



A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry

DETAILS

110 pages | | PAPERBACK

ISBN 978-0-309-15482-6 | DOI 10.17226/14402

AUTHORS

Jamie Parks; Paul Ryus; Kathryn Coffel; Albert Gan; Victoria Perk; Linda Cherrington; Jeffrey Arndt; Yuko Nakanishi; Transportation Research Board

BUY THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

TCRP REPORT 141

**A Methodology for Performance
Measurement and Peer
Comparison in the Public
Transportation Industry**

**Paul Ryus
Kathryn Coffel
Jamie Parks**
KITTELSON & ASSOCIATES, INC.
Portland, Oregon

Victoria Perk
CENTER FOR URBAN TRANSPORTATION RESEARCH, UNIVERSITY OF SOUTH FLORIDA
Tampa, Florida

**Linda Cherrington
Jeffrey Arndt**
TEXAS TRANSPORTATION INSTITUTE, THE TEXAS A&M UNIVERSITY SYSTEM
Houston, Texas

Yuko Nakanishi
NAKANISHI RESEARCH & CONSULTING, LLC
New York, New York

Albert Gan
LEHMAN CENTER FOR TRANSPORTATION RESEARCH, FLORIDA INTERNATIONAL UNIVERSITY
Miami, Florida

Subscriber Categories

Public Transportation • Administration and Management • Planning and Forecasting

Research sponsored by the Federal Transit Administration in cooperation with the Transit Development Corporation

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2010
www.TRB.org

TRANSIT COOPERATIVE RESEARCH PROGRAM

The nation's growth and the need to meet mobility, environmental, and energy objectives place demands on public transit systems. Current systems, some of which are old and in need of upgrading, must expand service area, increase service frequency, and improve efficiency to serve these demands. Research is necessary to solve operating problems, to adapt appropriate new technologies from other industries, and to introduce innovations into the transit industry. The Transit Cooperative Research Program (TCRP) serves as one of the principal means by which the transit industry can develop innovative near-term solutions to meet demands placed on it.

The need for TCRP was originally identified in *TRB Special Report 213—Research for Public Transit: New Directions*, published in 1987 and based on a study sponsored by the Urban Mass Transportation Administration—now the Federal Transit Administration (FTA). A report by the American Public Transportation Association (APTA), *Transportation 2000*, also recognized the need for local, problem-solving research. TCRP, modeled after the longstanding and successful National Cooperative Highway Research Program, undertakes research and other technical activities in response to the needs of transit service providers. The scope of TCRP includes a variety of transit research fields including planning, service configuration, equipment, facilities, operations, human resources, maintenance, policy, and administrative practices.

TCRP was established under FTA sponsorship in July 1992. Proposed by the U.S. Department of Transportation, TCRP was authorized as part of the Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA). On May 13, 1992, a memorandum agreement outlining TCRP operating procedures was executed by the three cooperating organizations: FTA, the National Academies, acting through the Transportation Research Board (TRB); and the Transit Development Corporation, Inc. (TDC), a nonprofit educational and research organization established by APTA. TDC is responsible for forming the independent governing board, designated as the TCRP Oversight and Project Selection (TOPS) Committee.

Research problem statements for TCRP are solicited periodically but may be submitted to TRB by anyone at any time. It is the responsibility of the TOPS Committee to formulate the research program by identifying the highest priority projects. As part of the evaluation, the TOPS Committee defines funding levels and expected products.

Once selected, each project is assigned to an expert panel, appointed by the Transportation Research Board. The panels prepare project statements (requests for proposals), select contractors, and provide technical guidance and counsel throughout the life of the project. The process for developing research problem statements and selecting research agencies has been used by TRB in managing cooperative research programs since 1962. As in other TRB activities, TCRP project panels serve voluntarily without compensation.

Because research cannot have the desired impact if products fail to reach the intended audience, special emphasis is placed on disseminating TCRP results to the intended end users of the research: transit agencies, service providers, and suppliers. TRB provides a series of research reports, syntheses of transit practice, and other supporting material developed by TCRP research. APTA will arrange for workshops, training aids, field visits, and other activities to ensure that results are implemented by urban and rural transit industry practitioners.

The TCRP provides a forum where transit agencies can cooperatively address common operational problems. The TCRP results support and complement other ongoing transit research and training programs.

TCRP REPORT 141

Project G-11
ISSN 1073-4872
ISBN 978-0-309-15482-6
Library of Congress Control Number 2010930498

© 2010 National Academy of Sciences. All rights reserved.

COPYRIGHT INFORMATION

Authors herein are responsible for the authenticity of their materials and for obtaining written permissions from publishers or persons who own the copyright to any previously published or copyrighted material used herein.

Cooperative Research Programs (CRP) grants permission to reproduce material in this publication for classroom and not-for-profit purposes. Permission is given with the understanding that none of the material will be used to imply TRB, AASHTO, FAA, FHWA, FMCSA, FTA, or Transit Development Corporation endorsement of a particular product, method, or practice. It is expected that those reproducing the material in this document for educational and not-for-profit uses will give appropriate acknowledgment of the source of any reprinted or reproduced material. For other uses of the material, request permission from CRP.

NOTICE

The project that is the subject of this report was a part of the Transit Cooperative Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical panel selected to monitor this project and to review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical panel and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The Transportation Research Board of the National Academies, the National Research Council, and the sponsors of the Transit Cooperative Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.

Published reports of the

TRANSIT COOPERATIVE RESEARCH PROGRAM

are available from:

Transportation Research Board
Business Office
500 Fifth Street, NW
Washington, DC 20001

and can be ordered through the Internet at
<http://www.national-academies.org/trb/bookstore>

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. www.TRB.org

www.national-academies.org

COOPERATIVE RESEARCH PROGRAMS

CRP STAFF FOR TCRP REPORT 141

Christopher W. Jenks, *Director, Cooperative Research Programs*
Crawford F. Jencks, *Deputy Director, Cooperative Research Programs*
Dianne Schwager, *Senior Program Officer*
Sagar Gurung, *Senior Program Assistant*
Eileen P. Delaney, *Director of Publications*
Doug English, *Editor*

TCRP PROJECT G-11 PANEL

Field of Administration

Jeanne Krieg, *Eastern Contra Costa Transit Authority, Antioch, CA (Chair)*
Mark R. Aesch, *Rochester Genesee Regional Transportation Authority, Rochester, NY*
Jerry R. Benson, *Utah Transit Authority, Salt Lake City, UT*
John Dockendorf, *Pennsylvania DOT, Harrisburg, PA*
Fred M. Gilliam, *Capital Metropolitan Transportation Authority, Austin, TX*
Ronald Kilcoyne, *Greater Bridgeport Transit Authority, Bridgeport, CT*
Anthony M. Kouneski, *AMK & Associates, Kensington, MD*
William Lyons, *Research and Innovative Technology Administration, Cambridge, MA*
Clarence W. "Cal" Marsella, *Denver Regional Transportation District, Denver, CO*
Theodore H. Poister, *Andrew Young School of Policy Studies, Atlanta, GA*
Alan M. Warde, *New York State DOT, Albany, NY*
Nigel H. M. Wilson, *Massachusetts Institute of Technology, Cambridge, MA*
Fred L. Williams, *FTA Liaison*
Martine A. Micozzi, *TRB Liaison*

AUTHOR ACKNOWLEDGMENTS

The research reported herein was performed under TCRP Project G-11 by Kittelson & Associates, Inc. (prime contractor), assisted by the Center for Urban Transportation Research at the University of South Florida; Texas Transportation Institute, the Texas A&M University System; Nakanishi Research & Consulting, LLC; and the Lehman Center for Transportation Research at Florida International University. Paul Ryus of Kittelson & Associates, Inc., was the principal investigator.

Victoria Perk of the Center for Urban Transportation Research led the project's agency outreach efforts and coordinated the project's domestic literature review, assisted by Mark Mistretta. Dr. Steve Polzin and Dr. Xuehao Chu of the Center for Urban Transportation Research provided review comments on early versions of the peer-grouping and performance-measurement methodology.

Linda Cherrington and Jeffrey Arndt of the Texas Transportation Institute developed performance measures and applications relating to small urban transit agencies and state departments of transportation (DOTs), coordinated the testing of the methodology by state DOTs, and contributed review comments throughout the project.

Dr. Yuko Nakanishi of Nakanishi Research & Consulting, LLC, prepared the sections on benchmarking in the private and public sectors, coordinated the testing of the methodology by several transit agencies in the northeastern United States, and contributed review comments throughout the project.

A number of staff from Kittelson & Associates, Inc., and Kittelson & Associates, LLC (Australia) contributed to the project. Kathryn Coffel coordinated the work activities of project team members, led the development of peer-comparison applications, and provided review comments throughout the project. Jamie Parks tested a variety of performance measures being considered for the peer-grouping methodology, helped develop a spreadsheet version of the methodology used for initial testing (with assistance from Jean Doig), prepared examples of transit agency reporting techniques, and coordinated some of transit agency tests of the methodology. Dr. Miranda Blogg coordinated the project's international literature review. Adam Vest, Severine Marechal, and Conor Semler coordinated some of the transit agency tests.

The Lehman Center for Transportation Research at Florida International University integrated the peer-grouping methodology into the online Florida Transit Information System (FTIS) software. Dr. Albert Gan led this effort, with Dr. Feng Gui performing the programming. Dr. Fabian Cevallos provided review comments on early versions of the methodology. The research team thanks the Florida Department of Transportation, which sponsors the FTIS software, for the partnership that allowed the project's peer-grouping methodology to be added to FTIS.

The project team thanks the numerous organizations and persons that participated in the project's outreach efforts, and particularly those who participated in the two rounds of testing of the project's peer-grouping and performance-measurement methodology. The project team also thanks Kjetil Vrenne, BEST Project Manager, Enable and Michael Skov, Senior Consultant, Movia for facilitating access to the Benchmarking in European Service of public Transport (BEST) database. Finally, the feedback provided by the TCRP Project G-11 panel throughout the project is gratefully acknowledged.



FOREWORD

By Dianne Schwager

Staff Officer

Transportation Research Board

TCRP Report 141: A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry is an important resource that will be of interest to transit managers, decision-makers, and others interested in using performance measurement and benchmarking as tools to (1) identify the strengths and weaknesses of their organization, (2) set goals or performance targets, and (3) identify best practices to improve performance.

This research developed and tested a methodology for performance measurement and peer comparison for (a) all fixed-route components of a public transit system, (b) the motor-bus mode specifically, and (c) major rail modes specifically (i.e., light rail, heavy rail, and commuter rail). This report complements *TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System*, which describes how to implement and use performance measurement on an ongoing basis at a transit agency.

This report describes eight steps for conducting a benchmarking effort. The steps are:

1. Understand the context of the benchmarking exercise,
2. Identify standardized performance measures appropriate to the performance question being asked,
3. Establish a peer group,
4. Compare performance within the peer group,
5. Contact best-practices peers in areas where one's performance can be improved,
6. Develop a strategy for improving performance based on what one learns from the best-practices peers,
7. Implement the strategy, and
8. Monitor changes in performance over time, repeating the process if the desired results are not achieved within the desired time frame.

The performance-measurement and peer-comparison methodology described in this report incorporates a variety of nationally available, standardized factors into the peer-selection process and describes ways for also incorporating policy objectives and other factors into the process. The methodology has been incorporated into a freely available, online software tool (the Florida Transit Information System, FTIS) that provides access to the full National Transit Database (NTD), allowing users to quickly identify a group of potential peer transit agencies, retrieve standardized performance data for them, and perform a variety of comparisons.

During the research the methodology was tested by transit agencies, which were typically able to learn how to use the software, create a peer group, and perform an analysis with 16 person-hours of effort or less. This project's testing efforts found that, for the most part, the NTD data used in analyses were reliable and that what errors did exist were readily spotted.

This report provides guidance on selecting performance measures appropriate to a particular performance question but does not prescribe a particular set of measures. This approach requires some thoughtfulness on the part of transit agencies in selecting measures, but also provides much-needed flexibility that allows the methodology to be applied to a wide variety of transit modes, transit agency sizes, and performance questions.

The methodology was not designed as a means of ranking transit agencies to determine the “best” agencies overall on a national basis or the best at a particular aspect of service. Rather, this report’s approach is that peer-grouping and performance measurement should serve as a starting point for a transit agency to ask questions about performance, identify areas of possible improvement, and contact top-performing peers. That course—a true benchmarking process—holds the greatest potential for producing long-term performance improvement.

A full-color PDF version of this report is available on the TRB website (www.trb.org) by searching for “TCRP Report 141.”

CONTENTS

1	Summary
4	Chapter 1 Introduction
4	Research Problem Statement
4	Research Objective and Scope
5	Research Approach
5	How to Use This Report
6	Chapter 2 Performance Measurement, Peer Comparison, and Benchmarking
7	Benchmarking in the Private Sector
7	Benchmarking in the Public Sector
8	Benchmarking in the Public Transit Industry
16	Levels of Benchmarking
17	Benchmarking Success Factors
19	Benefits of and Challenges with Transit Peer Comparisons
20	Lessons Learned
22	Chapter 3 Applications and Performance Measures
22	Applications
23	Performance Measures
30	Chapter 4 Benchmarking Methodology
30	Introduction
31	Step 1: Understand the Context of the Benchmarking Exercise
31	Step 2: Develop Performance Measures
34	Step 3: Establish a Peer Group
36	Step 4: Compare Performance
42	Step 5: Contact Best-Practices Peers
42	Step 6: Develop an Implementation Strategy
43	Step 7: Implement the Strategy
43	Step 8: Monitor Performance
44	Chapter 5 Case Studies
44	Overview
44	Altoona, Pennsylvania
46	Knoxville, Tennessee
49	Salt Lake City, Utah
56	Denver, Colorado
59	San Jose, California
62	South Florida

69	Chapter 6 Concluding Remarks
69	Value of Peer Comparison and Benchmarking
69	Key Findings and Conclusions
71	Accomplishment of Research Objectives
72	References
74	Appendix A FTIS Instructions
86	Appendix B Peer-Grouping Methodology Details
97	Appendix C Task 10 Working Paper

S U M M A R Y

A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry

Performance measurement is a valuable management tool that most organizations conduct to one degree or another. Many examples exist of report cards, dashboards, key performance indicators, and similar techniques for presenting performance results, and these are important first steps in efforts to improve an organization's performance. Taken in isolation, however, performance measures are capable of providing tremendous quantities of data but little in the way of context. To begin to provide real value, measures need to be compared to something else—for example, one's past performance, one's targeted performance, or comparable organizations' performance—to provide the context of "performance is good," "performance needs improvement," "performance is getting better," and so on. Once a need to improve performance has been identified, an important follow-up step is to identify and contact top-performing peers to learn from them and thereby improve one's own performance.

Performance measurement involves the collection, evaluation, and reporting of data that relate to how well an organization is performing its functions and meeting its goals and objectives. The measures used in the process ideally relate to the outcomes achieved by the organization; however, descriptive measures can also be used to provide context and help identify underlying reasons for changes in performance.

Peer comparison is an activity where an organization compares its performance to that of similar ("peer") organizations using a pre-determined set of performance measures. To provide meaningful results, the measures used in the comparison need to be consistently defined and reported among the different organizations included in the peer comparison.

Benchmarking is the process of systematically seeking out best practices to emulate. A peer comparison provides an informative, but passive, starting point to a performance analysis, but is unlikely to explain why particular organizations are successful in particular areas. Benchmarking involves direct contact with other organizations, delves into the reasons for their success, and seeks to uncover transferable practices applicable to the organization performing the analysis. A performance report is not the desired end product of a benchmarking effort; rather, performance measurement is a tool used to provide insights, raise questions, and identify other organizations from which one may be able to learn and improve.

Benchmarking was first used in the private sector in 1979 and has subsequently been embraced by business leaders and become the basis for many of the Malcolm Baldrige National Quality Award's performance criteria. It has been used in the U.S. public sector since the mid-1990s, particularly in municipal applications. Benchmarking in the public transit industry was the focus of several European research efforts in the early 2000s, and at least four international and one U.S. public transit benchmarking networks (voluntary associations of transit agencies that share data and practices with each other) now exist.

Despite this track record of success, benchmarking has yet to catch on to any significant degree within the U.S. public transportation industry. Past transit agency peer-comparison

efforts uncovered in this project's literature review and agency outreach rarely extended into the realm of true benchmarking. Commonly, transit agencies have conducted peer reviews as part of transit agency or regional planning efforts, although some reviews have also been generated as part of a management initiative to improve performance. In most cases, the peer-comparison efforts were one-time or infrequent events, rather than part of an ongoing performance measurement and improvement process.

One reason that benchmarking is not widely used in the U.S. public transportation industry is that many believe that no two transit agencies are alike and that the data available to measure transit agencies are not comparable or even reliable. Transit agencies that seem similar may have very different policy objectives or may operate in environments where public transportation has a vastly different competitive position relative to transportation alternatives. Such differences impact performance. However, if these and other issues could be overcome, the experience of other industries shows that benchmarking would be a valuable tool for the U.S. public transportation industry.

The performance-measurement and peer-comparison methodology described in this report addresses the issues described above by incorporating a variety of nationally available, standardized factors into the peer-selection process, and describing ways for also incorporating policy objectives and other factors into the process. The methodology has been incorporated into a freely available, online software tool (the Florida Transit Information System, FTIS) that provides access to the full National Transit Database (NTD), allowing users to quickly identify a group of potential peer transit agencies, retrieve standardized performance data for them, and perform a variety of comparisons. During real-world applications by transit agencies, users were typically able to learn how to use the software, create a peer group, and perform an analysis with 16 person-hours of effort or less. This project's testing efforts found that, for the most part, the NTD data used in analyses were reliable and that what errors did exist were readily spotted.

The range of performance questions that benchmarking and peer comparison can be applied to spans many aspects of a transit agency's functions. Applications can be divided into the following four general categories that describe the focus of a particular comparison effort, recognizing that there is room for overlap between the various categories:

1. *Administration* – questions related to the day-to-day administration of a transit agency, including (but not limited to) financial-performance questions asked by agency management, agency board members, and transit funding organizations.
2. *Operations* – questions related to a transit agency's daily operations.
3. *Planning* – long-term policy and service questions of interest to transit operators, metropolitan planning organizations, and state departments of transportation.
4. *Public and market focus* – questions that consider the viewpoint of the broad range of customers, including riders, non-riders, local jurisdictions, and policy-makers.

This report's methodology does not recommend one set of performance measures as being appropriate for the entire range of performance applications that exist. During testing of the methodology, even when transit agencies picked identical topics to study (e.g., relative subsidy levels), they selected different sets of performance measures that related to the outcomes of particular interest to them (in one case in this example, measures relating to a funding perspective and in the other, measures relating to an operating perspective). Therefore, this report provides guidance on selecting performance measures appropriate to a particular performance question and provides case study examples that can be used for inspiration, but does not prescribe a particular set of measures. This approach requires some thoughtfulness on the part of transit agencies that apply the methodology, but also provides

much-needed flexibility that allows the methodology to be applied to a wide variety of transit modes, transit agency sizes, and performance questions.

The methodology was not designed as a means of ranking transit agencies to determine the “best” agencies overall on a national basis or the best at a particular aspect of service. Rather, this report’s approach is that peer-grouping and performance measurement should serve as a starting point for a transit agency to ask questions about performance, identify areas of possible improvement, and contact top-performing peers. That course—a true benchmarking process—holds the greatest potential for producing long-term performance improvement.

This report describes eight steps for conducting a benchmarking effort, not all of which may be needed for a given analysis, depending on the performance question being asked and the time and resources available to conduct the effort. These steps are:

1. Understand the context of the benchmarking exercise,
2. Identify standardized performance measures appropriate to the performance question being asked,
3. Establish a peer group,
4. Compare performance within the peer group,
5. Contact best-practices peers in areas where one’s performance can be improved,
6. Develop a strategy for improving performance based on what one learns from the best-practices peers,
7. Implement the strategy, and
8. Monitor changes in performance over time, repeating the process if the desired results are not achieved within the desired timeframe.

This report complements *TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System*, which describes how to implement and use performance measurement on an ongoing basis at a transit agency.

CHAPTER 1

Introduction

Research Problem Statement

Performance measurement and peer comparison are important management tools that have been used by the private sector since the late 1970s. These tools are used to evaluate performance, identify opportunities for improvement, establish performance goals, and help guide expenditures and investments. They are a means to help organizations better understand and see themselves in relation to other, similar organizations. Performance measurement and peer comparison are often initial steps in an effort to assess strengths and weaknesses and develop strategies for changing business practices.

There are numerous challenges associated with developing a methodology for performance measurement and peer comparison for public transportation. Many would argue that no two public transportation systems are alike and that data used to measure public transportation performance are not comparable or even reliable. Public transportation systems that seem similar may have very different policy objectives or may operate in environments where public transportation has a vastly different competitive position relative to transportation alternatives. Such differences impact performance. Unless properly addressed, these matters raise questions about the value of performance measurement and peer comparison for public transit systems.

Another issue, specifically with peer comparison, is a concern on the part of public transit systems that the results of peer comparison may be misused and misconstrued. Most peer comparisons of public transportation systems have been conducted by grouping systems based on a narrow range of factors, focused principally or exclusively on characteristics of the public transit systems. This method can lead to groupings of systems that are, in fact, not comparable. However, other approaches for peer comparison exist and are worthy of consideration, including comparison based on organizational function and comparison that is purpose driven.

If an appropriate methodology is used, performance measurement and peer comparison of public transit systems can be extremely useful tools that can help managers identify the strengths and weaknesses of their organization, assist in setting goals or performance targets, and help identify best practices to improve performance. In addition to improved management and operations, performance measurement and peer comparison can assist public transit in demonstrating its ability to meet local or regional transportation goals that can include safe and efficient mobility as well as broader environmental, energy, and other goals.

Research Objective and Scope

The objective of this research was to develop and test a methodology for performance measurement and peer comparison for (a) all fixed-route components of a public transit system, (b) the motorbus mode specifically, and (c) major rail modes specifically (i.e., light rail, heavy rail, and commuter rail).

The scope of the project was as follows:

- The methodology should include performance measures composed of uniformly reported data that are as transparent as possible, credible, and relevant to the concerns of public transportation systems.
- The peer comparison approach should enable performance assessments of public transportation systems of different sizes, operating environments, and modes.
- The research should consider lessons learned from other industries and from international transit peer-comparison experience.
- The research should identify potential applications for the methodology and develop potential strategies for industry adoption of the methodology.
- The methodology should be able to be applied not only by individual public transit agencies, but also by state departments of transportation and other transit funding agencies.

Research Approach

The research plan consisted of the following tasks:

1. Prepare amplified research plan;
2. Prepare literature review and agency experience;
3. Identify comparison factors, performance measures, and applications;
4. Develop initial methodology;
5. Develop outreach plan;
6. Prepare interim report and conduct panel meeting;
7. Revise interim report/execute outreach plan;
8. Modify methodology;
9. Conduct small-scale application;
10. Conduct large-scale application;
11. Interpret results/recommendations; and
12. Prepare draft and final reports.

How to Use This Report

This report uses the term *peer comparison* to describe any activity where a transit agency compares its performance to that of similar (“peer”) agencies using a predetermined set of performance measures. *Benchmarking* adds the element of seeking out best practices to emulate. In other words, a peer comparison provides an informative, but passive, starting point to a performance analysis, but is unlikely to explain why particular transit agencies are successful in particular areas. Benchmarking involves direct agency contact, delves into the reasons for agency success, and seeks to uncover transferable practices applicable to the agency performing the analysis.

Chapter 2, Performance Measurement, Peer Comparison, and Benchmarking, will be of interest to transit managers, decision-makers, and others interested in learning more about benchmarking and its potential benefits. This chapter summarizes past and present benchmarking activities in the private and public sectors, along with benchmarking activities specific to the public transit industry inside and outside the United States. It describes the different levels of benchmarking, ranging from simple trend analysis to formal, long-term

cooperation and information-sharing with similar, like-minded transit agencies. The chapter goes on to describe the key factors required for a successful benchmarking effort, reviews the transit industry’s positive and negative perceptions of peer comparisons, and summarizes lessons learned.

Chapter 3, Applications and Performance Measures, describes a variety of applications for peer comparisons and provides lists of readily available, standardized measures that can be matched to specific performance questions. Transit managers will likely be most interested in the applications section of this chapter, while agency staff responsible for performing peer comparisons will want to read the entire chapter.

Chapter 4, Benchmarking Methodology, will be of interest to both transit management and agency staff responsible for performing self-reviews, peer-comparisons, or full-scale benchmarking efforts. The chapter describes the methodology developed and tested by this project as a series of steps. The full process involves eight steps; however, depending on the time and resources available for the effort and the nature of the performance question being investigated, not all of the steps may be needed.

Chapter 5, Case Studies, draws from the methodology testing conducted for this project. It provides examples of using peer comparison to address a variety of performance questions. Agencies can use this section for inspiration on how they might organize their own peer comparison.

Chapter 6, Concluding Remarks, includes recommended strategies for incorporating performance measurement and benchmarking into the business practice of the U.S. public transportation industry.

The report also includes the following appendixes:

- Appendix A provides detailed instructions for using the online software tool as part of a benchmarking effort.
 - Appendix B describes the development of the peer-grouping portion of the methodology and the process used to calculate likeness scores.
 - Appendix C provides the working paper for Task 10 of the project, summarizing the lessons learned from the real-world tests of the project’s benchmarking methodology.
-

CHAPTER 2

Performance Measurement, Peer Comparison, and Benchmarking

Performance measurement is a valuable management tool that most organizations conduct to one degree or another. Transit agencies measure performance for a variety of reasons, including:

- To meet regulatory and reporting requirements, such as annual reporting to the FTA's National Transit Database (NTD), to a state Department of Transportation (DOT), and/or to another entity that provides funding support to the agency;
- To assess progress toward meeting internal and external goals, such as (a) measuring how well customers and potential customers perceive the quality of the service provided, or (b) demonstrating how the agency's service helps support regional mobility, environmental, energy, and other goals; and
- To support agency management and oversight bodies in making decisions about where, when, and how service should be provided, and in documenting the impacts of past actions taken to improve agency performance (1).

Taken by themselves, performance measures provide data, but little in the way of context. To provide real value for performance measurement, measures need to be compared to something else to provide the context of "performance is good," "performance needs improvement," "performance is getting better," and so on. This context can be provided in a number of ways:

- By comparing performance against internal or external service standards or targets to determine whether minimum policy objectives or regulatory requirements are being met;
- By comparing current performance against the organization's past performance to determine whether performance is improving, staying the same, or getting worse, and to what degree; and
- By comparing the organization's performance against that of similar organizations to determine whether its perfor-

mance is better than, about the same as, or worse than that of its peers.

All of these methods of providing context are valuable and all can be integrated into an organization's day-to-day activities. This report, however, focuses on the second and third items, using peer comparisons and trend analysis as means to (a) evaluate a transit agency's performance, (b) identify areas of relative strength and weakness compared to its peers, and (c) identify high-performing peers that can be studied in more detail to identify and adopt practices that could improve the agency's own performance.

Benchmarking has been defined in various ways:

- "The continuous process of measuring products, services, and practices against the toughest competitors or those companies recognized as industry leaders" (David Kearns, chief executive officer, Xerox Corporation) (2).
- "The search for industry best practices that lead to superior performance" (Robert C. Camp) (2).
- "A process of comparing the performance and process characteristics between two or more organizations in order to learn how to improve" (Gregory Watson, former vice president of quality, Xerox Corporation) (3).
- "The process of identifying, sharing, and using knowledge and best practices" (American Productivity & Quality Center) (4).
- Informally, "the practice of being humble enough to admit that someone else is better at something and wise enough to try to learn how to match, and even surpass, them at it" (American Productivity & Quality Center) (4).

The common theme in all these definitions is that benchmarking is *the process of systematically seeking out best practices to emulate*. In this context, a performance report is not the desired end product; rather, performance measurement is a tool used to provide insights, raise questions, and identify

other organizations that one may be able to learn from and improve.

Benchmarking in the Private Sector

The process of private-sector benchmarking in the United States has matured to the point where the practice and benefits of benchmarking are well understood. Much of this progress can be attributed to the efforts of the Xerox Corporation, which, in 1979, decided to initiate competitive benchmarking in response to significant foreign competition. Since then, benchmarking has been embraced by business leaders and has become the basis for many of the Malcolm Baldrige National Quality Award's performance criteria.

The first use of documented private-sector benchmarking in the United States was performed by Xerox in 1979 (2, 5). Before this time, American businesses measured their performance against their own past performance, not against the performance of other companies. In the period between 1975 and 1979, Xerox's return on net assets dropped from 25% to under 5% due to its loss of patent protection and the subsequent inflow of foreign competition. Xerox was compelled to take action to stem this precipitous decline in market share and profitability. The CEO of Xerox, David Kearns, decided to analyze Xerox's competition to determine why they were gaining market share. Xerox discovered that its development time for new products was twice as long as its competitors'. Xerox also found out that its manufacturing cost was the same as the sales price of competing products. Xerox did not simply compare itself to one other entity. By observing and incorporating successful practices used by other businesses in its weak areas, Xerox was able to achieve a turnaround. For instance, Xerox examined Sears' inventory management practices and L.L. Bean's warehouse operations, and of course, carefully analyzed the development time and cost differences between itself and its competition (2).

In the early to mid 1980s, U.S. companies realized the many benefits of information-sharing and started developing networks. Examples of these networks, which eventually became benchmarking networks, are provided below:

- The General Motors Cross-Industry Study of best practice in quality and reliability was a 1983 study of business leaders in different industries to define those quality management practices that led to improved business performance.
- The General Electric Best-Practice Network, a consortium of 16 companies, met regularly to discuss best practice in noncompetitive areas. These companies were selected so that none competed against any other participant, thus creating an open environment for sharing sensitive information about business practices.
- Hewlett-Packard (HP) also had a wide variety of collaborative efforts with other businesses. For instance, HP helped Proctor & Gamble (P&G) understand policy deployment. The nature of this collaboration included inviting two P&G executives to work inside HP for a 6-month period to experience how HP's planning process worked. Ford and HP also engaged in numerous business practice sharing activities (6).

The criteria for the Malcolm Baldrige National Quality Award (7) were developed in the 1980s by 120 corporate quality executives who aimed to agree on the basic parameters of best practice. The inclusion of benchmarking elements in many of the evaluation criteria of the Malcolm Baldrige National Quality Award and Xerox's receipt of the award helped benchmarking and its benefits gain prominence throughout the private sector.

In 1992, the International Benchmarking Clearinghouse (IBC) was established to create a common methodology and approach for benchmarking. The IBC:

- Created a benchmarking network among a broad spectrum of industries and was supported by an information database and library,
- Conducted benchmarking consortium studies on topics of common interest to members,
- Standardized training materials around a simple benchmarking process and developed business process taxonomy that enabled cross-company performance comparison, and
- Accelerated the diffusion of benchmarking as an accepted management practice through the propagation of the *Benchmarking Code of Conduct* (8) governing how companies collaborate with each other during the course of a study.

In 1994, the Global Benchmarking Network (GBN) was established to bring together disparate benchmarking efforts in various nations, including the U.K. Benchmarking Centre, the Swedish Institute for Quality, the Informationszentrum Benchmarking in Germany, and the Benchmarking Club of Italy, along with U.S. benchmarking organizations.

Benchmarking in the Public Sector

A number of public agencies in the United States have implemented benchmarking programs. This section highlights a few of them.

New York City's COMPSTAT Program

New York City has employed the Comparative Statistics (COMPSTAT) program since the Giuliani mayoral era and many believe that this system has helped the city reduce crime

and make improvements in other areas as well. The program became nationally recognized after its successful implementation by New York City in the mid-1990s. The actual origin of COMPSTAT was within New York City Transit, when its police force began using comparative statistics for its law enforcement needs and saw dramatic declines in transit crime. The system significantly expanded after Mayor Rudolph Giuliani took office and decided to implement COMPSTAT on a citywide basis.

The internal benchmarking element within COMPSTAT is embodied in the unit versus unit comparison. Commanders can monitor their own performance, evaluate the effectiveness of their strategies, and compare their own success to that of others in meeting the established performance objectives. Timely precinct-level crime statistics reported both internally to their peers and to the public motivated commanders to improve their crime reduction and prevention strategies and come up with innovative ideas to fight crime.

COMPSTAT eventually became the best practice not only in law enforcement but also in municipalities. Elements of COMPSTAT have been implemented by cities across the United States to a greater or lesser extent. An example of a municipality that has implemented a similar system is Baltimore. City officials from Baltimore visited New York City, obtained information about COMPSTAT, and initiated CITISTAT, a similar performance evaluation system, described below, that facilitates continuous improvement (9).

Baltimore's CITISTAT

CITISTAT is used by the mayor of Baltimore as a management and accountability tool. The tenets of Baltimore's program are similar to that of New York City's:

- Accurate and timely intelligence,
- Effective tactics and strategies,
- Rapid deployment of resources, and
- Relentless follow-up and assessment.

Heads of agencies and bureaus attend a CITISTAT meeting every other week with the mayor, deputy mayors, and key cabinet members. Performance data are submitted to the CITISTAT team prior to each meeting and are geocoded for electronic mapping.

As with New York City's program, the success of CITISTAT has attracted visitors from many government agencies across the United States and from abroad (10).

District of Columbia's CapStat

The District of Columbia also developed a performance-based accountability process. CapStat identifies opportunities

to improve the performance and efficiency of DC's government and provide a higher quality of service to its residents. The mayor and city administrator have regular meetings with all executives responsible for improving performance for a specific issue, examine and interpret performance data, and develop strategies to improve government services. The effectiveness of the strategies is continuously monitored, and depending on the results, the strategies are modified or continued. CapStat sessions take place at a minimum on a weekly basis (11).

Philadelphia's SchoolStat Program

Philadelphia's SchoolStat program was modeled after COMPSTAT. During the 2005–2006 school year, Philadelphia began using the SchoolStat performance management system. All 270 principals, the 12 regional superintendents, and the chief academic officer attend the monthly meetings at which performance results are evaluated and strategies to improve school instruction, attendance, and climate are assessed. One major benefit of the program is that information and ideas are disseminated vertically and horizontally across the school district. Many performance improvements were seen in the program's first year in operation (12).

Air Force

In the Persian Gulf War, the Air Force shipped spare parts between its facility in Ohio and the Persian Gulf. The success of its rapid and reliable parts delivery system can be credited to the Air Force benchmarking of Federal Express's shipping methods (13).

Benchmarking in the Public Transit Industry

International Efforts

Benchmarking Networks

Several benchmarking networks, voluntary associations of organizations that agree to share data and knowledge with each other, have been developed internationally. There were four notable international public transit benchmarking networks in operation in 2009. Three of these were facilitated by the Railway Technology Strategy Centre at Imperial College London and shared common processes, although the networks catered to differing modes and city sizes. The two rail networks also shared common performance indicators.

The first of the networks now facilitated by Imperial College London, CoMET (Community of Metros), was initiated in 1994 when Hong Kong's Mass Transit Railway Corporation (MTR) proposed to metros in London, Paris, New York,

and Berlin that they form a benchmarking network to share information and work together to solve common problems. Since that time, the group has expanded to include 13 metros and suburban railways in 12 of the world's largest cities: Beijing, Berlin, Hong Kong, London, Madrid, Mexico City, Moscow, New York, Paris (Metro and RER), Santiago, São Paulo, and Shanghai. All of the member properties have annual ridership of over 500 million (14).

The Nova group, which started in 1997, focuses on medium-sized metros and suburban railways with ridership of under 500 million. As of 2009, its membership consisted of Bangkok, Barcelona, Buenos Aires, Delhi, Glasgow, Lisbon, Milan, Montreal, Naples, Newcastle, Rio de Janeiro, Singapore, Taipei, Toronto, and Sydney (15).

Imperial College London also facilitates the International Bus Benchmarking Group, which started in 2004 and had 11 members as of 2009: Barcelona, Brussels, Dublin, Lisbon, London, Montreal, New York, Paris, Singapore, Sydney, and Vancouver. The bus group shares the same basic benchmarking process as its rail counterparts, but uses a different set of key performance indicators (16, 17).

The fourth international benchmarking network that was active in 2009 was Benchmarking in European Service of public Transport (BEST). The program was initiated by Stockholm's public transit system in 1999. Originally conceived of as a challenge with the transit systems in three other Nordic capital cities—Copenhagen, Helsinki, and Oslo—it quickly evolved into a cooperative, non-competitive program with the goal of increasing public transport ridership. After a pilot program in 2000, BEST has reported results annually since 2001. In addition to the original four participants, Barcelona, Geneva, and Vienna have participated more-or-less continuously since 2001; Berlin and Prague have participated more recently; and London and Manchester also participated for a while. The program is targeted at regions with 1 to 3 million inhabitants that operate both bus and rail services, but does not strictly hold to those criteria. The network is facilitated by a Norway-based consultant (18).

Common features of these four benchmarking networks include:

- Voluntary participation by member properties and agreement on standardized performance measures and measure definitions;
- Facilitation of the network by an external organization (a university or a private consulting firm) that is responsible for compiling annual data and reports, performing case studies, and arranging annual meetings of participants;
- A set of annual case studies (generally 2–4 per year) on topics of interest to the participants;
- Confidentiality policies that allow the free flow of information within the network, but enforce strict confidentiality

outside the network, unless all participants agree to release particular information;

- An attitude that performance indicators are tools for stimulating questions, rather than being the output of the benchmarking process. The indicators lead to more in-depth analyses that in turn identify processes that produce higher levels of performance.

The three Imperial College–facilitated networks use relatively traditional transit performance measures as their “key performance indicators.” In contrast, BEST uses annual telephone citizen surveys (riders and non-riders) in each of its participating regions to develop its performance indicators. According to BEST's project manager, the annual cost to each participating agency is in the range of 15,000 to 25,000 euros, depending on how many staff participate in the annual seminar and on the number of case studies (“Common Interest Groups”) in which the agency participates. The cost also includes each agency's share of the telephone survey and the cost of compiling results. Annual costs for the other three networks were not available, but the CoMET project manager has stated that the “real, tangible benefits to the participants . . . have far outweighed the costs” (19, 20).

European Benchmarking Research

The European Commission has sponsored several studies relating to performance measurement and benchmarking.

