Evaluating Two Different Traffic Data Methods Based on Data Observed

Analysis of Provided Data – Final Report A

Publication No. FHWA-PL-021-040

November 2021

U.S. Department of Transportation

Federal Highway Administration

Notice

This document is disseminated under the sponsorship of the United States Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturers' names appear in this report only because they are considered essential to the object of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Technical Documentation Page

1. Report No. FHWA-PL-021-040	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Evaluating Two Different Traffic Data Methods Based on Data Observed, Analysis of Provided Data – Final Report A	5. Report Date November 2021 6. Performing Organization Code
7. Author(s) Robert Krile, Elizabeth Slone	8. Performing Organization Report No.
9. Performing Organization Name and Address Battelle 505 King Avenue Columbus, OH 43201	10. Work Unit No. (TRAIS) 11. Contract or Grant No. Contract 693JJ319D000006 Task Order 693JJ321F000215
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Highway Policy Information 1200 New Jersey Avenue SE Washington, DC 20590	13. Type of Report and Period Covered Final Report (July 2021 – October 2021) 14. Sponsoring Agency Code

15. Supplementary Notes
The project was managed by Task Manager Steven Jessberger and Chief of the Travel Monitoring and Surveys Division, Tianjia Tang, for the Federal Highway Administration (FHWA), who provided technical directions.

16. Abstract
Traditional traffic counts have long been utilized by U.S. DOT, the states, and MPOs to determine volumes of traffic. Recent developments in connected technologies have provided the potential for probe-based, or passive, measurement of traffic volumes that combines information on specific vehicles and advanced analytics algorithms to estimate AADT (Annual Average Daily Traffic). This task considers data provided from an evaluation of one such passive measurement system that produced AADT estimates from both the passive method and the traditional counts in 2019 for a sample of more than 800 roadways over 29 states, including a wide range of functional classifications and volumes. A statistical analysis was performed to determine if the data collected provided evidence that the mean (and median) AADT as determined by the passive method would be expected to be systematically different (higher or lower) than that of AADT based on traditional counts. Mean difference was evaluated through paired sample t-tests of base-10 logarithmically transformed AADT values, while median differences were evaluated through the Wilcoxon signed rank test. Overall, there was adequate evidence to reject the assumption of equality and instead conclude that the passive measurement mean (or median) AADT exceeded that of the traditional counts with at least 95% confidence [geometric mean difference estimated at 6.6 percent (p-value <0.001) / median difference estimated at 3.9 percent (p-value <0.001)]. More detailed comparisons were conducted by urban/rural status, functional classifications, and AADT volume ranges. For these comparisons, the statistical test results were adjusted for multiple comparisons to control the risk of erroneously concluding significant differences to no more than 5 percent. In this context, urban and rural, broad volume ranges (<5000, 5,000-50,000, and 50,000+), and especially toll roads provided strong evidence for higher values in the passive measurement. The comparisons by roadway Functional Class 1, 2, 3, 4 and 5, and for most smaller volume ranges (e.g., AADT range of 5,000), failed to reject the assumption of equal means (medians). The lack of significance for the smaller groups could be related to lower statistical power with fewer samples in each comparison. The overall significant difference between passive and traditional estimates was influenced by the very large observed difference for toll roads.

17. Key Words AADT, passive traffic volume counts	18. Distribution Statement No restrictions.

19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 25	22. Price

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AADT	Annual Average Daily Traffic
FHWA	Federal Highway Administration
TMAS	Travel Monitoring Analysis System
TMG	Traffic Monitoring Guide

1.0. INTRODUCTION

The Federal Highway Administration (FHWA) Office of Highway Policy Information has studied the use of alternate estimates of annual average daily traffic (AADT) from passive data sources compared to the practice of measured AADT from direct field counts. This work has been completed under a pooled fund study and has led to development of a set of hundreds of paired AADT observations over sites from many U.S. States, highway segment locations, area types (urban vs rural), and roadway functional classes.

For this task, FHWA requested a statistical analysis to determine if there is adequate evidence to conclude a difference in mean AADT between the two methods. This analysis is intended to evaluate the overall systemwide comparison as well as how roadway functional classes, area types, volume ranges, and locations (states) may affect any conclusions.

1.1. PURPOSE OF THIS REPORT

The purpose of this report is to detail the data review and quality control performed on the matched pair dataset, and then to detail the statistical methods performed, the results obtained, and the interpretations that may be made for the comparisons of interest.

2.0. DATASET AND QUALITY REVIEW

An initial dataset was provided by FHWA for this task that consisted of 1,333 records with both a passive measurement AADT and a continuous count station measured AADT for the calendar year 2019. The records included 29 different states, both rural and urban Area Types, and functional classes 1, 2, 3, 4, 5, and 7, as well as some toll roads. Reported AADT volumes ranged from as small as 200 vehicles to as many as 270,000. The records also included the TMG-designated Station ID and its latitude and longitude coordinates for the permanent count location, where available, and an indication of whether the volumes provided were for a single direction of traffic flow at that site or for bi-directional traffic (i.e., sum of two directions at a single site).

The data in the initial dataset for the continuous count stations were compared to the volumes provided by FHWA's Travel Monitoring Analysis System (TMAS). Three issues were found that required remediation.

1) Some of the station IDs provided indicated a functional classification and area type that were incongruent with the TMAS data submissions by the states for the same station.

2) There were duplicate records in some locations for the same station. These duplicates came in several forms, but a common one was two single direction records for a station and then also a bi-directional record for the same station.

3) In some cases, the benchmark AADT provided for the record was not consistent with the value that would be determined from the TMAS submission of that site.

Data remediation was completed by updating the functional classification and area type for those records where this was necessary, and by collapsing the duplicate records to a single record for a station. For those cases where the benchmark AADT did not match TMAS, the record was excluded from consideration.

The inclusion of both single direction and bi-directional sites provided an additional challenge since the lack of a second direction at each single direction traffic counting site would lead to an inaccurate overall AADT of that roadway. To remediate this issue, the single direction AADT values for both the passive and traffic count data were doubled at these sites.

The final outcome of this exercise is a dataset of 827 of the original 1,333 records provided. These records can be summarized by the relevant data values of geography (state) as shown in Table 1, and by functional class and area type as shown in Table 2. Table 3 provides the site totals in counts and percentages by two different traditional volume count ranges, a coarse division into three ranges, and a finer division into ten ranges.

