland, Oregon: Center for Urban Studies, Portland State University.

Smith, Robert L., Jr. 1983. *Innovation and incremental improvement in bus transit passenger data collection*. Washington, DC: U.S. Department of Transportation.

Strathman, James G., and Janet R. Hopper. 1991. Evaluation of automatic passenger counters: Validation, sampling, and statistical inference, *Transportation Research Record* 1308: 69–77.

Urban Mass Transportation Administration (UMTA).1988. *Sampling procedures for obtaining fixed route bus operating data under the Section 15 reporting system*. Circular UMTA-C-2710.1A, U.S. Department of Transportation.

About the Authors

Xuehao Chu (*xchu@cutr.usf.edu*) is a senior research associate at the Center for Urban Transportation Research, University of South Florida. He has his Ph.D. in economics from the University of California at Irvine. His research interests are in the areas of quantitative methods, transportation economics, and policy analysis. He has published widely in economics and transportation journals, including *Transport Economics and Policy*, *Transportation*, and *Transportation Research*. He is a referee of articles for many international journals, including *Journal of Political Economy* and *Transportation Science*. He served on the Editorial Board for *Transportation Research-Part A* during 2001–2003.

Ike Ubaka (*ike.ubaka@dot.state.fl.us*) is program manager for transit systems planning at the Florida Department of Transportation, where he is responsible for statewide transit systems planning. He has more than 18 years of experience in transit

planning, including service with the Los Angeles County Metropolitan Transportation Authority and the Orange County (California) Transportation Authority. Mr. Ubaka holds a master's degree in urban and regional planning from UCLA and a master of administration degree from the University of California, Riverside. He is a member of the American Institute of Certified Planners.