Citizens' Network Benchmarking Initiative. The Citizens' Network Benchmarking Initiative began as a pilot project in 1998, with 15 cities and regions of varying sizes and characteristics participating. Participation was voluntary, with the cities supplying the data and providing staff time to participate in working groups. The European Commission funded a consultant to assemble the data and coordinate the working groups. The goal of the pilot project was to test the feasibility of comparing public transport performance across all modes, from a citizen's point-of-view. During the pilot, 132 performance indicators were tested, which were refined to 38 indicators by the end of the process. The working groups addressed four topics; working group members for each topic made visits to the cities already achieving high performance in those areas, and short reports were produced for each topic area.

Following the pilot project, the program was expanded to 40 cities and regions. As before, agency participation was voluntary and the European Commission funded a consultant to assemble the data and coordinate the working groups. As Europe has no equivalent to the National Transit Database, the program's “common indicators” performance measures were intended to rely on readily available data and not require

aggregation into a more complex indicator. In the full program, some of the pilot indicators were abandoned due to lack of data or consistency of definition, while some new indicators were added. The program ended in 2002 when the funding for the consultant support ran out, although there appeared to be at least some interest among the participants in continuing the program (21).

Extending the Quality of Public Transport (EQUIP). A second initiative, EQUIP, occurred at roughly the same time as the Citizens' Network Benchmarking Initiative. EQUIP developed a *Benchmarking Handbook* (22) covering five modes: bus, trolleybus, tram/light rail, metro, and local heavy rail (i.e., commuter or suburban rail). The handbook consists of two volumes: (1) a methodology volume describing benchmarking in general and addressing sampling issues, and (2) an indicators volume containing 91 standardized indicators for measuring an agency's internal performance and service quality. Of these, 27 are considered "super-indicators" that provide an entry-level introduction to benchmarking. Ideally, each of these indicators would be collected for each of the five modes covered by the handbook.

EQUIP was tasked with developing methods that agencies could use for internal benchmarking, but the methodology lent itself to agencies submitting data to a centralized, potentially anonymous database that could be used for external comparisons, and then finally direct interaction with other agencies. During development, the methodology was tested on a network of 45 agencies in nine countries; however, the network did not continue after the conclusion of the project (23).

One challenge faced by EQUIP was that the full EQUIP process required collecting data that European agencies either were not already collecting or were not collecting in a standardized way, due to the absence of mandatory performance reporting along the lines of the NTD in the United States. Therefore, agencies would have incurred additional costs to collect and analyze the data. In addition, most European service is contracted out, with multiple companies sometimes providing service in the same city, so there can be competi-

tive reasons why a service provider may be reluctant to share data with others. In addition, the local transit authority needs to compile data from multiple operators. However, as the majority of the EQUIP measures are ones that U.S. systems already routinely collect for NTD reporting purposes, the EQUIP process appears transferable to the United States.

Benchmarking European Sustainable Transport (BEST). The European Union (EU) BEST program (different from the Nordic BEST benchmarking network described above) focused on developing or improving benchmarking capabilities for all transport modes in Europe (e.g., air, freight rail, public transport, and bicycling) at scales ranging from international to local. The program sponsored six conferences between 2000 and 2003 that explored different aspects of benchmarking, and also sponsored three pilot benchmarking projects in the areas of road safety, passenger rail transport, and airport accessibility (24).

Quality Approach in Tendering/contracting Urban Public Transport Operations (QUATTRO). The EU's QUATTRO program (25) developed a standardized performance-measurement process that subsequently was adapted into the EN 13816 standard (26) on the definition, targeting, and measurement of service quality on public transport. The standard describes a process for measuring service quality, recommends areas to be measured, and provides some general standardized terms and definitions, but does not provide specific targets for performance measures nor specific numerical values as part of measure definitions (e.g., the number of minutes late that would be considered "punctual" or on-time).

Both QUATTRO and EN 13816 describe a quality loop, illustrated in Figure 1, with four main components that measure both the service provider and customer points of view:

- **Service quality sought:** The level of quality explicitly or implicitly required by customers. It can be measured as the sum of a number of weighted quality criteria; the relative weights can be determined through qualitative analysis.

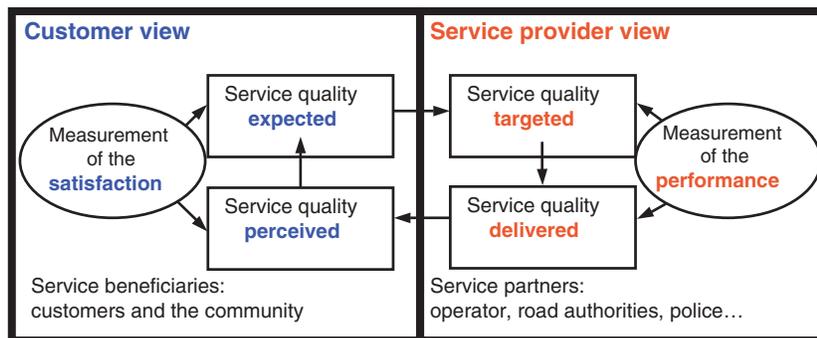


Figure 1. Quality Loop, QUATTRO and EN 13816.

- **Service quality targeted:** The level of quality that the service provider aims to provide for customers. It considers the service quality sought by customers as well as external and internal pressures, budgetary and technical constraints, and competitors' performance. The following factors need to be addressed when setting targets:
 - A brief statement of the service standard [e.g., “we intend our passengers to travel on trains which are on schedule (meaning a maximum delay of 3 minutes)”];
 - A targeted level of achievement (e.g., “98% of our passengers find that their trains are on schedule”); and
 - A threshold of unacceptable performance that, if crossed, should trigger immediate corrective action, such as (but not limited to) provision of alternative service or customer compensation.
- **Service quality delivered:** The level of quality achieved on a day-to-day basis, measured from the customer point of view. It can be measured using direct observation.
- **Service quality perceived:** The level of quality perceived by the customer, measured through customer satisfaction surveys. Customer perception depends on personal experience of the service or associated services, information received about the service from the provider or other sources, and the customer's personal environment. Perceived quality may bear little resemblance to delivered quality.

Transferring Knowledge between Industries

A presentation (27) at one of the Benchmarking European Sustainable Transport conferences focused on the process of looking outside one's own industry to gain new insights into one's business practices. As discussed later in this chapter, a common fear that arises when conducting benchmarking exercises, even among relatively close peers, is that some fundamental difference between peers (for example—in a transit context—relative agency or city size, operating environment, route network structure, agency objectives) will drive any observed differences in performance between the peers. As a result, some argue, it is difficult for a benchmarking exercise to produce useful results. When looking outside one's own industry, differences between organizations are magnified, as are the fears of those being measured. At the same time, benchmarking only within one's own industry can lead to performance improvements, but only up to the industry's current level of best practice. Looking outside one's industry, on the other hand, allows new approaches to be considered and adopted, resulting in a greater improvement in performance than would have been possible otherwise. In addition, competition and data confidentiality issues are lessened the further away one goes from one's own industry.

The value from an out-of-industry benchmarking effort comes from digging deeply into the portions of the organiza-

tions that share common issues rather than from looking at high-level performance indicators. For example, the BEST presentation (27) looked at revenue and risk management best practices that could be transferred to the freight transportation industry from such industries as travel, hospitality, energy, and banking. In the area of risk management, for example, common areas of risk include market risks due to changes in demand, changes in unit costs, over-capacity in the market, and insufficient capacity in the market. In the revenue-generation area, the freight transportation industry has adopted, among others, price forecasting, customer segmentation, and product differentiation practices from other industries.

International Databases

There is no international equivalent to the National Transit Database. The closest counterpart is the Canadian Transit Statistics database compiled annually by the Canadian Urban Transit Association (CUTA). The database is available only to CUTA members (28). The International Association of Public Transport (UITP) has produced a *Mobility in Cities Database* that provides 120 urban mobility indicators for 50 cities. The data cover the years 1995 and 2001. An interesting aspect of the database is that urban transport policies are also tracked as part of the database, both policies that were enacted between 1990 and 2001 and those planned to be enacted between 2001 and 2010 (29).

U.S. Efforts

Transit Agencies

Past transit agency peer-comparison efforts uncovered in the literature review and the initial agency outreach effort rarely extended into the realm of true benchmarking (i.e., involving contact with other agencies to gain insights into the results of comparisons and generating ideas for improvements). Commonly, agencies have conducted peer reviews as part of agency or regional planning efforts, although some reviews have also been generated as part of a management initiative to improve agency performance. In most cases, the peer-comparison efforts were one-time or infrequent events rather than part of an ongoing performance measurement and improvement process. Some examples of these efforts are described below.

The Ann Arbor Transportation Authority conducted a peer analysis at the end of 2006 that involved direct contact with 10 agencies (8 of which agreed to participate) to (a) provide more recent data than were available at the time through the NTD (year 2004 data) and (b) provide details about measures not available through the NTD (the presence of a downtown hub and any secondary hubs, and the presence of bike racks,

a trip planning system, and a bus tracking system). One impetus for the review was the agency's 25% increase in ridership during the previous 2 years.

The Central Ohio Transportation Authority (COTA) in Columbus regularly compares itself to Cleveland, Cincinnati, Dallas, Buffalo, and Austin using NTD data. Measures of particular interest include comparative cost information (often used in labor negotiations) and maintenance information. COTA also internally tracks customer-service-based measures such as the number of complaints per 100,000 passengers.

The Utah Transit Authority (UTA) commissioned a performance audit in 2005 (30). The audit was very detailed and included hundreds of performance measures, many of which went beyond the planning level and into the day-to-day level of operations. However, the audit also included a peer-comparison element that compared UTA's performance against peer agencies by mode (i.e., bus, light rail, and paratransit) in terms of boardings, revenue miles, operating costs, peak fleet requirements, and service area size and population. Peers were selected based on region (west of the Mississippi River), city size, and the existence of light rail systems built within the previous 30 years.

Some agencies use peer review panels or visits as tools to gain insights into transit performance and generate ideas for improvements. Peer review or "blue ribbon" panels tend to be more like an audit or management review, where peer representatives can be on-site at an agency for up to 3 or 4 days. Some form of peer-identification process is typically used to develop these panels. Those who have participated in such efforts have found it quite valuable to discuss details with their counterparts at other agencies. Visits to other agencies can also provide useful insights if the agencies have been selected on the basis of (a) having similar characteristics to the visiting agency and (b) strong performance in an area of interest to the visitors. Visits to agencies simply on the basis of reputation may be interesting to participants, but are not as likely to produce insights that can be adopted by the visiting agency.

APTA is developing a set of voluntary recommended practices and standards for use by its members. Some of these provide standard definitions for non-NTD performance measures. Other recommended practices address processes (for example, a customer comment process and comment-tracking database) that could lead to more widespread availability of non-NTD data, although the data would not necessarily be defined consistently between agencies. An example of a standard definition is APTA's *Draft Standard for Comparison of Rail Transit Vehicle Reliability Using On-Time Performance* (31), which defines two measures and an algorithm for determining the percentage of rail trips that are more than 5 minutes late as a result of a vehicle failure on a train.

Transit Finance Learning Exchange (TFLEX)

One U.S. benchmarking network currently in existence is the Transit Finance Learning Exchange, "a strategic alliance of transit agencies formed to leverage mutual strengths and continuously improve transit finance leadership, development, training practices and information sharing" (32). TFLEX was formed in 1999 and currently has thirteen members:

- Capital Metropolitan Transportation Authority – Austin,
- Central Puget Sound Regional Transit Authority (Sound Transit),
- Dallas Area Rapid Transit (DART),
- Hillsborough Area Regional Transit Authority (HART),
- Los Angeles County Metropolitan Transportation Authority (LACMTA),
- Massachusetts Bay Transportation Authority (MBTA),
- Orange County Transportation Authority (OCTA),
- Regional Transportation Authority of Northeastern Illinois (RTA),
- Regional Transportation Commission of Southern Nevada (RTC),
- Rochester-Genesee Regional Transportation Authority (RGRTA),
- San Joaquin Regional Transit District,
- Santa Monica Big Blue Bus, and
- Washington Metropolitan Area Transit Authority (WMATA).

As of 2009, annual membership fees were \$5,000 for agencies with annual operating budgets larger than \$50 million, and \$2,500 otherwise. To avoid losing members, membership dues have dropped in recent years due to the economic downturn.

TFLEX's original goal was to develop a standardized database of transit performance data to overcome the challenges, particularly related to consistency, of relying on NTD data. In most cases, the performance measures collected by TFLEX have been variants of data already available through NTD, rather than entirely new performance measures; the benefit is in the greater consistency of the TFLEX reporting procedures.

TFLEX has not entirely succeeded in meeting its goal of producing a standardized and regularly updated database of transit performance data for benchmarking purposes, for two primary reasons:

- First, collecting the data requires significant time and/or money commitments. It has been difficult in many cases for member agencies to dedicate resources to providing data for TFLEX. One successful strategy used in the past was to fly a TFLEX representative directly to member agencies to collect data. This approach worked well, but the

costs of the data collection effort were higher than can currently be supported.

- Second, developing standard definitions of core performance measures that everyone can agree to for reporting purposes is difficult. For instance, TFLEEx spent years trying to develop a standard definition of “farebox recovery” to ensure consistent reporting of this measure in TFLEEx data.

In general, TFLEEx’s experience has been that it is relatively easy to get data once, but updating the data on a regular basis is very difficult. Because of the financial difficulties faced by many transit agencies at the time of writing, funding and resources for TFLEEx data collection efforts have diminished considerably. Data for member agencies are still reported in a confidential section of the TFLEEx website, but the data typically come directly from NTD reporting at present, rather than representing the results of a parallel data collection effort.

While the database function of TFLEEx has subsided for the time being, the organization provides several other benefits for members. TFLEEx agencies meet for semi-annual workshops where best practices in transit financial management are shared. In addition, the TFLEEx website provides a forum to ask specific questions and receive feedback from other member agencies. TFLEEx leadership reported that questions to these forums typically elicit valuable responses. Although the current TFLEEx programs are not data-driven, the members still consider TFLEEx to be a valuable benchmarking tool simply through the ability to quickly share information with other member agencies. The semi-annual workshops are also seen as valuable ways to develop professional relationships to which one can turn for answers to many transit performance-related questions.

Despite the challenges that the group has faced in developing a reliable source of benchmarking data, TFLEEx leadership still feel that the need for these data within a transit agency is strong. As a result, they expect TFLEEx to undertake a renewed effort to collect benchmarking data in the next several years. Over the long term, the TFLEEx interviewees felt strongly that the NTD reporting procedures should be significantly revised to create a more standardized dataset. They felt there is a movement within the transit industry toward greater transparency that will allow for more standardized reporting in the future. One interviewee noted the success of public schools in developing benchmarks (e.g., test scores) that allow schools to be directly compared to one another as a model that the transit industry could hope to follow.

States

State DOTs are often interested in transit performance, as they are typically responsible for distributing federal grant

funding to rural, small urban, and medium urban systems. Funding formulas often include basic elements of performance (e.g., ridership, revenue miles operated), while some also include cost-related factors. For example, the Indiana DOT incorporates a 3-year trend of passenger trips per operating expense, vehicle miles per operating expense, and locally derived income per operating expense into its formula. The Texas DOT uses revenue miles per operating expense, riders per revenue mile, local investment per operating expense, and (for urban systems only) riders per capita in its formula (33). The North Carolina DOT (NCDOT) is considering developing minimum service standards for transit agencies in its state (34), but has not yet done so.

The Washington State DOT (WSDOT) has a strong focus on performance measurement, both internal and external. WSDOT’s Public Transit Division produces an annual summary on the state of public transportation in the state (35), which includes a statewide summary and sections for each of the state’s 28 local-government public transportation systems. The local-agency sections include a comparison of 10 key performance indicators to the statewide average for the group (e.g., urban fixed-route service), as well as 3-year trends and 5-year forecasts for each agency for a larger set of performance measures, broken out by mode. All of the reported measures are NTD measures. However, the summary also provides useful information about individual systems that go beyond NTD measures, including information on:

- Local tax rates and the status of local efforts to increase local public transportation taxes and/or expand transit districts;
- Type of governance (e.g., Public Transportation Benefit Area, City, County);
- Description of the makeup of the governing body (e.g., two county commissioners, two representatives from cities with populations greater than 30,000, etc.);
- Days and hours of service;
- Number and types of routes (e.g., local routes, commuter routes) that are operated;
- Base fare (adult and senior);
- Descriptions of transfer, maintenance, and other facilities; and
- Summary of the agency’s achievements in the previous year, objectives for the upcoming year, and objectives for the next 5 years.

The NCDOT commissioned a *Benchmarking Guidebook* (34), which was published in 2006. The guidebook’s purpose is to provide public transportation managers in North Carolina with step-by-step guidance for conducting benchmarking processes within their organizations. The state’s underlying goal is to help ensure that transit systems throughout the state serve their riders efficiently and effectively, and use the

state's public funding as productively as possible. The guidebook proposes a three-part benchmarking process:

1. Trend analysis—to be conducted at least annually by each transit system.
2. Peer group analysis—to be conducted at least annually by each transit system (comparing themselves to national peers) and by the NCDOT (comparing performance among groups of North Carolina peers).
3. Statewide minimum standards—10 measures that would be evaluated annually by NCDOT, with poorly performing transit systems provided help to improve their performance, and superior performance being recognized.

States frequently tabulate data that agencies within the state submit to the NTD, allowing access to the data up to a year earlier than waiting for the NTD. States have also frequently tabulated a limited set of data for smaller systems that prior to 2008 were not required to report to the NTD. However, due to the minimal amount of staff at smaller systems, data may not be reported consistently or at all. One state, for example, reported that their collection of rural and small city cost data was “substantially unreliable,” due to missing data and values they knew for certain were incorrect (36). Some state DOTs, such as Florida and Texas, have had universities audit NTD data submitted by transit agencies to ensure consistency with performance measure definitions and have had the universities perform agency training on measures that were particularly troublesome.

Regions

Peer comparisons are also performed at the regional level. For example, by state law, the Metropolitan Council in Minneapolis must perform a transit system performance audit every 4 years. The audit encompasses the 24 entities that provide service within the region. The 2003 audit included a peer comparison of the region as a whole to 11 peer regions (selected on the basis of area size and composition of transit services) and a comparison of bus and paratransit services by Metro Transit in Minneapolis/St. Paul to six peer cities. A trend analysis was performed for six key measures for both Metro Transit and the peer group average (37).

The Atlanta Regional Council conducted a Regional Transit Institutional Analysis (38) to examine how the region should best plan, fund, build, and operate public transit. A group of peer regions was constructed to assist in the analysis. Factors used to select peers consisted of urban area size (within 2 million of the Atlanta region's population), urban area growth rate, population density, annual regional transit trips, percent drive-alone trips, annual delay per traveler, and cost-of-living index. Boston, Portland, Los Angeles, and New York were also

included at the request of the council's board, as those areas are frequently cited as examples. The comparisons focused on non-NTD factors, such as the budget approval process, fare policy, responsibility for capital construction, funding allocation, bonding authority, and recent initiatives.

The state of Illinois' Office of the Auditor General periodically conducts performance audits of the transit funding and operating agencies in the Chicago region. The most recent audit was conducted in 2007 (39). A portion of the audit developed groups of five peers each for the following Chicago-area transit services: Chicago Transit Authority (CTA) heavy rail, CTA bus, Metra commuter rail, Pace bus, Pace demand response, and Pace vanpool. CTA and Metra peers were selected by identifying agencies operating in major cities with rapid rail service, while Pace's bus peers were selected by identifying agencies that operate in suburban portions of major cities and that operate from multiple garages. (Pace management noted that three of the five peers provide service within the major city, unlike Pace, and that Pace had the lowest service area population density of the group.) Pace operates the second-largest vanpool service in the country, so the other four largest vanpool operations were selected as the peers for that comparison. The peer comparisons looked at four major categories of NTD measures: service efficiency, service effectiveness, cost effectiveness, and passenger revenue effectiveness, plus a comparison of top-operator wage rates using data from other sources. Between 5 and 19 measures were compared among the peer groups, depending on the service being analyzed. For those areas where a service's performance was below that of its peers, the audit developed recommendations for improvements.

Research

The University of North Carolina at Charlotte produced annual rankings of transit system performance (40), derived from 12 ratios of NTD measures reflecting the resources available to operate service, the amount of service provided, and the resulting utilization of the service. Agencies were assigned to one of six groups based on mode operated (bus-only or multimodal) and population served. Each agency's value for a ratio was compared to the group mean for the ratio, resulting in a performance ratio. The 12 performance ratios for an agency were then averaged, and this overall performance ratio was used to rank systems within their own groups, as well as for all systems. Some of the criticisms of this effort included that there was no industry consensus or agreement on the measures that were used, that inconsistencies in the data exist, and that systems have different goals, regional demographics, and regional economies. It was also pointed out that a poor performance in one single category could overshadow good results in several other categories—for example, MTA-

New York City Transit ranked in the top three agencies overall for 6 of the 12 measures one year, yet ended up ranked 124 out of 137 agencies overall due to being in the bottom four agencies for 3 of the 12 measures. The performance ratio also had problems with autocorrelation among the component measures (1). The authors of the study did not believe that geographic or size differences affected their results, but did acknowledge that their findings did not shed light on the reasons why apparently similar systems differed so much in performance.

The National Center for Transit Research (NCTR) conducted a study on benchmarking (41) in 2004 for the Florida DOT. This study's objective was to develop a method of measuring commonly maintained performance statistics in a manner that would be broadly acceptable to the transit industry and thereby provide useful information that could help agencies improve their performance over time. The project focused on the fixed-route motorbus mode and was limited to NTD variables, with the intent of expanding the range of modes and variables in the future if the initial project proved successful. Peer groups were initially formed based on geographic region, and then subdivided on the basis of service area population, service area population density, total operating expense, vehicles operated in maximum service, and annual total vehicle miles. An additional group of the 20 largest transit systems from around the country was also formed, as the largest systems often did not have comparable peers within their region.

The NCTR study compared 22 performance measures in six performance categories: service supply/availability, service consumption, quality of service, cost efficiency, operating ratio, and vehicle utilization. For each measure, an agency was assigned points depending on where its performance value stood in relation to the group mean, with a value more than 2 standard deviations below the mean earning no points, a value more than 2 standard deviations above the mean earning 2 points, and between 0.5 and 1.5 points for values falling within ranges located between those two extremes. By adding up the point values for each measure, a total score can be developed for each agency, and the agencies can be ranked within their group based on their respective scores. In addition, composite scores and rankings can be developed for each performance category. According to the study's authors, the results of the process are not intended to indicate that one particular transit agency is "better" than another, but rather to serve as a tool that allows transit agencies to see where they fit into a group of relatively similar agencies.

NCHRP Report 569: Comparative Review and Analysis of State Transit Funding Programs (42) provides information to help states conduct peer analyses and other comparative assessments of their transit funding programs, using data from the Bureau of Transportation Statistics' Survey of State Funding for Public Transportation and other data sources. The re-

port presents the following framework for using the survey's data to construct peer groups and conduct peer analyses:

1. Determine the purpose of the analysis, or the types of measures to be compared (a common objective).
2. Determine the metrics for formulating peer groups (which similarities should be shared among the peers).
3. Develop the peer groups based on the metrics selected and their relative importance (i.e., determine weights).

The report provides examples of how the framework could be applied. In one sample analysis, the assumed objective was to compare state transit funding between "transit-dependent" and "non-transit-dependent" states. In a second example, peer groups were formed for the purposes of comparing state transit funding programs. These examples include suggestions for peer-grouping measures and suggestions for performance measures relevant to each example's objective that could be used for drawing comparisons.

TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System (1) describes a process for transit agencies to follow to set up an internal performance-measurement system, a necessary first step to any benchmarking effort. The report describes more than 400 performance measures that are used in the transit industry. Each measure is assessed on the basis of its performance category (availability, service delivery, community impact, travel time, safety and security, maintenance and construction, and economic/financial), its data collection needs, and its potential strengths and weaknesses for particular applications. A series of question-based menus guide readers from a particular agency objective to one or two relevant measures for that objective, considering the agency's size and data-collection capabilities. A recommended core set of measures for different agency sizes is also presented for agencies that want to start with a basic performance-measurement program prior to fine-tuning to reflect specific agency objectives. True benchmarking, involving contact with other agencies, is not covered in the report (only trend analyses and peer comparisons are described); however, the report can serve as a valuable resource to a benchmarking effort by providing a source of appropriate measures that can be applied to a particular benchmarking application.

Maintaining a customer focus is an important aspect of a successful benchmarking effort. Transit agencies often use customer satisfaction surveys to gauge how well customers perceive the quality of service being provided. *TCRP Report 47: A Handbook for Measuring Customer Satisfaction and Service Quality* (43) provides a recommended set of standardized questions that transit agencies could incorporate into their customer surveying activities. If more agencies adopted a standard core set of questions, customer satisfaction survey results could

be added to the mix of potential comparisons in a benchmarking exercise.

Levels of Benchmarking

Benchmarking can be performed at different levels of complexity that result in different levels of depth of understanding and direction for improvement. The European EQUIP project (22, 23), described previously, defined three levels of benchmarking complexity, which form a useful foundation for the discussions in this section. This report splits EQUIP's Level 3 (direct contact with other agencies) into two levels, one involving one-time or irregular contact with other agencies (this report's Level 3), and the other involving participation in a benchmarking network with a more-or-less fixed set of partner agencies (this report's Level 4).

Level 1: Trend Analysis

Are we performing better than last week/month/quarter/year?

The first level of evaluation is to track performance on a periodic basis, often year-to-year, but potentially also week-to-week, month-to-month, or quarter-to-quarter, using the same indicators in a consistent way. Trend analysis forms the core of an effective performance-measurement program and is essential for good management and stewardship of funds. A program can be tailored to measure an agency's success in meeting its goals and objectives, and each agency has the flexibility to choose exactly what to measure and how to measure it.

A trend analysis can show whether a transit agency is improving in areas of interest over time, such as carrying more rides, collecting more fare revenue, or decreasing complaints from the public. However, a trend analysis does not gauge how well an agency is performing relative to its potential. An agency could have increased its ridership substantially, but still be providing relatively few rides for the size of market it serves and the level of service being provided. To move to the next level of performance evaluation, a peer comparison should be conducted (Level 2).

Level 2: Peer Comparison

How are we performing in relation to comparable agencies?

There are a number of reasons why a transit agency might want to perform a peer comparison: for example, to support an agency's commitment to continual improvement, to validate the outcome of a past agency initiative, to help support the case for additional funding, to prioritize activities or actions as part of a strategic or short-range planning process, or to respond to external questions about the agency's operation. In a peer comparison, an agency compares its performance against other similar agencies that have contributed

similarly collected data to a centralized database, which may or may not be anonymous. No direct contact or sharing of knowledge occurs between agencies, other than knowledge that can be obtained passively (e.g., from documents or data obtained through an Internet search).

The set of performance measures that can be used in a peer comparison is much more limited than in a trend analysis, as the data for each measure must be available for all of the peer agencies involved in the comparison, and each transit agency must use the same definition for any given measure. As a result, most peer comparisons in the United States have relied on the NTD as it is readily available and uses standardized definitions. As discussed later in this chapter, the NTD does not provide measures for all performance topics of potential interest to a transit agency, nor do all reporting agencies consistently follow the FTA's performance measure definitions. Nevertheless, despite these handicaps, the industry consensus [as determined from this project's outreach efforts (44)] is that the NTD is the best source of U.S. transit data available and that the FTA is continually working to improve NTD data quality.

A critical element of a peer comparison is the selection of a credible peer group. If the peer group's characteristics are not sufficiently similar to that of the transit agency performing the comparison, any conclusions drawn from the comparison will be suspect, no matter how good the quality of the performance measure data used in the comparison. At the same time, it is unrealistic to expect that the members of a peer group will be exactly like the target agency. Data from standardized data sources can be used to form peer groups of comparable agencies [the peer-grouping methodology presented in Chapter 3 follows this approach, using the NTD, Census Bureau, the *2007 Urban Mobility Report* (45), and data developed by this project as the sources].

The transit agency's performance can then be compared to its peers in areas of interest, using standardized performance measures to identify areas where the agency performs as well as or better than the others and areas where it lags behind. It is unlikely that an agency will excel among its peers in all areas; therefore, the peer comparison process can help guide an agency in targeting its resources toward areas that show strong potential for improvement. A transit agency may discover, for example, that it is providing a comparable level of service but carrying fewer passengers than its peers. This knowledge can be used by itself to indicate that effectiveness may need to be improved, but it becomes more powerful when combined with more detailed data obtained directly from the peer agencies (Level 3).

Level 3: Direct Agency Contact

What can we learn from our peers that will help us improve our performance?

Level 3 represents the start of true benchmarking. At this level, the transit agency performing the comparison makes direct contact with one or more of its peers. More-detailed information and insights can be gained through this process than from a simple reliance on a database.

One reason for directly contacting other peers is that the measures required to answer a performance question of interest are simply not available from national databases. A variety of data that are not reported to the NTD (for example, customer satisfaction data) are often collected by peer agencies but are not necessarily summarized in standard reports. In other cases, performance measures may be reported to the NTD, but not at the desired level of detail—for example, an agency that is interested in comparing the cost-effectiveness of commuter bus routes will only find system-level data in the NTD, which aggregates all of a particular transit agency's bus services.

Another reason for directly contacting a peer is to gain insights into what the agency's top-performing peers are doing to achieve their superior performance in a particular area. These insights may lead to ideas on how these peer agencies' practices may be transferable to the transit agency performing the comparison, leading eventually to the agency being able to improve its performance in an area of relative weakness.

A third reason for contacting peers is to obtain background information about a particular transit agency (e.g., agency policies or board composition) and to verify or ask questions about unusually high or low results. These types of contacts help verify that the peer agency really is similar to the agency performing the comparison and that the performance results are reliable.

At Level 3, contact with other transit agencies occurs on a one-time or irregular basis, guided by specific agency needs, such as the need to update a transit development plan. Although benchmarking is occurring, a consistently applied and scheduled agency benchmarking program and an agency culture supporting continuous improvement may not yet exist. Because peer agencies are unlikely to change much over the short term (barring a major event such as a natural disaster or the opening of a new mode), the same set of peers can often be used over a period of years, resulting in regular contacts with peers. At some point, transit agencies may decide it would be valuable to institute a more formal information-sharing arrangement (Level 4).

Level 4: Benchmarking Networks

What knowledge can we share with each other in different areas that will help all of us improve our performance?

At the highest level of benchmarking, an agency implements a formal benchmarking program and establishes (or is taking steps to establish) an agency culture encouraging continuous improvement. The agency identifies similar, like-

minded agencies that have agreed to work together to regularly share data and experiences with each other for the benefit of all participants. The participants in the benchmarking network agree upon a set of data definitions and measures to be shared among the group, have a process set up that allows staff from different agencies to share their experiences with others, and may pool resources to fund investigations into performance topics of interest to the group. Much of the data-related portion of the process is similar to Level 3, but after the initial start-up, requires less effort to manage, as the peer group members have already been identified and a common set of measures and definitions has already been agreed upon.

Benchmarking Success Factors

The following is a summary of the key factors for successful peer comparison and full-fledged benchmarking programs that were identified from the project's literature review and agency outreach effort:

- The peer grouping process is perhaps the most important step in the benchmarking process. Inappropriate peers may lead to incorrect conclusions or stakeholder refusal to accept a study's results. For high-level performance comparisons, peers should have similar sizes, characteristics, and operating conditions. One should expect peers to be similar, but not identical. Different peer groups may be needed for different types of comparisons (2, 41, 44).
- Design the benchmarking study and identify the study's objectives before starting to collect data. Performance should be measured relative to the agency's goals and objectives. Common definitions of performance measures are essential (1–3, 44).
- A management champion is needed first to support the initial performance-measurement and benchmarking effort, and then later to implement any changes that result from the process. Without such support, time and resources will be wasted as no changes will occur (1, 3, 6, 44).
- Comparing trends, both internally and against peers, helps identify whether particularly high or low performance was sustainable or a one-time event, which leads to better interpretation of the results of a benchmarking effort (3, 44).
- Organizations should focus less on rankings in benchmarking exercises and more on using the information to stimulate questions and to identify ways they can adapt the best practices of others to their own activities. A “we can learn from anyone” attitude is helpful. Don't expect to be the best in every area (6, 20, 44).
- Consider the customer in any benchmarking exercise. Public transit is a customer-service business, and transit benchmarking should seek to identify ways to improve transit

performance and thereby improve ridership (1, 3, 6, 18, 22, 26, 44).

- A long-term approach to performance measurement and benchmarking is more likely to be successful than a series of independent studies performed at irregular intervals. Even established benchmarking programs should be monitored and reviewed over time to make sure they stay current with an organization's objectives and current conditions (1, 2, 20).

Confidentiality

The U.S. and European transit benchmarking networks (14–16, 18, 31), as well as the *Benchmarking Code of Conduct* (8), emphasize the importance of confidentiality, particularly in regard to information about business practices and the results of benchmarking comparisons. The networks also extend confidentiality, to one degree or another, to the inputs into the benchmarking process. All of these sources agree that information can be released if all affected parties agree to do so.

In an American transit context, confidentiality of inputs is not attainable in many circumstances because the NTD is available to all. However, certain types of NTD data (e.g., safety and security data) are not released to the public at present, while non-NTD data that may assist a benchmarking process, such as customer satisfaction survey results, are only available through the cooperation of other transit agencies, who may not wish the information to be broadly disseminated. On the other hand, as public entities, many U.S. transit agencies are subject to state “sunshine laws” that may require the release of information if requested to do so (e.g., by a member of the public or by the media). The public nature and standardization of the NTD (and, for Canadian agencies, the availability of standardized Canadian data) makes it easier for U.S. and Canadian transit agencies to perform peer comparisons than their counterparts in other parts of the world. At the same time, the public availability of the NTD makes it possible for others to compare transit performance in ways that transit agencies may not necessarily agree with.

Optimal Peer Group Size

A success factor that is rarely explicitly stated—North Carolina's *Benchmarking Guidebook* (34) being an exception—but that is generally implied through the way that peer groups are developed is that there are upper and lower limits to how many peers should be included in a peer group. Too many peers result in a heavy data collection burden and the possibility that peers are too dissimilar to draw meaningful conclusions. Too few peers makes it difficult to credibly judge how well a transit agency is performing, and in a worst case

could lead to accusations that the peers were hand-picked to make an agency look good. In general, anything below 4 peers is considered to be too few, while somewhere in the range of 10 to 20 peers is considered to be too many, depending on the application.

Benchmarking Networks

Transit benchmarking networks have had the greatest success, both in terms of longevity and documented results. Such networks also exist in the private sector. The advantages of benchmarking networks include:

- Participants agree upon common measures and data definitions—this provides standardization, focuses data collection on areas of interest to the group, and gives participants more confidence in the quality of the data and the results.
- Participants have already agreed that they share a sufficient number of characteristics in common—this helps reduce, if not eliminate, questions afterwards about how comparable a particular peer is.
- Cost-sharing is possible, allowing participants to get better-quality information at a lower cost than if they were to conduct a benchmarking exercise on their own.
- Networks facilitate the development and comparison of long-term performance trends.
- Agency staff grow professionally through exposure to and discussions with colleagues in similar positions at other participating agencies.
- Confidentiality, if desired.

Two key success factors for transit benchmarking networks in Europe have been the use of an external facilitator (e.g., a university or a private consultant) and ongoing financial support. The facilitator performs functions that individual transit agency staff may not have time or experience for, including compiling and analyzing data, producing reports, and organizing meetings (e.g., information-sharing working groups on a specific topic or an annual meeting of the network participants) (18, 20). The cost of the facilitator is shared among the participants. At least two European pilot benchmarking networks (13, 23) dissolved after EU funding for the research project (and the facilitator) ended.

Benchmarking networks are not easy to maintain: they require long-term commitments by the participating agencies to contribute resources to the effort, to the benefit of all. At the same time, both private- and public-sector experiences indicate that the knowledge transfer benefits and the staff networking opportunities provided by a benchmarking network provide a valuable return on the agency's investment.

Benefits of and Challenges with Transit Peer Comparisons

Benefits of Transit Peer Comparisons

Most of the participants in this project's outreach effort (44) agreed that peer comparisons should be used as one tool in a set of various management tools for measuring performance. From a manager's perspective, it is always valuable to have a sense of where performance lies relative to other similar agencies. Useful information can be revealed even if a given methodology might have some flaws and not be "perfect" or "ideal." In addition, even if not necessarily used by outside agencies to determine funding levels or otherwise measure accountability, peer comparisons can be used as a way to foster competition and motivate transit agencies to improve their performance. When used internally, such comparisons can provide insight into areas where an agency is performing relatively well among its peers or where some improvements might be needed. However, nearly all those contacted stated that peer comparisons should not be used as the only benchmark for a transit agency's performance. The general consensus is that they are very good diagnostic tools but are typically not complex enough (by nature) to facilitate a complete understanding of performance.

Most transit agencies use the NTD for peer comparisons, and most expressed a general satisfaction with being able to use the data relatively easily to facilitate comparisons. While there are certainly limitations to the NTD (see the next section), it was noted that it has less ambiguity relative to other data sources due to the somewhat standard definitions and reporting requirements. Comments such as "it's what we've got" and "it's the best of the worst" were heard in the discussions. More than one individual stated that the NTD is "better" and more reliable than the comparable data used on the highway side by the Federal Highway Administration (thus making the point that all large federal databases have their own sets of problems and issues).

Also, peer comparisons can be used to support requests for more resources for a transit agency. This might be an easier task when an agency's performance is considered better than its peers. However, with the proper presentation, an agency's relatively poorer performance might also be used to show a need for more resources (e.g., when an agency's local funding levels are shown to be much lower than its peers).

Overall, the outreach participants have learned a great deal from their experiences with peer comparisons, and they can be considered tools that are valuable regardless of the outcome. When a transit agency compares favorably to its peers, it can provide a sense of validation for current efforts. When an agency appears not so favorably, lessons can be learned about what areas need more attention.

Challenges with Transit Peer Comparisons

While most outreach participants agreed that transit peer comparisons are useful tools, many challenges to the process were acknowledged. Outreach participants noted that making true "apples to apples" comparisons are difficult and that, in designing a methodology, it is hard to "be all things to all people."

At least one participant believes that all statistical comparisons among transit systems are "fatally flawed" due to the basic settings of the various systems or their board policies, which result in substantial differences that make clear comparisons nearly impossible. Alternatively, as one participant stated, "No one said this has to be easy." There will always be arguments that "we're so unique, we'll never be like so-and-so," or "it's so different here," yet most agree that such complaints should not thwart the careful development and use of such comparisons.