Table 1. Summary of Data Sites by State

State	Count of Records
Alabama	20
Arkansas	2
Arizona	12
California	17
Colorado	14
Connecticut	2
Delaware	1
Florida	32
Georgia	20
Idaho	18
Illinois	98
Kentucky	8
Maryland	31
Maine	25
Minnesota	4
North Carolina	42
North Dakota	29
Nebraska	7
New Jersy	38
New Mexico	3
Nevada	7
Ohio	52
Oklahoma	12
Oregon	74
Pennsylvania	10
Rhode Island	1
South Carolina	19
Texas	151
Virginia	78
Total	827

Table 2. Summary of Data Site Counts and Percentages by Functional Class and Area Type

Functional Class	By Site Counts			By Percentages		
	Rural	Urban	Total	Rural	Urban	Total*
1	89	94	183	10.8%	11.4%	22.1%
2	7	41	48	0.8%	5.0%	5.8%
3	132	95	227	16.0%	11.5%	27.4%
4	87	29	116	10.5%	3.5%	14.0%
5	49	10	59	5.9%	1.2%	7.1%
6	3	0	3	0.4%	0.0%	0.4%
7	2	0	2	0.2%	0.0%	0.2%
TOLL	70	114	184	8.5%	13.8%	22.2%
Missing	5	0	5	0.6%	0.0%	0.6%
Total*	444	383	827	53.7%	46.3%	

* Percentage total sums may not match sums of individual entries due to rounding.

Table 3. Summary of Data Site Counts and Percentages by Reference Traditional AADT Ranges, Coarse and Fine

Coarse AADT Volume Ranges	Sites	Percentage	Fine AADT Volume Ranges	Sites	Percentage*
<5,000	172	20.8%	<500	8	1.0%
			500 to 1,999	68	8.2%
			2,000 to 4,999	96	11.6%
5,000 to 54,999	515	62.3%	5,000 to 9,999	132	16.0%
			10,000 to 19,999	188	22.7%
			20,000 to 34,999	118	14.3%
			35,000 to 54,999	77	9.3%
55,000+	140	16.9%	55,000 to 84,999	51	6.2%
			85,000 to 124,999	47	5.7%
			125,000+	42	5.1%
Total	827		Total	827	

* Percentages do not sum to 100 due to rounding.

3.0. STATISTICAL METHODS

The primary objective of the analysis was to determine if the dataset provided adequate evidence of a difference in the two methods for AADT estimation using the passive data and the site counts. A secondary objective was to evaluate differences within sub-samples of the original dataset, including geographic locations (states), area types, functional classes, and volume ranges.

A statistical hypothesis test framework is relevant to this analysis. The form of the hypothesis test and the accompanying assumptions are provided in section 3.1.

3.1. HYPOTHESIS TEST

A hypothesis test to evaluate the differences in two different data methods for determining AADT can be defined as:

Null Hypothesis H_0 – There is no difference between the two methods in mean AADT

$\mu_p = \mu_t$

Alternative Hypothesis H_A – This is sufficient evidence to reject the Null Hypothesis and conclude that the mean AADT is not identical between the two methods.

$\mu_p \neq \mu_t$

Where

μ_p is the mean of the AADT by the passive measurement

μ_t is the mean of the AADT by the traditional count measurement

The statistical hypothesis test is conducted under a probabilistic framework where it is possible to control the likelihood of the test reaching the correct conclusion based on assumptions about the data collected, the form of the statistical tests, and specifics about the true populations of interest.

To evaluate the AADT data, both a passive and a traditional count have been obtained for every site. This allows a generally more powerful statistical analysis of the differences between measurements within the same site rather than a comparison of data collected independently for two separate treatments. In the original hypothesis test framework, this is equivalent to testing a modified hypothesis:

H_0 – There is no difference between the two methods in mean AADT

$\mu_p - \mu_t = 0$

Alternative Hypothesis H_A – This is sufficient evidence to reject the Null Hypothesis and conclude that the mean AADT is not identical between the two methods.

$$\mu_p - \mu_t \neq 0$$

3.2. STATISTICAL ANALYSIS ASSUMPTIONS

The statistical inference that can be made with this set of data is dependent on a number of assumptions. These should be evaluated for their validity. If issues are found, the analysis interpretation may be altered to accommodate such violations. In some cases, failure to meet assumptions can be directly tested and alternative methodologies employed that mitigate the failed assumptions.

For this analysis of differences in AADT measurement by two methods, the following assumptions are considered:

- The inference is assumed to be applicable to a representative population. Since this dataset consists of roadways in a large number of states and of varying area types and functional classes, the conclusions may be considered broadly applicable. However, unless the sites were selected in a probabilistic manner to be representative of all roadways within the U.S., the statistical conclusions cannot be extended to cover this general outcome.

- The observations are assumed to be independent of each other. In this case, the two methods are each executed independently of each other for each site, and information for the accuracy at one site is not incorporated into the estimated values at another site.

- The two methods should not include an interaction between the particular site and the difference in measured AADT. If the data are found to violate this assumption, remediation may be possible through a transformation.

- The data for the two methods should not include significant outliers that could indicate errors in the measurements or a lack of uniformity in the populations to be compared. Additionally, the statistical test to be performed should conform to assumptions about the form of the underlying data distributions such as normality and differential skewness. These assumptions are ones that can be evaluated and may be mitigated in some cases. For instance, the data may be transformed by performing the analysis on the original values after taking their logarithms.

3.3. STATISTICAL TESTS

From the format of the data being paired observations of two different treatment levels, a paired difference analysis is appropriate. Specifically, if certain assumptions can be validated, the one-sample t-test can be used to evaluate if there is a difference in means between the two types of measurement. If these assumptions cannot be fully supported, a less restrictive non-parametric test can be performed such as the Sign Test or the Wilcoxon Signed Rank Test.

3.4. REPORTED RESULTS

The reported results for the one-sample t-test were the mean of the difference of the passive to the traditional method AADT. If logarithmic transformations were necessary, the reported difference was that of the geometric means rather than the traditional arithmetic means. For a non-parametric test, the reported result was the estimated difference in population medians. In all three cases, the results only reflected differences of the center of the distribution. Other tests would be required to determine whether other features (e.g., variability, skewness) of the two methods differ from each other. In addition to the estimate of the difference, a 95% confidence interval for the range of the difference was provided.

The test also included a p-value to indicate the probability that a result as extreme or more extreme than that observed could have occurred simply by chance if the Null Hypothesis was true (i.e., there is no difference in central value [mean, median] between the two populations). By standard convention, a p-value less than 0.05 indicated a statistically significant difference, though users may prefer a stronger (or weaker) level of evidence to justify rejecting the Null Hypothesis and concluding the means are truly different.

