UNC SCHOOL OF GOVERNMENT

PUBLIC MANAGEMENT BULLETIN NO. 20 | OCTOBER 2020

Performance Management: The Value of Service Delivery Outcomes

Obed Pasha and William C. Rivenbark

CONTENTS

Introduction

Outcome measures are fundamental to performance measurement systems in local government, representing the primary means for reporting on the service dimension of effectiveness. The performance measurement process begins with the department's mission, followed by service delivery goals and SMART (specific, measurable, aggressive, result-oriented, and time-bound) objectives. The next step is to identify the performance measures (outputs, efficiencies, and outcomes) that align with each SMART objective.[1]

There are several advantages to this process. First, performance measures provide feedback on the overall success of the department because they derive from the department's mission, goals, and objectives. This process also increases the likelihood of moving beyond output measures in local government, where SMART objectives often define performance expectations in the form of outcomes. In other words, meaningful performance measurement systems monitor performance accountability over time, providing information on the critical dimensions of efficiency and effectiveness.

Obed Pasha is an assistant professor of public management at the School of Government. He specializes in performance management and strategic planning.

William C. Rivenbark is professor of public administration and government and director of the MPA Program at the School of Government. He specializes in performance and financial management.

1. David N. Ammons, *Performance Measurement for Managing Local Government* (Irvine, CA: Melvin & Leigh, 2020), 40-49.

In response to the reinventing government movement of the early 1990s,[2] local officials began to pay more attention to how to use performance data for making decisions and for performance improvement. This change in focus is often referred to as the paradigm shift from performance measurement to performance management in local government, which only increased the importance of outcome measures.

Research has clearly demonstrated that local officials are more likely to use performance data for decision-making and performance improvement when they move from reporting on output measures to outcome measures of service delivery.[3] Therefore, the value of service delivery outcomes is embedded within the concepts of both performance measurement and performance management in local government.

This bulletin explores how outcome measures may play an even broader role in local government beyond performance accountability and performance improvement, where higher levels of performance are more likely to be sustained over time when monitored with outcomes of service delivery. We begin our exploratory research with an overview of the Federal Transit Administration (FTA) grant program for small transit intensive cities (STIC), in which grant funds are awarded based on six performance measures that include service provision measures (outputs) and utilization measures (outcomes). We present data that show how these small transit cities increased both their outputs and outcomes in the short term to access the grant funds, which would be expected. These data also show that these same transit cities sustained their high levels of performance in regard to service outcomes over the long term even though their service outputs began to decrease, demonstrating the utility of outcome measures beyond decision-making and process improvement.

Grant Program

The Federal Transit Administration funds local public transit authorities according to the following three categories: small transit systems with area populations of under 200,000; mid-sized systems with area populations between 200,000 and 1 million; and large transit systems with populations of over 1 million. The FTA formula for funding mid-sized and large transit systems is tied to demographic indicators such as population and density in addition to the service indicators of vehicle miles, passenger miles, and ridership. Therefore, the transit systems from these two categories that offer more services to highly and more densely populated areas

2. David Osborne and Ted Gaebler, “Reinventing Government,” *Journal of Leisure Research* 27, no. 3 (1995): 302.

3. William C. Rivenbark, Roberta Fasiello, and Stefano Adamo, “Exploring Performance Management in Italian Local Government: The Necessity of Outcome Measures and Citizen Participation,” *The American Review of Public Administration* 49, no. 5 (2019): 545–53; David N. Ammons and Dale J. Roenigk, “Benchmarking and Organizational Learning in Local Government,” *Journal of Public Administration Research and Theory* 25, no. 1 (2015): 309–35; Alexander Kroll, “Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research,” *Public Performance and Management Review* 38, no. 3 (2015): 459–86; Donald P. Moynihan and Sanjay K. Pandey, “The Big Question for Performance Management: Why Do Managers Use Performance Information?,” *Journal of Public Administration Research and Theory* 20, no. 4 (2010): 849–66; David N. Ammons and William C. Rivenbark, “Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina Benchmarking Project,” *Public Administration Review* 68, no. 2 (2008): 304–18.

receive more funding. The FTA formula for funding small transit systems is different, being tied only to the demographic indicators of population and density. This approach presents a challenge to small transit systems that operate in less populated and less dense areas but provide higher levels of service in terms of miles and ridership.