One major issue that can cause problems in transit comparisons is the peer selection process itself. Who is selecting the peers? When a transit agency self-selects, there can be a bias against including agencies that might be performing better. Several participants noted that managers might ignore, manipulate, or otherwise skew information that does not make the system look good. It can be relatively easy to present data in a way that an agency wants it presented. In addition, several of those with direct experience developing peer groups for analysis indicated that they were often told to include certain agencies in the analysis that were clearly not appropriate peers. Often the motivation for adding such agencies to the analysis included a sense of competition with the other community or a desire to be more like another community (perhaps in other ways besides just transit; i.e., "We are always compared to such-and-such community in other ways, so we should add them to the peer group"). Including communities that are not necessarily appropriate peers might be helpful if the purpose of the exercise is to see what the differences are; however, it will not be as instructive if the purpose is to benchmark existing performance.

Because much of the information used in the typical transit peer comparisons is statistical in nature, a lack of the appropriate technical knowledge among those either conducting the analysis or interpreting it can cause problems. As one participant noted, "Do you have 'numbers' people on staff?" Without a thorough understanding of how the numbers are derived and what they mean, and without being able to properly convey that information to those who will interpret and use the results, "weaknesses can be magnified" and the overall usefulness of the process is reduced.

While the NTD, as a relatively standardized database, is the source of most information used in typical transit comparisons, there is some limited utility of the data. The following

are some of the issues that outreach participants see with the use of NTD as related to transit peer comparisons:

- Despite standardized definitions, some transit agencies still report some items differently and/or not very well, particularly in the service area and maintenance categories (and other factors not necessarily in the NTD, such as on-time performance, can also be measured quite differently among agencies);
- NTD provides only a “once a year” (or year-end) picture of performance;
- Data lags (e.g., data for report year 2006 data become nationally available at the beginning of calendar year 2008);
- Only one service area is reported for an agency, but service area can vary greatly by mode (especially with demand-response service), thus leading to issues with any per-capita measures;
- Missing or otherwise incomplete data, particularly for smaller agencies; and
- Limited information for contracted services, although such services sometimes represent a significant portion of the service operated.

In addition, many participants noted that other relevant factors that should be included in any comparison are not found in the NTD. To paraphrase one participant, it should be remembered that NTD is just a “compromise” offset against the burden to the agencies of reporting requirements.

Another negative, according to a few participants, is that the typical transit comparisons focus too much on the information that is most easily measured or focus on too few measures. While some might argue that such a focus is appropriate, especially if a method is expected to be widely applied, others believe it will not result in the best and most meaningful comparisons, thus reducing the effort to simply a “paper exercise.” Some believe in the “less is more” approach, while others believe that “more is more.”

Media Issues

For many in the project’s outreach effort, media reactions to transit peer comparisons have not been very controversial. Unless there is some major negative issue, the media will often ignore the information or simply report anecdotal information. In some areas where peer comparisons are very favorable, the agencies often promote the information to the media as a way to gain additional support for transit services in the community.

Alternatively, dealing with the media can sometimes be a challenge for some transit agencies. There might be questions about the peer selection process and why some agencies were included (or excluded) from the analysis. As one participant

stated, the media will always “slant to the negative,” and so whoever might be presenting the information must really understand the methods and numbers and be able to convey them appropriately to the audience. The agency representatives should be comfortable enough with the data and results and be ready and able to explain the meaning and relevance of the information. In addition, if something looks “different,” it is important to remember that “different” does not necessarily mean “bad.” Some participants added that having a set methodology to follow (determined external to the agency) can be a way to show that an objective process was used.

Lessons Learned

After 30 years, benchmarking is well established in the private sector, and its benefits are widely recognized. Public sector adoption of benchmarking is more recent (generally since the mid-1990s), but many examples of successful benchmarking programs can already be found in the United States. There has been significant interest in Europe in public transit benchmarking, particularly since the late 1990s, and there are currently four well-established international benchmarking networks catering to different modes and city sizes. However, although a few in the U.S. public transit industry have recognized the benefits of benchmarking and have taken steps toward incorporating it into agency activities, it is not yet a widespread practice.

U.S. and Canadian transit agencies wishing to conduct peer comparisons or full-scale benchmarking efforts have a significant advantage not available to their counterparts in the rest of their world, namely the existence of standardized databases (the NTD and Canadian Transit Statistics, respectively) that provide access to a wide array of consistently defined variables that have been reported by a large number of agencies over a considerable period of time. Although NTD data are still perceived by many in U.S. transit agencies as being unreliable—and certainly there is still room for improvement—the testing conducted by this project found that the NTD is usable for a wide variety of benchmarking applications.

The Florida DOT has sponsored for a number of years the Florida Transit Information System (FTIS) software, which is a freely available, powerful tool for accessing and analyzing data from the complete NTD. The peer-grouping methodology described in this report has now been added to FTIS, making peer comparisons quicker to perform than ever and allowing for a greater depth of analysis.

Peer comparison is best applied as a diagnostic tool that helps agency management identify areas for improvement, particularly when one takes the approach from the start that one always has room for improvement. The results of peer comparisons should be used to stimulate questions about the reasons behind the performance results, which in turn can

lead to ideas that can result in real performance improvements. Many international transit agencies have found that the contacts they make with peer agencies as a result of a benchmarking process provide the greatest benefit, rather than any set of numerical results. However, the numerical analysis remains an important intermediate step that allows one to identify best-practice peers.

Management support is vital for performance measurement in general, but particularly so for a benchmarking process. Resources need to be allocated to make the peer agency contacts, and both resources and a management champion are needed to support any initiatives identified through the benchmarking process designed to improve agency performance. During times of economic hardship, management support is particularly vital to keeping an established program running; however, it is also exactly at these times when benchmarking can be a particularly vital tool for identifying potential improvements and efficiencies that can help a transit agency continue to perform its mission using fewer overall resources.

Benchmarking networks represent the highest level of benchmarking and are particularly useful for (a) compiling standardized databases of measures not available elsewhere and (b) coordinating contacts between organizations on topics of mutual concern. Networks can also help spread the cost of data analysis and collection over a group of agencies, thus reducing the costs for all participants compared to each participant performing their own separate analyses. The use of

an external facilitator has been a common success factor for transit benchmarking networks. However, joining a network is not a requirement to successfully perform benchmarking—agencies can still learn a lot from conducting their own individual efforts.

Finally, while it is desirable to have peer transit agencies share as many characteristics as possible as the agency performing the comparison, it should also be kept in mind that all transit agencies are different in some respect and that one will never find exact matches to one's own agency. The need for similarity is more important when higher-level performance measures that can be influenced by a number of factors are being compared (e.g., agency-wide cost per boarding) than when lower-level measures are being compared (e.g., miles between vehicle failures). Keep in mind that a number of successful benchmarking efforts have occurred across industries by focusing comparisons only on the areas or functions that the organizations have in common.

In summary, performance measurement, peer comparison, and benchmarking are tools that a transit agency can apply and benefit from right now. Potential applications are described in Chapter 3. Some of the potential issues identified earlier in this section are addressed by this project's methodology, while others simply require awareness of the potential presence of the issue and tools for dealing with the issue (Chapter 4). There is also room for improvements to the process in the future; Appendix C provides recommendations on this subject.

CHAPTER 3

Applications and Performance Measures

Applications

The range of questions where benchmarking and peer comparison are valuable spans all aspects of a transit agency's functions. Applications can range from the very detailed, such as a comparison of mean time between farebox failures, to broad public policy goals, such as a planning effort to develop a balanced, multi-modal regional transportation system. Peer-comparison applications are divided below into four general categories that describe the overall focus of a particular analysis, recognizing that there is room for overlap between the various categories.

1. *Administration* – questions related to the day-to-day administration of a transit agency, including (but not limited to) financial-performance questions asked by agency management, agency board members, and transit funding organizations.
2. *Operations* – questions related to a transit agency's daily operations.
3. *Planning* – long-term policy and service questions of interest to transit operators, metropolitan planning organizations, and state departments of transportation.
4. *Public and market focus* – questions that consider the viewpoint of the broad range of customers, including riders, non-riders, local jurisdictions, and policy-makers.

Administration

Performance questions falling into the agency administration category can be raised at all levels of management and oversight, including department managers, top-level transit agency managers, transit board members, oversight and funding agencies, and legislative bodies. Historically, peer comparison has been most widely applied in the United States to the financial aspects of transit agency administration (and financial questions were the most common per-

formance topic picked by participating agencies in this project's methodology testing), but peer comparison can also be applied to other aspects of transit agency administration, particularly aspects relating to labor costs and labor utilization.

Examples of performance questions relating to agency administration include:

- How efficient are our bus and rail operator work schedules?
- How comparatively cost-effective is our operation?
- What percentage of transit revenue comes from advertising?
- What is the typical subsidy level for an area our size?
- How does our absenteeism compare to peer agencies?
- What is the farebox recovery ratio for peer agencies' long-distance regional commuting routes?
- How do our state's small urban operators compare to their peers in terms of cost-effectiveness, cost-efficiency, and productivity?
- How relatively cost-efficient are the transit agencies that we fund?
- How does our employee compensation compare to other transit agencies?

Operations

These are performance questions asked by those responsible for the day-to-day operations of the transit agency to help ensure that the service provided meets the agency's stated goals. These kinds of questions can also be asked when looking for ways to improve specific departmental operations. Transit operators typically ask these questions in support of continuous process improvement and short-term (e.g., 1-year) planning efforts. Examples of these questions include:

- How cost-effective are our vehicle and non-vehicle maintenance programs?

- How often do our buses break down on the road, compared to our peers?
- How does the average speed of our buses while in service compare to our peers?
- What is our vehicle fuel economy compared to our peers?
- What are vehicle accident rates for other agencies our size?
- How do other agencies compare to ours in terms of demand-response ridership trends?

Planning

Planning performance questions are typically longer-term in nature and have policy and funding implications. They can be asked by transit agencies as part of their own internal planning or by external agencies [e.g., cities, Metropolitan Planning Organizations (MPOs), states] in support of long-range or modal plans. Planning questions can be hypothetical in nature and involve looking at peers that have characteristics that a transit agency expects or wants to have in the future. Examples of these kinds of questions include:

- How well do peer agencies with dedicated local funding sources perform in terms of ridership and financial performance, compared to ours?
- What mix of funding sources are used by transit agencies that have just reached the 200,000 population threshold?
- How much does it cost per hour for relatively new light-rail systems to provide service?
- What mix of transit services do peer regions provide?

Public and Market Focus

In Chapter 2, the importance of integrating the customer perspective into benchmarking efforts was identified. Public transit has multiple customer types, both those who use the service directly and those who benefit indirectly (for example, through improved air quality, reduced congestion, or land use and infrastructure improvements designed to support transit). Public and market-focus applications look at the viewpoints of a broad range of customers. Examples of performance questions in this area include:

- How does our service quality compare to that of our peers?
- How do we compare to our peers in terms of customer service and satisfaction?
- How do we compare to our peers in terms of how much transit service is provided?
- How does our level of investment in transit compare to peer regions?
- How do our fares compare to fares of other agencies?

Performance Measures

Performance measures are used in peer-comparison and benchmarking processes to (a) provide quantitative information about the selected performance topic, (b) provide context about the peer agencies, and (c) screen out potential peers based on specific transit agency characteristics.

Once a performance topic has been picked, it is necessary to identify measures that can be used to compare a transit agency to its peers in a standardized, credible way. Some performance measures used in a peer comparison quantify outcomes, while other, descriptive, measures provide context about peer agencies or are used to screen out transit agencies with particular characteristics from consideration as potential peers.

The performance measures selected for any given peer comparison will vary, depending on the performance question being asked. For example, a question about the cost-effectiveness of a transit agency's operations would focus on financial outcome measures, while a question about the effectiveness of an agency's maintenance department could use measures related to maintenance outcomes (e.g., maintenance expense per vehicle mile), agency investments (e.g., average fleet age), and performance outcomes (e.g., revenue miles between failures). In addition, some descriptive measures would often be incorporated into the review to provide context about the individual peer agencies.

Because each performance question is unique, it is not possible to provide a standard set of measures. Instead, the remainder of this section provides lists of standardized measures that are available from (or derivable from) the FTIS software tool, categorized by type of measure (descriptive versus performance ratio) and subject area (e.g., maintenance performance, agency characteristics). Scan through these lists and read the accompanying text on definitions and limitations of the measures to identify readily available measures that relate to the performance question. The case studies in Chapter 5 can also be used to identify performance measure examples for selected performance questions. *TCRP Report 88 (1)* provides definitions and further information about these and other measures.

The lists in this section also provide a selection of standardized measures available outside FTIS, along with other measures commonly collected by transit agencies. These measures can be used for peer comparisons on topics where the NTD lacks measures. However, refer to Chapter 4 for cautions about the extra time and effort required when incorporating non-NTD measures into a peer comparison. Note also that the NTD only provides detail at the agency and mode levels and that it will be necessary to obtain data directly from peer agencies if a finer level of detail is desired. Chapter 4 also discusses things to consider when requesting data directly from peer agencies.

Table 1. Cost-efficiency measures.

Directly Available from FTIS
Operating cost per revenue hour
Operating cost per revenue mile
Vehicle miles (hours) per revenue mile (hour)
Operating cost per peak vehicle in service

Outcome Measures

Outcome measures describe the performance achieved by the transit agency, given a set of inputs. Many of these measures are performance ratios that compare an out-come (e.g., ridership) to an input (e.g., revenue hours). These ratios can often be derived from two or more NTD variables, and FTIS provides a set of “Florida Standard Variables” that includes common performance ratios as direct outputs. Outcome measures are organized into the following nine categories:

- Cost-efficiency,
- Cost-effectiveness,
- Productivity,
- Service utilization,
- Resource utilization,
- Labor administration,
- Maintenance administration,
- Perceived service quality, and
- Safety and security

Cost-Efficiency

Cost-efficiency measures (Table 1) assess an agency’s ability to provide service outputs within the constraints of service inputs. According to *TCRP Report 88 (1)*, “These types of measures are very common and are utilized by virtually all transit systems when evaluating system-wide performance. However, these measures should be viewed with caution, because they do not measure a transit system’s ability to meet the needs of its passengers. These measures only evaluate how efficiently a system can put service on the street, irrespective of where the service is going or how much it is utilized.” Four cost-efficiency measures are directly available from FTIS. *Operating cost per revenue hour* and *operating cost per revenue mile* measure how much it costs to provide a unit of service. *Vehicle miles (hours) per revenue mile (hour)* assesses how much vehicle usage occurs in revenue

Table 2. Cost-effectiveness measures.

Directly Available from FTIS	Derivable from FTIS
Farebox recovery ratio	Operating ratio
Operating cost per boarding	Subsidy per boarding
Operating cost per passenger-mile	
Operating cost per service area capita	

service (as opposed to traveling to or from a garage, or other non-revenue service). *Operating cost per peak vehicle in service* looks at how much it costs annually to operate each vehicle used in peak service.

Cost-Effectiveness

Cost-effectiveness measures (Table 2) compare the cost of providing service to the outcomes resulting from the provided service. As with the cost-efficiency measures, many of these measures are commonly used by the transit industry. *Farebox recovery ratio* measures how much of a transit agency’s operating costs are covered by farebox revenue. As noted in Chapter 4, some agencies may have significant directly generated revenue that does not come from the farebox (e.g., service contracts with universities or advertising); therefore, the *operating ratio* (directly generated non-tax revenue divided by operating costs) may be a better measure in those situations. Operating ratio is not directly provided by FTIS, but can be derived from other measures available through FTIS. (When doing a mode-specific analysis, a portion of the agency’s non-farebox revenue will need to be allocated to each mode—for example, in proportion to ridership.) *Operating cost per boarding* looks at how much it costs to serve one unlinked trip, while *subsidy per boarding* (derivable from FTIS) measures the difference between the average cost to provide a trip and the average fare paid. *Operating cost per passenger-mile* relates costs to passenger loads, while *operating cost per service area capita* relates costs to the number of people within the agency’s service area. (Because service area populations are reported inconsistently by transit agencies, this variable should be used with caution.)

Productivity

Productivity measures (Table 3) look at how many passengers are served per unit of service—hours, miles, vehicles, or employee full-time equivalents.

Table 3. Productivity measures.

Directly Available from FTIS	Derivable from FTIS
Boardings per revenue hour	Boardings per vehicle operated in maximum service
Boardings per revenue mile	
Boardings per employee full-time equivalent	

Table 4. Service utilization measures.

Directly Available from FTIS	Not Available from FTIS
Annual boardings (unlinked trips)	Annual linked trips
Annual passenger miles	Annual linked trips per service area capita
Average trip length	
Annual boardings per service area capita	

Service Utilization

Service utilization measures (Table 4) look at how passengers use the service that is provided. *Annual boardings (unlinked trips)* is one of the most basic performance indicators for a transit agency; however, it overstates the number of person-trips made by transit each day, as each transit vehicle boarding is counted as a separate trip (i.e., a one-way trip involving a transfer between vehicles is counted as two unlinked trips). *Annual linked trips* measures the number of actual person-trips made using transit, which is useful for comparing transit usage to the usage of other modes. Annual linked trips can be calculated as annual unlinked trips minus annual transfers (which may be available from agency farebox data, depending on the type of fare media used, or may have been estimated from rider surveys). *Annual passenger miles* reflects both how many people use transit and the length of their trips; *average trip length* can be calculated as annual passenger miles divided by annual unlinked trips. *Average boardings per service area capita* is a useful measure for comparing transit usage between regions, but can be influenced by the service pattern used by an agency. (Agencies with timed-transfer hubs, grid networks, or multiple modes, for example, may have more boardings than agencies using radial networks that have an equivalent number of people making transit trips.) Using linked trips, if possible, addresses this issue. Since service area population is reported inconsistently to the NTD, urban area population can be used as a substitute, but only when the agencies being compared have similar service patterns (e.g., when they are the only agencies providing service to their regions).

Resource Utilization

Resource utilization measures (Table 5) investigate how well the agency's resources—vehicles, employees, consumables, and so on—are used. Most of these measures are self-explanatory. *Peak-to-base ratio* compares the number of vehicles operated during the highest peak period to the number of vehicles operated midday. It can be derived from FTIS for larger agencies (those operating 150 or more vehicles, not including demand response and vanpool vehicles).

Labor Administration

Labor administration measures (Table 6) include an array of measures that are applicable to both day-to-day transit agency management and to labor negotiations (e.g., comparisons of wages and benefits). A number of these measures can be derived from other FTIS measures. The relative proportions of administrative, vehicle operator, vehicle maintenance, and non-vehicle maintenance staff costs to total operating costs can be compared. *Pay-to-platform hours* compares vehicle operators' total regular paid working time (including reporting and turn-in time, minimum work guarantees, and other time allowances) to platform time worked (i.e., time spent operating the vehicle). *Percent of labor hours that are overtime* looks at the contribution of overtime to overall costs. Some overtime may be beneficial to a transit agency's bottom line, as it can be less than the total wages and benefits required to hire someone else to do the work, but excessive

Table 5. Resource utilization measures.

Directly Available from FTIS	Derivable from FTIS
Vehicle hours per vehicle operated in peak service	Revenue hours per vehicle operated in peak service
Vehicle miles per vehicle operated in peak service	Revenue miles per vehicle operated in peak service
Revenue hours per employee full-time equivalent	Peak-to-base ratio
Vehicle miles per gallon of fuel consumed	
Vehicle miles per kilowatt-hour of power consumed	

Table 6. Labor administration measures.

Derivable from FTIS	Not Available from FTIS
Cost of staff type/operating costs	Employee absenteeism rate
Pay-to-platform hours	Staff turnover rate
Percent of labor hours that are overtime	
Percent of operating costs that are wages (and benefits)	

Table 7. Maintenance administration measures.

Directly Available from FTIS	Derivable from FTIS
Vehicle (car) miles between failures	Labor cost per vehicle hour
Number of vehicle system failures	Maintenance category cost/total maintenance cost
Maintenance cost as a percentage of operating costs	Average annual maintenance cost per vehicle operated in maximum service
	Vehicle maintenance cost/vehicle (car) mile
	Maintenance full-time equivalents (FTEs)/vehicle operated in maximum service
	Non-vehicle maintenance cost/track mile

overtime and overtime required to cover other employees' absences is not cost-efficient (1). *Percent of operating costs that are wages (and benefits)* measures how much employee compensation contributes to total operating costs. Other employee-related data that are not available from FTIS, but may be available from peer agencies' human resources departments, are *employee absenteeism rate* (impacts the costs required to pay other employees to do the work scheduled for the absent employees) and *staff turnover rate* (reflects costs required to train new staff and inefficiencies when other employees are covering for staff who have left).

Maintenance Administration

Maintenance administration measures (Table 7) focus on the performance of the transit agency's vehicle maintenance function, and also provide insights into the overall condition of the vehicle fleet. *Vehicle (car) miles between failures* is a measure of how often vehicles break down while in service, while *number of vehicle system failures* looks at the total number of failures. It should be kept in mind that these measures do not tell the whole story about maintenance quality, as fleet age and overall agency investment in maintenance activities (e.g., *maintenance cost as a percentage of operating costs*) also play a role. *Labor cost per vehicle hour* is an indicator of how much maintenance work is required relative to the amount of time that vehicles are operated. Cost data are available for several maintenance categories (labor, parts, consumables), which can be compared to the overall maintenance budget. Finally, the *average annual maintenance cost per vehicle operated in maximum service* can be derived from FTIS.

Perceived Service Quality

Perceived service quality measures (Table 8) describe the transit agency's service as perceived by customers. (Delivered service quality—taking the agency point of view—is discussed later in the descriptive measures section.) Except for *average system speed* (revenue miles per revenue hours), which is provided directly by FTIS, the NTD does not provide any measures of perceived service quality. However, a number of useful measures may be obtainable from peer agencies. *On-time performance* is a measure of reliability; however, it is not defined consistently by transit agencies [i.e., what constitutes “on-time” and the location(s) where it is measured]. If archived automatic vehicle location data are available, *excess wait time* (the number of extra minutes passengers had to wait past the scheduled departure time) is an alternative measure of reliability that avoids the “on-time” definition issue. *Passenger load* data may be available from archived automatic passenger counter data, or (with considerable manual effort) from data-collection sheets used for NTD passenger-mile reporting. Many transit agencies conduct customer satisfaction surveys, and questions relating to *overall satisfaction* are often asked in a consistent manner (although the scale used to measure satisfaction may vary from survey to survey). Many transit agencies also track complaints and compliments, but because the process to submit comments may be easier at some agencies than at others, it may be necessary to analyze the total volume of comments in conjunction with analyzing (for example) the *number of complaints (compliments) per 1,000 boardings* to get an accurate picture of relative satisfaction or dissatisfaction with service. *Call center response time* is a measure of how conveniently passengers

Table 8. Perceived service-quality measures.

Directly Available from FTIS	Not Available from FTIS
Average system speed	On-time performance
	Excess wait time
	Passenger loading
	Overall satisfaction
	Number of complaints per 1,000 boardings
	Number of compliments per 1,000 boardings
	Call-center response time
	Missed trips

Table 9. Safety and security measures.

Derivable Available from FTIS	Not Available from FTIS
Casualty and liability cost per vehicle mile	Collisions per 1,000 miles Collisions per 1,000 boardings Incidents per 1,000 boardings

can request information or book a demand response trip by telephone. Finally, *missed trips* tracks how many scheduled demand response trips were missed due to a problem on the part of the transit agency or its service contractor.

Safety and Security

Safety and security measures (Table 9) look at performance related to accidents, crimes, and quality-of-life incidents that can impact passengers' perceptions of the transit agency. Except for *casualty and liability cost per vehicle mile*, these measures are not available through FTIS because safety and security data are not publicly released by the FTA. However, as discussed in Chapter 3, when peer agencies are willing to share their NTD viewer password with the target agency, safety and security data reported to the NTD can be readily incorporated into a peer comparison. It should be kept in mind that there are consistency issues in how crime data are reported by transit agencies, depending on, for example, whether or not a transit agency has its own police force, how frequently arrests are made for lesser incidents, and how incidents are coded in police reports (46).

Descriptive Measures

Descriptive measures provide context about a particular transit agency. While they are not direct indicators of transit agency performance (i.e., outcomes), they are nevertheless valuable components of a performance-measurement process. Descriptive measures are particularly useful for diagnosing why outcome measure results vary between transit agencies. They can also be used as screening tools to make sure the

selected peer agencies match the target agency in specific characteristics relevant to the performance question being asked. Finally, descriptive measures can provide additional information to stakeholders in the benchmarking process that confirms that the selected peer agencies are reasonably similar to the target agency.

Many descriptive measures are available from FTIS. These measures usually come directly from NTD reporting data, but also include selected measures available from (or derivable from) other standardized national databases. These measures are organized into five categories:

- Urban area characteristics,
- Transit service characteristics,
- Transit agency characteristics,
- Delivered service quality, and
- Transit investment.

Urban Area Characteristics

Urban area characteristics measures available from FTIS (Table 10) describe the region's population characteristics, geographic size, land use patterns, demographic characteristics, congestion level, and presence of a state capital. These measures were derived from Census Bureau or *Urban Mobility Report* (45) data or were developed by TCRP Project G-11. See Appendix B for definitions of these measures; urban areas themselves are defined by the Census Bureau. Other standardized measures that are available outside of FTIS relate to climate (available from the National Oceanic and Atmospheric Administration) and cost-of-living index (for example, from the Council for Economic and Community Research).

Table 10. Urban area characteristics measures.

Directly Available from FTIS	Not Available from FTIS
Urban area population	Annual rainfall
Urban area size	Mean January high temperature
Urban area population density	Mean July high temperature
Urban area population growth rate	Cost-of-living index
Census block density	
Population dispersion	
Employment dispersion	
Percent residents in transit-supportive areas	
Percent college students	
Percent low income residents	
Annual delay per traveler	
Freeway lane-miles per capita	
State capital (yes/no)	

Table 11. Service characteristics measures.

Directly Available from FTIS	Not Available from FTIS
Service area population	Percent of population served by fixed-route transit
Service area size	
Service area type	
Annual vehicle miles operated	
Annual revenue hours operated	
Miles of track	
Number of stations	
Percent of service operated as fixed-route	
Percent of service that is demand-response	
Average fare	

Transit Service Characteristics

The service characteristics measures available from FTIS (Table 11) describe the size and population of a transit agency's service area, the type of service provided by a transit agency (e.g., service to the entire region vs. service to a portion of the region's suburbs combined with commuter trips into the central city), the amount of hours and miles of service provided, the amount of transit infrastructure provided, the amount of service that is contracted, the amount of total service that is demand-response, and the average fare. Except for service type (developed by TCRP Project G-11), these measures are taken directly from the NTD or are derived from other NTD measures (for example, *average fare* is defined as *annual fare revenue* divided by *annual unlinked trips*). Note that service area population and size are not currently reported consistently by transit agencies. Analysts with access to transit route network data in a geographic information systems (GIS) compatible format can combine these data with census data to estimate the *percent of a region's population served by fixed-route transit*.

Transit Agency Characteristics

The transit agency characteristics measures available from FTIS (Table 12) describe the organization type (e.g., public

agency that directly operates all service); the institutional structure (e.g., independent agency with an appointed board of directors); the demand-response provider type (e.g., social service agency); the number of employee full-time equivalents (FTEs) in vehicle operations, vehicle maintenance, non-vehicle maintenance, and general administration; and the amount of revenue from various sources. All of these measures are taken directly from the NTD. A transit agency's *service philosophy* (i.e., service coverage emphasis vs. efficiency emphasis) is a potential screening measure. It can often be identified from an Internet search, by looking at an agency's goals and objectives, or by looking at the system's route map.

Delivered Service Quality

Delivered service quality measures (Table 13) describe the transit agency's service as delivered by the agency. The service quality perceived by passengers was discussed earlier in the outcome measures section (Table 8). Except for *service span*, which applies to the agency as a whole, the NTD does not provide any direct measures of delivered service quality. A few measures are derivable from NTD data and are provided directly by FTIS, including *average system peak headway* (derived from directional route miles, average system speed, and the number of vehicles operated in maximum service), *revenue miles per urban square*

Table 12. Agency characteristics measures.

Directly Available from FTIS	Not Available from FTIS
Organizational type	Service philosophy (coverage vs. efficiency)
Institutional structure	
Demand-response provider type	
Number of employee FTEs by category	
Revenue by source	

Table 13. Delivered service quality measures.

Directly Available from FTIS	Derivable from FTIS
Service span	Percent of fleet with ramps/low-floor
Average system peak headway	
Revenue miles per urban area sq. mi	
Revenue miles (hours) per capita	

Table 14. Transit investment measures.

Directly Available from FTIS	Derivable from FTIS
Average fleet age	Operating funding per capita
Spare ratio	Operating subsidy per capita
Local revenue	Capital funding per capita
State revenue	
Federal revenue	

mile, and *revenue hours per capita* (measures of coverage). Finally, *percent of fleet with ramps/low floor* is a measure of ADA accessibility that can be derived from the NTD vehicle fleet data available through FTIS.

Transit Investment

Transit investment measures (Table 14) look at local, state, and federal investments in transit service and infrastructure

and the agency's investment in transit vehicles. These measures can also compare the total transit investment to the number of people within an agency service area or region. (As discussed previously, per-capita measures based on service area population should be used with caution, as this service area population is reported inconsistently to the NTD.) *Average fleet age* is based on the active vehicles in the fleet. *Spare ratio* is the difference between the number of vehicles available and the number of vehicles operated in maximum service, divided by the number of vehicles operated in maximum service. Low spare ratios may indicate potential problems in scheduling preventative maintenance and lack of vehicle capacity to respond to increased demand for service, while high spare ratios may indicate an inefficient use of capital and maintenance funds. Local, state, and federal operating and capital revenue amounts are available through FTIS, both as aggregate amounts and broken down by source.

CHAPTER 4

Benchmarking Methodology

Introduction

This chapter describes a step-by-step process, in eight steps, for conducting a trend-analysis, peer-comparison, or full-fledged benchmarking effort. Not all of the steps described below will be needed for every effort.

The first step of the process is to understand the context of the benchmarking exercise. What is the source of the issue being investigated? What is the timeline for providing an answer? Will this be a one-time exercise or a permanent process? The answers to these questions will determine the level of effort for the entire process and will also guide the selection of performance measures and screening criteria in subsequent steps.

In Step 2, performance measures are developed that relate to the performance question being asked. A benchmarking program that is being set up as a regular (e.g., annual) effort will use the same set of measures year after year, while one-time efforts to address specific performance questions or issues will use unique groups of measures each time. This report's peer-grouping methodology screens potential peers based on a number of common factors that influence performance results between otherwise similar agencies; however, additional screening factors may be needed to ensure that the final peer group is appropriate for the performance question being asked. These secondary screening factors are also identified at this stage.

In all applications except a simple trend analysis of the target agency's own performance, a peer group is established in Step 3. This report's basic peer-grouping methodology has been implemented in the freely available Web-based FTIS software. Instructions for using FTIS are provided in Appendix A. FTIS identifies a set of potential peers most like the agency performing the peer-review (the "target agency"), based on a variety of criteria. The software provides the results of the screening and the calculations used in the process so that users can inspect the results for reasonableness. Once the

initial peer group has been established, the secondary screening factors can be applied to reduce the initial peer group to a final set of peers.

After a peer group is identified, Step 4 compares the performance of the target agency to its peers. A mix of analysis techniques is appropriate—not just looking at a snapshot of the agencies' performance for the most recent year, but also looking at trends in the data. This effort identifies both the areas where the transit agency is doing well relative to its peers (but might be able to better) and areas where the agency's performance lags behind its peers. Ideally, the process does not focus on producing a "report card" of performance (although one can be useful for supporting the need for performance improvements), but instead is used to raise questions about potential reasons behind the performance and to identify peer group members that the target agency can learn from.

In a true benchmarking application, the process moves on to Step 5, where the target agency contacts its best-practices peers. The intent of these contacts is to (a) verify that there are no external factors unaccounted for that explain the difference in performance and (b) identify practices that could be adopted to improve one's own performance. A transit agency can skip this step, but it loses the value of learning what its peers have tried previously and thus risks spending resources unnecessarily in re-inventing the wheel.

If a transit agency seeks to improve performance in a given area, it moves on to Step 6, developing strategies for improving performance, and Step 7, implementing the strategies. The specifics of these steps depend on the particular performance-improvement need and the agency's resources and operating environment.

Once strategies for improving performance have been implemented, Step 8 monitors results on a regular basis (monthly, quarterly, or annually, depending on the issue) to determine whether the strategies are having a positive effect on performance. As the agency's peers may also be taking steps to

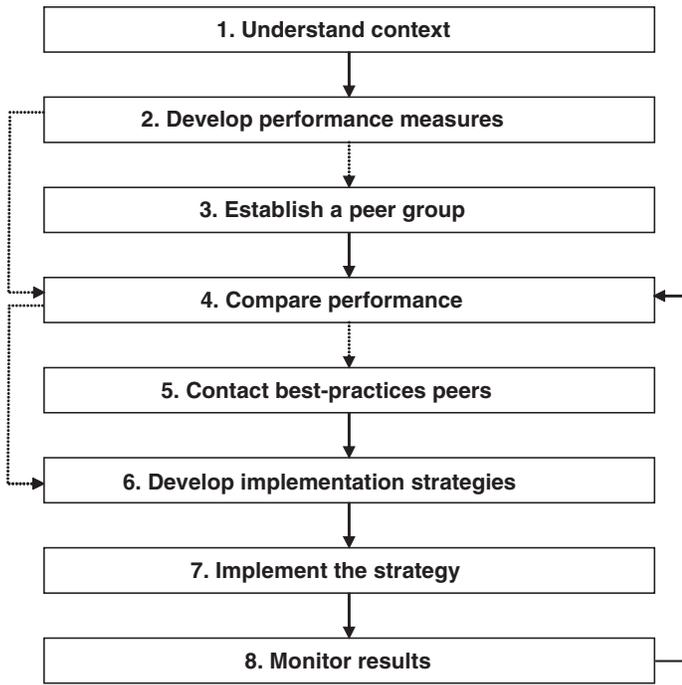


Figure 2. Benchmarking steps.

improve performance, the transit agency should periodically return to Step 4 to compare its performance against its peers. In this way, a cycle of continuous performance improvement can be created.

Figure 2 summarizes the steps involved in the benchmarking methodology. Places in the methodology where a step can be skipped or the process can end (depending on the application) are shown with dotted connectors.

Step 1: Understand the Context of the Benchmarking Exercise

The first step of the process is to clearly identify the purpose of the benchmarking effort since this determines the available timeframe for the effort, the amount and kind of data that can and should be collected, and the expected final outcomes. Examples of the kinds of benchmarking efforts that could be conducted, in order of increasing effort, are:

- Immediate one-time request, such as a news media inquiry following a proposed increase in fares.
- Short-range one-time request, such as a management focus to increase the fuel efficiency of the fleet in response to rising fuel costs.
- Long-range one-time request, such as a regional planning process that is relating levels of service provision to population and employment density, or a state effort to develop a process to incorporate performance into a formula-based distribution of grant funding.

- Permanent internal benchmarking process, where agency performance will be evaluated broadly on a regular (e.g., annual) basis.
- Establishment of a benchmarking network, where peer agencies will be sought out to form a permanent group to share information and knowledge to help the group improve its collective performance.

The level of the benchmarking exercise should also be determined at this stage since it determines which of the remaining steps in the methodology will need to be applied:

- Level 1 (trend analysis): Steps 1, 2, and 4, and possibly Steps 6–8 depending on the question to be answered.
- Level 2 (peer comparison): Steps 1–4, and possibly Steps 6–8 depending on the question to be answered.
- Level 3 (direct agency contact): Steps 1–5, and frequently Steps 6–8.
- Level 4 (benchmarking networks): Steps 1–3 once, Steps 4 and 5 annually, Step 6 through participation in working groups, and Steps 7 and 8 at the discretion of the agency.

Step 2: Develop Performance Measures

Step 2a: Performance Measure Selection

The performance measures used in a peer comparison are, for the most part, dependent on the performance question being asked. For example, a question about the cost-effectiveness of an agency's operations would focus on financial outcome measures, while a question about the effectiveness of an agency's maintenance department could use measures related to maintenance activities (e.g., maintenance expenses), agency investments (e.g., average fleet age), and maintenance outcomes (e.g., revenue miles between failures). Additional descriptive measures that provide context about peer agencies are also valuable to incorporate into a review.

Because each performance question is unique, it is not possible to provide a standard set of measures to use. Instead, use Chapter 3 of this report to identify 6 to 10 outcome measures that are the most applicable to the performance question, plus additional descriptive measures as desired. In addition, Chapter 5 provides case-study applications of the methodology that include examples of performance measures used for each application.

Performance measures not directly available or derivable from the NTD (or from the other standardized data included with FTIS) will require contacting the other transit agencies in the peer group. If peer agency data are needed, be sure to budget plenty of time into the process to contact the peers, to obtain the desired information from them, and to compile

the information. Examples of common situations where outside agency data might be required are:

- Performance questions involving specific service types (e.g., commuter bus routes);
- Performance questions involving customer satisfaction;
- Performance questions involving quality-of-service factors such as reliability or crowding; and
- Performance questions requiring detailed maintenance data.

Significant challenges exist whenever non-standardized data are needed to answer a performance question. Agencies may not collect the desired information at all or may define desired measures differently. If data are available, they may not be compiled in the desired format (e.g., route-specific results are provided, but the requesting agency desires service-specific results). Therefore, the target agency should plan on performing additional analysis to convert the data it receives into a useable form. It is often possible to obtain non-standard data from other agencies, but it does take more time and effort.

Benchmarking networks are a good way for a group of transit agencies to first agree on common definitions for non-NTD measures of interest and then to set up a regular data-collection and reporting process that all can benefit from.

Step 2b: Identify Secondary Screening Measures

This report's recommended peer-grouping methodology incorporates a number of factors that can influence one transit agency's performance relative to another. However, it does not account for all potential factors. Depending on the performance question being asked, a secondary screening might need to be performed on the initial peer group produced by the methodology. These measures should be selected prior to forming peer groups to avoid any perception later on that the peer group was hand-picked to produce a desired result. Examples of factors that might be considered as part of a secondary screening process include:

- **Institutional structure** (e.g., appointed board vs. a directly elected board): Available from NTD form B-10. (All NTD forms with publicly released data are viewable through the FTIS software.)
- **Service operator** (e.g., directly operated vs. purchased service): Although this factor is included in the peer-grouping methodology, it is not a pass/fail factor. Some performance questions, however, may require a peer group of agencies that purchase or do not purchase service. In other situations, the presence, lack, or mix of contracted service could help explain performance results, and therefore, this factor would not be desirable for secondary screening.