A p-value greater than 0.05 led to a conclusion that the Null Hypothesis is not rejected. It did not confirm the Null Hypothesis as true.

More than one statistical comparison was performed in this evaluation. If a p-value of less than 0.05 is required for adequate evidence of a single statistical difference, the application of the same standard to multiple comparisons produces a higher than five percent risk that at least one of the significant comparisons was due to random chance and erroneously concluded a significant difference in population means (or medians) when such a difference does not truly exist. To control the risk of this type of false discovery of significant results, the p-values for the entire set of comparisons in the evaluation were adjusted by the Bonferroni-Holm method so that the collective risk of any of the comparisons falsely concluding significance was controlled at no more than five percent.

For each of the t-test comparisons performed, the differences between the log transformed AADTs of passive and traditional count methods were compared to the overall mean of log differences. This set of values, known as residuals, was examined by a quantile-quantile plot and formally tested to determine if the values could be considered to follow a Normal distribution. The Shapiro-Wilk test was conducted to assess the normality of the residuals. If this test did not provide evidence of non-normality (i.e., $p \geq 0.05$), the t-test assumption of normality in the residuals was not rejected. If the tests did suggest a lack of normality but the number of samples was adequately large (i.e., ≥ 30), the parametric t-test was still appropriate. If the normality assumption was untenable for a small dataset, other transformations or remediations were evaluated and/or a non-parametric signed rank test was used to assess median difference in AADT. The non-parametric test was performed on all comparisons as a reference and to establish consistency, but the t-test results were primary if considered valid.

Based on the number of sites evaluated for each comparison, and the level of variability in the data, it was possible to identify the approximate statistical power of each test. The statistical power is the probability of correctly concluding the alternative hypothesis when it is in fact true.

The statistical power is a function of where the true differences lie in the region of the alternative hypothesis. This statistical power can provide evidence of sensitivity of the testing procedure and may help evaluate the interpretation of the results. It cannot be used to determine veracity of the conclusions reached for a particular sample. The number of sites to evaluate were not identified in advance to provide a particular statistical power. Consequently, identification of the approximate statistical power will be important to understand the magnitude of true differences that this evaluation was probably likely to be able to detect. If those differences are large, a lack of significant results may be less an indication of real lack of differences and more a lack of sensitivity on the part of the evaluation to be able to detect a difference of interest. Conversely, if the differences are very small, a large number of significant results may reflect true population differences, but the differences may not be meaningful in a practical sense. For instance, statistically significant AADT differences of 1 vehicle in 1000 may not be reason to reject an alternative measurement if this level of difference does not meaningfully impact the purposes to which the data might be subsequently applied (e.g., pavement studies, revenue projections).

All statistical tests were performed in SAS® version 9.4.

4.0. RESULTS

Before performing any statistical tests, some simple descriptive statistics were obtained for the passive and traditional count data as well as the paired differences at each site (passive – traditional). These results are shown in Table 4.

Table 4. Descriptive Statistics for Passive, Count, and Paired Difference (Passive – Count)

Type	Mean	Std. Dev.	Minimum	1st Quartile	Median	3rd Quartile	Maximum
Passive	32,628.72	41,658.15	388	6,482	17,400	38,464	266,878
Traditional	30,930.43	39,428.25	237	6,270	15,655	36,770	212,866
Difference (Passive-Traditional)	1,698.29	7,652.87	-33,300	-399	400	2,668	60,302

From the descriptive statistics, it is apparent that the count data for the two sources vary over several orders of magnitude. A plot of the difference data as a function of the reference counts is provided in Figure 1.

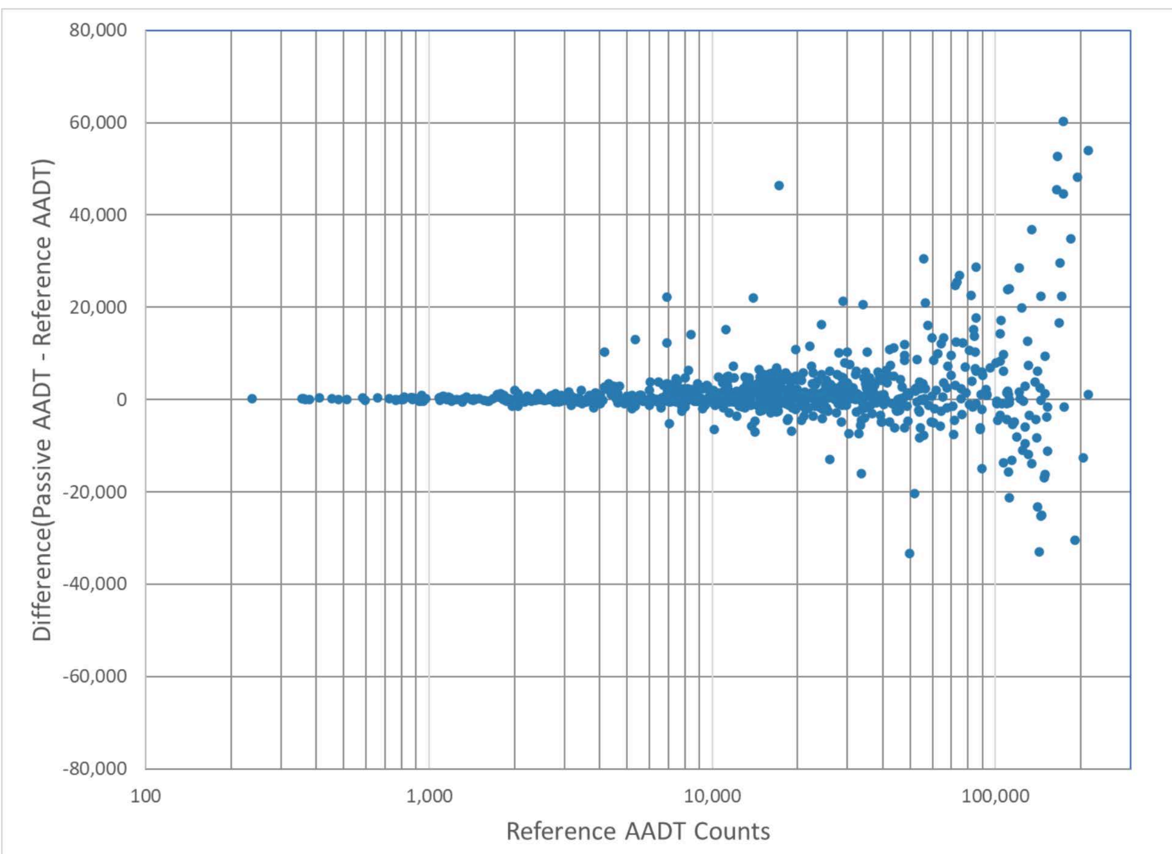

Figure 1: Passive – Reference Count Differences as a Function of the Reference Counts (n=827).