The FTA Small Transit Intensive Cities program was established in 2005 in response to this problem. It provides smaller systems with more funds when they demonstrate mean service demands greater than the mean service demands of mid-sized systems. This additional funding formula was based on the following two assumptions: mid-sized transit systems have higher service demands than small transit systems, and small transit systems should receive additional federal funds when they exceed mean service demands of their larger counterparts. Table 1 contains the six indicators selected by the FTA to guide this new funding process. These indicators can be classified into the categories of service provision (revenue hours per capita and revenue miles per capita), service utilization (unlinked passenger trips per capita and passenger miles per capita), and system productivity (passenger miles per revenue hour and passenger miles per revenue mile).

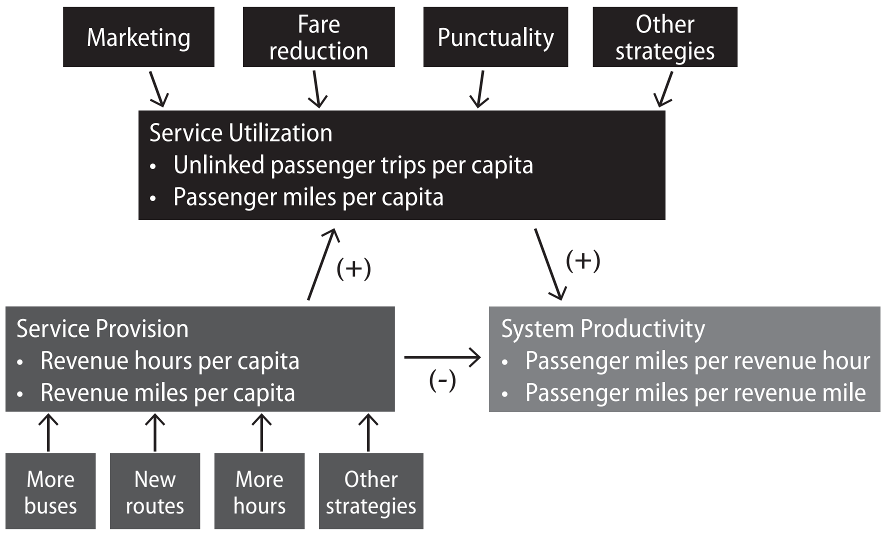
Table 1. Performance indicators and Definitions

Performance Indicators	Definition
Revenue hours per capita	The cumulative number of hours all vehicles (buses, streetcars, trains, paratransit, etc.) were in operation for a given service area population
Revenue miles per capita	The cumulative number of miles all vehicles traveled for a given service area population
Unlinked passenger trips per capita	The total number of passenger-boarding on each vehicle for a given service area population
Passenger miles per capita	The cumulative number of miles traveled by passengers for a given service area population
Passenger miles per revenue hour	The number of miles traveled by passengers for a given hour that transit vehicles were in service
Passenger miles per revenue mile	The number of miles traveled by passengers for a given mile that transit vehicles were in service

To be eligible for these additional funds, the transit system must be registered with the FTA and must serve a population area of under 200,000. STIC's funding formula is then designed to award small transit systems with more funds as they exceed the mean of each additional indicator. A small transit system exceeding the mean of mid-sized systems on any one indicator received an additional $180,000 in 2013, for example, and could receive over $1 million if it met the criteria for all six indicators. The amount received through this additional funding mechanism has transformative potential for small transit systems, given that their average 2013 budget was approximately $7 million. Moreover, the number of small transit systems receiving STIC grant funds that met the criterion of at least one indicator increased from 79 in 2005 to 128 in 2013, representing 58 percent of all small transit systems.