- **Service philosophy** [e.g., providing service to as many residents and worksites as possible (coverage) vs. concentrating service where it generates the most ridership (efficiency)]: Determined from an Internet inspection of agency goals and/or route networks.
- **Service area type** (e.g., being the only operator in a region): This report's peer grouping methodology considers eight different service area types in forming peer groups, but allows peers to have somewhat dissimilar service areas. Some performance questions, however, may require exact matches. Service area information is available through FTIS; the Internet can also be used to compare agencies' system maps.
- **Funding sources**: Available from NTD form F-10.
- **Vehicles operated in maximum service**: Available from NTD form B-10.
- **Peak-to-base ratio**: Derivable for larger agencies (at least 150 vehicles in maximum service, excluding vanpool and demand response) from NTD form S-10.
- **FTA population categories for grant funding**: An agency may wish to compare itself only to other agencies within its FTA funding category (e.g., <50,000 population, 50,000–200,000 population, 200,000 to 1 million population, >1 million population), or a funding category it expects to move into in the future. Service area populations are available on NTD form B-10, while urban area populations are available through FTIS.
- **Capital facilities** (e.g., number of maintenance facilities): Available from NTD form A-10.
- **Right-of-way types**: Available from NTD form A-20.
- **Service days and span**: Available from NTD form S-10.

Some of the case studies given in Chapter 5 provide examples of secondary screening.

Step 2c: Identify Thresholds

The peer-grouping methodology seeks to identify peer transit agencies that are similar to the target agency. It should not be expected that potential peers will be identical to the target agency, and the methodology allows potential peers to be different from the target agency in some respects. However, if a potential peer is substantially different in one respect from the target agency, it needs to be quite similar in several other respects for the methodology to identify it as a potential peer.

The methodology testing determined that not all transit agencies were comfortable with having no thresholds on any given peer-grouping factor—some thought suggested peers were too big or too small in comparison to their agency, for example, despite considerable similarity elsewhere. This report discourages setting thresholds for peer-grouping fac-

tors (e.g., the size that constitutes “too big”) when not needed to address a particular performance question. However, it is also recognized that the credibility and eventual success of a benchmarking exercise depends in great measure on how its stakeholders (e.g., staff, decision-makers, board, or the public) perceive the peers used in the exercise. If the peers are not perceived to be credible, the results of the exercise will be questioned. Users of the methodology at the local level are in the best position to gauge the factors that might make peers not appear credible to their stakeholders.

If thresholds are to be used, users should review the methodology’s peer-grouping factors to determine (a) whether a threshold is needed and (b) what it should be. As with screening measures, it is important to do this work in advance in order to avoid perceptions later on that the peer group was hand-picked.

Step 3: Establish a Peer Group

Overview

The selection of a peer group is a vital part of the benchmarking process. Done well, the selection of an appropriate, credible peer group can provide solid guidance to the agency, point decision-makers towards appropriate directions, and help the agency implement realistic activities to improve its performance. On the other hand, selecting an inappropriate peer group at the start of the process can produce results that are not relevant to the agency’s situation, or can produce targets or expectations that are not realistic for the agency’s operating conditions. As discussed above, the credibility of the peer group is also important to stakeholders in the benchmarking process—if the peer group appears to be hand-picked to make the agency look good, any recommendations for action (or lack of action) that result from the process will be questioned.

Ideally, between eight and ten transit agencies will ultimately make up the peer group. This number provides enough breadth to make meaningful comparisons without creating a burdensome data-collection or reporting effort. Some agencies have more unique characteristics than others, and it may not always be possible to come up with a credible group of eight peers. However, the peer group should include at least four other agencies to have sufficient breadth. Examples of situations where the ideal number of peers may not be achievable include:

- Larger transit agencies generally, as there is a smaller pool of similar peers to work with;
- Largest-in-class transit agencies (e.g., largest bus-only operators), as nearly all potential peers will be smaller or will operate modes that the target agency does not operate;

- Transit agencies operating relatively uncommon modes (e.g., commuter rail), as there is a smaller pool of potential peers to work with; and
- Transit agencies with uncommon service types (e.g., bus operators that serve multiple urban areas), as again there is a small pool of potential peers.

The peer-grouping methodology can be applied to a transit agency as a whole (considering all modes operated by that agency), or to any of the specific modes operated by an agency. Larger multi-modal agencies that have difficulty finding a sufficient number of peers using the agency-wide peer-grouping option may consider forming mode-specific peer groups and comparing individual mode performance. Mode-specific groups are also the best choice for mode-specific evaluations, such as an evaluation of bus maintenance performance.

Larger transit agencies that have difficulty finding peers may also consider looking internationally for peers, particularly to Canada. Statistics Canada provides data for most of the peer-grouping methodology’s demographic screening factors, including population, population density, low-income population, and 5-year population growth for census metropolitan areas. Many Canadian transit agency websites provide basic budget and service data that can be integrated into the peer-grouping process, and Canadian Urban Transit Association (CUTA) members have access to CUTA’s full *Canadian Transit Statistics* database (28).¹

For ease of use, this report’s basic peer-grouping methodology has been implemented in the Web-based FTIS software, which provides a free, user-friendly interface to the full NTD. However, the methodology can also be implemented in a spreadsheet, and was used that way during the initial testing of the methodology.

Detailed instructions on using FTIS to perform an initial peer grouping are provided in Appendix A, and full details of the calculation process used by the peer-grouping methodology are provided in Appendix B. The following subsections summarize the material in these appendices.

Step 3a: Register for FTIS

The NTD component of FTIS is accessed at <http://www.ftis.org/INTDAS/NTDLogin.aspx>. The site is password protected, but a free password can be requested from this page. Users typically receive a password within one business day.

¹ An important difference that impacts performance ratios derived from CUTA ridership data is that U.S. ridership data are based on vehicle boardings (i.e., *unlinked trips*), while CUTA ridership data are based on total trips regardless of number of vehicles used (i.e., *linked trips*). Thus, a transit trip that includes a transfer counts as two rides in U.S. data, but only one ride in CUTA data. *Unlinked trips* is the sum of *linked trips* and *number of transfers*. Some larger Canadian agencies also report unlinked trip data to APTA.

Step 3b: Form an Initial Peer Group

The initial peer-grouping portion of the methodology identifies transit agencies that are similar to the target agency in a number of characteristics that can influence performance results between otherwise similar agencies. “Likeness scores” are used to determine the level of similarity between a potential peer agency and the target agency both with respect to individual factors (e.g., urban area population, modes operated, and service areas) and for the agencies overall. Appendix A provides detailed instructions on using FTIS to form an initial peer group.

Transit agencies should not expect that their peers will be exactly like themselves. The methodology allows peers to differ substantially in one or more respects, but this must be compensated by a high degree of similarity in a number of other respects. (Agencies not comfortable with having a high degree of dissimilarity in a given factor can develop and apply screening thresholds, as described in Step 2c.) The goal is to identify a set of peers that are *similar enough* to the target agency that credible and useful insights can be drawn from the performance comparison to be conducted in Step 4.

The methodology uses the following three screening factors to help ensure that potential peers operate a similar mix of modes as the target agency:

- **Rail operator (yes/no).** A rail operator is defined here as one that operates 150,000 or more rail vehicle miles annually. (This threshold is used to distinguish transit agencies that operate small vintage trolley or downtown streetcar circulators from large-scale rail operators.) This factor helps screen out rail-operating agencies as potential peers for bus-only operators.
- **Rail-only operator (yes/no).** A rail-only operator operates rail and has no bus service. This factor is used to screen out multi-modal operators as peers for rail-only operators.
- **Heavy-rail operator (yes/no).** A heavy-rail operator operates the heavy rail (i.e., subway or rapid transit) mode. This factor helps identify other heavy-rail operators as peers for transit agencies that operate this mode.

As discussed in more detail in Appendix A, bus-only operators that wish to consider rail operators as potential peers can export a spreadsheet containing the peer-grouping results and then manually recalculate the likeness scores, excluding these three screening factors.

Depending on the type of analysis (rail-specific vs. bus-specific or agency-wide) and the target agency’s urban area size, up to 14 peer-grouping factors are used to identify transit agencies similar to the target agency. All of these peer-grouping

factors are based on nationally available, consistently defined and reported measures. The factors are:

- **Urban area population.** Service area population would theoretically be a preferable variable to use, but it is not yet reported in a consistent way to the NTD. Instead, the methodology uses a combination of urban area population and service area type—discussed below—as a proxy for the number of people served.
- **Total annual vehicle miles operated.** This is a measure of the amount of service provided, which reflects service frequencies, service spans, and service types operated.
- **Annual operating budget.** Operating budget is a measure of the scale of a transit agency’s operations; agencies with similar budgets may face similar challenges.
- **Population density.** Denser communities can be served more efficiently by transit.
- **Service area type.** Agencies have been assigned one of eight service types, depending on the characteristics of their service (e.g., entire urban area, central city only, commuter service into a central city).
- **State capital (yes/no).** State capitals tend to have a higher concentration of office employment than other similarly sized cities.
- **Percent college students.** Universities provide a focal point for service and often directly or indirectly subsidize students’ transit usage, thus resulting in a higher level of ridership than in other similarly sized communities.
- **Population growth rate.** Agencies serving rapidly growing communities face different challenges than either agencies serving communities with moderate growth rates or agencies serving communities that are shrinking in size.
- **Percent low-income population.** The amount of low-income population is a factor that has been correlated with ridership levels. Low-income statistics reflect both household size and configuration in determining poverty status and are therefore a more robust measure than either household income or automobile ownership.
- **Annual roadway delay (hours) per traveler.** Transit may be a more attractive option for commuters in cities where the roadway network is more congested. This factor is only used for target agencies in urban areas with populations of 1 million or more.
- **Freeway lane miles (thousands) per capita.** Transit may be more competitive with the automobile from a travel-time perspective in cities with relatively few freeway lane-miles per capita. This factor is only used for target agencies in urban areas with populations of 1 million or more.
- **Percent service demand-responsive.** This factor helps describe the scale of agency’s investment in demand-response service (including ADA complementary paratransit service)

as compared with fixed-route service. This factor is only used for agency-wide and bus-mode comparisons.

- **Percent service purchased.** Agencies that purchase their service will typically have different organization and cost structures than those that directly operate service.
- **Distance.** This factor serves multiple functions. First, it serves as a proxy for other factors, such as climate, that are more difficult to quantify but tend to become more different the farther apart two agencies are. Second, agencies located within the same state are more likely to operate under similar legislative requirements and have similar funding options available to them. Finally, for benchmarking purposes, closer agencies are easier to visit and stakeholders in the process are more likely to be familiar with nearby agencies and regions. This factor is not used for rail-mode-specific peer grouping due to the relatively small number of rail-operating agencies.

Likeness scores for most of these factors are determined from the percentage difference between a potential peer's value for the factor and the target agency's value. A score of 0 indicates that the peer and target agency values are exactly alike, while a score of 1 indicates that one agency's value is twice the amount of the other. For example, if the target agency was in a region with an urbanized area population of 100,000 while the population of a potential peer agency's region was 150,000, the likeness score would be 0.5, as one population is 50% higher than the other. For the factors that cannot be compared by percentage difference (e.g., state capital or distance), the factor likeness scores are based on formulas that are designed to produce similar types of results—a score of 0 indicates identical characteristics, a score of 1 indicates a difference, and a score of 2 or more indicates a substantial difference. Appendix A provides the likeness score calculation details for all of the peer-grouping factors.

The total likeness score is calculated from the individual screening and peer-grouping factor likeness scores as follows:

$$\text{Total likeness score} = \frac{\text{Sum}(\text{screening factor scores}) + \text{Sum}(\text{peer grouping factor scores})}{\text{Count}(\text{peer grouping factors})}$$

A total likeness score of 0 indicates a perfect match between two agencies (and is unlikely to ever occur). Higher scores indicate greater levels of dissimilarity between two agencies. In general, a total likeness score under 0.50 indicates a good match, a score between 0.50 and 0.74 represents a satisfactory match, and a score between 0.75 and 0.99 represents potential peers that may be usable, but care should be taken to investigate potential differences that may make them unsuitable. Peers with scores greater than or equal to 1.00 are undesirable due to a large number of differences with the target agency,

but may occasionally be the only candidates available to fill out a peer group.

A total likeness score of 70 or higher may indicate that a potential peer had missing data for one of the screening factors. (A factor likeness score of 1,000 is assigned for missing data; dividing 1,000 by the number of screening factors results in scores of 70 and higher.) In some cases, suitable peers may be found in this group by manually re-calculating the total likeness score in a spreadsheet and removing the missing factor from consideration, if the user determines that the factor is not essential for the performance question being asked. Missing congestion-related factors, for example, might be more easily ignored than a missing total operating budget.

Step 3c: Performing Secondary Screening

Some performance questions may require looking at a narrower set of potential peers than found in the initial peer group. For example, one case study described in Chapter 5 involves an agency that did not have a dedicated local funding source and was interested in comparing itself to peers that did have one. Another case study involves an agency in a region that was about to reach 200,000 population (thus moving into a different funding category) and wanted to compare itself to peers that were already at 200,000 population or more. Some agencies may simply want to make sure that no peer agency is “too different” to be a potential peer for a particular application. Data contained in FTIS can often be used to perform these kinds of screenings. Some other kinds of screening, for example based on agency policy or types of routes operated (e.g., commuter bus or BRT), will require Internet searches or agency contacts to obtain the information.

The general process to follow is to first identify how many peers would ideally end up in the peer group. For the sake of this example, this number will be eight. Starting with the highest-ranked potential peer (i.e., the one with the lowest total likeness score), check whether the agency meets the secondary screening criteria. If the agency does not meet the criteria, replace it with the next available agency in the list that meets the screening criteria. For example, if the #1-ranked potential peer does not meet the criteria, check the #9-ranked agency next, then #10, and so forth, until an agency is found that meets the criteria. Repeat the process with the #2-ranked potential peer. Continue until a group of eight peers that meets the secondary screening criteria is formed, or until a potential peer's total likeness score becomes too high (e.g., is 1.00 or higher).

Table 15 shows an example of the screening process for Knoxville Area Transit, using “existence of a dedicated local funding source” as a criterion. The top 20 “most similar” agencies to Knoxville are shown in the table in order of their total likeness score. The table also shows whether or not each agency has a dedicated local funding source. In this case,

Table 15. Example secondary screening process for Knoxville Area Transit.

Agency	City	State	Likeness Score	Dedicated Local Funding?	Use as Peer?
Knoxville Area Transit	Knoxville	TN	0.00		
1 Winston-Salem Transit Authority	Winston-Salem	NC	0.25	Yes	✓
2 South Bend Public Transportation Corporation	South Bend	IN	0.36	Yes	✓
3 Birmingham-Jefferson County Transit Authority	Birmingham	AL	0.36	Yes	✓
4 Connecticut Transit - New Haven Division	New Haven	CT	0.39	No	
5 Fort Wayne Public Transportation Corporation	Fort Wayne	IN	0.41	Yes	✓
6 Transit Authority of Omaha	Omaha	NE	0.41	Yes	✓
7 Chatham Area Transit Authority	Savannah	GA	0.42	Yes	✓
8 Stark Area Regional Transit Authority	Canton	OH	0.44	Yes	✓
9 The Wave Transit System	Mobile	AL	0.46	No	
10 Capital Area Transit	Raleigh	NC	0.48	No	
11 Capital Area Transit	Harrisburg	PA	0.48	No	
12 Shreveport Area Transit System	Shreveport	LA	0.49	No	
13 Rockford Mass Transit District	Rockford	IL	0.50	No	
14 Erie Metropolitan Transit Authority	Erie	PA	0.52	No	
15 Capital Area Transit System	Baton Rouge	LA	0.52	No	
16 Western Reserve Transit Authority	Youngstown	OH	0.53	Yes	✓
17 Central Oklahoma Transportation & Parking Auth.	Oklahoma City	OK	0.53	No	
18 Des Moines Metropolitan Transit Authority	Des Moines	IA	0.55	No	
19 Mass Transportation Authority	Flint	MI	0.56	Yes	
20 Escambia County Area Transit	Pensacola	FL	0.57	No	

seven of Knoxville's top eight peers have a dedicated local funding source. Connecticut Transit–New Haven Division does not, so it would be replaced by the next-highest peer in the list that does—in this case, Western Reserve Transit Authority. Although it is the 16th-most-similar agency in the list, it still has a good total likeness score of 0.53.

Although not needed in this example, some user judgment might be needed about the extent of dedicated local funding that would qualify. Some local funding sources might only provide 1% or less of an agency's total operating revenue, for example.

Step 4: Compare Performance

The performance measures to be used in the benchmarking effort were specified during Step 2a. Now that a final peer group has been identified, Step 4 focuses on gathering the data associated with those performance measures and analyzing the data.

Step 4a: Gather Performance Data

NTD Data

Performance measures that are directly collected by the NTD or can be derived from NTD measures can be obtained through FTIS. The process for doing so is described in detail

in Appendix A. NTD measures provide both descriptive information such as operating costs and revenue hours and outcome measures such as ridership. Many useful performance measures, however, are ratios of two other measures. For example, cost per trip is a measure of cost-effectiveness, cost per revenue hour is a measure of cost-efficiency, and trips per revenue hour is a measure of productivity. None of these ratios is directly reported by the NTD, but all can be derived from other NTD measures. FTIS provides many common performance ratios, and any ratio derivable from NTD data can be calculated by exporting it from FTIS to a spreadsheet.

One potential concern that users may have with NTD data is the time lag between when data are submitted and when data are officially released, which can be up to 2 years. Rapidly changing external conditions—for example, fuel price increases or a downturn in the economy—may result in the most recent conditions available through the NTD not being reflective of current conditions. There are several ways that these data lag issues can be addressed if they are felt to be a concern:

1. Request NTD viewer passwords directly from the peer agencies. These passwords allow users to view, but not alter, data fields in the various NTD forms. As long as agencies are willing to share their viewer passwords, the agency perform-

ing the peer comparison has access to the most up-to-date information available.

2. Request data from state DOTs. Many states require their transit agencies to report NTD data to them at the same time they report it to the FTA.
3. Review trends in NTD monthly data. The following variables are available on a monthly basis, with only an approximate 6-month time lag: unlinked passenger trips, revenue miles, revenue hours, vehicles operated in maximum service, and number of typical days operated in a month.
4. Review trends in one's own data. Are unusual differences between current data and the most-recent NTD data due to external, national factors that would tend to affect all peers (in which case conclusions about the target agency's performance relative to its peers should still be valid), or are they due to agency- or region-specific changes?

With either of the first two options, it should be kept in mind that data obtained prior to their official release from the NTD may not yet have gone through a full quality-control check. Therefore, performing checks on the data as described in Step 4b (e.g., checking for consistent trends) is particularly recommended in those cases.

Peer Agency Data

Transit agencies requesting data for a peer analysis from other agencies should accompany their data request with the following: (a) an explanation of how they plan to use the data and whether the peer agency's data and results can or will be kept confidential, and (b) a request for documenting how the measures are defined and, if appropriate, how the data for the measures are collected.

Transit agencies may be more willing to share data if they can be assured that the results will be kept confidential. This avoids potential embarrassment to the peer agency if they turn out to be one of the worst-in-group peers in one or more areas, and also saves them the potential trouble of having to explain differences in results to their stakeholders if they do not agree with the study's methodology or result interpretations. In one of the case studies conducted for this project, for example, one agency was not interested in sharing customer-satisfaction data because they disagreed with the way the target agency calculated and used a publicly reported customer-satisfaction index. The potential peer did not want to be publicly compared to the target agency using the target agency's methodology.

Confidentiality can be addressed in a peer-grouping study by identifying which transit agencies were selected as peers but not publicly identifying the specific agency associated with a specific data point in graphs and reports. This information would, of course, be available internally to the agency (to help them identify best-in-group peers), but conclusions about

where the target agency stands relative to its peers can still be made and supported when the peer agency results are shown anonymously. The graphs that accompany the examples of data-quality checks in Step 4b give examples of how information can be presented informatively yet confidentially.

It is important to understand how measures are defined and—in some cases—how the data were collected. For example, on-time performance is a commonly used reliability measure. However, there are wide variations in how transit agencies define “on-time” (e.g., 0 to 5 minutes late vs. 1 minute early to 2 minutes late) that influence the measure's value, since a more generous range of time that is considered “on-time” will result in a higher on-time performance value (1). In addition, the location where on-time performance is measured—departure from the start of the route, a mid-route point, or arrival at the route's terminal—can influence the measure results.

For a peer agency's non-NTD data to be useful for a peer-comparison, the measure values need to be defined similarly, or the measure values need to be re-calculated from raw data using a common definition. The likelihood of having similar definitions is highest when an industry standard or recommended practice exists for the measure. For example, at the time of writing, APTA was developing a draft standard on defining rail transit on-time performance (32), while *TCRP Report 47* (43) provides recommendations on phrasing customer-satisfaction survey questions. The likelihood of being able to calculate measures from raw data is highest when the data are automatically recorded and stored (e.g., data from automatic passenger counter or automated vehicle location equipment) or when a measure is derived from other measures calculated in a standardized way.

Normalizing Cost Data

Transit agencies will often want to normalize cost data to (a) reflect the effects of inflation and (b) reflect differences in labor costs between regions. Adjusting for inflation allows a trend analysis to clearly show whether an agency's costs are changing at a rate faster or slower than inflation. Adjusting for labor costs differences makes it easier to draw conclusions that differences in costs between agencies are due to internal agency efficiency differences rather than external cost differences. Some of the case studies in Chapter 5 provide examples of performing inflation and cost-of-living adjustments; the general process is described below.

The consumer price index (CPI) can be used to adjust costs for inflation. CPIs for the country as a whole, regions of the country, and 26 metropolitan areas are available from the Bureau of Labor Statistics (BLS) website (<http://www.bls.gov/cpi/data.htm>). FTIS also provides the national CPI. To adjust costs for inflation, multiply the cost by (base year CPI)/

(analysis year CPI). For example, the national CPI was 179.9 for 2002 and 201.6 for 2006. To adjust 2002 prices to 2006 levels for use in a trend analysis, 2002 costs would be multiplied by $(201.6/179.9)$ or 1.121.

Average labor wage rates can be used to adjust costs for differences in labor costs between regions since labor costs are typically the largest component of operating costs. These data are available from the Bureau of Labor Statistics (http://www.bls.gov/oes/oes_dl.htm) for all metropolitan areas. The “all occupations” average hourly rate for a metropolitan area is recommended for this adjustment because the intent here is to adjust for the general labor environment in each region, over which an agency has no control, rather than for a transit agency’s actual labor rates, over which an agency has some control.

Identifying differences in a transit agency’s labor costs, after adjusting for regional variations, can be an important outcome of a peer-comparison evaluation. Although it is possible to drill down into the BLS wage database to get more-specific data—for example, average wages for “bus drivers, transit and intercity”—the ability to compare agency-controllable costs would be lost because the more-detailed category would be dominated by the transit agency’s own workforce. The “all occupations” rates, on the other hand, allow an agency to (a) investigate whether it is spending more or less for its labor relative to its region’s average wages, and (b) adjust its costs to reflect differences in a region’s overall cost of living (which impacts overall average wages within the region).

To adjust peer agency costs for differences in labor costs, multiply the cost by $(\text{target agency metropolitan area labor cost})/(\text{peer agency metropolitan area labor cost})$. For example, Denver’s average hourly wage rate in 2008 was \$22.67,

while Portland’s was \$21.66. If Denver RTD is performing the analysis and wants to adjust TriMet costs to reflect the higher wages in the Denver region, it would multiply TriMet costs by $(22.67/21.66)$, or 1.047.

Step 4b: Analyze Performance

Data Checking

Before diving into a full analysis of the data, it is useful to create graphs for each measure to check for potential data problems, such as unusually high or low values for a given agency’s performance measure for a given year, and for values that bounce up and down with no apparent trend. The following figures give examples of these kinds of checks.

Figure 3 illustrates outlier data points. Peer 4 has an obvious outlier for the year 2003. As it is much higher than the agency’s other values (including prior years, if one went back into the database) and is much higher than any other agency’s values, that data point could be discarded. The rest of Peer 4’s data show consistent trends; however, since this agency had an outlier and would be the best-in-group performer for this measure, it would be worth a phone call to the agency to confirm the validity of the other years’ values. Peer 5 also has an outlier for the year 2004. The value is not out of line with other agencies’ values, but is inconsistent with Peer 5’s overall trend. In this case, a phone call would find out whether the agency tried (and then later abandoned) something in 2004 that would have improved performance, or whether the data point is simply incorrect.

In Figure 4, Peer 2’s values for the percent of breaks and allowances as part of total operating time are nearly zero and

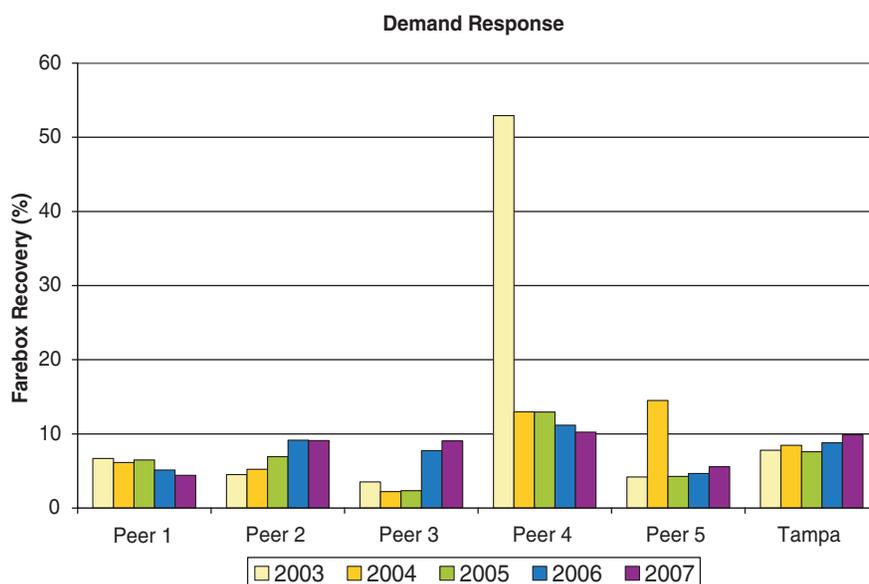


Figure 3. Outlying data points example.

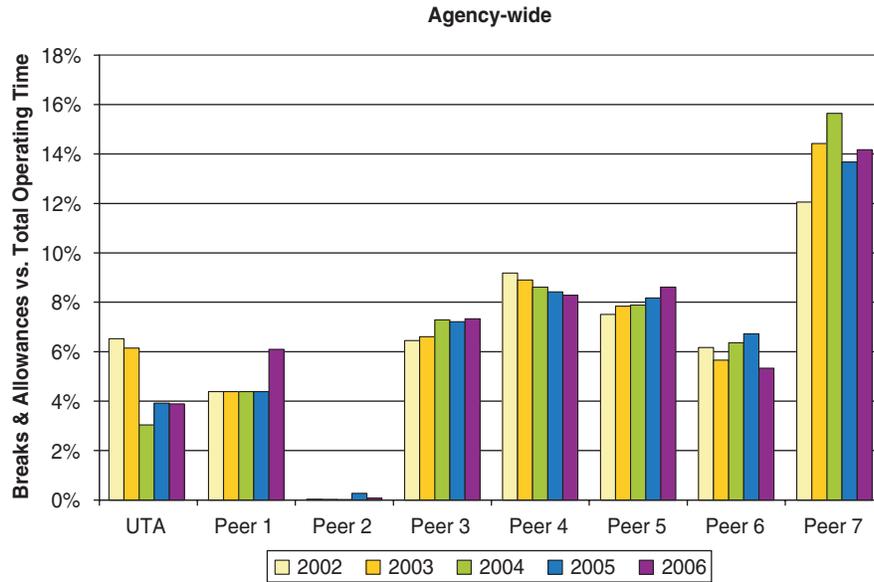


Figure 4. Outlying peer agency example.

far below those of the other agencies in the peer group. It might be easy to conclude that this is an error, as vehicle operators must take breaks, but this would be incorrect in this case. According to the NTD data definitions, breaks that are taken as part of operator layovers are counted as platform time, whereas paid breaks and meal allowances are considered straight time and are accounted for differently. Therefore, Peer 2’s values could actually be correct (and are, as confirmed by a phone call). Peer 7 is substantially higher than the others and may be treating all layover time as break time. The conclusion to be drawn from this data check is that the measure being used will not provide the desired information (a comparison of schedule efficiency). Direct agency contacts would need to be made instead.

Figure 5 shows a graph of spare ratio (the number of spare transit vehicles as a percentage of transit vehicles used in maximum service). As Figure 5(a) shows, spare ratio values can change significantly from one year to the next as new bus fleets are brought into service and old bus fleets are retired. It can be difficult to discern trends in the data. Figure 5(b) shows the same variable, but calculated as a three-year rolling average (i.e., year 2007 values represent an average of the actual 2005–2007 values). It is easier to discern from this version of the graph that Denver’s average spare ratio (along with Peer 1, Peer 3, and Peer 4) has held relatively constant over the longer term, while Peer 2’s average spare ratio has decreased over time and the other two peers’ spare ratios have increased over time. In this case, there is no apparent problem with the

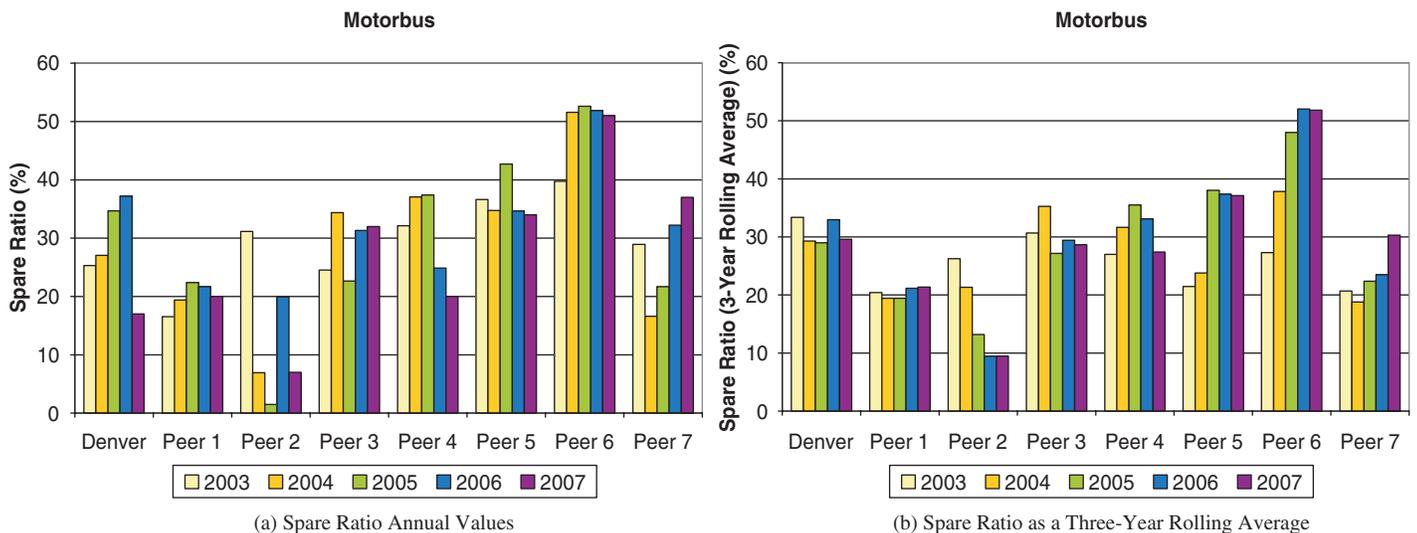


Figure 5. Data volatility example.

data, but the data check has been used to investigate a potentially better way to analyze and present the data.

Data Interpretation

For each measure selected for the evaluation, the target agency's performance is compared to the performance of the peer agencies. Ideally, this evaluation would look both at the target agency's current position relative to its peers (e.g., best-in-class, superior, average, inferior), and the agency's trend. Even if a transit agency's performance is better than most of its peers, a trend of declining performance might still be a cause for concern, particularly if the peer trend was one of improving performance. Trend analysis also helps identify whether particularly good (or bad) performance was sustained or was a one-time event, and can also be used for forecasting (e.g., agency performance is below the agency's target at present, but if current trends continue is forecast to reach the agency's target in 2 years).

Graphing the performance-measure values is a good first step in analyzing and interpreting the data. Any spreadsheet program can be used, and FTIS also provides basic graphing functions. It may be helpful to start by looking at patterns in the data. In Figure 6, for example, it can be seen that the general trend in the data for all peers, except Peer 7, has been an increase in operating costs per boarding over the 5-year period, with Peers 3–6 experiencing steady and significant increases each year. Denver's cost per boarding, in comparison, has consistently been the second-best in its peer group during this time, and Denver's cost per boarding has increased by about half as much as the top-performing peer. Most of

Denver's peers also experienced a sharp increase in costs during at least one of the years included in the analysis, while Denver's year-to-year change has been relatively small and, therefore, more predictable. This analysis would indicate that Denver has done a good job of controlling cost per boarding, relative to its peers.

Sometimes a measure included in the analysis may turn out to be misleading. For example, farebox recovery (the portion of operating costs covered by fare revenue) is a commonly used performance measure in the transit industry and is readily available through FTIS. When this measure is applied to Knoxville, however, Knoxville's fare recovery ratio is by far the lowest of its peers, as indicated in Figure 7(a). Given that Knoxville's performance is among the best in its peer group in a number of other measures, an analyst should ask why this result occurred. Clues to the answer can be obtained through a closer inspection of the NTD data.

NTD form F-10, available within FTIS, provides information about each agency's revenue, broken down by a number of sources. For 2007, this form shows that Knoxville earned nearly as much revenue from "other transportation revenue" as it did from bus fares. A visit to the agency's website, where budget information is available, confirms that the agency receives revenue from the University of Tennessee for operating free shuttle service to the campus and sports venues. Therefore, farebox recovery is not telling the entire story about how much of Knoxville's service is self-supporting. As an alternative, all directly generated non-tax revenue used for operations can be compared to operating costs (a measure known as the *operating ratio*). This requires more work, as non-fare revenue should be allocated among the various

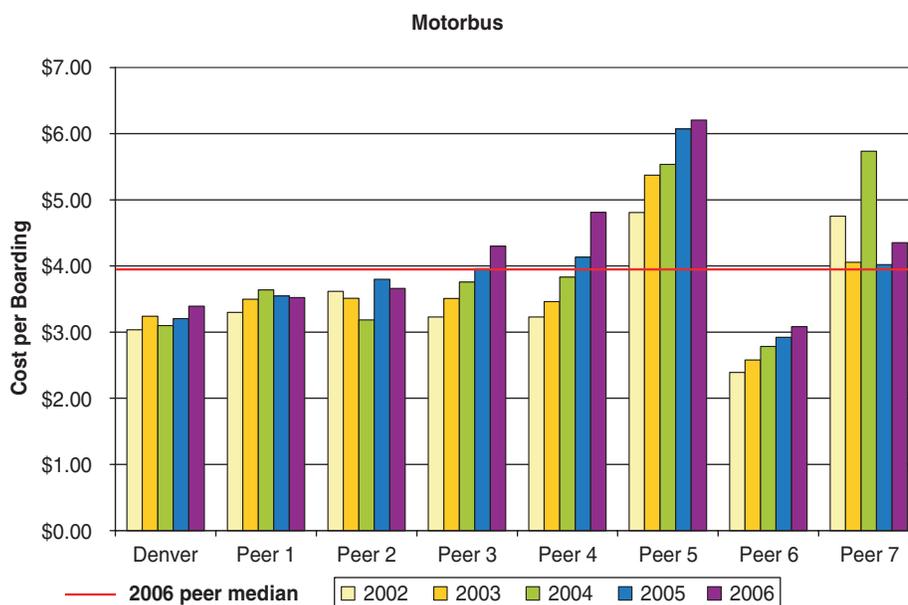


Figure 6. Pattern investigation example.

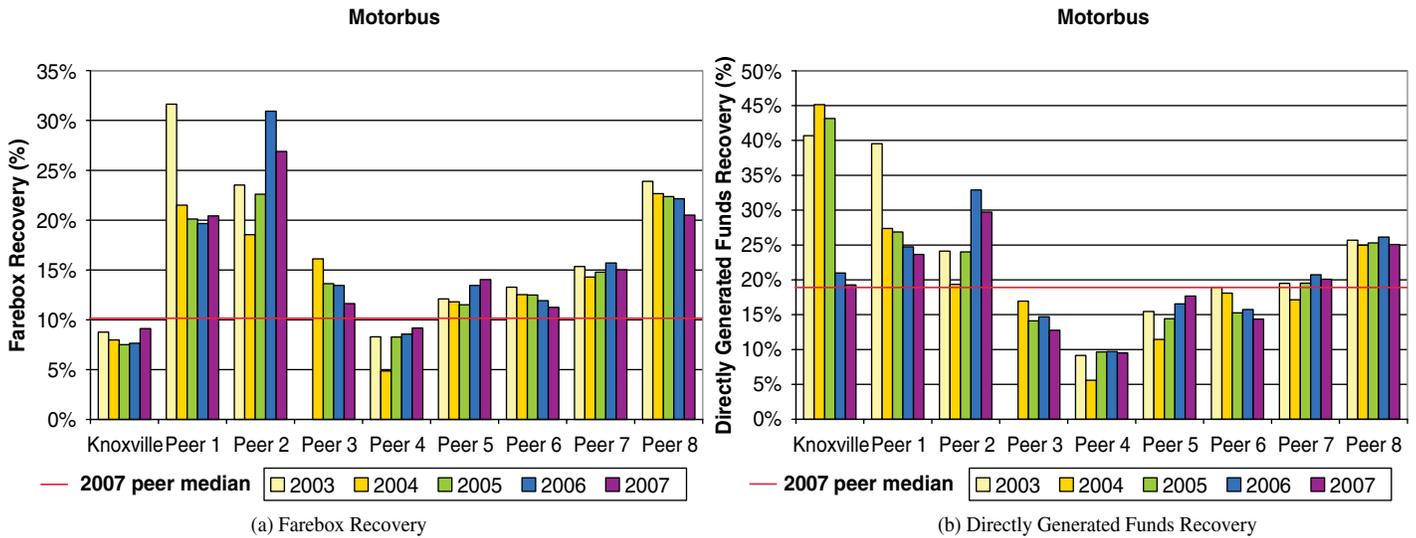


Figure 7. Data interpretation example #1.

modes operated (it is only reported on a system-wide basis), but all of the required data to make this allocation is available through FTIS, and the necessary calculations can be readily performed within a spreadsheet.