This pattern of much larger variability of differences for roadways with higher AADTs precludes performing the statistical hypothesis testing on the count data directly. To do so would lead to a flawed conclusion that the mean difference of 1,698 as estimated in Table 1 was representative of the populations at large. In fact, a mean difference of 1,698 in AADT is approximately appropriate for sites only in a limited range of AADTs in the middle of the overall range.

With data of the pattern in Figure 1, transformation of the individual AADT observations by the logarithm may be beneficial. Applying a base-10 logarithm to each AADT value before taking the differences results in the updated plot shown in Figure 2.

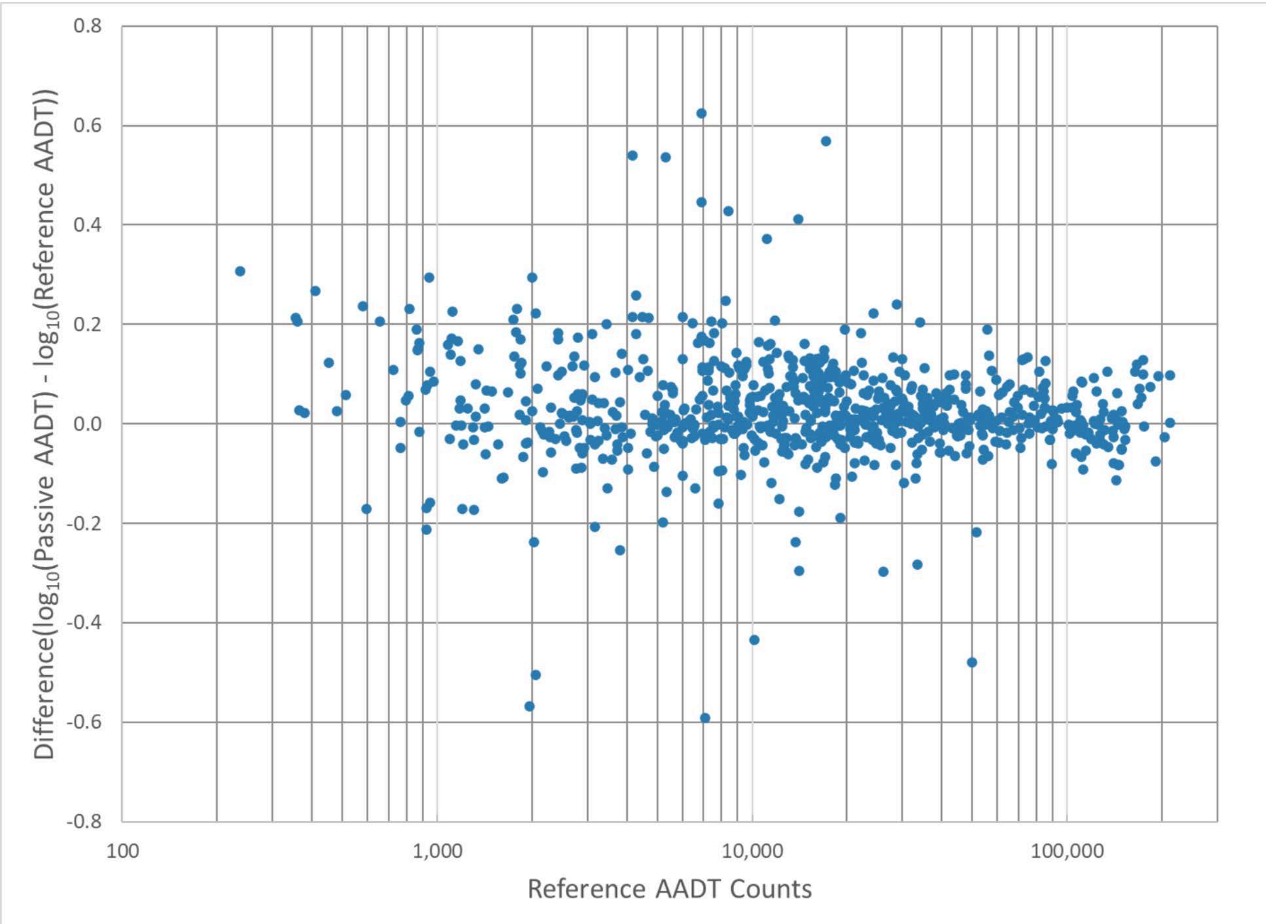

Figure 2: Overall Log_{10}(Passive) – log_{10}(Reference Count) Differences as a Function of the Reference Counts (n=827).

The logarithmic transformation of the AADT data before differencing greatly reduces variability as a function of the AADT itself. In Figure 2, there are some data observations that appear to be outliers, both on the high side and the low side. Additionally, the sites with the lowest AADT values (i.e., below 1,000) provide some evidence of a differential log difference, with the passive measurement producing a larger relative count. If this effect is real, as opposed to random variability in the data, it may slightly bias the estimated outcome, so its impact will be considered subsequently. Analysis of the results as a function of the AADT count range will be included in the results, with volumes separated into either 3 coarse categories or 10 finer categories as previously documented in Table 3.

Figure 3 depicts the log differences from Figure 2 separated into sites identified as rural and urban.