The small transit systems focused primarily on service provision and service utilization measures, as these underlie the two system-productivity measures (passenger miles per revenue hour and passenger miles per revenue mile). The small transit systems used the strategies of more buses, new routes, and more hours, for example, to increase the service provision measures of revenue hours per capita and revenue miles per capita. In other words, additional resources (inputs) were used to impact these output measures. The small transit systems then turned to strategies such as marketing, fare reduction, and punctuality to increase the service utilization measures unlinked passenger trips per capita and passenger miles per capita, representing outcome measures of service delivery. Figure 1 contains a logic model for transit systems to provide more insight into these relationships.

Figure 1. Logic Model for Transit Systems



Case

We compared the performance of 221 small transit systems with 155 mid-sized transit systems from 1997 to 2013 to determine the impact of the Small Transit Intensive Cities program. More specifically, we examined the respective performance trends of the treatment group (small transit systems) against those of the control group (mid-sized systems) eight years before and after the implementation of STIC (intervention) in 2005. One aspect of this methodology is important to interpreting the results. If the pre-intervention trends of the treatment and control groups were approximately the same, any divergence in the post-intervention trends between the treatment and control groups can be attributed to STIC. The appendix contains the statistical model and analysis that support our observations. Again, our purpose is to explore how service utilization measures may have played an even broader role in local government behavior, where higher levels of performance are likely to be sustained over time when monitored with outcomes of service delivery.

Figure 2 shows the service provision measure of average revenue hours per capita (output) for both the small transit systems (solid line) and the mid-sized transit systems (dotted line) from 1997 to 2013, with the x-axis representing time and the y-axis representing performance. The first step is to understand the variance between the two groups in the pre-intervention phase. The performance gap between the small and mid-sized transit systems was increasing on an annual basis from 2000 to 2004 after remaining fairly consistent during the prior years; however, this annual increase can be characterized as gradual with the exception of 2004 to 2005, which brings us to the implementation of STIC.