Figure 7(b) shows the results of these calculations, where it can be seen that Knoxville used to be at the top of its peer group in terms of operating ratio but is now in the middle of the group, as the university payments apparently dropped substantially in 2006. A comparison of the two graphs also shows that Knoxville is the only agency among its peers (all of whom have dedicated local funding sources) to get much directly generated revenue at present from anything except fares.

A final example of data interpretation is shown in Figure 8, comparing agencies' annual casualty and liability costs, normalized by annual vehicle miles operated. This graph tells several stories. First, it can be clearly seen that a single serious accident can have a significant impact on a transit agency's casualty and liability costs in a given year because many agencies are self-insured. Second, it shows how often the peer group experiences serious accidents. Third, it indicates trends in casualty and liability costs over the 5-year period. Eugene, Peer 3, and Peer 6 were the best performers in this group over the study period, while Peer 7's costs were consistently higher than the group as a whole.

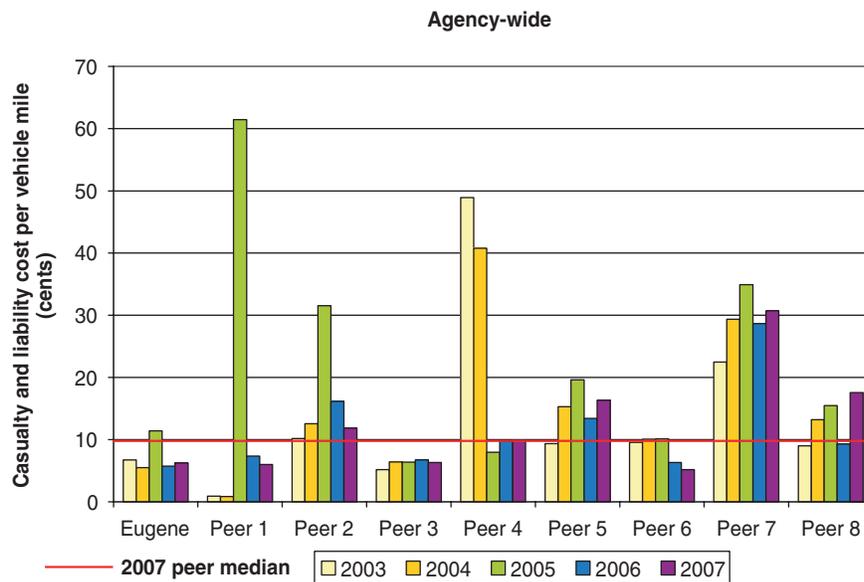


Figure 8. Data interpretation example #2.

Results Presentation

The results of the data analysis will need to be documented for presentation to the stakeholders in the process. The exact form will depend on the audience, but can include any or all of the following:

- An executive summary highlighting the key findings,
- A summary table presenting a side-by-side comparison of the numeric results for all the measures for all the peers,
- Graphs, potentially including trend indicators (such as arrows) or lines indicating the group average,
- A combination of graph and table, with the table providing the numeric results to accompany the graph,
- A combination of graph and text, with the text interpreting the data shown in the graph,
- Multiple graphs, with one or more secondary graphs showing descriptive data that support the interpretation of the main graph, and
- Graphics that support the interpretation of text, tables, and/or graphs.

Peer group averages can be calculated as either means or medians. Means are more susceptible to being influenced by a transit agency with particularly good or particularly poor performance, while medians provide a good indicator of where the middle of the group lies.

The case studies in Chapter 5 and the material in Appendix C give a variety of examples of how performance information can be presented. *TCRP Report 88 (1)* contains a section providing guidance on presenting performance results, and publications available on the European benchmarking network websites (14–16, 47) can also be used as examples.

Step 5: Contact Best-Practices Peers

At this point in the process, a transit agency knows where its performance stands with respect to its peers, but not the reasons why. Contacting top-performing peers addresses the “why” aspect and can lead to identifying other transit agencies’ practices that can be adopted to improve one’s own performance.

In most cases a transit agency will find one or more areas where it is not the best performer among its peers. An agency with superior performance relative to most of its peers, and possessing a culture of continuous improvement, would continue the process to identify what it can learn from its top-performing peers to improve its already good performance. When an agency identifies areas of weakness relative to its peers, it is recommended that it continue the benchmarking process to see what it can learn from its best-performing peers.

For Level 1 and 2 benchmarking efforts, it is possible to skip this step and proceed directly to Step 6, developing an implementation strategy. However, doing so carries a higher

risk of failure since agencies may unwittingly choose a strategy already tried unsuccessfully elsewhere or may choose a strategy that results in a smaller performance improvement than might have been achieved with alternative strategies. Step 5 is the defining characteristic of a Level 3 benchmarking effort, while the working groups used as part of a Level 4 benchmarking effort would automatically build this step into the process. Step 5 would also normally be built into the process when a benchmarking effort is being conducted with an eye toward changing how the agency conducts business.

The kind of information that is desired at this step is beyond what can be found from databases and online sources. Instead, executive interviews are conducted to determine how the best-practices agencies have achieved their performance, to identify lessons learned and factors that could inhibit implementation or improvement, and to develop suggestions for the target agency. There are several formats for conducting these interviews, which can be tailored for the specific needs of the performance review.

- Blue ribbon panels of expert staff and/or top management from peer agencies are appropriate to bring in for one-time-only or limited-term reviews, such as a special management focus on security or a large capital project review.
- Site visits can be useful for hands-on understanding of how peer agencies operate. The staff involved could range from line staff to top management, depending on the specific issues being addressed.
- Working groups can be established for topic-specific discussions on performance, such as a working group on preventative maintenance practices. Line staff and mid-level management in the topic area would be most likely to be involved.

The private sector has also used staff exchanges as a way of obtaining a deeper understanding of another organization’s business practices by having one or two select staff become immersed in the peer organization’s activities for an extended period of time.

Involving staff from multiple levels and functions within the transit agency helps increase the chances of identifying good practices or ideas, helps increase the potential for staff buy-in into any recommendations for change that are made as a result of the contacts, helps percolate the concept of continuous improvement throughout the transit agency, and helps provide opportunities for staff leadership and professional growth.

Step 6: Develop an Implementation Strategy

In Step 6, the transit agency develops a strategy for making changes to the current agency environment, with the goal of improving its performance. Ideally, the strategy development

process will be informed by a study of best practices, which would have been performed in Step 5. The strategy should include performance goals (i.e., quantify the desired outcome) and provide a timeline for implementation, and should identify any required funding. The strategy also needs to identify the internal (e.g., business practices or agency policies) or external (e.g., regional policies or new revenue sources) changes that would be needed to successfully implement the strategy. Top-level management and transit agency board support is vital to getting the process underway. However, support for the strategy will need to be developed at all levels of the organization: lower-level managers and staff also need to buy into the need for change and understand the potential benefits of change. Therefore, the implementation strategy should also include details on how information will be disseminated to agency staff and external stakeholders and should include plans for developing internal and external stakeholder support for implementing the strategy.

Step 7: Implement the Strategy

TCRP Report 88 (1) identified that once a performance evaluation is complete and a strategy is identified, the process can often halt due to lack of funding or stakeholder support. If actual changes designed to improve performance are not implemented at the end of the process, the peer review risks becoming a paper exercise, and the lack of action can reduce stakeholder confidence in the effectiveness of future performance evaluations. If problems arise during implementation,

the agency should be prepared to address them quickly so that the strategy can stay on course.

Step 8: Monitor Performance

As noted in Step 6, the implementation strategy should include a timeline for results. A timeline for monitoring should also be established to make sure that progress is being made toward the established goals. Depending on the goal and the overall strategy timeline, the reporting frequency could range from monthly to annually. If the monitoring effort indicates a lack of progress, the implementation strategy should be revisited and revised if necessary. Hopefully, however, the monitoring will show that performance is improving.

In the longer term, the transit agency should continue its peer-comparison efforts on a regular basis. The process should be simpler the second time around because many or all of the agency's peers will still be appropriate for a new effort, points of contact will have been established with the peers, and the agency's staff will now be familiar with the process and will have seen the improvements that resulted from the first effort. The agency's peers hopefully will also have been working to improve their own performance, so there may be something new to learn from them—either by investigating a new performance topic or by revisiting an old one after a few years. A successful initial peer-comparison effort may also serve as a catalyst for forming more-formal performance-comparison arrangements among transit agencies, perhaps leading to the development of a benchmarking network.

CHAPTER 5

Case Studies

Overview

This chapter presents six real-world applications of this report's peer-comparison and performance-measurement methodology. These case studies have been selected as examples of the variety of applications, transit agency sizes, and modes that the methodology can be applied to, but are in no way comprehensive. Agencies considering performing peer comparisons similar to the ones shown here should not feel constrained by the case studies' choices of performance measures and screening criteria. Every agency's goals, objectives, and reasons for performing a peer comparison will be different, resulting in different choices. Each case study includes a description of the context of the study, which helps in understanding the choices that were made.

These case studies are based on studies that were performed during the course of the research to test different drafts of the peer-grouping methodology. As a result, applying the final peer-grouping methodology described in this report and implemented in FTIS may not result in exactly the same peer group members or likeness scores presented in this chapter. The focus here is on the *process* of conducting a peer comparison.

The following case studies are included in this chapter:

- Altoona, PA: An application of state performance indicators to a small urban bus operator and an example of exploring the causes of performance results.
- Knoxville, TN: An example of applying secondary screening criteria to help answer a "what-if" question at a medium-sized bus operator.
- Salt Lake City, UT: A comparison of bus and light rail operator schedule efficiency at a large multimodal transit agency.
- Denver, CO: A financial performance comparison for a large multimodal transit agency, illustrating the normalization of cost data.

- San Jose, CA: A maintenance performance comparison for a light rail operator.
- South Florida: A comparison of transit investments and outcomes for a commuter rail operator receiving significant funding from a state department of transportation.

Altoona, Pennsylvania

Context

Altoona Metro Transit serves the Altoona, Pennsylvania, urban area, which had a population of just over 80,000 in 2007. The agency operates fixed-route bus service and contracts its demand-response service. In 2007 it operated about 581,000 vehicle miles and had an operating budget of \$3.7 million.

This case study was developed on behalf of the Pennsylvania DOT. PennDOT is required by its state legislature (Act 44 of 2007) to report four performance indicators annually for all urban and rural transit operators in Pennsylvania: *cost per revenue hour*, *fare revenue per revenue hour*, *boardings per revenue hour*, and *cost per boarding*. In addition, PennDOT includes performance factors in its operating grant funding formula. Similar case studies were developed for the other nine small-urban transit operators in Pennsylvania that report to the NTD, giving PennDOT a picture of how Pennsylvania small-urban operators compare to their peers in the areas of interest to the state legislature.

Performance Question

How do Pennsylvania's small-urban transit systems compare to their peers in the areas focused on by state legislation?

Performance Measures

In this case, the set of performance measures had already been decided by the state legislature, and are listed above. The

Table 16. Altoona peer group candidates.

Agency	City	State	Likeness Score
Sheboygan Transit System	Sheboygan	WI	0.26
Sioux City Transit System	Sioux City	IA	0.27
Ohio Valley Regional Transportation Authority	Wheeling	WV	0.31
Wausau Area Transit System	Wausau	WI	0.35
Battle Creek Transit	Battle Creek	MI	0.38
Belle Urban System - Racine	Racine	WI	0.39
City of Anderson Transportation System	Anderson	IN	0.42
Springfield City Area Transit	Springfield	OH	0.43

performance question is basic and the agencies involved in the full case study were spread across the state, so no secondary screening measures were necessary.

Peer Grouping

FTIS was used to develop a set of peers for Altoona, using this report's methodology. Table 16 shows which peers were identified through this process.

All of the likeness scores are very good (0.50 or less), so no further investigation of the peers was performed. As identified above, no secondary screening was needed.

Performance Results

FTIS was used to retrieve the desired performance data from the NTD. All of the desired performance measures are ratios of NTD measures, and three of the four are provided directly by FTIS as part of its set of Florida Standard Variables (these are labeled as *operating expense per revenue hour*, *operating expense per passenger trip*, and *passenger trips per revenue hour* in FTIS). The fourth desired measure, *fare revenue per revenue hour*, can be calculated from three of the Florida Standard Variables in a spreadsheet as follows: multiply *average fare* by *passenger trips* to get *total farebox revenue*, and then divide the result by *revenue hours*. The advantage of using the Florida Standard Variables is that FTIS provides agency-wide totals (all modes combined) and service totals (directly operated and purchased transportation combined) for the Florida Standard Variables. If the raw NTD measures were retrieved from FTIS, the analyst would need to manually sum the individual mode and service results to get the same agency-wide total.

A spreadsheet's pivot table function was used to organize the data for each measure by year and agency. A 2007 peer-group median value was also determined for each measure within the spreadsheet. Finally, the spreadsheet's charting functions were used to develop comparative graphs for each measure of interest, as shown in Figure 9.

Interpreting Results

Altoona has the highest *operating expense per revenue hour* in its peer group, more than \$25 per hour above the peer group median in 2007 [Figure 9(a)]. Altoona's trend of a sharp increase in this measure over the 5-year period is consistent with its peers. At the same time, Altoona generates the second-highest *fare revenue per revenue hour* in its peer group [Figure 9(b)]. Altoona generated nearly \$5 per revenue hour more than the peer group median. Altoona's upward trend in this measure is consistent with its peers. A peer of note in this category is Sioux City, which more than doubled its *fare revenue per revenue hour* between 2003 and 2007 while maintaining ridership levels over the longer term.

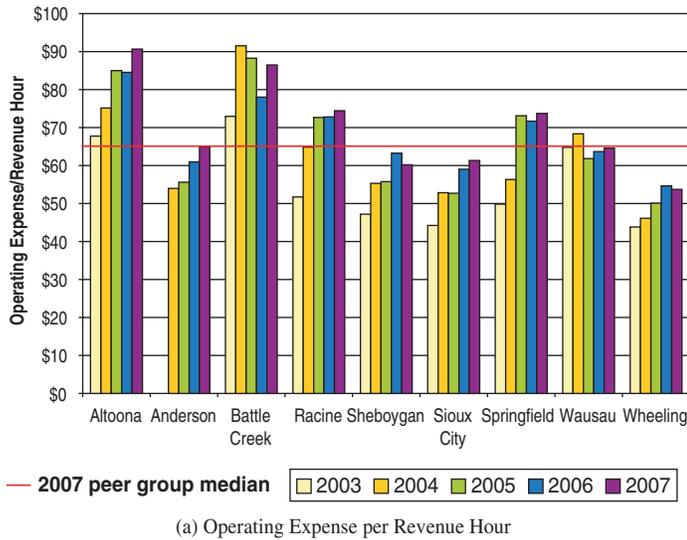
Looking at the other two measures, Altoona is slightly above the group median for *boardings per revenue hour* [Figure 9(c)] and at the group median for *cost per boarding* [Figure 9(d)]. Altoona's small upward trend for *boardings per revenue hour* is better than most of its peers, which generally held steady or dropped from 2003 to 2007.

Asking Questions

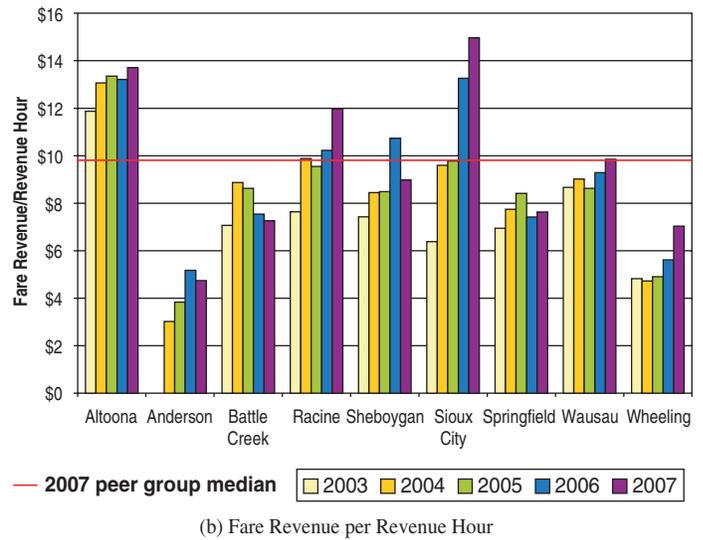
Altoona's relatively high hourly operating cost stands out as an area to investigate more closely to see if any clues can be found that would indicate the source(s) of the high costs, which could then be the focus of efforts to lower those costs. FTIS' data-exploration functions, such as its cross-table feature, can be used to quickly go through a list of possible causes.

As a first step, demand-response costs can be compared to motorbus costs to try to narrow the cause down by mode. Altoona has the second-lowest demand-response *cost per boarding* and is at the group median for *cost per revenue hour*, so demand-response can be eliminated as a significant contributor.

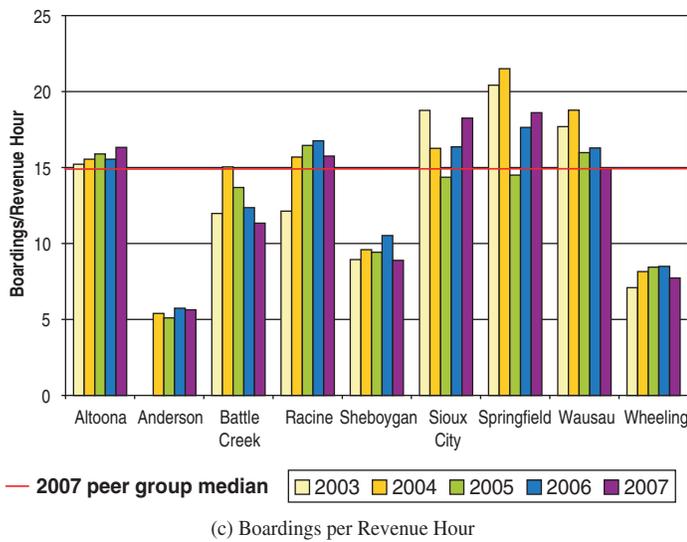
Next, Florida Standard Variables relating to costs can be investigated for the motorbus mode specifically. Altoona's *average bus fleet age* (by far the highest at 16 years), *vehicle miles per gallon* (lowest), *vehicle system failures* (second highest), and *maintenance cost per revenue mile* (second highest) all suggest



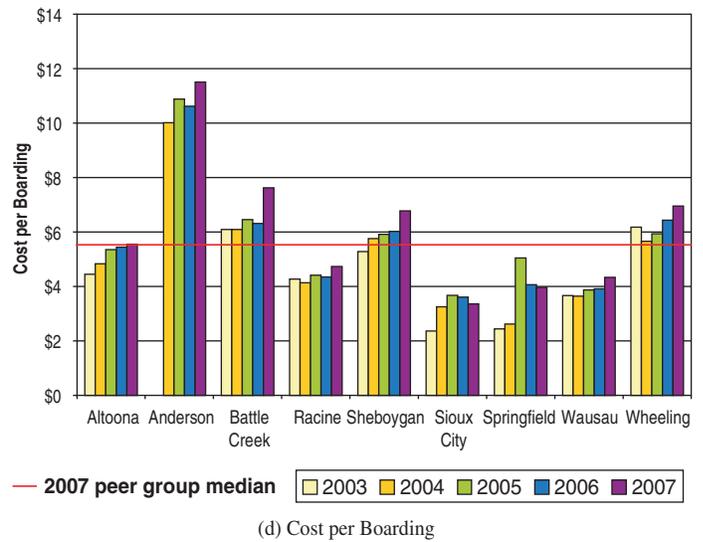
(a) Operating Expense per Revenue Hour



(b) Fare Revenue per Revenue Hour



(c) Boardings per Revenue Hour



(d) Cost per Boarding

Figure 9. Performance results for Altoona.

that the cost of maintaining an old fleet is contributing to the high operations costs. From a state DOT perspective, channeling grant funding to Altoona for vehicle replacement could pay off with ongoing maintenance cost savings.

Data available on NTD form F-30, relating to agency expenses, can be used to dig deeper into possible causes for the higher costs, particularly when the data are normalized by revenue hours. Here, fleet maintenance costs also stand out in terms of *maintenance wage cost per revenue hour* (highest), *fuel costs per revenue hour* (highest), and *other materials/supplies costs per revenue hour* (second-highest). At the same time, other cost factors are uncovered: *fringe benefit costs per revenue hour* are \$3.35 higher than the peer group median, *non-vehicle operations staff wage costs per revenue hour* are \$4.30 higher, and *administrative staff wage costs per revenue hour* are \$0.80 higher. These data do not indicate by themselves that these costs are “too

high,” as no context is available from the data to make that determination, but merely that the costs are higher and that it could be worthwhile for Altoona to investigate them further.

Knoxville, Tennessee

Context

The Knoxville urban area had approximately 452,000 residents in 2007. The urban area is served by Knoxville Area Transit, which operates both motorbus and demand-response service, including service contracted by the University of Tennessee (which is operated fare-free). The agency operated 3.2 million vehicle miles in 2007 and had a budget of \$14.3 million. The largest source of operations funding for the agency is the city’s general fund.

Performance Question

How does Knoxville’s performance compare to similarly sized transit agencies that have a dedicated local funding source, both in terms of the amount of service that can be delivered and the cost-effectiveness of that service?

Performance Measures

There are three types of measures that need to be considered:

- Measures that address the service-delivery question,
- Measures that address the cost-effectiveness question, and
- Measures that screen for the presence of dedicated local funding.

To address the service-delivery question, the tables in Chapter 3 relating to transit investment and delivered service quality are consulted, and the following measures are selected: *operating expense per capita*, *operating subsidy per capita*, and *revenue hours per capita*.

To address the cost-effectiveness question, the tables in Chapter 3 relating to cost-effectiveness, cost-efficiency, and productivity are consulted, and the following measures are selected: *cost per revenue hour*, *boardings per revenue hour*, *cost per boarding*, and *boardings per capita*. The *farebox recovery ratio* would also be a common measure to include in this kind of analysis, but because Knoxville’s university-subsidized service is fare-free, this measure would not be particularly informative in this case. Instead, a measure

that looks at the *percentage of operating costs that are subsidized* is used, as this accounts for all of an agency’s directly generated non-tax revenue.

Finally, information from NTD form F-10 will be used to identify potential peers that do not have a dedicated local funding source.

Peer Grouping

FTIS was used to develop an initial set of potential peers for Knoxville (Table 17) using this report’s methodology. A secondary screening process was then used to eliminate peers without a dedicated local funding source (shown in ~~strikeout~~ type), as was illustrated in Chapter 4 (methodology Step 3c and Table 15).

Performance Results

Service Delivery

FTIS was used to retrieve the desired performance data from the NTD (Figure 10). None of the selected service-delivery performance measures are provided directly from FTIS, but they can be derived from other variables available through FTIS. Following the guidance in Chapters 3 and 4, urban area population from the American Community Survey (ACS) was used for “per capita” measures, as all of the agencies in the peer group are the sole agencies in their respective urban areas. ACS population estimates include university students based on a “2-month rule”—if they are staying in their university residence for at least 2 months at the time of survey contact, they

Table 17. Knoxville peer group candidates.

Agency	City	State	Likeness Score
Winston-Salem Transit Authority	Winston-Salem	NC	0.25
South Bend Public Transportation Corporation	South Bend	IN	0.36
Birmingham-Jefferson County Transit Authority	Birmingham	AL	0.36
Connecticut Transit—New Haven Division	New Haven	CT	0.39
Fort Wayne Public Transportation Corporation	Fort Wayne	IN	0.41
Transit Authority of Omaha	Omaha	NE	0.41
Chatham Area Transit Authority	Savannah	GA	0.42
Stark Area Regional Transit Authority	Canton	OH	0.44
The Wave Transit System	Mobile	AL	0.46
Capital Area Transit	Raleigh	NC	0.48
Capital Area Transit—	Harrisburg	PA	0.48
Shreveport Area Transit System	Shreveport	LA	0.49
Rockford Mass Transit District	Rockford	IL	0.50
Erie Metropolitan Transit Authority	Erie	PA	0.52
Capital Area Transit System	Baton Rouge	LA	0.52
Western Reserve Transit Authority	Youngstown	OH	0.53

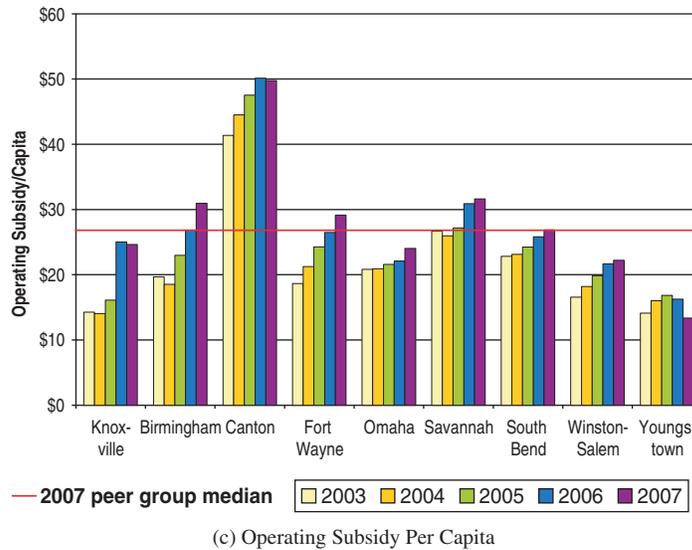
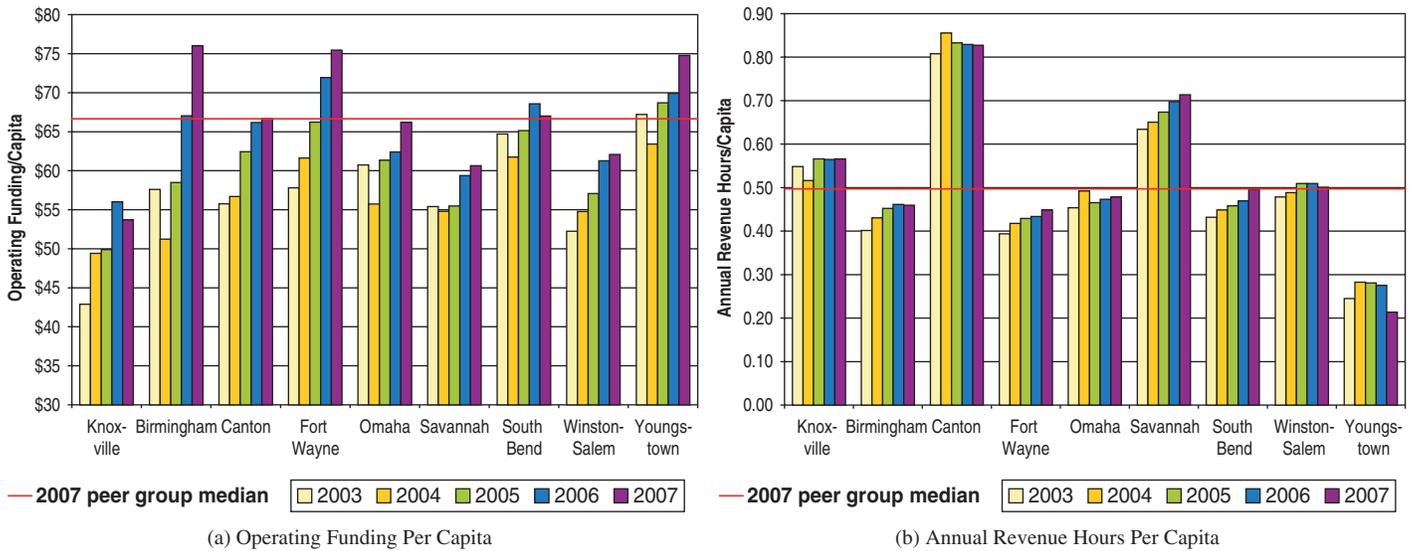


Figure 10. Service delivery performance results for Knoxville.

are counted as living in the community where the university is located. This is different from the decennial census procedure, where persons are counted based on their “usual residence,” which may not be their current residence (48). The ACS’s population-counting methodology (and, therefore, the per-capita measures based on those population estimates) reasonably accounts for Knoxville’s student population, as well as the student populations of other communities in the peer group, such as South Bend.

Operating funding per capita is a ratio of total operating expenses (a Florida Standard Variable) and urban area population (a TCRP Project G-11 variable). Similarly, revenue hours per capita divides the Florida Standard Variable revenue hours by urban area population. Operating subsidy per capita sub-

tracts total farebox revenue and total directly generated park-and-ride/other/auxiliary revenue (both from NTD form F-10) from total operating expenses and divides the result by urban area population.

Service Cost and Productivity

Three of the five measures, cost per revenue hour, boardings per revenue hour, and cost per boarding, are available directly from FTIS as Florida Standard Variables. Boardings per capita is calculated from passenger trips (a Florida Standard Variable) and urban area population (a TCRP Project G-11 variable). Percent of operating costs subsidized is calculated by subtracting total farebox revenue and total directly generated

park-and-ride/other/auxiliary revenue (both from NTD form F-10) from *total operating expenses* and dividing the result by *total operating expenses*. Figure 11 shows the service cost and productivity results for Knoxville.

Interpreting Results

Service Delivery

On a per-capita basis, the Knoxville region's investment in transit is the lowest in the peer group. Although it grew from 2003 to 2007, so did the peer regions' investments, as shown in Figure 10(a). Despite the relatively low investment, the amount of service Knoxville has been able to put on the street (*revenue hours per capita*) is slightly above the peer group median. Knoxville's *revenue hours per capita* held steady during 2003–2007, while the peer group trend was a slight increase, as seen in Figure 10(b). In terms of *operating subsidy per capita*, Knoxville is slightly below the group median. Knoxville's subsidy increased sharply in 2006 due to a reduction in the revenue received from its contract with the university to provide shuttle service. The peer group trend for subsidy has been higher to sharply higher [Figure 10(c)].

Service Cost and Productivity

Knoxville has the lowest *cost per revenue hour* of any agency in the peer group. Knoxville's costs are increasing, as are those of the peers [Figure 11(a)]. Knoxville's *boardings per revenue hour* are slightly below the peer median [Figure 11(b)]; its long-term trend is generally upward, though, while there is no clear trend among the peers (some are decreasing, some are steady, and some are increasing). There is a wide spread of *cost per boarding* values within the peer group and no clear peer trend [Figure 11(c)]; Knoxville is slightly below the group median value and held costs steady during 2003–2007. Knoxville's *boardings per capita* and *percent of service subsidized* [Figures 11(d) and (e)] are both at the group median and both values have increased over time. Savannah, South Bend, and Winston-Salem are the other top performers in the peer group that Knoxville could consider looking to for ideas to further improve its service.

Answering Questions

In terms of building support for a dedicated local funding source, the *operating funding per capita* measure indicates that all of Knoxville's peer cities have invested more in transit operations than Knoxville, while the *revenue hours per capita* measure indicates that Knoxville is doing a good

job converting revenue into service on the street. Knoxville's *subsidy per capita* is currently a little below the peer group average; adding a new tax-supported revenue source would tend to increase this value, but fare revenue derived from the new service would tend to decrease it. Determining the overall impact of new funding and new service on this measure would require more detailed analysis. Both the *cost per revenue hour* and *cost per boarding values* support an argument that Knoxville has done a good job relative to its peers of controlling costs. *Boardings per capita* is at the group median and would be expected to increase with new service. Neither of the other cost-related measures would argue against seeking additional funds, compared to looking first internally for opportunities for cost savings. However, the cost data also highlight the importance of Knoxville Area Transit's relationship with the University of Tennessee, and the agency could also look to see what it could do to strengthen that partnership.

Salt Lake City, Utah

Context

Utah Transit Authority serves the Salt Lake City and Provo urban areas. It operates light rail, motorbus, and vanpool service, and started commuter rail service in 2008. Demand-response service is partially directly operated and partially contracted. UTA operated 30.1 million vehicle miles in 2007 and had an operating budget of \$136.8 million. The Salt Lake City urban area had a 2007 population of 944,000.

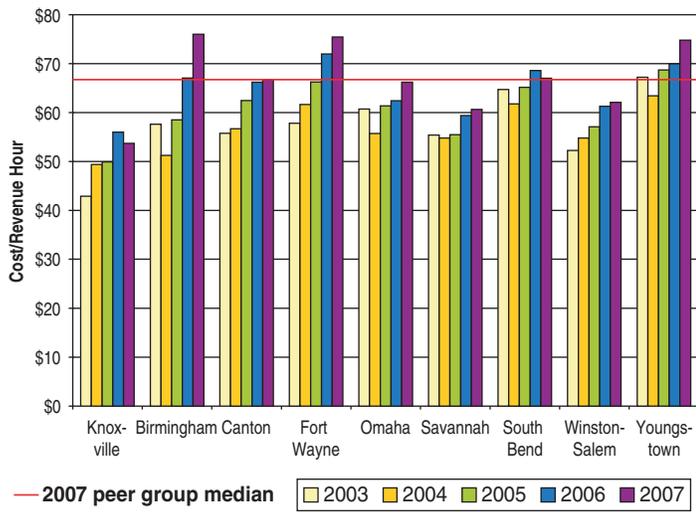
Performance Question

How efficient are UTA's motorbus and light rail operator work schedules?

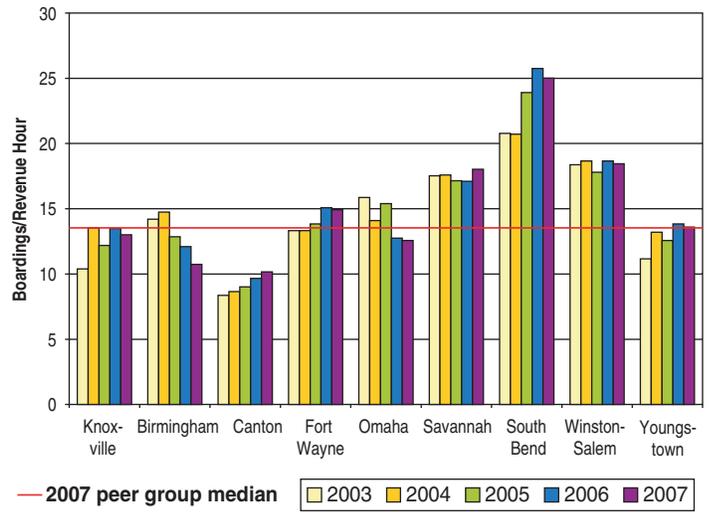
Performance Measures

The following measures are derived from the tables in Chapter 3 relating to labor administration and resource utilization, using data specific to operating employees:

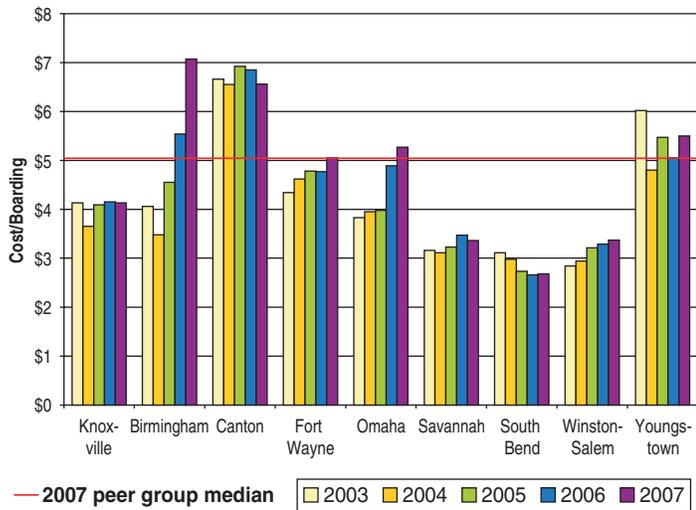
- *Operator wages as a percent of total operating expenses,*
- *Operator wages and fringe benefits as a percent of total operating expenses,*
- *Pay-to-platform hours,*
- *Premium hours as a percent of total operating hours,*
- *Vehicle revenue hours per operating employee full-time equivalent, and*
- *Boardings per operating employee full-time equivalent.*



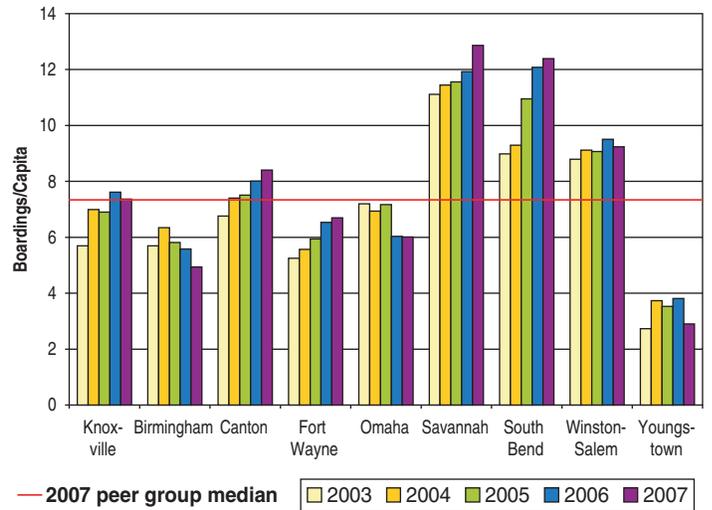
(a) Cost per Revenue Hour



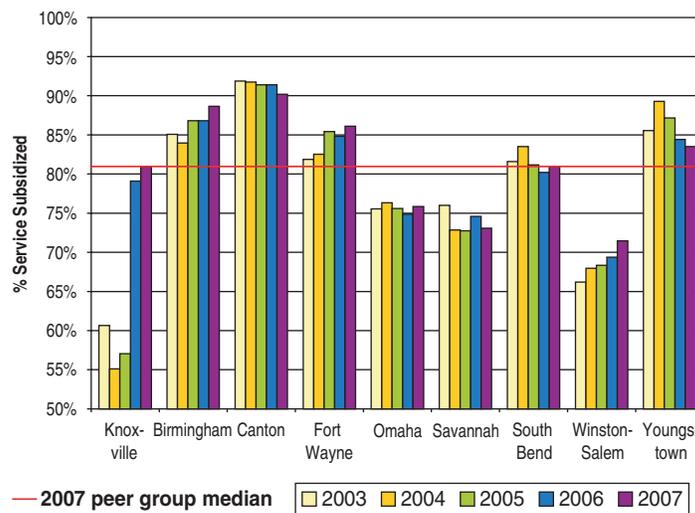
(b) Boardings per Revenue Hour



(c) Cost per Boarding



(d) Boardings Per Capita



(e) Percent of Service Subsidized

Figure 11. Service cost and productivity performance results for Knoxville.