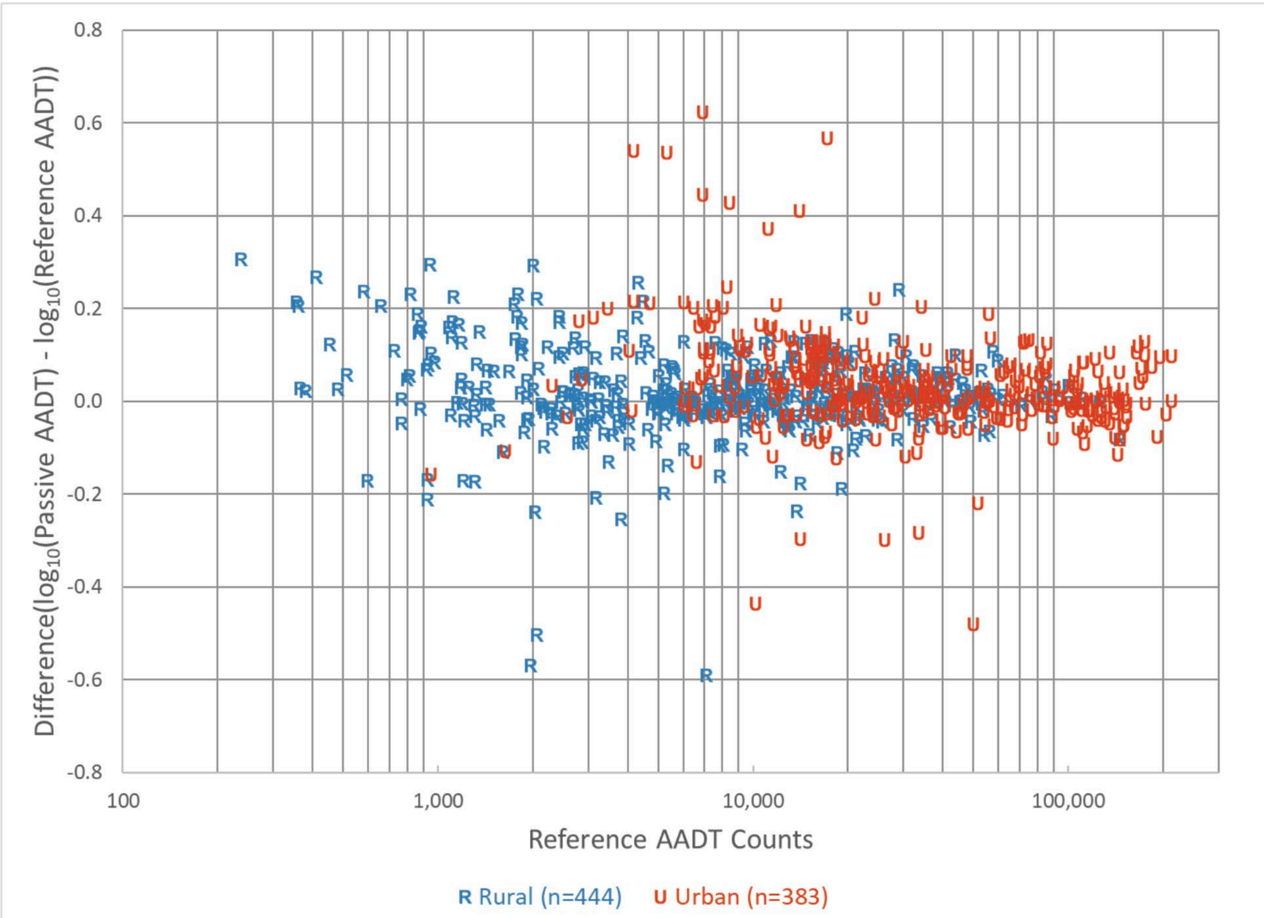

Figure 3: Rural and Urban Log$_{10}$(Passive) – log$_{10}$(Reference Count) Differences as a Function of the Reference Counts.

Figure 4 depicts the log differences from Figure 2 separated into sites identified as Functional classifications 1, 2, 3, 4, and 5. It also shows toll road sites, which were not provided a functional classification. This graph excludes 10 sites, including 5 rural sites in South Carolina with no functional classification provided and 5 total rural sites for functional classifications 6 and 7, which were too few to be analyzed separately. These sites are included in Figures 1 through 3.

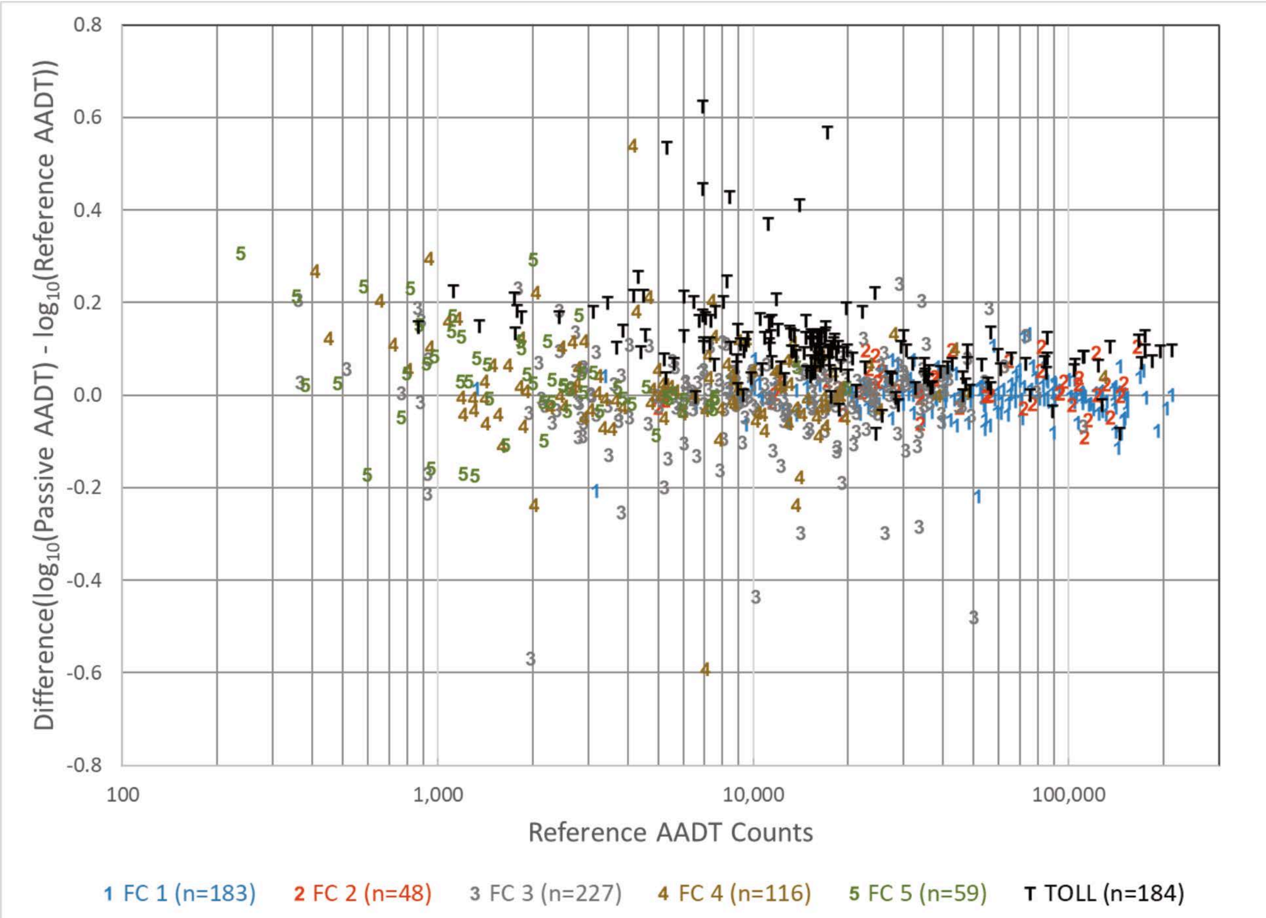

Figure 4: Functional Classification and Toll Log_{10}(Passive) – log_{10}(Reference Count) Differences as a Function of the Reference Counts.