Figure 2. Revenue Hours per Capita

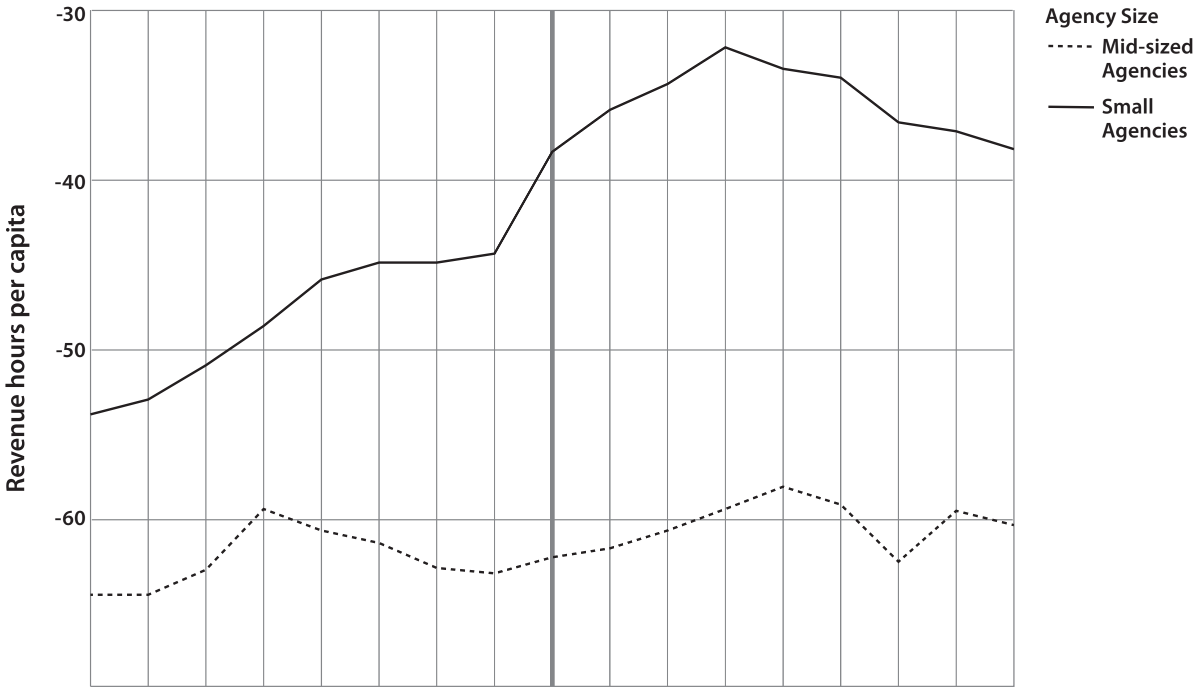
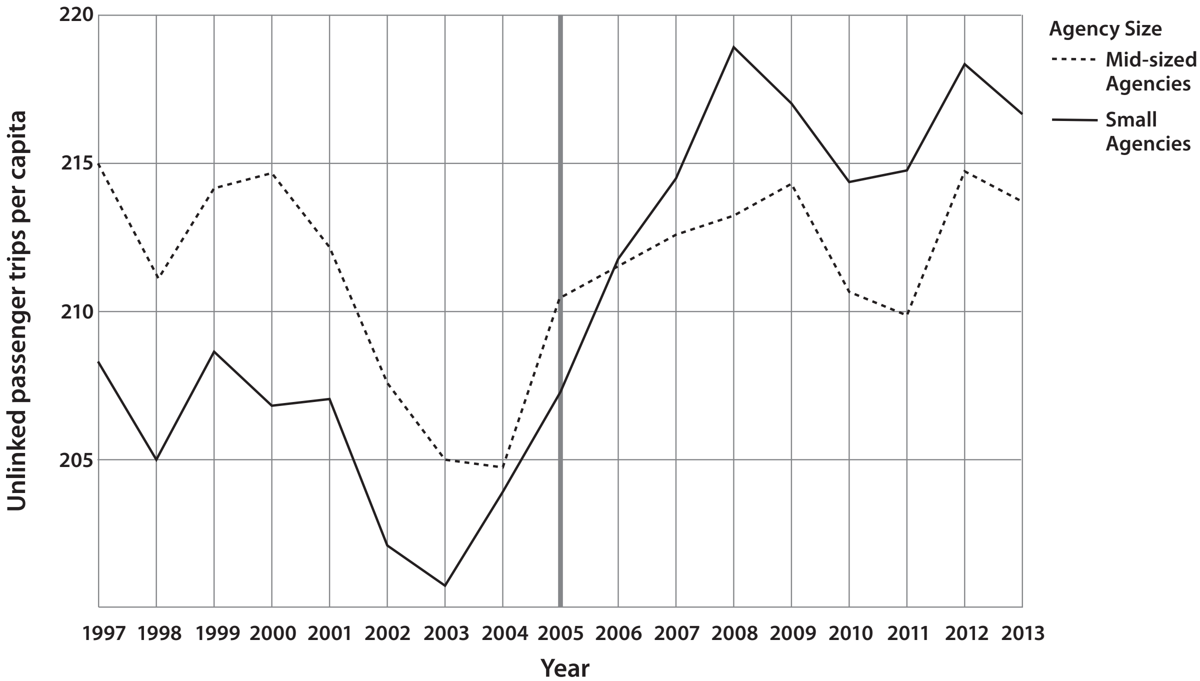