Table 18. UTA light rail peer group candidates.

Agency Name	Location	State	Total Likeness Score
Denver Regional Transportation District	Denver	CO	0.52
Santa Clara Valley Transportation Authority	San Jose	CA	0.59
Sacramento Regional Transit District	Sacramento	CA	0.65
Maryland Transit Administration	Baltimore	MD	0.79
Tri-County Metropolitan Transportation District of Oregon	Portland	OR	0.88
Bi-State Development Agency	St. Louis	MO	0.94
Metro Transit	Minneapolis	MN	0.95

Three cost-effectiveness and cost-efficiency measures are also selected to provide context about overall mode efficiency: *revenue hours per vehicle hours*, *cost per boarding*, and *cost per revenue hour*.

UTA desires that the peer agencies operate bus and light rail service, provide region-wide service, and be located in regions with growing populations that have similar land-use characteristics.

Peer Grouping

FTIS was used to develop two sets of peers for UTA, one for the light rail mode and one for the motorbus mode.

Light Rail

Table 18 shows the initial set of potential light rail peers that was identified, based on selecting all peers with likeness scores of 1.00 or less. Based on UTA's screening criteria, Baltimore is eliminated on the basis of also operating heavy rail, while Minneapolis is eliminated because (a) its light rail line opened during the 2003–2007 period planned to be studied and (b) other agencies provide service to its suburbs (determined from the “service type” measure in FTIS' peer-grouping results). The five peer agencies in the group are less than the recommended ideal number of eight to ten, but

exceeds the four-agency minimum. Larger, multimodal agencies typically have fewer agencies with similar characteristics available to consider as peers.

Motorbus

Table 19 shows the initial set of potential motorbus peers that was identified, based on selecting all peers with likeness scores of 1.00 or less. Based on UTA's screening criteria, North County Transit District is eliminated because it only provides suburban service and its (diesel) light rail line opened in 2008, San Francisco MUNI is eliminated because its service area is limited to its region's central city, Buffalo is eliminated because its region is losing population, and Jacksonville is eliminated because it does not operate light rail. All of these eliminated agencies' likeness scores are over 0.75 (i.e., are in the “consider with caution” category), so eliminating them is reasonable.

Performance Results

Data Retrieval

FTIS was used to retrieve the desired performance data from the NTD. *Operator wages as a percent of total operating expenses* and *operator wages and fringe benefits as a percent of total operating expenses* are derivable from data on NTD form F-30.

Table 19. UTA motorbus peer group candidates.

Agency Name	Location	State	Total Likeness Score
Santa Clara Valley Transportation Authority	San Jose	CA	0.58
Sacramento Regional Transit District	Sacramento	CA	0.63
Denver Regional Transportation District	Denver	CO	0.68
Tri-County Metropolitan Transportation District of Oregon	Portland	OR	0.74
North County Transit District	Oceanside	CA	0.78
Charlotte Area Transit System	Charlotte	NC	0.80
Bi-State Development Agency	St. Louis	MO	0.85
San Francisco Municipal Railway	San Francisco	CA	0.90
Niagara Frontier Transportation Authority	Buffalo	NY	0.94
Jacksonville Transportation Authority	Jacksonville	FL	1.00

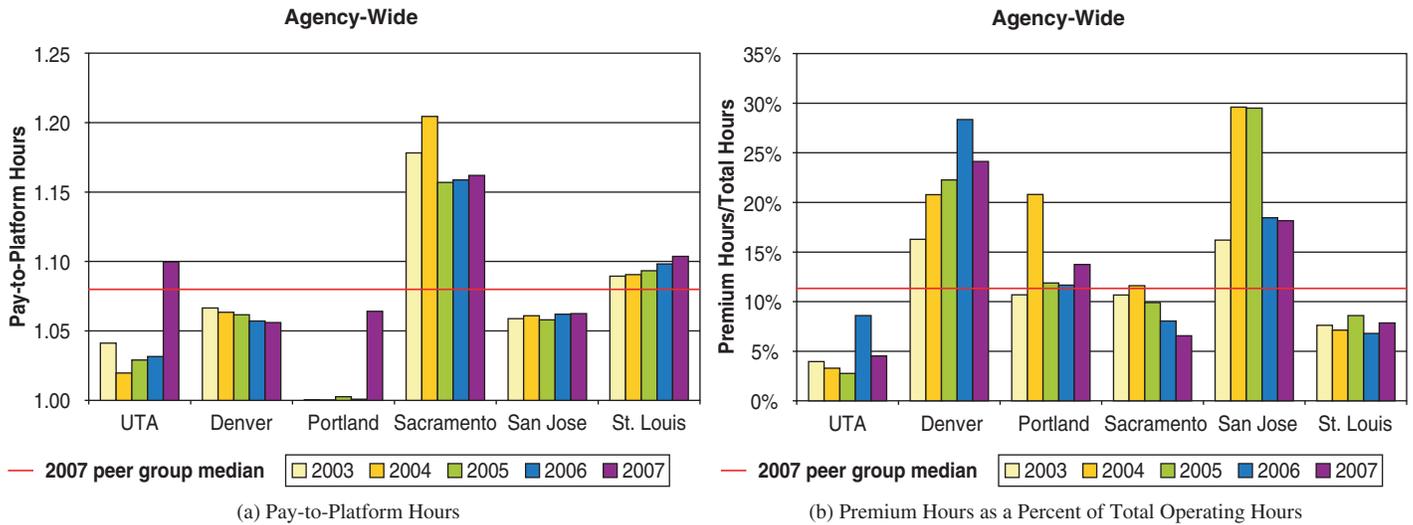


Figure 12. Agency-wide performance measure results for UTA.

(Note that fringe benefit costs need to be proportioned between vehicle operators and other operating staff.) *Pay-to-platform hours* and *premium hours as a percent of total operating hours* are derivable from data on NTD form F-50. *Vehicle revenue hours per operating employee full-time equivalent, annual boardings per operating employee full-time equivalent, cost per boarding, and cost per revenue hour* are directly available from FTIS as Florida Standard Variables. *Vehicle hours per revenue hour* are derivable from the Florida Standard Variables *vehicle hours* and *revenue hours*.

Denver Regional Transportation District (RTD) is the only agency in the motorbus peer group to use a significant amount of purchased transportation; in 2007, about 47% of RTD’s motorbus revenue hours were contracted. Because many of the detailed wage-related variables are not reported to the NTD for purchased transportation, only Denver’s directly operated service is included in the comparison. However, the broader

cost-efficiency variables can be compared: for example, in 2007, RTD’s purchased transportation *revenue hours per vehicle hour* was 85%, *cost per boarding* was \$4.33, and *cost per revenue hour* was \$60.08.

Agency-Wide Results

The NTD data used to derive two of the measures in this case study, *pay-to-platform hours* and *premium hours as a percent of total operating hours* are only reported on an agency-wide basis (i.e., mode-specific data are not available). Figure 12 shows the performance results for these two measures.

Light Rail

Figure 13 shows the performance results for the light rail mode.

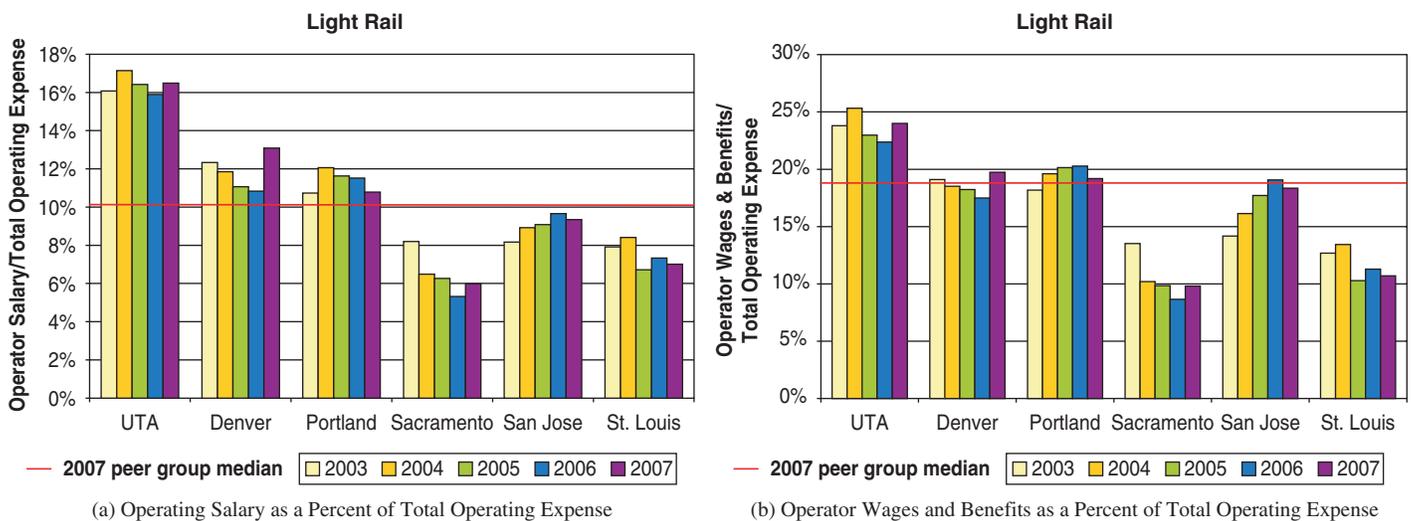
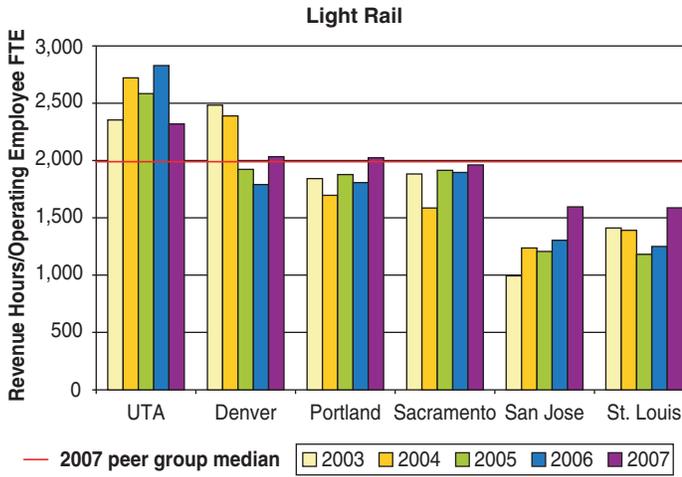
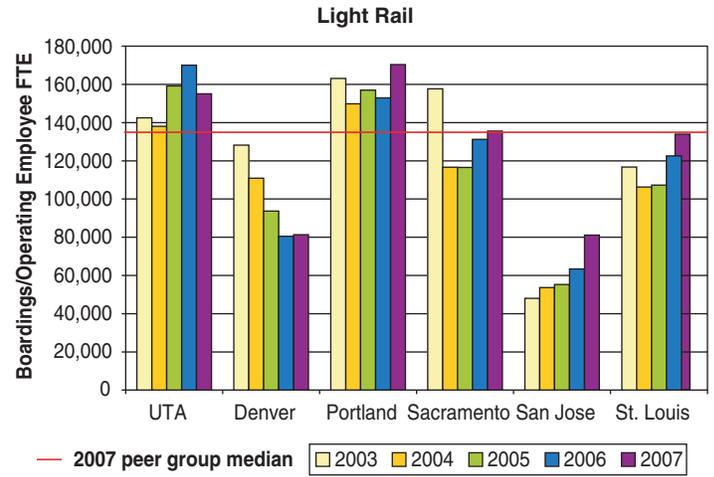


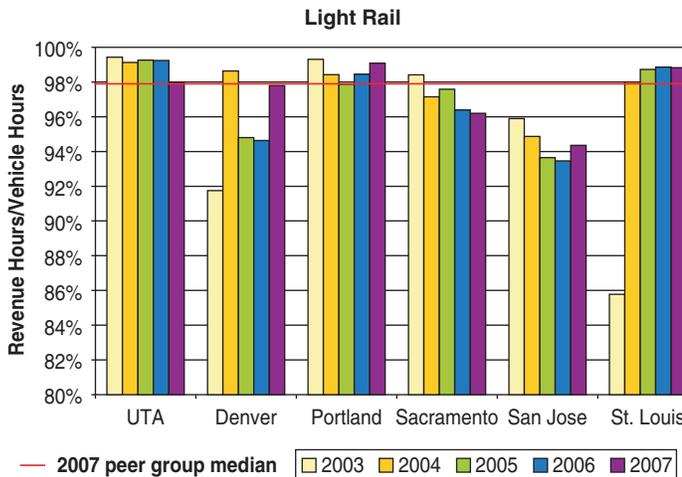
Figure 13. Light rail performance measure results for UTA.



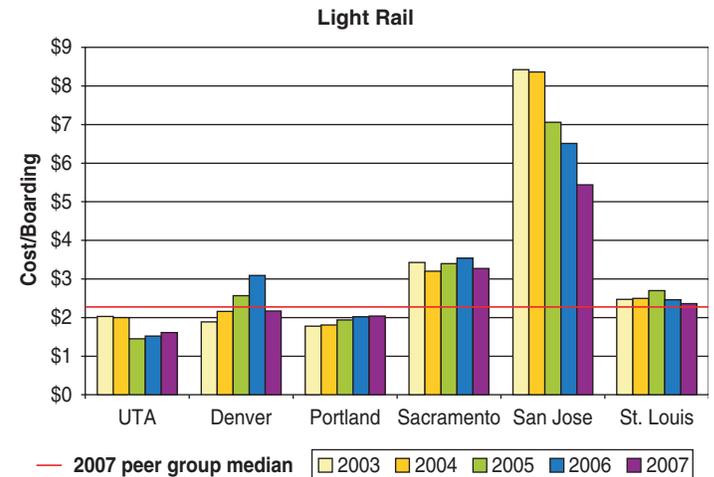
(c) Revenue Hours per Operating Employee Full-Time Equivalent



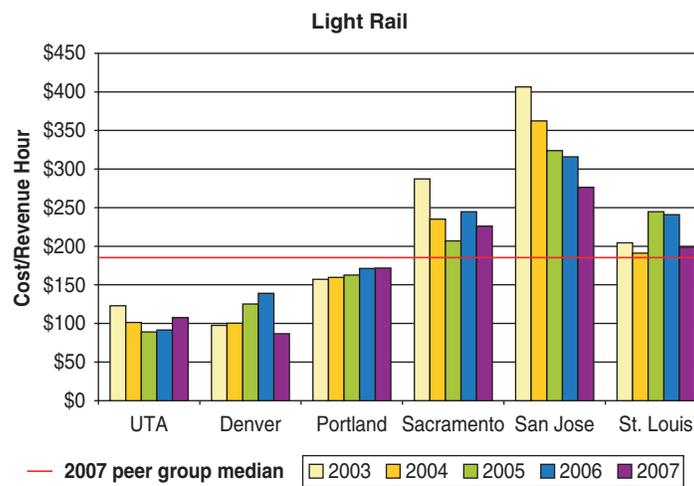
(d) Annual Boardings per Operating Employee Full-Time Equivalent



(e) Revenue Hours per Vehicle Hours



(f) Cost per Boarding



(g) Cost per Revenue Hour

Figure 13. (Continued).

Motorbus

Figure 14 shows the performance results for the motorbus mode.

Interpreting Results

Agency-Wide

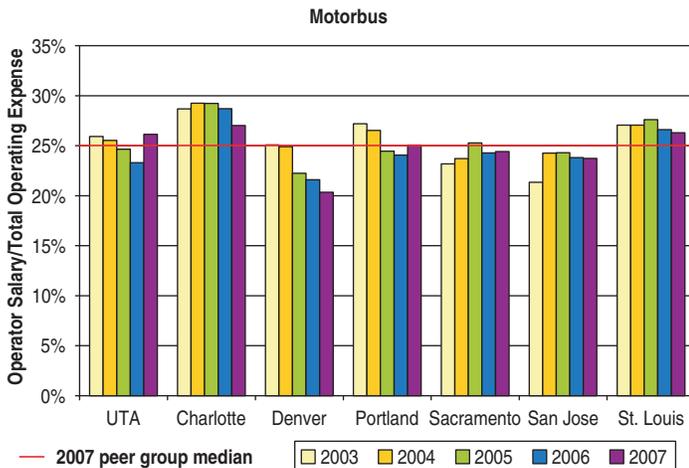
UTA's *pay-to-platform hours* ratio had been the second-lowest in the group, but rose sharply in 2007 and is now above the group median [Figure 12(a)]. (Portland's low values for this measure in most years are explained by its union contract, which allows operator breaks to occur as part of layover and recovery time, and thus are treated as platform time, rather than as separately paid break time.) UTA's *percentage of hours worked that were overtime* is the lowest in the group [Figure 12(b)], which is not necessarily good or bad, but more a reflection of the agency's philosophy regarding overtime.

However, having a low overtime rate means that more operators are needed to work the same number of hours, which can result in higher fringe benefit costs.

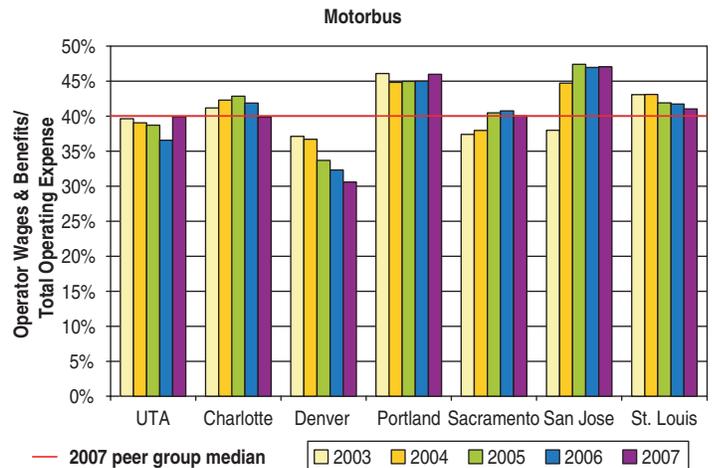
Light Rail

Operator wages [Figure 13(a)] and the combination of operator wages and benefits [Figure 13(b)] form a greater proportion of overall operating costs at UTA than at any other agency in the peer group. This result is not necessarily good or bad, but does indicate that increases in costs in these categories will translate more significantly to increased operating costs at UTA than at its peer agencies.

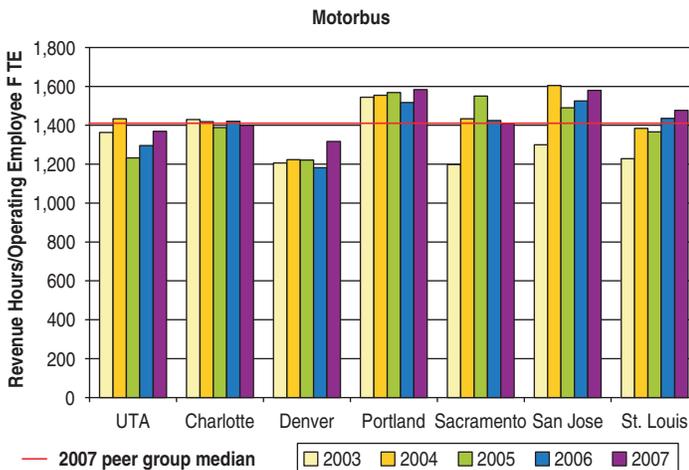
UTA operates more *revenue hours per employee FTE* than any of its peers [Figure 13(c)], although this ratio dropped in 2007 to its lowest level in the 5-year analysis period while the ratio rose at all of UTA's peer agencies. UTA's *boardings per employee FTE* is second-highest in the peer group [Figure 13(d)]. Here,



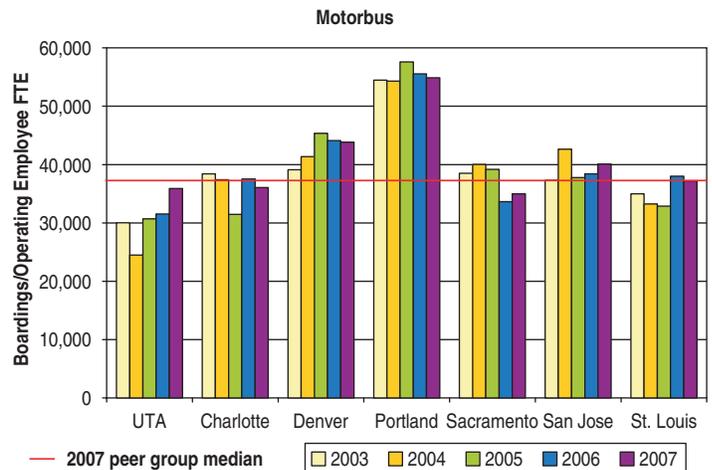
(a) Operating Salary as a Percent of Total Operating Expense



(b) Operator Wages and Benefits as a Percent of Total Operating Expense



(c) Revenue Hours per Operating Employee Full-Time Equivalent



(d) Annual Boardings per Operating Employee Full-Time Equivalent

Figure 14. Motorbus performance measure results for UTA.

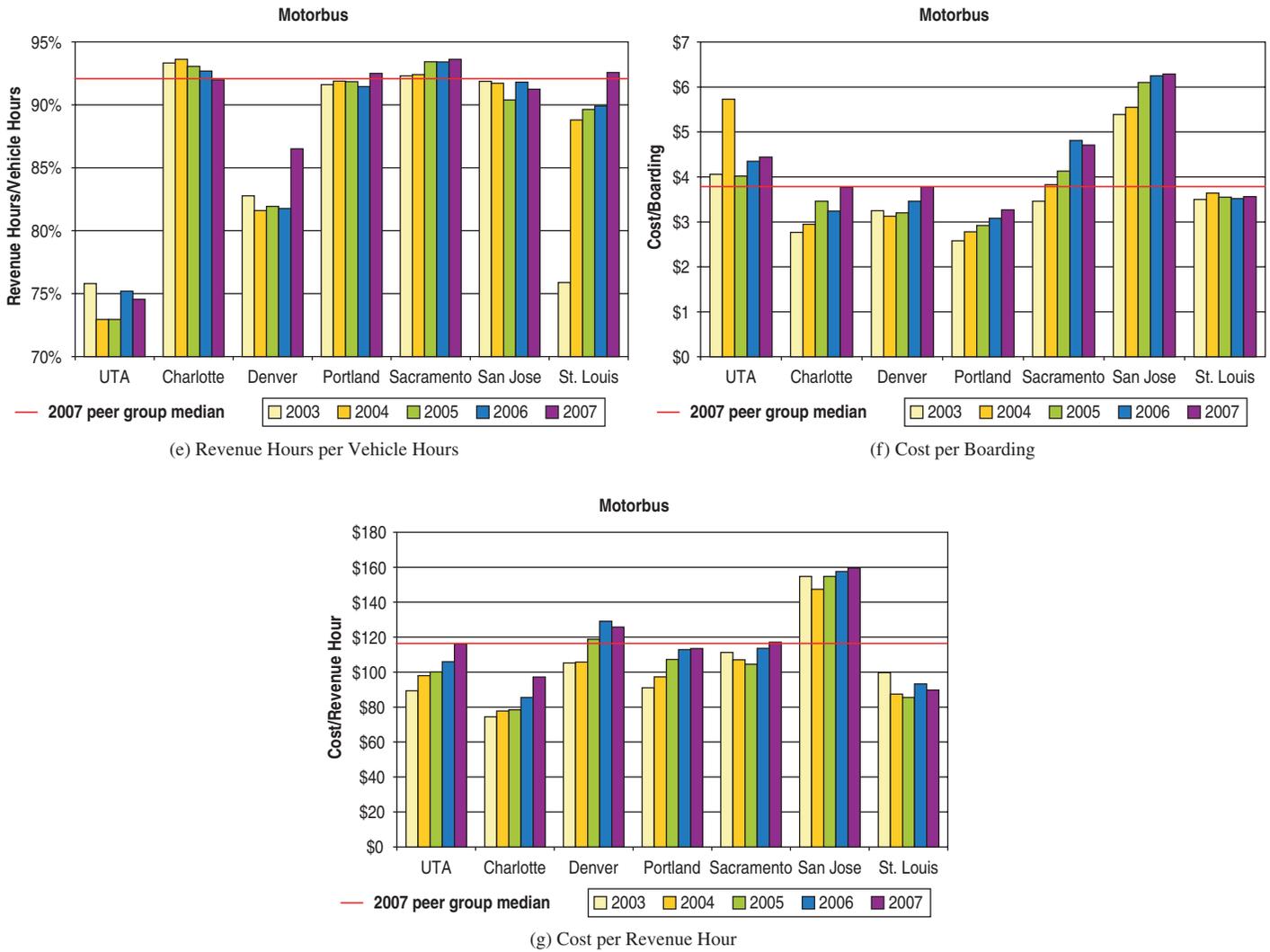


Figure 14. (Continued).

too, UTA’s result dropped in 2007 while rising at all the other peer agencies.

Through 2006, UTA was the peer group leader for *revenue hours per vehicle hour*, but dropped to the group median in 2007 [Figure 13(e)]. UTA’s *cost per revenue hour* is second-lowest in the peer group, but increased in 2007 while peer costs held steady or declined [Figure 13(f)]. UTA’s *cost per boarding* is lowest in the peer group. It increased slightly in 2007, and there was no consistent peer trend [Figure 13(g)].

Motorbus

Vehicle operator wages are above the group median [Figure 14(a)], while the combination of wages and fringe benefits are at the group median [Figure 14(b)]. The same comments that applied to these measures for light rail also apply here.

In terms of both *revenue hours per operating employee FTE* [Figure 14(c)] and *trips per operating employee FTE* [Fig-

ure 14(d)], UTA was second-lowest in the peer group. There was a fairly narrow range of values among the peer group for *revenue hours per operating employee FTE*; Portland and Denver stand out for *trips per operating employee FTE*, with the other agencies in a relatively narrow range. UTA’s performance in both categories is improving.

UTA’s *revenue hours per vehicle hour* are by far the lowest in the peer group [Figure 14(e)]. UTA’s service area is more spread out than any of the other peers, with the possible exception of Denver, so significant deadheading may be required to serve longer-distance commute trips. (This case study focuses on scheduling efficiency; however, a comparison of *farebox recovery ratio* would provide clues as to whether UTA is recouping the cost of providing longer-distance service.) UTA’s *cost per boarding* is above the peer-group median; the cost has increased in recent years, consistent with the peers [Figure 14(f)]. Finally, UTA’s *cost per revenue hour* is at the peer-group median, but is increasing at a faster rate than its peers [Figure 14(g)].

Asking Questions

On the light rail side, UTA's performance was generally among the top in its peer group and UTA appears to have historically scheduled its employees efficiently. However, in the final year, UTA's pay-to-platform hours increased substantially, from well below average to above average, which had an impact on costs. UTA would want to look into the reasons for the increase to see if actions could be taken to reverse it. On the bus side, UTA's performance lags its peers in many areas. With the peer data now in hand, UTA could use the motorbus results to dig deeper into its own data; for example, by comparing efficiency by service type (e.g., urban bus service vs. commuter bus service). As noted previously, a comparison of farebox recovery and other related financial indicators would help answer the question of whether UTA is recouping the cost of the extra deadhead time it incurs. UTA could also analyze its garage locations and the impacts of future commuter rail service on commuter bus routes to see if deadheading could or will be reduced in the future.

UTA's light rail cost indicators moved up in 2006–2007, which was opposite the peer-group trend. Although UTA can obviously track its own current-year costs, NTD data typically have a 2-year time lag, which makes it difficult to apply peer-group trend insights to near-term decision-making. Peer-group information that is as up to date as one's own would be more useful in that regard. However, now that UTA has identified its peer group, it could contact its peers to either (a) request their NTD viewer passwords (to obtain NTD data submitted, but not yet released by FTA) or (b) request the desired data directly. Ideally, if the peer group members agreed to share their current cost information with each other on a regular basis, all could benefit from having up-to-date peer trend information to work with.

Because this performance question focused on schedule efficiency, no adjustments were made to costs to reflect either inflation or differences in wage rates between regions. However, since average wages in Salt Lake City were among the lowest of comparably sized Western metropolitan areas, an adjustment for wage rates would provide insights into how much of UTA's relatively low *cost per revenue hour* and *cost per boarding* for light rail is due to efficient operation and how much is due to the region's lower labor costs. This kind of adjustment is illustrated next as part of the Denver case study.

Denver, Colorado

Context

The Denver RTD serves the Denver and Boulder urban areas. RTD provides light rail, motorbus, demand-response, and vanpool service. About 47% of the motorbus revenue

hours, nearly all of the demand-response revenue hours, and all of the vanpool service are contracted. RTD operated and purchased 54 million vehicle miles in 2007 and had an operating budget of \$320 million. The Denver urban area had a 2007 population of slightly over 2 million.

Performance Question

How comparatively cost-efficient is RTD's overall operation in terms of a fairly calculated and compared cost per revenue hour of service?

Performance Measures

The performance question identifies one measure, *cost per revenue hour*. Two other measures, *cost per boarding* (cost-effectiveness) and *boardings per revenue hour* (productivity) will also be included to provide a more rounded comparison. Costs will be adjusted for both inflation and regional wage rates, and the two cost-related measures will be compared on both an adjusted and an unadjusted basis in order to look at the impact of including those two factors in the analysis. No secondary screening criteria were identified.

Peer Grouping

FTIS was used to develop an agency-wide peer group for RTD. A peer group of eight was desired, likeness score values permitting. The candidate peers that were identified are shown in Table 20.

Five of the potential peers have likeness scores over 0.75, which suggests the need for a closer look at the suitability of the peers. Based on the peer-grouping and service area data supplied by FTIS, the following are noted:

- Denver RTD has the largest service area by far of the peer group, even after accounting for the fact that—like most agencies—it reports its district size (which in this case includes large unpopulated and unserved areas) rather than its actual service area [determined as $\frac{3}{4}$ mile from bus routes and rail stations, according to the NTD reporting instructions (49)].
- Houston is very comparable to Denver in operating budget and revenue miles operated and is the only regional transit agency in its urban area.
- Metro Transit's urban area contains multiple transit operators, unlike Denver, where RTD serves the entire region. However, Metro Transit's service area population is comparable to that of other peer agencies, and both Denver and St. Paul are state capitals (which provide concentrations

Table 20. Denver peer group candidates.

Agency	City	State	Likeness Score
Santa Clara Valley Transportation Authority	San Jose	CA	0.50
Utah Transit Authority	Salt Lake City	UT	0.59
Tri-County Metropolitan Transportation District of Oregon	Portland	OR	0.66
Metropolitan Transit Authority of Harris County, Texas	Houston	TX	0.77
Metro Transit	Minneapolis	MN	0.88
Dallas Area Rapid Transit	Dallas	TX	0.88
Sacramento Regional Transit District	Sacramento	CA	0.88
Bi-State Development Agency	St. Louis	MO	0.93

of office employment). Metro Transit's light rail service started during the analysis period.

- DART serves just the Dallas sub-region of the Dallas–Ft. Worth–Arlington urbanized area, but its budget is similar in size to RTD's.
- Sacramento's revenue miles operated and budget are one-third and one-half of Denver's, respectively, but the urban areas have similar population densities—Sacramento's urban area population is larger than both San Jose's and Salt Lake City's—and both Denver and Sacramento are state capitals. Other transit operators serve about 20% of the population within the Sacramento urban area.
- St. Louis is a comparably sized urban area and Bi-State Development Agency is the only multimodal transit operator in its urban area, although it only operates about half the amount of service that Denver does.

Keeping in mind the principle that peers should be similar but should not be expected to be exactly the same, Houston, Dallas, and St. Louis can readily be included as peers. Minneapolis and Sacramento differ more substantially from Denver, but also have notable similarities. Therefore, they will be retained as peers, but the differences will be kept in mind when interpreting the results.

Performance Results

Cost Adjustments for Inflation and Cost of Living

All three performance measures are directly available from FTIS as Florida Standard Variables. Inflation and wage data required to calculate adjusted costs were obtained from the BLS at the websites identified in Chapter 4.

Inflation data specific to metropolitan areas are available for seven of the nine agencies. For the two regions without detailed inflation data, Sacramento and Salt Lake City, average inflation data for urban areas in the Western United States will be used instead. Table 21 shows the CPI (all consumers) values for each urban area, by year. It can be seen that different regions experienced different levels of inflation between 2003 and 2007, with Denver experiencing the lowest percent inflation.

The process described in Chapter 3 was used to develop factors that convert prior-year costs into 2007 equivalents. The results are shown in Table 22.

Average hourly wage data across all occupations is available for all nine transit agencies' urban areas. As described in Chapter 3, it is possible to drill down into the BLS wage database to get more-specific data—for example, average wages for

Table 21. Consumer Price Index values for Denver peer group.

	2003	2004	2005	2006	2007	% change
Dallas	176.2	178.7	184.7	190.1	193.2	9.6%
Denver	186.8	187.0	190.9	197.7	202.0	8.1%
Houston	163.7	169.5	175.6	180.6	183.8	12.3%
Minneapolis	182.7	187.9	193.1	196.2	201.2	10.1%
Portland	186.3	191.1	196.0	201.1	208.6	12.0%
Sacramento	188.6	193.0	198.9	205.7	212.2	12.5%
Salt Lake City	188.6	193.0	198.9	205.7	212.2	12.5%
San Jose	196.4	198.8	202.7	209.2	216.0	10.0%
St. Louis	173.4	180.3	186.2	189.5	193.2	11.4%

Table 22. Inflation cost adjustments for Denver peer group.

	2003	2004	2005	2006	2007
Dallas	1.096	1.081	1.046	1.016	1.000
Denver	1.081	1.080	1.058	1.022	1.000
Houston	1.123	1.084	1.047	1.018	1.000
Minneapolis	1.101	1.071	1.042	1.025	1.000
Portland	1.120	1.092	1.064	1.037	1.000
Sacramento	1.125	1.099	1.067	1.032	1.000
Salt Lake City	1.125	1.099	1.067	1.032	1.000
San Jose	1.100	1.087	1.066	1.033	1.000
St. Louis	1.114	1.072	1.038	1.020	1.000

“bus drivers, transit and intercity.” However, the more-detailed category would be dominated by the transit agencies’ own workforces. The intent here is to (a) investigate whether Denver is spending more or less for its labor relative to its region’s average wages and (b) adjust costs to reflect differences in a region’s overall cost of living (which impacts overall average wages within the region). Table 23 shows the average hourly wage values for each urban area by year. It can be seen that different urban areas experienced varying amounts of wage growth between 2003 and 2007 and that there is a relatively wide spread in the cost of living (as reflected by the regional average wage) among the peer agencies.

The process described in Chapter 4 was used to develop factors that reflect how much higher or lower each region’s wages are compared to Denver. These factors are applied to each region’s cost data to produce adjusted costs that reflect the approximate cost each agency would have experienced if their region’s average wages and cost of living were the same as Denver’s. The results are shown in Table 24.

Performance Comparison Graphs

Figure 15 shows the performance results, based on costs adjusted for regional differences in inflation, labor market con-

ditions, and cost of living. For illustrative purposes, results based on unadjusted costs are also presented.

Interpreting Results

Cost-Efficiency

Looking at the *adjusted cost per revenue hour* first, Denver has the best performance among its peers [Figure 15(a)]. The trend data indicate that Denver’s costs held steady relative to inflation during 2003–2007. There is no apparent peer trend: some agencies’ costs increased at a faster rate than inflation, while other agencies’ costs increased at a slower rate (indicated by a declining trend in cost per revenue hour, as measured in 2007 dollars).

If the comparison had been performed with unadjusted data, Denver would have been second-best in the peer group, just behind Houston [Figure 15(b)]. However, Houston’s cost performance is influenced by the fact that average wages in Houston are 10% lower than in Denver. Region-wide wages are something that is out of the control of a transit agency, whereas one objective of performing a peer comparison is to find things that are under an agency’s control that can be improved. Using adjusted costs as a basis of comparison helps

Table 23. Mean hourly wages (all occupations) for Denver peer group.

	2003	2004	2005	2006	2007	% change
Dallas	\$18.35	\$18.86	\$19.23	\$19.68	\$20.57	12.1%
Denver	\$19.65	\$20.05	\$20.49	\$21.15	\$21.93	11.6%
Houston	\$18.05	\$18.51	\$18.71	\$19.09	\$19.72	9.3%
Minneapolis	\$19.92	\$20.59	\$21.07	\$21.63	\$22.31	12.0%
Portland	\$18.50	\$18.97	\$19.35	\$20.07	\$20.85	12.7%
Sacramento	\$19.19	\$19.81	\$20.11	\$20.98	\$21.64	12.8%
Salt Lake City	\$16.51	\$16.91	\$17.49	\$18.22	\$19.04	15.3%
San Jose	\$25.99	\$26.84	\$27.88	\$28.84	\$29.67	14.2%
St. Louis	\$17.88	\$18.03	\$18.22	\$18.72	\$19.53	9.2%

Table 24. Labor market and cost-of-living adjustments for Denver peer group.

	2003	2004	2005	2006	2007
Dallas	1.071	1.063	1.066	1.075	1.066
Denver	1.000	1.000	1.000	1.000	1.000
Houston	1.089	1.083	1.095	1.108	1.112
Minneapolis	0.986	0.974	0.972	0.978	0.983
Portland	1.062	1.057	1.059	1.054	1.052
Sacramento	1.024	1.012	1.019	1.008	1.013
Salt Lake City	1.190	1.186	1.172	1.161	1.152
San Jose	0.756	0.747	0.735	0.733	0.739
St. Louis	1.099	1.112	1.125	1.130	1.123

to eliminate some of these external factors. A comparison of the adjusted and unadjusted data also indicates that operating costs in Dallas and Sacramento are relatively high regardless of the cost basis used, while much of San Jose's relatively high operating costs can be explained by that region's high cost of living.