Tables 5 and 6 show the full set of statistical test results for the log-transformed differences between the passive and traditional counts. In each table, the results are shown for the overall complete set of 827 sites; the rural and urban sites; subsets of the sites by functional classes 1 through 5, and the toll roads (which were not provided with a functional class); and by the coarse and fine volume ranges of the traditional count AADTs. Note that the results for the functional classes and toll roads exclude 3 records for functional class 6 and 2 records for functional class 7, which were sample sizes too small to provide separate analyses. Additionally, five records in South Carolina were known to be rural but no functional classification was provided so these do not appear in any of the functional classification analyses.

The consequence of logarithmic transformation of the AADT data is that the original statistical results of Table 5 are in terms of a difference in log-transformed volumes. To interpret these results, it is necessary to back-transform the statistical estimates by exponentiation. The resulting mean differences will be ratios of geometric means between the passive and traditional measures, or ratios of medians for the non-parametric tests. Table 6 provides the same statistical results (i.e., p-values) but shows the statistical estimates in this ratio form rather than as differences in log-transformed volumes. The Table 6 results are also directly interpretable as percentages, by subtracting 1 from each ratio estimate and multiplying by 100. This percentage interpretation is utilized in the final section of the report. The Table 6 results are the primary form for subsequent discussion.

For each subset of interest, both parametric and non-parametric test results are provided. The p-value column refers to each single test, with a statistically significant outcome denoted by a plus sign (+) and yellow shading. A multiple-comparison adjusted p-value is also provided that controls the risk of an erroneous conclusion of a significant difference to no more than five percent across all tests within the class of tests. A stronger level of evidence is required for significance at the adjusted p-value level as evidenced by a smaller number of significant results indicated. For the parametric tests, the Shapiro-Wilk column provides the p-value for the test of whether the residual data from the parametric analysis can safely be assumed to be Normally distributed. Most of the results in Table 6 suggest this assumption is not supported. For subsets of data with sample sizes exceeding 30, this lack of normality does not invalidate the conclusions of the t-test. The only line of Table 6 (and 5) with a sample size less than 30 is for the 8 sites with AADT < 500. This subset does not fail the normality test and therefore the parametric result is acceptable for all analyses in this table. After back transforming the mean of the log differences, and their corresponding 95% confidence intervals, in Table 5, the outcome is an estimated ratio of geometric means between the passive and traditional counts. For instance, Table 6 indicates that geometric mean AADT for passive counts is an average of 1.066 times that of the traditional counts. Since this estimate is based on a sample of data, it is subject to uncertainty, and an appropriate 95% confidence interval for this ratio extends from 1.049 to 1.083. While it cannot be known if this interval truly brackets the ratio of population geometric means, the procedure used here would be expected to produce a similar interval that would bracket the true value 95 times in 100. The ratios of geometric means do not reflect the multiple comparison adjustments and are consistent with the first reported p-value.

The non-parametric results provide an alternative evaluation of the original hypothesis where the inference is based on differences in the medians of the populations of passive and traditional counts, rather than the differences in log-transformed counts. The results for statistical significance are very similar as shown in Table 6, with a few differences. For instance, sites with traditional AADT between 500 and 1,999 are not significantly different (based on adjusted p-value of 0.295) in their geometric means, but the median of the passive counts is significantly larger (at p-value of 0.036) than the traditional counts. It is notable, though, that the actual estimates of ratios of geometric means (1.102) and medians (1.127) are still similar. The ratios of medians are also reported with 95% confidence intervals based on rank statistics from the original data. It appears that ratios of the passive and traditional population geometric means from Table 6 are consistently more extreme (i.e., ratios greater than 1 are larger and ratios less than 1 are smaller) than the corresponding ratios of medians. This effect will be discussed in the interpretation section.

Table 5. Statistical Analysis Results for Both Parametric and Non-Parametric Comparisons of Differences in Base-10 Logarithmically Transformed AADTs for Passive and Traditional Counts