Figure 2 then shows that the performance gap between the two groups increased significantly between 2005 and 2008 where, on average, small transit systems increased their revenue hours per capita at a much faster rate in response to the new STIC funding formula, as compared to the average of mid-sized transit systems. However, the trend began to decline after 2008 for the treatment group. One plausible explanation for this decline is that small transit systems began to reduce their inputs during this period, responding to the great recession. Another plausible explanation is that these same systems realized that they could decrease their inputs while still operating well above the average of mid-sized transit systems and still receiving the additional grant funds. We also found that the pre-intervention and post-intervention performance gaps for the service provision measure of average revenue miles per capita (output) were basically the same for both the treatment and control groups. Again, the average revenue miles per capita for small transit systems experienced a sharp increase from 2004, the year before STIC, to 2008 before decreasing over the next several years.

Figure 3 shows the service utilization measure of average unlinked passenger trips per capita (outcome) for both the small transit systems (solid line) and the mid-sized transit systems (dotted line) from 1997 to 2013, where the pre-intervention trends were basically the same for the treatment and control groups. However, average unlinked passenger trips per capita for small transit systems increased substantially from approximately 200 in 2003, two years before the intervention, to approximately 220 in 2008 before experiencing some minor fluctuations over the next several years. In contrasting this outcome measure to the previous two output measures, we can make two observations. Like the output measures, the intervention of STIC did have an impact on the outcome measure of average unlinked passenger trips per capita. Unlike the output measures, small transit systems were more likely to sustain higher levels of performance over the next several years.

Figure 3. Unlinked Pasenger Trips per Capita



The comparison between the service utilization measure of average passenger miles per capita (outcome) and the previous two output measures is even more telling. Figure 4 shows that the performance gap between the treatment and control groups for this outcome measure remained basically the same between 1997 and 2004. However, the average passenger miles per capita for small transit systems steadily increased from 2005 to 2013, representing an overall increase of 8.5 percent. In contrast, the average passenger miles per capita for mid-sized transit systems experienced an increase of approximately 3 percent during the same period.

The first observation that can be made from these trends is that the service utilization measure of average passenger miles per capita for small transit systems remains below mid-sized transit systems, impacting the treatment group's ability to access additional funds from STIC. The second is that small transit systems have continued to emphasize this critical outcome measure since 2004, the year before the STIC intervention.

Figure 4. Passenger Miles per Capita

Lessons Learned

These results present several lessons that can be applied to performance measurement and performance management in local government. We begin with the process of designing meaningful performance measurement systems. Without question, performance should be placed within the context of SMART (specific, measurable, aggressive, result-oriented, and time-bound) objectives. As discussed earlier, this approach facilitates the identification of outcome measures. SMART objectives also include performance targets, including the desired level of performance (for example, 75 percent of individuals who complete the job training program will be employed within six months). The Federal Transit Administration grant program included requirements beyond the six performance measures, such as performance targets, which small transit systems had to exceed to obtain additional grant funding. Therefore, establishing SMART objectives increases the probability of obtaining a higher level of performance across organizational programs and services.

Performance management, as previously discussed, represents the actual use of performance data to make management and policy decisions for service improvement. The information presented here suggests that local officials used output measures in the short term to make management and policy decisions in regard to providing more services, which may or may not lead to higher levels of performance. However, this case also would suggest that local officials used outcome measures in the short and long term to implement strategies for higher levels of performance regarding the service utilization measures of unlinked passenger trips per capita and passenger miles per capita, which also ultimately increased the system productivity measures of passenger miles per revenue hour and passenger miles per revenue mile. In other words, the FTA grant program demonstrates once again that local officials are more likely to use performance data for service improvement when they move beyond output measures to outcome measures of service delivery.