Cost-Effectiveness

Again looking at the adjusted data first, Denver is slightly below the group median for *cost per boarding* [Figure 15(c)]. Denver's value increased during 2003–2007, while five of the eight peers showed a decrease. The relative placement of the agencies does not change much when the unadjusted data are compared [Figure 15(d)]. This is due in part to the fact that *cost per boarding* measures both a service input and a service outcome, while *cost per revenue hour* compares two service inputs. Denver's shift from best-in-class for *cost per revenue hour* to middle-of-the-pack for *cost per boarding* suggests that other agencies generate more boardings per revenue hour. This tentative conclusion will be confirmed by the productivity measure.

Productivity

The final graph only comes in one version, as *boardings per revenue hour* does not involve any cost data. This graph shows that Denver has the second-lowest productivity among the peer group [Figure 15(e)]. Denver's trend of a slight decline in this measure is consistent with five of the eight peer agencies. The two leaders in this category are Minneapolis and Portland.

Asking Questions

The results show that Denver has done a good job relative to its peers at controlling the costs of providing transit service.

However, the service that Denver provides has not been as productive as that of most of its peers. One possible explanation is that because Denver has a larger service area than any of its peers, it provides relatively more long-distance routes, which would be expected to have lower productivity due to the amount of time that passengers spend on the bus. This theory could be tested in at least two ways.

First, Denver could look more in-depth at UTA's and Houston's results. Those two systems are similar to Denver in terms of regional coverage and the operation of longer-distance bus routes, yet they had better productivity. Second, Denver could use its own in-house data to remove operating costs, revenue hours, and boardings for routes serving outlying communities (e.g., routes originating outside the Denver urban area). The results for the remainder of the system could then be compared to the results of the six remaining peers with more compact service areas since there would now be more of an apples-to-apples comparison of service area sizes.

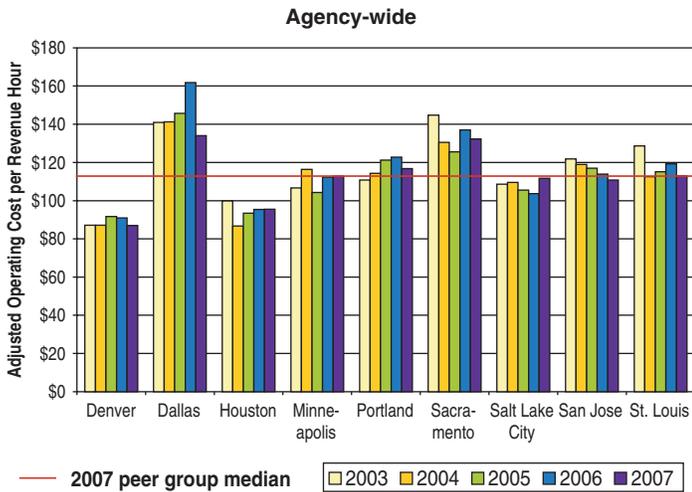
San Jose, California

Context

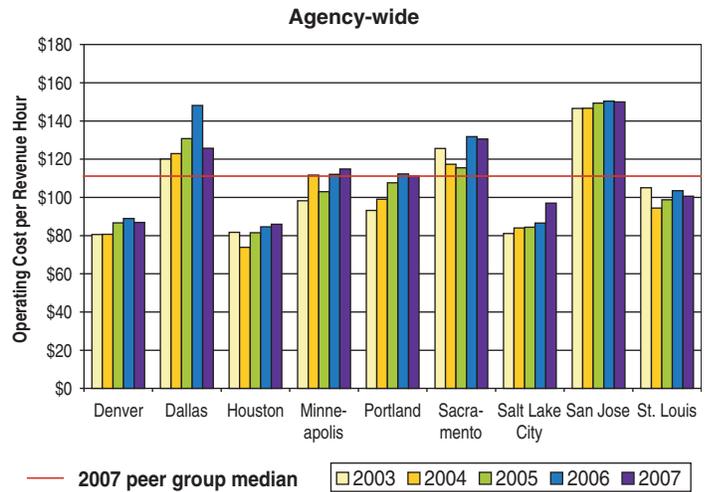
The Santa Clara Valley Transportation Authority (VTA) serves Santa Clara County, located at the south end of San Francisco Bay and containing the Bay Area's largest city, San Jose. VTA directly operates light rail and motorbus service and purchases about 3% of its motorbus revenue hours and all of its demand response service. VTA operated 25 million vehicle miles in 2007 and had an operating budget of \$282 million. The San Jose urban area had a 2007 population of 1.58 million.

Performance Question

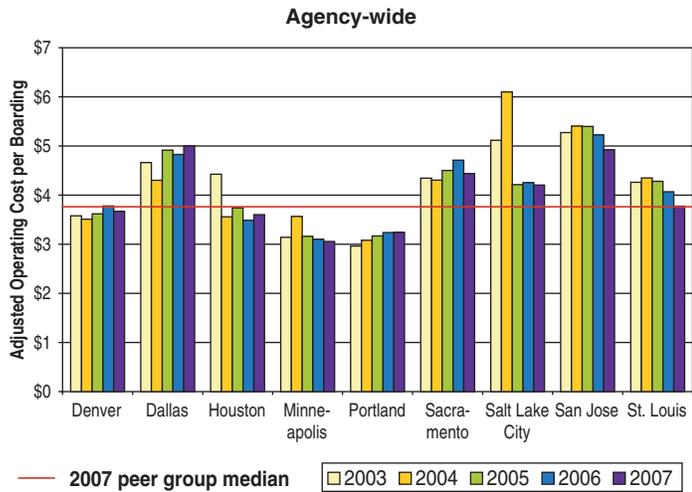
How effective are VTA's light rail vehicle maintenance and non-vehicle maintenance programs?



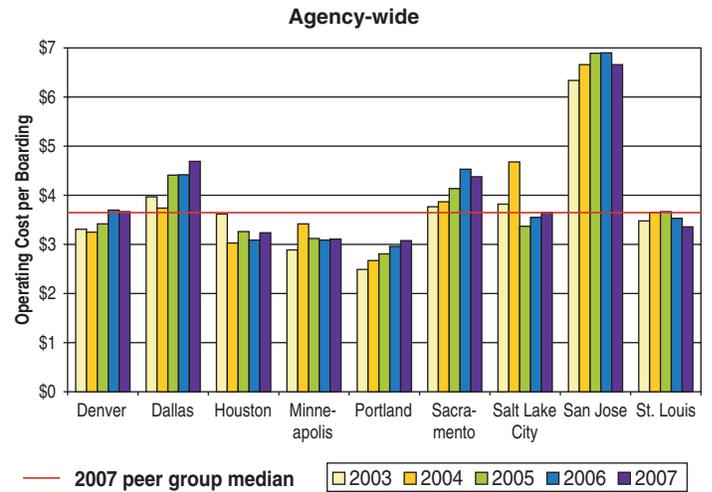
(a) Operating Cost per Revenue Hour (Adjusted Costs)



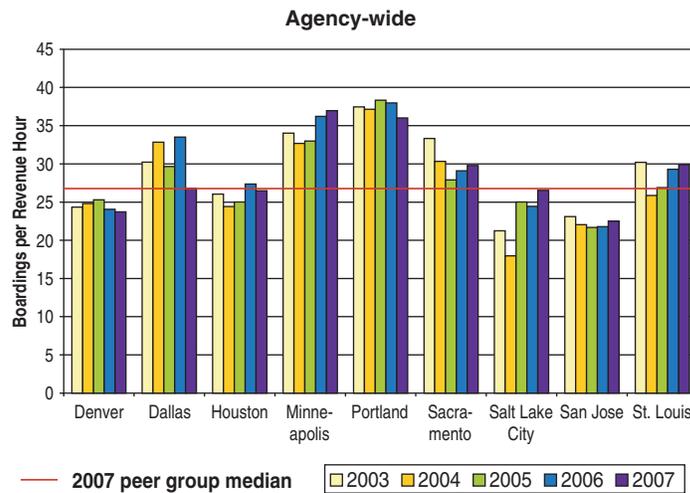
(b) Operating Cost per Revenue Hour (Unadjusted Costs)



(c) Operating Cost per Boarding (Adjusted Costs)



(d) Operating Cost per Boarding (Unadjusted Costs)



(e) Boardings per Revenue Hour

Figure 15. Performance results for Denver.

Performance Measures

The following measures are selected or derived from the tables in Chapter 3 relating to maintenance administration, service characteristics, and transit investment:

- *Percent of maintenance costs that are labor,*
- *Average fleet age,*
- *Spare ratio,*
- *Miles of track,*
- *Vehicle maintenance cost per vehicle operated in maximum service,*
- *Vehicle maintenance cost per car mile,*
- *Car miles between failures,*
- *Maintenance costs as a percentage of total operating costs, and*
- *Non-vehicle maintenance cost per track mile.*

The first four measures are descriptive measures that provide context about each light rail operator. The remaining measures are outcome measures. No secondary screening measures were identified. As noted in the Denver case study, San Jose's cost of living is higher than in many other parts of the country. Therefore, vehicle maintenance cost comparisons will be adjusted to account for wage differences between regions.

Peer Grouping

FTIS was used to develop a light rail peer group for VTA. A peer group of eight was desired, likeness score values permitting. Table 25 shows the candidate peers that were identified.

MBTA was dropped as a peer on the basis of being an operator of streetcars rather than modern light rail vehicles (LRVs). San Francisco Muni also operates some historic streetcars, but the majority of its fleet consists of modern LRVs. A significant portion of Muni's system operates underground, unlike the others in its peer group, so that fact will need to be considered when non-vehicle maintenance costs (e.g., stations and right-of-way) are compared.

Table 25. San Jose peer group candidates.

Agency	City	State	Likeness Score
Denver Regional Transportation District	Denver	CO	0.49
Sacramento Regional Transit District	Sacramento	CA	0.55
Maryland Transit Administration	Baltimore	MD	0.57
Utah Transit Authority	Salt Lake City	UT	0.58
Tri-County Metropolitan Transportation District of Oregon	Portland	OR	0.61
San Diego Trolley, Inc.	San Diego	CA	0.70
San Francisco Municipal Railway	San Francisco	CA	0.71
Massachusetts Bay Transportation Authority	Boston	MA	0.82

Performance Results

Cost Adjustments for Labor Market and Cost of Living

Wage data required for cost adjustments were obtained from the BLS at the website identified in Chapter 3. The process used to adjust wage data is similar to the one used in the Denver case study (except that San Jose is used as the reference point this time) and will not be repeated here.

Data Retrieval

Average fleet age, spare ratio, total maintenance costs, total operating costs, and number of vehicle system failures are Florida Standard Variables. *Total rail track miles* is available from NTD form A-30. Maintenance vehicle labor costs come from two variables on NTD form F-30 (*vehicle maintenance other salaries/wages* and *vehicle maintenance fringe benefits*); two other variables on NTD form F-30 provide the same information for non-vehicle maintenance labor costs. Finally, *car miles* is available from NTD form S-10. All of the desired performance measures can be determined directly from these variables or as ratios of these variables.

Data Issues

The following data issues were noted when the data were retrieved from FTIS:

- Denver did not report *vehicle system failures* in 2003–2006. It did report them in 2007, but the resulting *average distance between failures* was 20 times greater than any other peer in 2007. Therefore, Denver's 2007 data were discarded from the analysis. Similarly, Salt Lake City's *average distance between failures* was four times greater than any other peer in 2003–2005 and was discarded.
- San Diego's light rail data were reported by the San Diego Metropolitan Transit System (MTS) in 2007, which is a separate NTD reporter from the former San Diego Trolley, Inc., which reported in earlier years. Some key variables

needed for this case study's performance measures were not reported by MTS in 2007.

- San Francisco Muni reported the exact same number of *light rail vehicle failures* (2,002) each year from 2003–2005.
- *Track miles* was not an NTD reporting variable until 2005.

Performance Comparison Graphs

Figure 16 shows the performance results. Note that some of the graphs show a 2006 median value to maximize the number of peers included in the calculation of the median because some variables could not be calculated for San Diego for 2007.

Interpreting Results

Descriptive Measures

VTA has the second-youngest fleet among its peers [Figure 16(a)] and the third-most track miles [Figure 16(b)]. Labor makes up slightly more than 50% of the total maintenance budget, which is a higher ratio than all but one peer [Figure 16(c)]. However, this result is unsurprising, given the urban area's high average hourly wage rate. VTA has by far the largest *spare ratio* of any of its peers [Figure 16(d)], with 50% more LRVs available as spares than are operated in maximum service (i.e., a spare ratio of 150%).

Outcome Measures

Even after adjusting for labor costs, VTA has the highest *maintenance cost per vehicle in maximum service* [Figure 16(e)] and second-highest *maintenance cost per car mile operated* [Figure 16(f)], although both measures have been trending downward while generally holding steady or increasing at VTA's peers. VTA's non-vehicle maintenance costs, on the other hand, are at the peer group median after adjusting for labor costs [Figure 16(g)]. Maintenance costs make up slightly more than half of total operating costs [Figure 16(h)], which is second-highest in the peer group.

In terms of *distance between light rail car failures*, VTA has been at or near the best-in-class within its peer group throughout the analysis period [Figure 16(i)]. However, there appears to be a wide variation in how agencies report light rail car failures to the NTD, so it may not be possible to conclude much from this measure.

Asking Questions

The data suggest that VTA's high *spare ratio* may be a key driver of the agency's relatively high maintenance costs, after controlling for labor cost differences among the peer regions. VTA received 70 new low-floor LRVs during the analysis pe-

riod. VTA's own maintenance records could be used to compare the maintenance costs of the two fleets to confirm whether or not this theory is true. If true, and if the agency anticipated keeping its older high-floor vehicles to support future service expansions, it could contact its peers that have also purchased low-floor vehicles to learn from their experiences maintaining mixed high- and low-floor fleets. The objective of the contacts would be to try to identify whether VTA is performing more maintenance than needed on low-usage vehicles to keep them in good working order.

To effectively draw solid conclusions from the maintenance data, agency contacts would be needed to provide more context about maintenance activities and needs. For example, VTA would be interested in finding out the types of non-vehicle maintenance performed at its peer agencies and the ages of various components of its peers' light rail infrastructure. Because VTA's peers appear to report vehicle failures differently, agency contacts would also be needed to find out what definitions they used before firm conclusions could be drawn from the *car miles between failures* measure.

South Florida

Context

The Florida Department of Transportation (FDOT) contributed capital and operating funds to double-track the Tri-Rail commuter rail line operated by the South Florida Regional Transportation Authority (SFRTA) and was SFRTA's largest source of funds in 2005, 2006, and 2007. Consequently, FDOT is interested in comparing Tri-Rail's performance to that of similar commuter rail operations to make sure that the value of its investment is maximized.

SFRTA contracts commuter rail service in a single corridor running from Palm Beach County through Broward County into northern Miami-Dade County. It also contracts motor-bus and demand-response service that feeds its commuter rail stations. In 2007, SFRTA's commuter rail service operated just over 2 million vehicle miles and had an operating budget of \$33.5 million. The Miami urbanized area had a 2007 population of 5.23 million and contains all three counties that SFRTA operates in.

Performance Question

Compare Tri-Rail's level of service, investment in public transportation, and cost-effectiveness to that of its peers.

Performance Measures

The following measures are selected or derived from the tables in Chapter 3 relating to cost-efficiency, cost-effectiveness,

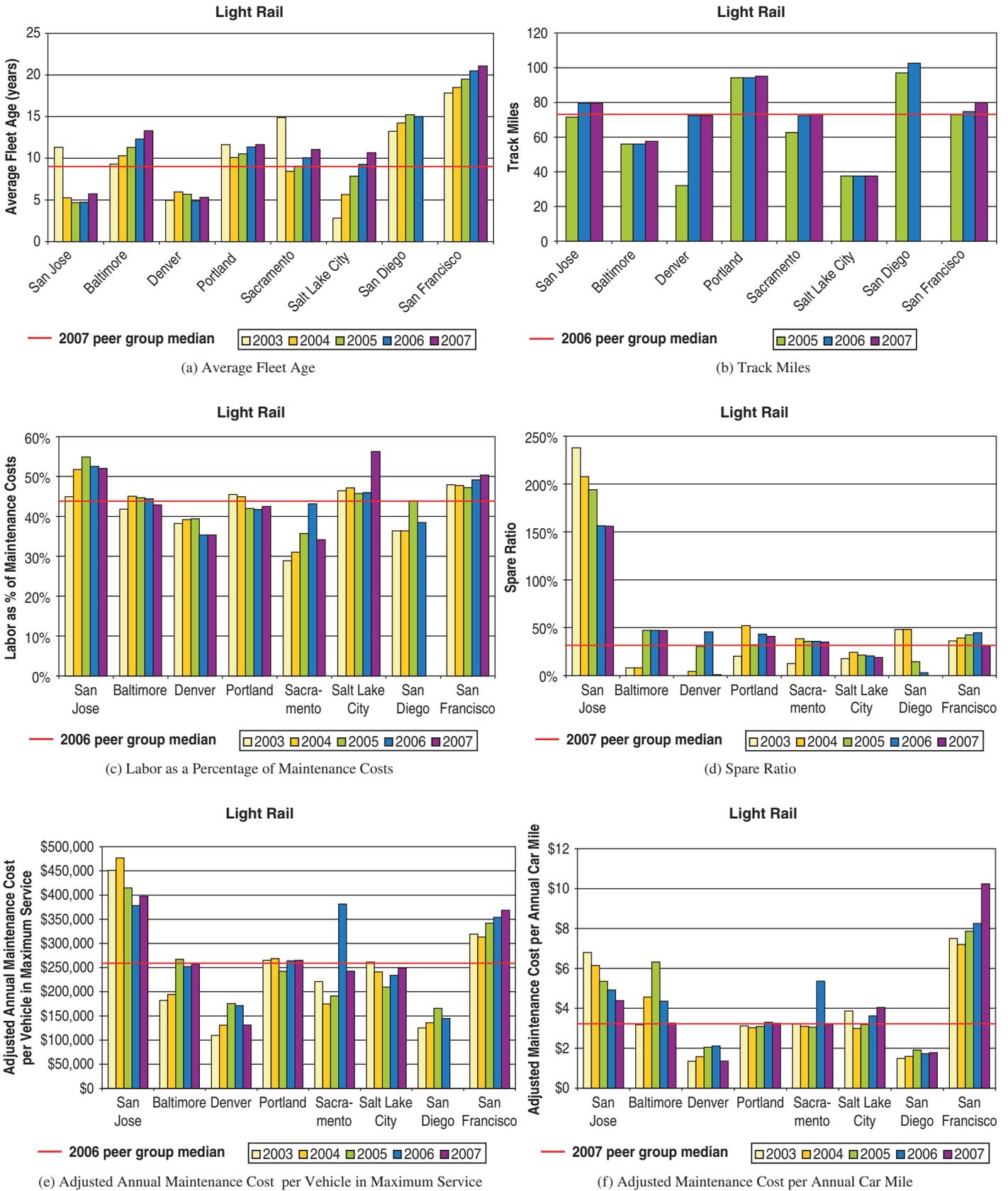
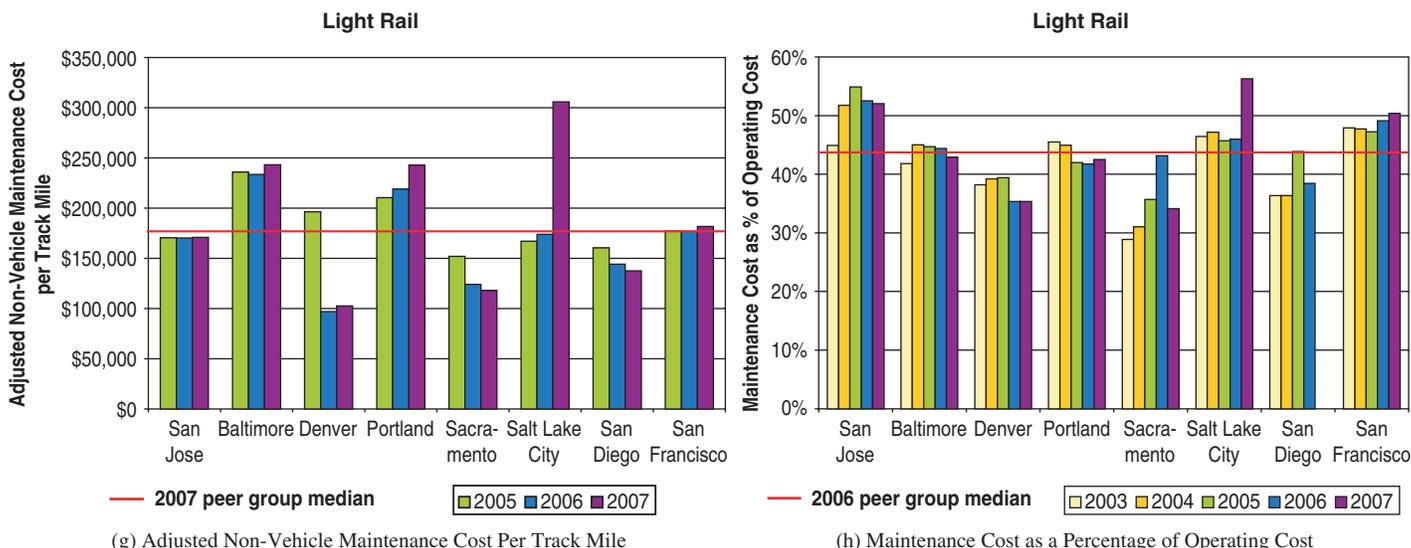
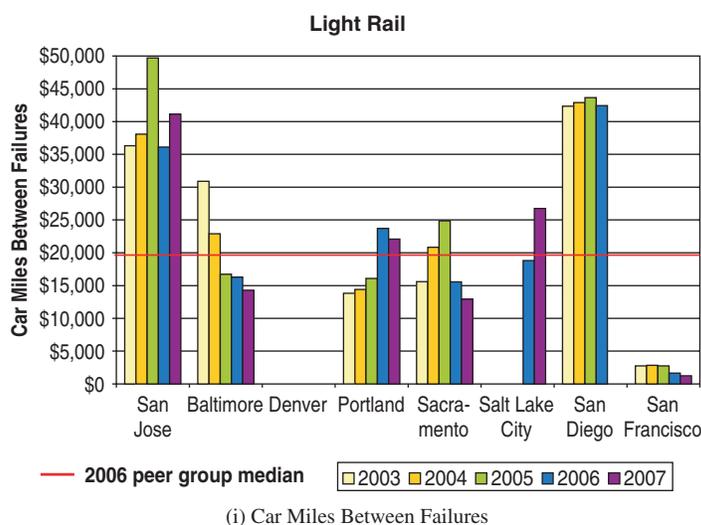


Figure 16. Performance results for San Jose.



(g) Adjusted Non-Vehicle Maintenance Cost Per Track Mile

(h) Maintenance Cost as a Percentage of Operating Cost



(i) Car Miles Between Failures

Figure 16. (Continued).

resource utilization, service utilization, perceived service quality, and delivered service quality:

- Operating cost per revenue hour,
- Operating cost per revenue mile,
- Operating cost per passenger mile,
- Operating funding per capita,
- Revenue miles per capita,
- Vehicle hours per vehicles operated in peak service,
- Service span,
- Average system speed,
- Average trip length, and
- Average system peak headway.

A comparison of *miles of track* to *directional route miles* was also investigated (to describe the prevalence of double tracking); however, all the agencies in the peer group reported

the same number of *miles of track* as *directional route miles*, so this was not possible. (According to the NTD reporting guidelines, *miles of track* should be one-half the *directional route miles* for commuter rail lines operating on single track, plus the length of any sidings/passing tracks and yard tracks, so the two values should never be equal, even when a route is fully double-tracked.) Individual agency contacts would therefore be required to determine the prevalence of double tracking.

The *average peak headway* is derived by FTIS from other NTD measures. [FTIS divides *directional route miles* by *average speed (revenue miles per revenue hours)* to give the average time for a train to make a round-trip, then divides this result by the *number of trains operated in peak service* to give an *average peak headway* in hours, and finally multiplies the result by 60 to give a value in minutes.] Since many commuter rail lines tend to have strongly directional peak service and some also

operate a variety of service patterns, the reported value will often not correspond to the peak-direction headway experienced at a given station. However, the measure is still useful as a comparative indicator of the relative frequency of service operated by different systems. (Note that a direct comparison of rail schedules using the Internet would also have difficulty accounting for multiple service patterns, variations in headways during the peak period, and the relative amounts of peak-direction and off-peak-direction service.)

For this comparison, FDOT wishes to focus on other commuter rail operators that operate a single route like SFRTA does, or a single route with two branches.

Peer Grouping

FTIS was used to develop a commuter rail peer group for Tri-Rail. Table 26 identifies the candidate peers that were identified.

Commuter rail operators in the Baltimore and Los Angeles regions were screened out by the criterion that peers should not operate multiple routes (excepting a trunk-and-branch operation like that operated by Virginia Railway Express). Trinity Railway Express (TRE) is somewhat unique in that it is jointly operated by the transit agencies in Dallas and Ft. Worth. Therefore, data used for TRE's likeness score calculation had to be manually combined in a spreadsheet from the individual data reported by the two agencies.

Two commuter rail lines had likeness scores that warranted further investigation. Coaster is the smallest operator in the peer group in terms of metropolitan area population and operating budget, but the San Diego region experiences a similar level of congestion as Miami, as measured by *annual hours of delay per traveler*. A significant portion of TRE's likeness score came from the difference between its parent agencies' service types (primary agency serving a region's central city) and SFRTA's service type (suburban service connecting to a

central city). If TRE were treated as a stand-alone agency with the same service type as SFRTA, its likeness score would be a satisfactory 0.63. Therefore, both of these commuter rail lines were retained as peers.

Performance Results

Cost Adjustments for Labor Market and Cost of Living

Because there is a wide range of average wages among the urban areas represented in the peer group, operating costs were adjusted to reflect wage differences between regions. As in the previous two case studies, the wage data were obtained from the BLS at the website identified in Chapter 4. A similar process was used to make the adjustments as in the previous two case studies. In three cases, Chicago, Miami, and San Francisco, wage data are available for subareas within the urbanized area. In these cases, the subarea containing a transit agency's headquarters was used.

Data Retrieval

All of the desired performance measures are available as Florida Standard Variables or as ratios of Florida Standard Variables except for *urban area population* (TCRP Project G-11 variable), *weekday A.M. peak number of vehicles/trains in operation* (NTD Form S-10), and *rail total track miles* (NTD Form A-20).

Following this report's guidance, *urban area population* was used instead of *service area population*. (Note that both Tri-Rail and South Shore report the population of the entire Miami and Chicago urban areas, respectively, as their service area populations, even though those lines serve only relatively small portions of their urban areas. Because the other peers all operate single lines or a trunk with two branches, the portion of the urban area served by each commuter rail line should

Table 26. Tri-Rail peer group candidates.

Agency	City	State	Likeness Score
South Florida Regional Transportation Authority (Tri-Rail)	Pompano Beach	FL	0.00
Central Puget Sound Regional Transit Authority (Sounder)	Seattle	WA	0.47
Peninsula Corridor Joint Powers Board (Caltrain)	San Carlos	CA	0.47
Virginia Railway Express (VRE)	Alexandria	VA	0.65
Northern Indiana Commuter Transportation District (South Shore)	Chesterton	IN	0.67
Southern California Regional Rail Authority (Metrolink)	Los Angeles	CA	0.79
North County Transit District (Coaster)	Oceanside	CA	0.80
Maryland Transit Administration (MTA)	Baltimore	MD	0.95
Trinity Railway Express (TRE)	Dallas-Ft. Worth	TX	1.05
Dallas Area Rapid Transit (TRE)	Dallas	TX	1.27
Fort Worth Transportation Authority (TRE)	Fort Worth	TX	1.41

be relatively similar.) As noted previously, while service area population would be the theoretically preferable basis of comparison, consistently reported service area data are not available from the NTD, the NTD service area definition does not include commuter rail's park-and-ride market area in any event, and the detailed census data required to develop station-area population estimates may be up to 10 years old. While not perfect, urban area population is sufficient for developing insights that can be followed up later on with a more-detailed analysis, if necessary.

As was the case with the peer-grouping data, performance data for TRE had to be combined from the data separately submitted by the Dallas and Ft. Worth transit agencies.

Descriptive Measure Graphs

Figure 17 presents descriptive measures for the peer group.

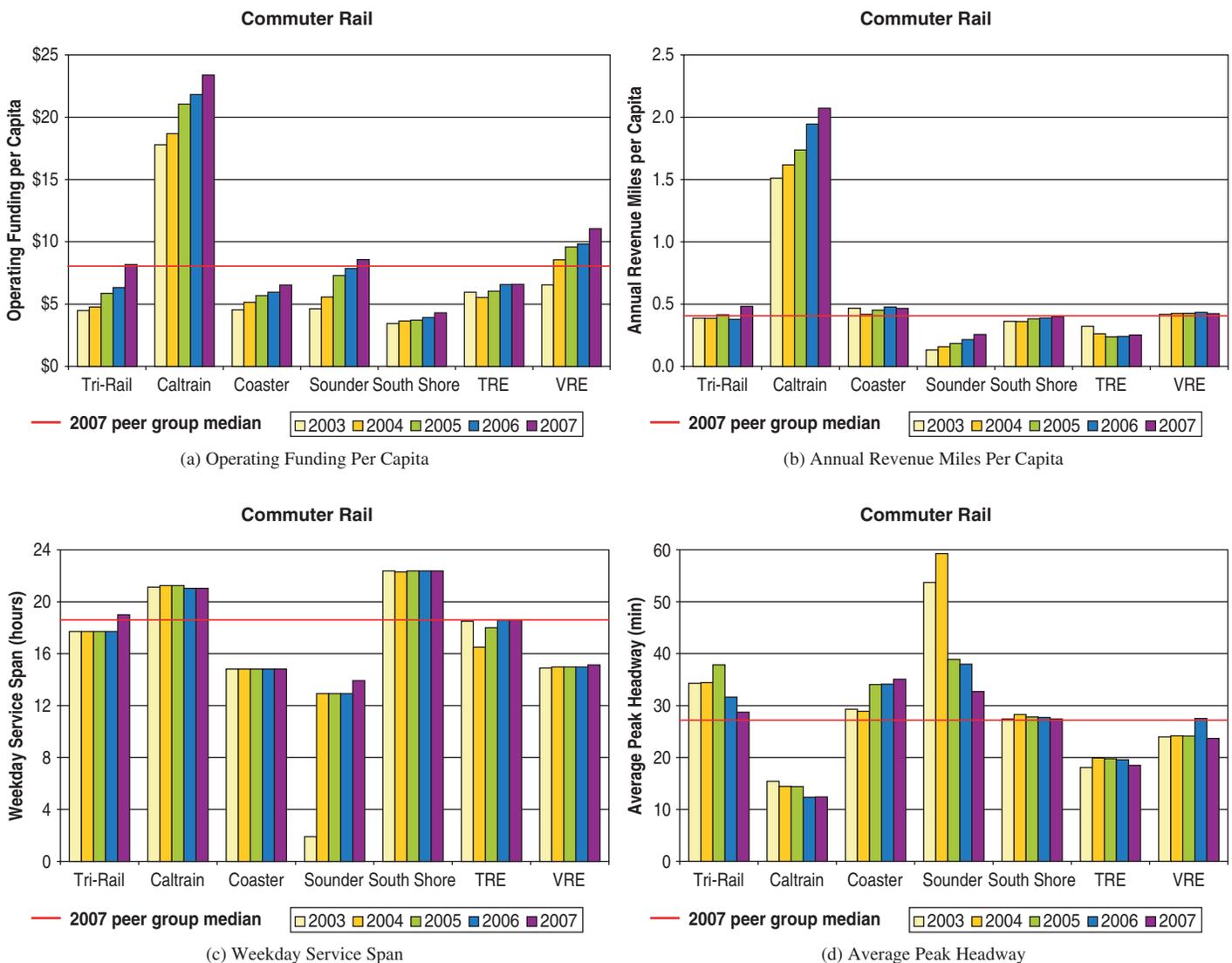


Figure 17. Descriptive measure results for Tri-Rail.

Outcome Measure Graphs

Figure 18 presents outcome measure results for the peer group.

Interpreting Results

Descriptive Measures

Tri-Rail's *operating funding per capita* is at the peer group median [Figure 17(a)]. Except for Caltrain, which is much higher than the rest of the peer group, Tri-Rail operated the most *revenue miles per capita* [Figure 17(b)]. Tri-Rail's values for both of these measures increased at the same rate as or faster than its peers between 2003 and 2007. Tri-Rail's *weekday service span* is a little above the peer group median [Figure 17(c)], while its *average peak headway* is a little longer than the median [Figure 17(d)]. *Weekday service span* has increased

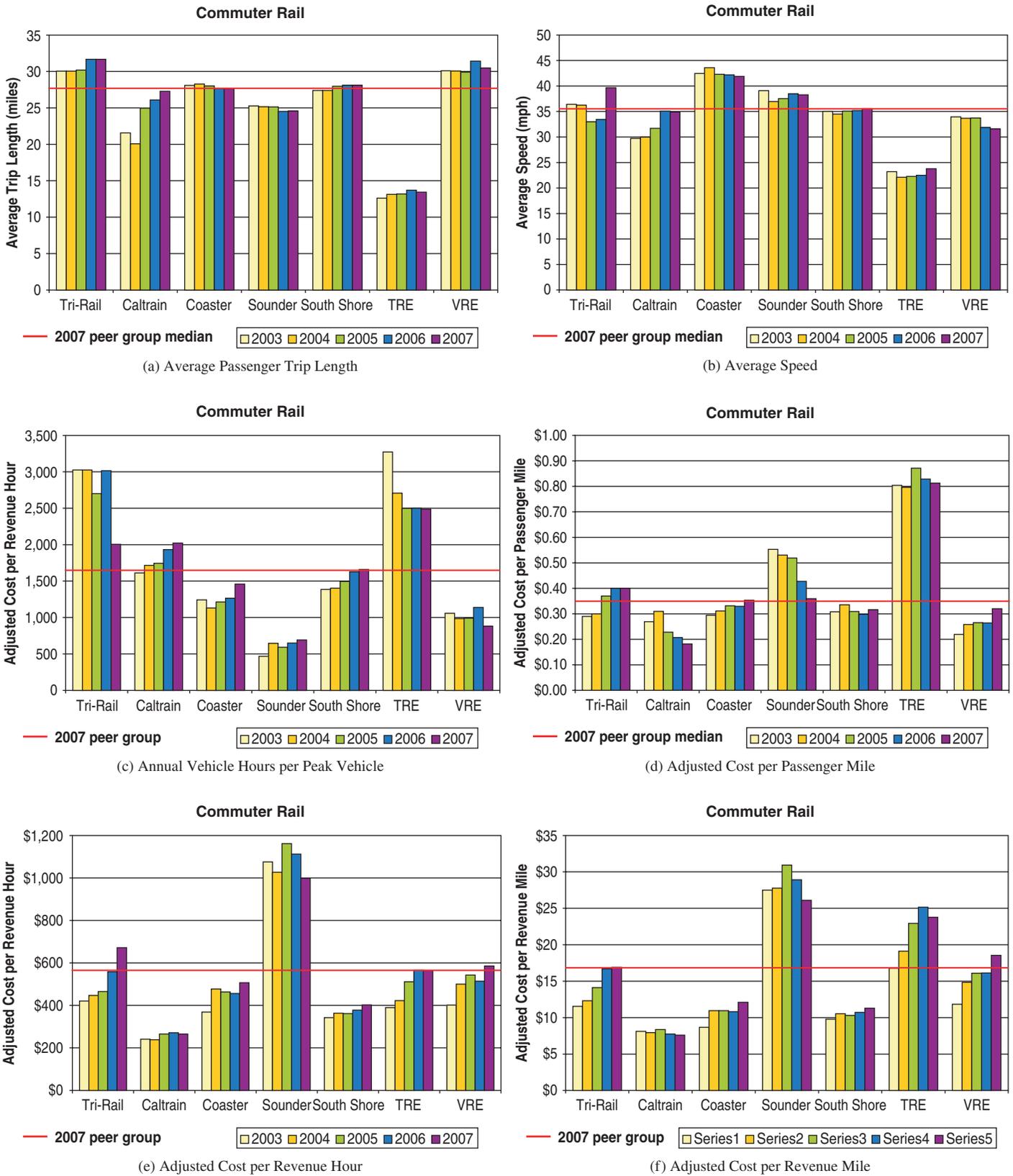


Figure 18. Outcome measure results for Tri-Rail.

(compared to a peer group trend of holding steady) and *average peak headway* has gotten shorter (compared to a peer group trend of steady to shorter).

Outcome Measures

Tri-Rail's passengers take the longest trips of any in the peer group [Figure 18(a)] and travel at the second-fastest average speed [Figure 18(b)]. Speeds increased significantly in 2007, while the long-term trend for the peers has been one of little change, except for Caltrain.

Tri-Rail gets good utilization out of its vehicles (in terms of *vehicle hours per vehicle operated in maximum service*), although it dropped sharply in 2007 from being consistently the best in this category to third in the peer group, opposite the peer group trend [Figure 18(c)].

Looking at cost-related measures, TriRail has the second-highest *adjusted cost per passenger-mile* [Figure 18(d)] and *adjusted cost per revenue hour* [Figure 18(e)] in the peer group, and these values have increased more between 2003 and 2007 than for those of any of its peers. (Note that *cost per passenger-mile* values are fairly tightly clustered for five of the agencies, including Tri-Rail, with one outlier above and one below.) Tri-Rail's *adjusted cost per revenue mile*, on the other hand, is at the peer group median [Figure 18(f)], although it, too, has gone up significantly since 2003 (although it held steady

between 2006 and 2007). If unadjusted costs had been used, Tri-Rail's position relative to its peers would have been the same, but VRE's unit costs would have increased more than Tri-Rail's due to the Washington, DC, region's relatively high wages (\$26.37 in 2007 vs. \$18.75 for Ft. Lauderdale) and much greater increase in average wages from 2003 to 2007 (19.6%, compared to 12.5% for the Ft. Lauderdale region).

Asking Questions

The peer-comparison results show that, on a per-capita basis, the state's and region's investment in commuter rail service is on a par with Tri-Rail's peers. The aspects of Tri-Rail's quality of service that can be assessed through the NTD were as good as or better than its peers. Two of Tri-Rail's cost-effectiveness and cost-efficiency values, on the other hand, were higher than most of its peers, and all three cost-related measures increased significantly during the analysis period. Since Tri-Rail began a service expansion during this period, associated with its double-tracking project, a question to be investigated further would be: What aspects of the service expansion, if any, are contributing to the significant unit cost increases? Caltrain stands out as the best-in-class in the peer group, even with its region's high cost of living, and could be an agency that Tri-Rail could look to for inspiration for cost-saving ideas and ideas for operating varying service patterns on double track.