Subset	Parametric t-test Results for Log Differences				Non-Parametric Signed Rank Test Results		
	p-value	Adj p-value	Mean Log Difference (95% CI)	Shapiro - Wilk	p-value	Adj p-value	Median Log Difference (95% CI)
All (n=827)	<.001 +	<.001 +	0.028 (0.021, 0.035)	<.001 +	<.001 +	<.001 +	0.017 (0.012, 0.022)
Rural (n=444)	<.001 +	0.001 +	0.018 (0.009, 0.027)	<.001 +	<.001 +	<.001 +	0.012 (0.005, 0.016)
Urban (n=383)	<.001 +	<.001 +	0.039 (0.028, 0.050)	<.001 +	<.001 +	<.001 +	0.030 (0.018, 0.035)
FC 1 (n=183)	0.632	1.000	-0.002 (-0.008, 0.005)	<.001 +	0.962	1.000	0.001 (-0.004, 0.006)
FC 2 (n=48)	0.013 +	0.292	0.017 (0.004, 0.029)	0.263	0.011 +	0.253	0.010 (0.001, 0.030)
FC 3 (n=227)	0.207	1.000	-0.008 (-0.021, 0.005)	<.001 +	0.845	1.000	-0.002 (-0.010, 0.008)
FC 4 (n=116)	0.035 +	0.781	0.023 (0.002, 0.044)	<.001 +	0.016 +	0.347	0.011 (-0.004, 0.026)
FC 5 (n=59)	0.016 +	0.344	0.033 (0.007, 0.060)	0.015 +	0.012 +	0.258	0.020 (-0.003, 0.046)
TOLL (n=184)	<.001 +	<.001 +	0.109 (0.095, 0.124)	<.001 +	<.001 +	<.001 +	0.097 (0.088, 0.105)
AADT<500 (n=8)	0.008 +	0.177	0.149 (0.053, 0.245)	0.182	0.008 +	0.172	0.164 (0.026, 0.306)
500-1,999 (n=68)	0.013 +	0.295	0.042 (0.009, 0.076)	<.001 +	0.002 +	0.036 +	0.052 (0.009, 0.085)
2,000-4,999 (n=96)	0.043 +	0.955	0.026 (0.001, 0.051)	<.001 +	0.044 +	0.964	0.010 (-0.008, 0.026)
5,000-9,999 (n=132)	<.001 +	0.013 +	0.038 (0.017, 0.060)	<.001 +	<.001 +	<.001 +	0.015 (0.001, 0.030)
10,000-19,999 (n=188)	<.001 +	<.001 +	0.036 (0.022, 0.050)	<.001 +	<.001 +	<.001 +	0.037 (0.018, 0.051)
20,000-34,999 (n=118)	0.027 +	0.594	0.015 (0.002, 0.028)	<.001 +	0.002 +	0.048 +	0.013 (0.007, 0.023)
35,000-54,999 (n=77)	0.662	1.000	0.004 (-0.013, 0.020)	<.001 +	0.031 +	0.673	0.011 (0.000, 0.019)
55,000-84,999 (n=51)	<.001 +	0.001 +	0.033 (0.018, 0.048)	0.111	<.001 +	0.002 +	0.021 (0.005, 0.050)
85,000-124,999 (n=47)	0.273	1.000	0.007 (-0.006, 0.021)	0.318	0.325	1.000	0.004 (-0.004, 0.010)
AADT≥125,000 (n=42)	0.328	1.000	0.009 (-0.010, 0.028)	0.335	0.526	1.000	0.001 (-0.013, 0.024)
AADT<5,000 (n=172)	<.001 +	0.004 +	0.038 (0.018, 0.058)	<.001 +	<.001 +	<.001 +	0.025 (0.009, 0.047)
5,000-54,999 (n=515)	<.001 +	<.001 +	0.027 (0.018, 0.036)	<.001 +	<.001 +	<.001 +	0.018 (0.013, 0.024)
AADT≥55,000 (n=140)	<.001 +	0.006 +	0.017 (0.008, 0.026)	0.041 +	0.001 +	0.031 +	0.006 (0.001, 0.019)

+ statistically significant at the 0.05 level (yellow)

Table 6. Back-Transformed Statistical Analysis Results for Both Parametric and Non-Parametric Comparisons of Differences in Base-10 Logarithmically Transformed AADTs for Passive and Traditional Counts

Subset	Parametric t-test Results				Non-Parametric Signed Rank Test Results		
	p-value	Adj p-value	Ratio of Geometric Means (95% CI)	Shapiro - Wilk	p-value	Adj p-value	Ratio of Medians (95% CI)
All (n=827)	<.001 +	<.001 +	1.066 (1.049, 1.083)	<.001 +	<.001 +	<.001 +	1.039 (1.028, 1.053)
Rural (n=444)	<.001 +	<.001 +	1.043 (1.022, 1.064)	<.001 +	<.001 +	<.001 +	1.029 (1.012, 1.038)
Urban (n=383)	<.001 +	<.001 +	1.093 (1.066, 1.121)	<.001 +	<.001 +	<.001 +	1.072 (1.042, 1.085)
FC 1 (n=183)	0.632	1.000	0.996 (0.982, 1.011)	<.001 +	0.962	1.000	1.002 (0.991, 1.014)
FC 2 (n=48)	0.013 +	0.119	1.039 (1.008, 1.070)	0.263	0.011 +	0.103	1.023 (1.003, 1.072)
FC 3 (n=227)	0.207	1.000	0.981 (0.953, 1.011)	<.001 +	0.845	1.000	0.995 (0.978, 1.018)
FC 4 (n=116)	0.035 +	0.319	1.053 (1.004, 1.106)	<.001 +	0.016 +	0.142	1.025 (0.991, 1.060)
FC 5 (n=59)	0.016 +	0.141	1.080 (1.015, 1.149)	0.015 +	0.012 +	0.106	1.047 (0.992, 1.111)
TOLL (n=184)	<.001 +	<.001 +	1.287 (1.245, 1.329)	<.001 +	<.001 +	<.001 +	1.251 (1.225, 1.274)
AADT<500 (n=8)	0.008 +	0.177	1.410 (1.130, 1.760)	0.182	0.008 +	0.172	1.460 (1.063, 2.025)
500-1,999 (n=68)	0.013 +	0.295	1.102 (1.021, 1.190)	<.001 +	0.002 +	0.036 +	1.127 (1.020, 1.215)
2,000-4,999 (n=96)	0.043 +	0.955	1.062 (1.002, 1.126)	<.001 +	0.044 +	0.964	1.023 (0.982, 1.062)
5,000-9,999 (n=132)	<.001 +	0.013 +	1.092 (1.040, 1.148)	<.001 +	<.001 +	<.001 +	1.034 (1.002, 1.072)
10,000-19,999 (n=188)	<.001 +	<.001 +	1.086 (1.051, 1.123)	<.001 +	<.001 +	<.001 +	1.088 (1.042, 1.126)
20,000-34,999 (n=118)	0.027 +	0.594	1.035 (1.004, 1.067)	<.001 +	0.002 +	0.048 +	1.031 (1.017, 1.055)
35,000-54,999 (n=77)	0.662	1.000	1.009 (0.970, 1.048)	<.001 +	0.031 +	0.673	1.025 (1.001, 1.044)
55,000-84,999 (n=51)	<.001 +	0.001 +	1.079 (1.042, 1.118)	0.111	<.001 +	0.002 +	1.049 (1.012, 1.122)
85,000-124,999 (n=47)	0.273	1.000	1.017 (0.986, 1.049)	0.318	0.325	1.000	1.009 (0.992, 1.024)
AADT≥125,000 (n=42)	0.328	1.000	1.022 (0.978, 1.067)	0.335	0.526	1.000	1.002 (0.970, 1.058)
AADT<5,000 (n=172)	<.001 +	0.004 +	1.092 (1.043, 1.143)	<.001 +	<.001 +	<.001 +	1.058 (1.020, 1.115)
5,000-54,999 (n=515)	<.001 +	<.001 +	1.064 (1.043, 1.085)	<.001 +	<.001 +	<.001 +	1.041 (1.030, 1.056)
AADT≥55,000 (n=140)	<.001 +	0.006 +	1.041 (1.019, 1.063)	0.041 +	0.001 +	0.031 +	1.014 (1.002, 1.044)

+ statistically significant at the 0.05 level (yellow)

5.0. INTERPRETATION AND CONCLUSIONS

The dataset provided was suited to a paired statistical comparison of passive to traditional counts after logarithmic transforming of the counts provided. The overall conclusions are provided below.