The grant program discussed here presents another lesson that has not been adequately addressed in the academic and professional literature. Outcome measures have the potential to play an even broader role in local government beyond performance accountability and performance improvement, where higher levels of performance are likely to be sustained over time when monitored with outcomes of service delivery. The performance data from the small transit systems clearly show that higher levels of performance as measured with outcomes were maintained even after the outputs began to decline. A plausible explanation for this dynamic is that output measures typically support decisions regarding input changes, as demonstrated by this case with the implementation of more buses, routes, and hours. Outcome measures require strategies that focus on changing human behavior within the context of transit systems to increase and maintain unlinked passenger trips and miles per capita over extended periods of time.

Summary

This bulletin explores how outcome measures may play a role in sustaining higher levels of performance once these levels are obtained by a program or service. Based on the Federal Transit Administration's Small Transit Intensive Cities Program, implemented in 2005, we present performance data that show how small transit systems increased their outputs in the short term to obtain additional grant proceeds but decreased these same outputs in the long term. However, these same systems increased their outcomes in the short term and continued to operate at the higher performance levels in the following years. We reached three conclusions as a result. The first two expand on the current literature, where SMART (specific, measurable, aggressive, result-oriented, time-bound) objectives are critical to meaningful performance measurement systems and outcome measures increase the likelihood that local officials embrace performance management for decision-making and performance improvement. The third, which comes directly from this transit case, is that higher levels of service are likely to be sustained over time when outcome measures are used to identify and implement strategies of performance improvement.

References

Ammons, David N. *Performance Measurement for Managing Local Government*. Irvine, CA: Melvin & Leigh, 2020.

Ammons, David N., and Rivenbark, William C. "Factors Influencing the Use of Performance Data to Improve Municipal Services: Evidence from the North Carolina Benchmarking Project." *Public Administration Review* 68, no. 2 (2020): 304–18.

Ammons, David N., and Roenigk, Dale J. "Benchmarking and Interorganizational Learning in Local Government." *Journal of Public Administration Research and Theory* 25, no. 1 (2015): 309–35.

Kroll, Alexander. "Drivers of Performance Information Use: Systematic Literature Review and Directions for Future Research." *Public Performance and Management Review* 38, no. 3 (2015): 459–86.

Moynihan, Donald P., and Pandey, Sanjay K. "The Big Question for Performance Management: Why Do Managers Use Performance Information?" *Journal of Public Administration Research and Theory* 20, no. 4 (2010): 849–66.

Osborne, David, and Gaebler, Ted. "Reinventing Government." *Journal of Leisure Research* 27, no. 3 (1995): 302.

Rivenbark, William C., Fasiello, Roberta, and Adamo, Stephano. "Exploring Performance Management in Italian Local Government: The Necessity of Outcome Measures and Citizen Participation." *The American Review of Public Administration* 49, no. 5 (2019): 545–53.

Appendix. Statistical Model and Results

The statistical model estimating the change in the level of mean performance of transit systems after the intervention of the Small Transit Intensive Cities program is specified as follows:

$$100 * ln(Y_{it}) = \alpha + \eta D_i + \pi P_t + z(D_i * P_t) + \partial X_{it} + \lambda_i + \delta_t + \varepsilon_{it.}$$

In the equation, $ln(Y_{it})$ is the natural log of performance of transit systems i in year t, where t is a continuous variable centered at the intervention point of 2005 (t = 0). The minimum value for t = -8 for 1997 and the maximum value for t = 8 for 2013. The dichotomous indicator D_i is coded as D_i = 1 if agency i is in the treatment group (ever treated) and D_i = 0 if agency i is in the control group (never treated). All small transit systems with populations below 200,000 are in the treatment group, and all transit systems with populations between 200,000 and 1 million are in the control group.

The variable α estimates the mean performance of transit systems in the comparison group in the pre-intervention period. Post-intervention is represented by P_t, a dichotomous indicator coded as P_t = 0 for all observations of the systems before 2005 (intervention year) and P_t =1 for all observations of the systems occurring in 2005 and beyond. The variable η estimates the difference in mean performance between the treatment and the control groups, while z estimates the difference in mean performance for untreated systems before and immediately after STIC. The coefficient of interest is z, estimating the change in the level of mean performance for systems in the treatment group (small transit systems) after the intervention when compared to the change in the level of the comparison group (mid-sized transit systems).

In the equation, X_{it} is a vector of control variables consisting of the natural logs of population density (in thousands of individuals per square mile), fleet size (number of vehicles available for service), and fare-to-expense ratio as a proxy for external dependence. These variables help control for the change in task difficulty and resources over time and are expected to have a positive impact on performance. Data on control variables are derived from the National Transit Database. Other controls for demographics could not be used because transit agencies frequently traverse city, county, and, in some cases, state boundaries.

The year fixed effects and agency-level fixed effects are represented by λ_i and δ_t, respectively, and ε_{it} represents the unobserved random error. A two-way autoregressive fixed-effects model is adopted, where the standard errors of each agency are clustered to produce unbiased standard errors and correctly sized confidence intervals. Agency-level fixed effects help control for department-specific covariates such as cultural factors and management practices across agencies. Year fixed effects account for annual shocks in the time series data, such as changes in federal policies and regulations, economic shocks, and changes in national attitudes toward public transportation.

Table 2 shows that revenue hours per capita and revenue miles per capita increased by 9.542 and 9.702 percent, respectively, for the treatment group in the post-intervention period compared to the comparison group. Unlinked passenger trips per capita and passenger miles per capita also increased by 11.20 and 15.55 percent, respectively. Similarly, an increase in passenger miles per revenue hour and passenger miles per revenue mile are significant at 6.591 and 6.201 percent, respectively. STIC is shown to improve performance even when it was implemented to support existing service levels. Although STIC was created as a needs-based grant program by the Federal Transit Administration, the funding structure incentivized performance improvement of grantees due to process similarities with performance-based grants systems. Thus, these results support earlier studies that found a positive impact of such systems on organizational performance.

Table 2. Difference-in-Differences Estimates of STIC Impact on Small and Mid-sized Transit Systems

	Service Provision		Service Utilization		System Productivity	
	Revenue hours per capita	Revenue miles per capita	Unlinked passenger trips per capita	Passenger miles per capita	Passenger miles per revenue hour	Passenger miles per revenue mile
	(1)	(2)	(3)	(4)	(5)	(6)
Post-STIC	-3.766 (3.178)	-3.054 (3.380)	-12.310*** (4.206)	-7.569 (5.100)	-4.606 (3.643)	-4.679 (3.393)
Treat x post-STIC	9.542*** (2.648)	9.702*** (2.651)	11.20*** (3.833)	15.55*** (4.627)	6.591* (3.638)	6.201* (3.499)
Population density	0.053** (0.026)	0.042 (0.026)	0.075** (0.030)	0.072** (0.032)	0.020 (0.020)	0.032** (0.016)
Fleet size	38.320*** (4.842)	42.850*** (4.699)	41.000*** (8.640)	51.890*** (9.653)	15.120** (6.660)	10.530 (6.760)
Fare ratio	-2.198 (3.515)	-0.987 (3.243)	8.162** (3.458)	8.766 (5.390)	10.85*** (3.321)	9.307*** (3.157)
Constant	-255.100*** (32.080)	2.536 (32.440)	7.488 (46.260)	115.500** (50.070)	363.300*** (28.840)	104.800*** (28.220)
Fixed effects						
Year	Yes	Yes	Yes	Yes	Yes	Yes
Agency	Yes	Yes	Yes	Yes	Yes	Yes
N	5400	5403	5388	5252	5247	5250
R^2 *(within)*	0.192	0.208	0.141	0.139	0.052	0.047
Agencies	379	379	379	379	379	379

*Note: Dependent variables: 100*ln (performance indicator). Robust standard errors in parentheses (Huber-White).*
$*p < 0.05$
$**p < 0.01$
$***p < 0.001$