There is significant statistical evidence (adjusted p-value < 0.001) that counts provided by passive measurement were an average of 6.6 percent higher (95% confidence interval [4.9%, 8.3%]) compared to those from traditional methods. This conclusion was reached with recognition that the paired sample t-test did not produce residual values (observed difference minus estimated geometric mean) that could be assumed to be normally distributed. Due to the very large number of observations in the comparison (n=827), this condition of normality is not necessary to have confidence in the estimated difference and its significance. For completeness, though, the comparison was also conducted using the non-parametric signed rank test which confirmed the statistical significance (adjusted p-value < 0.001) while showing an estimated median difference of 3.9 percent (95% confidence interval [2.8%, 5.3%]).

The lack of normality in the residuals suggests the possibility of non-uniformity in the underlying distribution of log differences. To assess this possibility, the paired sample t-tests and non-parametric signed rank tests were performed on subsets of the overall stations with the following outcomes:

- Rural sites provided significant statistical evidence (adjusted p-value < 0.001) that counts provided by passive measurement were an average of 4.3 percent higher (95% confidence interval [2.2%, 6.4%]) compared to those from traditional methods.

- Urban sites provided significant statistical evidence (adjusted p-value < 0.001) that counts provided by passive measurement were an average of 9.3 percent higher (95% confidence interval [6.6%, 12.1%]) compared to those from traditional methods.

From the perspective of urban and rural sites, greater geometric mean AADT overcounts were found by passive measurement in urban sites than rural sites, though both sub-groups evidenced statistically significant overcounts compared to traditional methods. The samples for these comparisons were all large enough that the lack of normality in the residuals did not preclude using the t-test results. However, the non-parametric comparisons were also made and showed the same statistical significance, though of slightly smaller magnitude.

Five of seven functional classes contained adequate data to compare passive and traditional counts. For all five individual classes, no significant differences (i.e., all adjusted p-values > 0.05) were found between passive and traditional counts. The smaller sample sizes for these individual class comparisons produced wider confidence intervals, but in all cases the sample sizes were still large enough (i.e., >30) to justify using parametric test results even though the residuals did not support Normality (except for functional class 2). In each case, the non-parametric test comparison of medians confirmed the lack of a statistically significance conclusion. Despite the lack of statistical significance, it is still noted that the estimated ratio of geometric means for passive measurement exceeded 1.0 for functional classifications 2, 4, and 5, was below 1.0 for functional classification 3, and was almost identically equal to 1.0 for functional classification 1. This suggests functional classification-based sensitivity to the geometric mean accuracy of the passive counts. The toll sites, which were not identified by functional classification, were analyzed separately, and resulted in a highly significant difference (adjusted p-value < 0.001) with t-test result estimating a 28.6 percent greater geometric mean AADT from the passive method than traditional counts (95% confidence interval [24.5%, 32.9%]). The non-parametric results agreed closely for the median of the log differences.

When analyzed by volume ranges of the traditional counts, the results showed significantly higher AADTs for the passive method in all three coarse volume ranges compared to the traditional counts, but the magnitude of the differences diminished with traditional count.

- For AADT less than 5,000, the passive counts were estimated to be 9.2 percent higher (95% confidence interval [4.3%, 14.3%]).
- For AADT of 5,000 to less than 55,000, the passive counts were estimated to be 6.4 percent higher (95% confidence interval [4.3%, 8.5%]).
- For AADT of 55,000 and above, the passive counts were estimated to be 4.1 percent higher (95% confidence interval [1.9%, 6.3%]).

The non-parametric signed rank test results were similarly significant for all three volume ranges and diminishing with increased traditional volume, but the estimated ratio of medians was of a smaller magnitude than the parametric test ratio of geometric means. When the volume range comparisons were expanded to ten finer volume ranges, the results were far less clear, with only three of ten volume ranges showing significantly higher counts for the passive method by the t-test comparison, and only five of the ten by the non-parametric analysis. In every case, though, the estimated ratio of geometric means, though not statistically significant under the multiple comparison adjusted p-values, was nevertheless a value greater than 1.0, corresponding to higher AADT for the passive method.

Though the results clearly establish evidence of statistically significantly higher AADT estimates from the passive method as compared to the traditional counts, the exact magnitude of this difference was shown to be sensitive to whether applied to rural or urban roads, toll roads, and broad traditional count ranges. To the extent that the sites selected for this comparison are generally representative of many roadways in multiple states across the U.S., the conclusion of higher AADT is likely sound, but without application of some sort of weighting measures, the specific value of 6.6 percent (95% confidence interval [4.9%,8.3%]) is only strictly applicable for a system of roads with similar proportions of road types and volume ranges as identified in Tables 2 and 3.

A statistical power analysis provides some insight into why the aggregate results may be showing statistically significant differences, while the smaller subgroup analyses do not. For sample sizes of 800 (consistent with the 827 for this analysis) and based on a true standard deviation in the paired log-10 AADT differences of 0.1, consistent with what was observed in the overall dataset, true passive to traditional population differences of as small as 3 percent would be highly likely to produce sample data that correctly concluded, with 95% probability, that a significant difference exists. Therefore, while it cannot be known if the 6.6 percent national difference is close to the true value, it well exceeds the statistical power threshold of 3 percent. Some of the sub-sample analyses consisted of more on the order of 100 observations or less. From a statistical power perspective, true population differences of as much as 9 percent would be required to have high probability of producing a sample that correctly confirmed, with 95% probability, that a significant difference exists, under the same assumed standard deviation of 0.1 for the paired log-10 AADT differences. Since this 9-percent difference is larger than many of the observed differences, even though it cannot be directly compared to them, it still provides rationale for why the smaller sub-samples estimated similar differences to the national and large sample (i.e., urban/rural, coarse volume range) results, but failed to conclude with high statistical significance that the geometric mean passive counts exceeded the traditional counts.

U.S. Department of Transportation
Federal Highway Administration

U.S. Department of Transportation
Federal Highway Administration
Office of Highway Policy Information
1200 New Jersey Ave., SE
Washington, D.C. 20590
https://www.fhwa.dot.gov/policyinformation
November 2021
FHWA-PL-021-040