CHAPTER 6

Concluding Remarks

Value of Peer Comparison and Benchmarking

The literature review summarized in Chapter 2 showed that peer comparison and benchmarking are commonly used management tools in other industries, including portions of the public sector. The integration of benchmarking into many of the Malcolm Baldrige National Quality Award's performance criteria speaks to its long-term value as a core business practice.

During the last half of the 1990s and the first decade of the 21st century, benchmarking has been a focus of European efforts to improve the quality of transit service delivered to customers. At least four international benchmarking networks, funded by member contributions, have been in existence for extended periods of time; their longevity also speaks to those agencies' belief in the value of benchmarking.

Despite having a significant advantage not available to transit agencies in most of the rest of the world—the existence of a database of relatively standardized transit data in the form of the NTD—the U.S. transit industry has been slow to adopt benchmarking as a business practice. The development of this report and the incorporation of its peer-grouping methodology into the FTIS software tool will hopefully remove a barrier to its adoption. However, additional work will be required for the U.S. public transit industry to fully realize benchmarking's potential. This issue is discussed next.

Key Findings and Conclusions

Transit Agencies

- **NTD data quality.** Each transit agency has a responsibility to make sure that its NTD data are reported consistent with the NTD definitions. Some transit agencies view NTD reporting as a burden, and this project's outreach efforts identified that industry confidence in NTD data quality was still lacking. However, as the Chapter 5 case studies show, agencies can quickly obtain a number of useful insights

about their own performance by using FTIS to work with NTD data. However, consistent data collection is essential for a transit agency's reporting effort to generate any value.

- **Performance improvement.** Benchmarking can validate an agency's strengths and may reveal opportunities for improvement; in either case, the approach from the start should be that the agency is committed to looking for (and implementing) ways to improve its performance. Many successful benchmarking applications start with the premise that everyone has room for improvement. A performance comparison should be a starting point for asking questions, and looking for ways to improve one's performance is preferable to an exercise where a performance "report card" is the final outcome.
- **First steps.** The FTIS tool makes it possible to quickly form a peer group and retrieve NTD-based performance data. A small-scale peer-grouping exercise focusing on key cost and productivity outcomes can demonstrate the value of performing peer comparisons and can help build internal agency support for a larger-scale, permanent effort.
- **Permanent internal performance-measurement program.** As demonstrated by some of the Chapter 5 case studies, an initial NTD-based performance comparison can lead to insights and questions that require more-detailed internal data. Having a structure set up to routinely collect, analyze, report, and store this information will both support a transit agency's day-to-day activities as well as support less-frequent benchmarking efforts. *TCRP Report 88 (1)* provides guidance on developing such a program. The more transit agencies there are that have good performance-measurement programs, the easier it will be for all to share and obtain non-NTD performance data.
- **Success stories.** Early adopters of benchmarking, such as members of the TFLEx benchmarking network in the United States and individual transit agency general managers who have incorporated peer comparisons into their agency activities, can spread the word about the tangible

(e.g., performance improvements) and less-tangible (e.g., staff professional growth) benefits of their work. Forums include professional conferences, committees, and workshops; the APTA General Manager's Workshop; and working groups of transit agencies that have decided to form a benchmarking network. Transit agencies that have implemented and use benchmarking and performance-measurement programs could also consider applying for the Malcolm Baldrige National Quality Award.

- **Benchmarking and performance-measurement champions.** Management support is necessary to provide the resources to begin a program, to protect those resources during tight financial times, and to implement programs designed to improve performance. Sharing past success stories, internal and external, with transit agency stakeholders (e.g., the agency board, lower-level management, and front-line employees) can help build long-term support for a program and actions that are developed as a result of having the program.

State and Regional Transportation and Funding Agencies

- **Local transit agency NTD data.** The public transit function within these agencies should be familiar with local transit agencies and should know whether a change in a performance trend is due to something that has changed locally or whether it is a sign of a possible data problem. Some states, such as Texas and Florida, contract with universities to check NTD data and provide training in areas where data problems occur. In addition, for those state DOTs that incorporate performance results into grant-allocation formulas, having a data-checking process will help in obtaining transit agency acceptance that the data used by the distribution process are reliable.
- **Training efforts.** If the state DOT's review of its transit agencies finds that many are lagging their out-of-state peers in particular areas, the state can use this information to develop training activities in those areas that will benefit a large number of agencies.
- **Transit agency benchmarking programs.** The North Carolina DOT, for example, has developed a benchmarking guidebook (34) for use by its state's transit agencies. This activity helps support the regional or state funder's goal of having its transit agencies serve riders efficiently and effectively and helps ensure that public money directly provided by the state is used responsibly. Funding agencies could consider providing incentives each year to local transit agencies that have developed and use such programs.
- **Annual reports on transit performance.** These reports can highlight performance-improvement success stories and the need for action in certain areas (such as dealing with aging infrastructure). These reports can also incorporate non-NTD measures that are of interest at a regional or

state level, providing an additional information source that benefits all. The Washington State DOT's annual public transportation report (35) is a good model.

- **Service area population and size values.** This research has shown the value of using per-capita performance measures and the desire of practitioners for reliable service area data. However, tracking regional population is not a normal transit agency function, and as a result the service area population and size values are not reported consistently to the NTD. MPOs, on the other hand, have the data and tools to readily perform these calculations.

Standards Development

- **Standard definitions for important non-NTD measures.** APTA serves as a standards development organization for the U.S. transit industry and is a logical organization for developing standard definitions of measures (such as those relating to transit reliability) that relate to customer quality of service and that would be useful to benchmark but are not available through the NTD. Such standards are more likely to be accepted when they are a result of a transit industry consensus.
- **Performance measurement and benchmarking as standard practices.** Defining standard practices for performance measurement and benchmarking that transit agencies can routinely undertake would elevate their prominence within the industry. TCRP research provides the tools for implementing such programs.

National Transit Database

- **Transit industry outreach.** Industry confidence in the quality of NTD data is crucial for obtaining support for conducting benchmarking efforts and taking actions based on their results—if transit agency management is not convinced of the NTD's data quality, they will not devote resources to an effort that relies on that data. This research's outreach efforts found that there is still considerable skepticism in the transit industry about the reliability of NTD data, while this research's testing efforts found that, for the most part, the data needed for an analysis were reliable and that what errors did exist were readily spotted.
- **NTD data quality.** NTD measures that most frequently had errors during the course of this research's testing efforts were *service area population*, *service area size*, *vehicle system failures*, and *route miles vs. track miles*. Each would be valuable for developing outcome or descriptive measures as part of a benchmarking effort if the data quality could be improved.

Future Steps

- **Pilot benchmarking network projects.** International experience shows that benchmarking networks can provide greater knowledge benefits and cost-sharing oppor-

tunities than when individual agencies conduct their own peer-comparison activities. However, that experience also shows that having an external organization to manage the data collection and analysis process and to facilitate the exchange of information within the network is an important long-term success factor. Pilot projects could:

- Recruit agencies of various sizes and/or modes operated that already have established performance-measurement programs to participate in a benchmarking network. Having a performance-measurement program already in place (a) demonstrates a transit agency’s commitment to performance measurement and (b) reduces the time required to begin benchmarking and demonstrate results.
- Fund a facilitator for each network for the first few years. Since several pilot European networks dissolved after the pilot funding ended, it would be important for pilot projects to seek ways to minimize costs, while still providing real benefits, to maximize the potential for the benchmarking network continuing on its own after the pilot period.

- Document success stories (i.e., tangible and intangible benefits realized) from the pilot networks to encourage greater use of benchmarking by others.

- **Confidential clearinghouse for performance data.** The literature review found that many organizations, both inside and outside the transit industry, are more willing to share data and practices when they are assured that the data will be kept confidential. The clearinghouse could contain standardized non-NTD measures of value to benchmarking efforts as well as more-detailed versions of NTD measures (e.g., summarized by service type—commuter bus, suburban bus, BRT—rather than by mode). Transit agency participation would be voluntary, but only contributors to the clearinghouse would be able to access data.

Accomplishment of Research Objectives

Table 27 lists the original objectives of the research and how they have been accomplished.

Table 27. Accomplishment of research objectives.

Research Objective	How the Objective Was Accomplished
The methodology should include performance measures composed of uniformly reported data that are as transparent as possible, credible, and relevant to the concerns of public transportation systems.	The peer-grouping methodology incorporates uniformly reported data from the NTD, the U.S. Census Bureau, and the <i>Urban Mobility Report (45)</i> , as well as data developed by this research. The performance-measurement component of the methodology incorporates uniformly reported inflation and labor-cost data from the U.S. Bureau of Labor Statistics and transit data from the NTD. Guidance is provided on obtaining uniform non-NTD data and checking for data consistency. For transparency, the inputs to the peer-grouping process are provided with the peer-grouping results in FTIS; the entire methodology is described in Appendix B. The methodology underwent two rounds of review with the project panel and industry stakeholders, followed by two rounds of real-world transit agency testing.
The peer comparison approach should enable performance assessments of public transportation systems of different sizes, operating environments, and modes.	The methodology accommodates assessments of any non-rural transit system operating any mode reported to the NTD. Performance can be compared for the agency as a whole or for an individual mode. Operating environment (service area type) is a factor used directly in the peer-grouping process. Guidance is provided on screening for other factors, such as the agency’s operating philosophy (e.g., coverage vs. efficiency).
The research should consider lessons learned from other industries and from international transit peer-comparison experience.	These lessons learned are summarized in Chapter 2 of this report.
The research should identify potential applications for the methodology and develop potential strategies for industry adoption of the methodology.	Applications are summarized in Chapter 3, while recommended strategies are provided in Chapter 6.
The methodology should be able to be applied not only by individual public transit agencies but also by state departments of transportation and other transit funding agencies.	Five state DOTs and the Chicago Regional Transportation Authority were included in the methodology testing. Two of the case studies in Chapter 5 demonstrate DOT applications.

References

1. Kittelson & Associates, Inc., Urbitrans, Inc., LKC Consulting Services, Inc., MORPACE International, Inc., Queensland University of Technology, and Yuko Nakanishi. *TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System*. Transportation Research Board of the National Academies, Washington, D.C., 2003.
2. Camp, Robert C. *The Search for Industry Best Practices That Lead to Superior Performance*. ASQ Quality Press, Milwaukee, Wis., 1989.
3. Watson, Gregory H. *Strategic Benchmarking Reloaded with Six Sigma*. John Wiley & Sons, Inc., Hoboken, N.J., 2007.
4. American Productivity & Quality Center. *What Is Benchmarking?* Houston, Tex. http://www.apqc.org/portal/apqc/ksn/BenchmarkingMethodology.pdf?paf_gear_id=contentgearhome&paf_dm=full&pageselect=contentitem&docid=108227, accessed September 22, 2009.
5. Hillkirk, John. *Xerox: American Samurai*. Macmillan, New York, 1986.
6. Bogan, Christopher E. and Michael J. English. *Benchmarking for Best Practice: Winning Through Innovative Adaptation*. McGraw-Hill, Inc., New York, 1994.
7. National Institute of Standards and Technology. Frequently Asked Questions about the Malcolm Baldrige National Quality Award. http://www.nist.gov/public_affairs/factsheet/baldafaqs.htm. Accessed September 22, 2009.
8. APQC. *Benchmarking Code of Conduct*. Houston, Tex. http://www.apqc.org/portal/apqc/ksn/Code_of_Conduct_electronic.pdf?paf_gear_id=contentgearhome&paf_dm=full&pageselect=contentitem&docid=119399. Accessed September 22, 2009.
9. McDonald, Phyllis P. *Managing Police Operations: Implementing the NYPD Crime Control Model Using COMPSTAT*. Wadsworth Publishing, Belmont, Calif., 2002.
10. City of Baltimore. CITISTAT website. www.ci.baltimore.md.us/news/citistat/index.html. Accessed September 22, 2009.
11. District of Columbia. CapStat website. http://capstat.oca.dc.gov/?portal_link=rt. Accessed September 22, 2009.
12. Patusky, Christopher, Leigh Botwinik, and Mary Shelley. *The Philadelphia SchoolStat Model*. IBM Center for the Business of Government, Washington, D.C., 2007.
13. Best Practices Benchmarking & Consulting, Inc. Public Sector Benchmarking: A Research Study. Interview with Carolyn Burstein. Federal Quality Institute, April 1993.
14. Railway and Transport Strategy Centre, Centre for Transport Studies, Imperial College London. CoMET website. www.comet-metros.org. Accessed September 22, 2009.
15. Railway and Transport Strategy Centre, Centre for Transport Studies, Imperial College London. Nova website. www.nova-metros.org. Accessed September 22, 2009.
16. Railway and Transport Strategy Centre, Centre for Transport Studies, Imperial College London. International Bus Benchmarking Group website. www.busbenchmarking.org. Accessed September 22, 2009.
17. Randall, E. R., B. J. Condry, and M. Trompet. *International Bus System Benchmarking: Performance Measurement Development, Challenges, and Lessons Learned*. Paper 07-3079. Presented at the 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
18. AB Storstockholms Lokaltrafik. *Benchmarking in European Service of Public Transport: Final Report, 2000–2004*. Stockholm, Sweden, 2004.
19. E-mail correspondence with Kjetil Vrenne, BEST Project Manager, Enable, Oslo, Norway, February 13, 2008.
20. Anderson, Richard. Metro Benchmarking Yields Tangible Benefits. *European Rail Outlook*, March 2006, pp. 22–25.
21. Citizens' Network Benchmarking Initiative. *Results of the Common Indicators, Statistical Indicators on Local and Regional Passenger Transport in 40 European Cities and Regions*. European Commission Directorate-General for Energy and Transport, Brussels, Belgium, 2002.
22. University of Newcastle upon Tyne, ASM Brescia SpA, Universität für Bodenkultur, Erasmus University, European Transport and Telematics Systems, and Viatek Ltd. *EQUIP Part II: The Indicators: Guide to Completion of the Handbook*. European Commission Directorate-General for Energy and Transport, Brussels, Belgium, 2000.
23. University of Newcastle upon Tyne, Viatek Ltd., ASM Brescia SpA, Universität für Bodenkultur, European Transport and Telematics Systems, and Erasmus University. *EQUIP Summary Report*. European Commission Directorate-General for Energy and Transport, Brussels, Belgium, 2000.
24. OGM. *Benchmarking European Sustainable Transport: Final Publishable Report*. Brussels, Belgium, July 29, 2003.
25. European Commission. *QUATTRO Final Report: Synthesis and Recommendations*. Brussels, Belgium, 1998.
26. European Committee for Standardisation. *European Standard EN 13816: Transportation–Logistics and Services–Public Passenger Transport–Service Quality Definition, Targeting and Measurement*. Brussels, Belgium, 2000.
27. Nielsen, Søren. *How Apples Can Learn from Pears*. Presented at the 4th Benchmarking European Sustainable Transport Conference, Brussels, Belgium, October 4, 2001.

28. Canadian Urban Transit Association. *Quick Facts on 2005 Conventional Transit Services*. Toronto, Ontario, 2005.
 29. UITP. *Mobility in Cities Database*. Brussels, Belgium, 2006.
 30. Booz Allen Hamilton, Inc. *Utah Transit Authority: FY01–FY05 Performance Audit*. Salt Lake City, Utah, December 2005.
 31. American Public Transportation Association. *Draft Standard for Comparison of Rail Transit Vehicle Reliability Using On-Time Performance*. APTA RT-SS-VIM-020-08, Washington, D.C., March 28, 2008.
 32. Transit Finance Learning Exchange website. www.tflex.org/about.asp. Accessed October 19, 2009.
 33. Texas Department of Transportation. *Implementation of Equity- and Performance-Based Allocation of Transit Funds in Texas*. Austin, Tex., August 2006.
 34. Institute for Transportation Research and Education. *Benchmarking Guidebook for North Carolina Public Transportation Systems*. North Carolina State University, Raleigh, N.C., June 2006.
 35. Washington State Department of Transportation, Public Transit Division. *Washington State Summary of Public Transportation—2006*. Olympia, Wash., September 2007.
 36. E-mail correspondence with Sharon Peerenboom, Small City and Rural Program Manager, Oregon Department of Transportation, Salem, Ore., April 28, 2006.
 37. Metropolitan Council. *2003 Transit System Performance Audit*. St. Paul, Minn., 2003.
 38. Atlanta Regional Council. *Regional Transit Institutional Analysis*. Atlanta, Ga., 2005.
 39. State of Illinois, Office of the Auditor General. *Performance Audit—Mass Transit Agencies of Northeastern Illinois: RTA, CTA, Metra, and Pace*. Springfield, Ill., March 2007.
 40. Hartgen, David T. and Mark W. Horner. *Transportation Publication Report 163: Comparative Performance of Major U.S. Bus Transit Systems: 1988–1995*. University of North Carolina at Charlotte, May 1997.
 41. Perk, Victoria and Nilgün Kamp. *Benchmark Rankings for Transit Systems in the United States*. National Center for Transit Research at the Center for Urban Transportation Research, University of South Florida, Tampa, Fla., December 2004.
 42. ICF International. *NCHRP Report 569: Comparative Review and Analysis of State Transit Funding Programs*. Transportation Research Board of the National Academies, Washington, D.C., 2006.
 43. MORPACE International, Inc. and Cambridge Systematics, Inc. *TCRP Report 47: A Handbook for Measuring Customer Satisfaction and Service Quality*. TRB, National Research Council, Washington, D.C., 1999.
 44. Perk, Victoria. *Working Paper 1a: Summary of Agency Outreach Efforts*. TCRP Project G-11. Center for Urban Transportation Research, University of South Florida, Tampa, Florida, March 25, 2008.
 45. Schrank, David and Tim Lomax. *2007 Urban Mobility Report*. Texas Transportation Institute, Texas A&M University System, College Station, Tex., September 2007.
 46. Ketola, H. N. and D. Chia. *TCRP Web Document 18: Developing Useful Transit-Related Crime and Incident Data*. TRB, National Resource Council, Washington, D.C., April 2000. http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_webdoc_18.pdf. Accessed November 4, 2009.
 47. Benchmarking in European Service of public Transport (BEST) website. <http://www.best2005.net>. Accessed October 29, 2009.
 48. U.S. Census Bureau. American Community Survey: About the Data (Methodology): Data Collection & Processing. <http://www.census.gov/acs/www/AdvMeth/CollProc/CollProc1.htm>. Accessed November 4, 2009.
 49. Federal Transit Administration. *National Transit Database 2009 Annual Reporting Manual*. Washington, D.C., 2009. http://www.ntdprogram.gov/ntdprogram/pubs/ARM/2009/html/2009_Reporting_Manual_Table_of_Contents.htm. Accessed November 4, 2009.
-

APPENDIX A

FTIS Instructions

Introduction

TCRP Project's G-11 peer-comparison and performance-measurement methodology has been incorporated into a Web-based software tool, the Integrated National Transit Database Analysis System (INTDAS) component of the FTIS. Despite its name, FTIS provides access to the complete publicly available portion of the NTD as well as to standardized national data added by TCRP Project G-11. Some data collected by the Federal Transit Administration for the NTD, such as safety and security data, are not released publicly and therefore are not available through FTIS.

One can obtain an initial peer grouping with FTIS with just a few clicks of a mouse, although more work is usually necessary to conduct a secondary screening that will narrow the list of potential peers to one appropriate for the particular peer-comparison application being conducted. Once a final peer group is set, FTIS can be used to quickly find and export a variety of NTD-based performance measures as well create tables and graphs of the data. During the methodology testing conducted by TCRP Project G-11, transit agencies were able to perform a peer comparison in 16 person-hours or less, including the time required to learn to use the software.

FTIS is sponsored by the Florida Department of Transportation's Public Transit Office and is maintained by Florida International University (FIU). It is freely available to the public; however, a one-time free registration is required to gain access. The main FTIS page is at www.ftis.org; INTDAS is accessed at <http://www.ftis.org/INTDAS/NTDLogin.aspx>.

These instructions provide a step-by-step description of how to use FTIS to form a peer group and obtain NTD performance data for the group. The instructions are intended to be used in combination with (a) the step-by-step description of the complete benchmarking methodology provided in Chapter 4 of this report, (b) online FTIS help, and (c) one's spreadsheet software's instructions. Screen shots shown in these instructions reflect the online version of FTIS as of October 2009; the

actual layout of screens may be somewhat different depending on any software updates that may have occurred since that time. In addition, screen content will vary depending on the user's selections.

- Users should be aware that FTIS automatically logs users out after a period of inactivity (as of November 2009, this was 15 minutes) to free up slots for other users. If a user is logged in and comes back after a period of inactivity, the user will be taken back to the log-in screen or, sometimes, a server error may occur. In either case, the user could lose some work, so saving the peer group one is working with and exporting analysis results on a regular basis is recommended.

Computer Requirements

FTIS is designed for the Internet Explorer browser, Version 6 or later. Other popular web browsers may also work, but are not supported by FIU staff. A screen resolution of 1152 by 864 pixels or greater is recommended.

A spreadsheet program is recommended for in-depth data analysis. FTIS data can be exported into several spreadsheet-readable formats.

FTIS Within the Overall Benchmarking Process

As described in Chapter 4, a complete transit benchmarking effort consists of eight steps. FTIS is used in Step 3: Establish a Peer Group and Step 4: Compare Performance. Before starting to use FTIS, the context of the benchmarking effort should be well-understood (Step 1) and a set of performance measures appropriate to the particular benchmarking application should have already been identified (Step 2). Figure A1 shows the five FTIS-related sub-steps described in

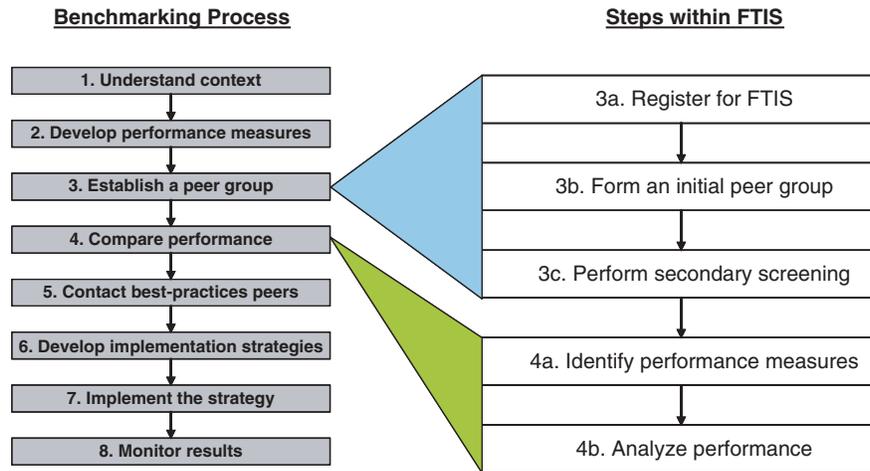


Figure A1. Relationship of FTIS-implemented steps to the overall benchmarking process.

these instructions and their relationship to the overall benchmarking process.

Step-By-Step Instructions

Step 3a: Register for FTIS

The INTDAS component of FTIS is accessed at <http://www.ftis.org/INTDAS/NTDLogin.aspx>. Its login screen is shown in Figure A2. The site is password-protected; one should enter one's password in the box shown with ❶ in Figure A2. If one is not already signed up as a user, request a free password by

clicking on the "Access Request Form" link ❷ just above the password box. Users typically receive a password within one business day.

Step 3b: Form an Initial Peer Group

Identify the Target Agency

After logging in, the INTDAS front page will appear, as shown in Figure A3.

Click the "Select Peers" tab ❶ at the top of the screen shown in Figure A3 to move to the "Peers" screen, shown in Figure A4.

Figure A2. INTDAS login screen.

Figure A3. INTDAS front page.

Enter the following settings on the “Peers” page shown in Figure A4:

- Select “TCRP G-11 Method” ①.
- Select an agency by first choosing the state ② and then the agency name ③ from the list of states. The number next to

each agency’s name is the ID number assigned to that agency by the NTD.

- Peer comparisons may be performed using agency-wide data (across all modes) or for a single specific mode. Select the desired type of comparison ④. If the comparison is to focus only on data for a single mode (e.g., light

Figure A4. Peer selection screen.

6 3 5

<< Back Excel Save Peers Select Groups 2

4	No.	NTD ID	Agency Name	Location	State	Total Likeness Score	Rail	Likeness Score	Rail Only	Likeness Score	Heavy Rail	Likeness Score	Urban Area Population	Likeness Score	Total Vehicle Miles Operated	Likeness Score	Total Operating Budget
1	<input checked="" type="checkbox"/>	0	7 Lane Transit District	Eugene	OR	0	No	0	No	0	No	0	236,543	0	4,091,403	0	\$29,543,481
	<input checked="" type="checkbox"/>	1	25 Salem Area Mass Transit District	Salem	OR	0.40	No	0.00	No	0.00	No	0.00	225,188	0.05	3,873,992	0.06	\$22,673,924
	<input checked="" type="checkbox"/>	2	2 Spokane Transit Authority	Spokane	WA	0.47	No	0.00	No	0.00	No	0.00	343,291	0.45	8,729,676	1.13	\$45,618,425
	<input checked="" type="checkbox"/>	3	9027 Fresno Area Express	Fresno	CA	0.52	No	0.00	No	0.00	No	0.00	614,477	1.60	5,211,560	0.27	\$35,541,343
	<input checked="" type="checkbox"/>	4	12 Municipality of Anchorage - Public Transportation Department	Anchorage	AK	0.64	No	0.00	No	0.00	No	0.00	242,621	0.03	3,157,728	0.30	\$22,778,276
	<input checked="" type="checkbox"/>	5	6051 Corpus Christi Regional Transportation Authority	Corpus Christi	TX	0.65	No	0.00	No	0.00	No	0.00	297,698	0.26	4,048,382	0.01	\$18,766,534
	<input checked="" type="checkbox"/>	6	9004 Golden Empire Transit District	Bakersfield	CA	0.71	No	0.00	No	0.00	No	0.00	457,984	0.94	3,922,291	0.04	\$18,428,809
	<input checked="" type="checkbox"/>	7	9078 Central Contra Costa Transit Authority	Concord	CA	0.74	No	0.00	No	0.00	No	0.00	578,168	1.44	4,383,852	0.07	\$28,020,735
	<input checked="" type="checkbox"/>	8	9079 SunLine Transit Agency	Thousand Palms	CA	0.74	No	0.00	No	0.00	No	0.00	294,187	0.24	2,413,751	0.70	\$18,487,875
	<input type="checkbox"/>	9	6019 ABQ Ride	Albuquerque	NM	0.76	No	0.00	No	0.00	No	0.00	671,413	1.84	6,486,822	0.59	\$35,682,431
	<input type="checkbox"/>	10	7002 Transit Authority of Omaha	Omaha	NE	0.76	No	0.00	No	0.00	No	0.00	651,814	1.76	4,279,299	0.05	\$19,239,981
	<input type="checkbox"/>	11	6006 Mass Transit Department - City of El Paso	El Paso	TX	0.77	No	0.00	No	0.00	No	0.00	698,227	1.95	8,687,902	1.12	\$40,640,177
	<input type="checkbox"/>	12	5033 Interurban Transit Partnership	Grand Rapids	MI	0.82	No	0.00	No	0.00	No	0.00	551,855	1.33	6,294,044	0.54	\$28,558,706

Figure A5. Candidate peer group screen.

rail), then select that mode from the adjacent drop-down menu 5.

- Select the data year used for the peer-grouping variables. Typically, the most recent year available would be selected 6.
- For most applications, check the “American Community Survey” option for urban area population 7. However, if a forward-looking application is desired, where an agency is compared against larger peers to see where it might be in the future, a population can also be manually entered in the box provided 8.
- Click the “Find Peers” button 9 to open a new page (Figure A5) containing a table showing the results of the peer grouping, including the variables and data used to develop the peer grouping.

Select Potential Peers

The selected (“target”) agency is listed in the top row of the table shown in Figure A5 1, with other agencies shown sorted by their “total likeness score” 2 indicating their level of similarity with the target agency. Lower values indicate a greater level of similarity with the target agency.

This listing can be opened in Microsoft Excel format (whether or not Excel is installed on the computer), and then saved in a variety of spreadsheet-readable formats. To do so, click the “Excel” button 3 at the top of the page, select “Save As” from the “File” menu, name the export file, and select the file type to be exported. Performing this step allows digging deeper later on into the reasons why a particular agency was (or wasn’t) highly ranked as a potential peer, without having

to re-create the process in FTIS. The Excel file can also be used to document the peer-grouping process.

Optionally, use the checkboxes in the left-hand column 4 to select agencies to include in the peer group. To save a peer group, click the “Save Peer Group” button at the top of the page 5.

A pop-window will appear, as shown in Figure A6. Either a user-defined number of top peers can be saved to the group (select button 1), or just the peers that were previously checked (select button 2). Provide a name for the peer group 3 to identify it within FTIS. Finally, click the “Save” button to close the pop-up window 4.

- A minimum of four peers is recommended for an analysis, with 8–10 being a good upper number of peers to end up with.
- The “total likeness score” indicates how similar a potential peer agency is to the target agency. A score ≤0.50 indicates a very good match, while a score of between 0.50 and 0.75 indicates a reasonably good match. Agencies with scores

1 Save top 5 peers as a peer group

2 Save all checked systems as a peer group.

3 Group Name: FTIS example

4 Save Cancel

Figure A6. Save peer group window.

greater than 0.75 may still be acceptable matches but should be investigated more carefully as they may have significant differences in some areas that may make them unsuitable for a particular benchmarking application.

- Agencies in Alaska, Hawaii, and Puerto Rico will have relatively few peer agencies in their immediate vicinity, so their potential peers' likeness scores will tend to be higher simply because they are located further away. The "total likeness score" ranges described above may need to be expanded for agencies in these regions.
- Try to save more than 8–10 peers at this step if a secondary screening will be performed in the next step, in order to allow for some potential peers being filtered out during the secondary screening. However, if potential peers' total likeness scores are too high (e.g., higher than 1.00), indicating a great deal of dissimilarity with the target agency, it may not be possible to end up with the ideal number of peers. This is more likely to occur with very large agencies than with smaller agencies.

Click the "<< Back" button (shown as 6 in Figure A5) to return to the "Peers" page (don't use the browser's "Back" button).

Select Agencies

Select the "Select Groups" tab, shown as 1 in Figure A7, where the just-saved group will now appear in the list of system groups 2. (If one has never used FTIS before, it will be the only group listed.)

Click the desired group's name to select it. A list of the agencies in the group will appear in the "Select Systems" box in the lower-left corner of the window 3. Select all of the agencies by clicking the "All" 4 button; they will be copied over to the "Selected Systems" window 5.

Step 3c: Performing Secondary Screening

Overview

Some performance questions may require looking at a narrower set of potential peers than the TCRP Project G-11 methodology produces. For example, one case study described in this report involved an agency that did not have a dedicated local funding source and was interested in comparing itself to peers that did have that source of funding. Another case study involved an agency in a region that was about to reach 200,000 population (thus moving into a different funding category) and wanted to compare itself to peers that were already at 200,000 population or more. Some agencies may simply want to make sure that no peer agency is "too different" to be a potential peer for a particular application. Data contained in FTIS can often be used to perform these kinds of screenings. Other kinds of screening, for example based on agency policy or types of routes operated (e.g., commuter bus or BRT), will require Internet searches or agency contacts to obtain the information. Any desired screening factors should have already been determined during Step 2.

The general process to follow is to first identify how many peers would ideally end up in the peer group. For the sake

Figure A7. Select groups screen.

of this example, this number will be eight. Starting with the highest-ranked potential peer (i.e., the one with the lowest total likeness score), check whether the agency meets the secondary screening criteria. If the agency does not meet the criteria, replace it with the next available agency in the list that meets the screening criteria. For example, if the #1-ranked potential peer does not meet the criteria, check the #9-ranked agency next, then #10, and so on, until an agency is found that meets the criteria. Repeat the process with the #2-ranked potential peer. Continue until a group of eight peers that meets the secondary screening criteria is formed, or until a potential peer’s total likeness score becomes too high (e.g., is 1.00 or higher).

Using NTD Forms for Screening

The NTD forms are a quick way to check screening criteria that use NTD data. From the “Select Groups” form illustrated in Figure A7, first change the years setting ⑥ to show just one year of data (e.g., 2007). Next, click the “Forms” button ⑦ to call up the NTD forms for that year for each agency in the initial peer group, as illustrated in Figure A8.

Use the tabs ① at the top of the screen (Figure A8) to switch between forms. The navigation buttons ② ③ will scroll along with the form. The “Next” and “Last” buttons ② let one move between agency forms. More than one copy of certain forms may exist for a given year and provider (for example, forms describing providers of contracting service and the fleet composition) and forms will be provided for multiple years if more than one year was selected in the “Select Groups” window). Use the “Close” button ③ to return to the “Select Groups” window.

The following lists the available forms and common screening factors available on those forms. Note that data on certain forms are not released by the FTA.

- **B-10, Transit Agency Identification:** Organization type, institutional structure, vehicles operated in maximum service, service area size and population (often not reported consistently between agencies).
- **B-30, Contractual Relationship:** (One form per contractor.) Type of relationship with the reporting agency, contracted vehicles operated in maximum service, contract costs and revenues, demand-response provider type.
- **F-10, Sources of Funds—Funds Expended and Funds Earned:** Modes operated, revenues by source, funds expended on operations and capital.
- **F-20, Use of Capital:** Capital funds expended by mode, type of capital expense, and purpose (existing service vs. expansion of service).
- **F-30, Operating Expense:** (One form per mode and service type.) Operating funds expended by function (operations, vehicle maintenance, non-vehicle maintenance, and general administration) and expense class.
- **F-40, Operating Expense:** Similar to F-30, but reporting agency totals.
- **F-50, Operators’ Wages:** Platform time, straight time, premium time, and non-operating time in dollars and hours. (Only agencies with 150 or more directly operated vehicles in maximum annual service, excluding demand-response and vanpool vehicles, are required to report this information,

Sources of Funds-Funds Expended and Funds Earned (F-10)

Year NTD ID Location State

Sources of Directly Generated Funds	Mode Code	Service Type	Funds Earned During Period	Funds Expended on Operations	Funds Expended on Capital
01. Passenger Fares	VP	DO	\$378,692		
	DR	PT	\$50,935		
	MB	DO	\$6,082,747		
	DR	DO	\$154,149		
02. Total All Directly Operated Modes			\$6,615,588	\$6,615,588	
03. Total all Purchased Transportation Modes			\$50,935	\$50,935	
04. Park and Ride Parking Revenue					
05. Other Transportation Revenue					

Figure A8. Sample forms window.

and FTA stopped releasing these data beginning with the 2008 reporting year.)

- **A-10, Stations and Maintenance Facilities:** Number of stations by ADA accessibility (yes/no), number of maintenance facilities by size and type (owned vs. leased).
- **A-20, Transit Way Mileage:** Rail miles of track and number of grade crossings by right-of-way type, non-rail miles of exclusive right-of-way.
- **A-30, Revenue Vehicle Inventory:** (One form per fleet.) Number of vehicles in fleet, fleet age, average mileage, standing capacity, ADA features.
- **S-10, Transit Agency Service:** (One form per mode and service type) Vehicles operated and available for maximum service; service start/end times, vehicle miles and hours, revenue miles and hours, and ridership by average weekday/Saturday/Sunday; number of vehicles operated A.M. peak/midday/P.M. peak/other.
- **R-10, Employees:** (One form per mode and service type) Full- and part-time employees by function.
- **R-20, Maintenance Performance:** (One form per mode and service type) Number of major and other mechanical failures, labor hours for maintenance and inspection.
- **R-30, Energy Consumption:** (One form per mode and service type) Amount of energy consumed by fuel/power type.

Other forms shown in the window either summarize data from other forms or are used for information (e.g., safety and security information) that is not released by the FTA (therefore, the forms appear blank in FTIS).

For example, someone wanting to screen for peers that use a dedicated local sales tax as a funding source could go to form F-10, Sources of Funds: Funds Expended and Funds Earned. Scrolling down the window shown in Figure A9, one would discover that the agency used for this example does receive local sales tax funding.

Edit the Peer Group

Once a final peer group has been established, go back to the “Select Peers” tab (Figure A4) and re-enter the same information entered previously. After the “Candidate Peer Group” screen appears (Figure A5), check the boxes ❹ that correspond with the final peer group members and click “Save Peer Group” ❺. In the pop-up window (Figure A6), choose to save the checked systems as a peer group ❷ and give the peer group a new name (one not used previously) ❸. Save the peer group and go back to the “Groups” window (Figure A7) as before. Follow the same steps as before to load the new saved peer group from the list of groups and copy the agencies in the group over to the “Selected Systems” window.

Step 4a: Identify Performance Measures

Specify Analysis Years, Modes, and Service Types

Once the final peer group has been loaded, the remaining options in the “Select Groups” window filter the data that will be retrieved during the next step.

15.	Contra account for expenses				
16.	Net Contributed Services				
17.	Subsidy from Other Sectors of Operations				
18.	Bond and Loan Payments				
19.	Funds Dedicated to Transit at their Source				
		Rate			
20.	Dedicated Taxes: Income taxes				
21.	Sales taxes		\$47,198,519	\$33,084,166	\$8,932,695
22.	Property taxes				
23.	Gasoline taxes				
24.	Other taxes*				
Describe:					
<input type="text"/>					
25.	Bridge, tunnel and highway Tolls				
26.	High occupancy tolls				
27.	Other dedicated funds				
28.	Total Funds Dedicated to Transit at Their Source		\$47,198,519	\$33,084,166	\$8,932,695
29.	Other Directly Generated Funds				
30.	Total Directly Generated Funds		\$57,285,599	\$39,750,689	\$8,932,695
31.	Bond and Loan Payments				

First

Prev

Next ❷

Last

Close ❸

1

of

9

⬆ Top

Figure A9. Dedicated tax information on NTD form F-10.

The screenshot shows the FTIS software interface with the following sections:

- Select Years (1):** From: 2003, To: 2007. Includes a "Delete" button.
- Select Individual Modes (2):** A list of modes including "All individual modes", AG Automated Guideway, CC Cable Car, CR Commuter Rail, DR Demand Response, FB Ferryboat, HR Heavy/Rapid Rail, IP Inclined Plane, JT Jitney, LR Light Rail, MB Motorbus, MO Monorail, OR Other, PB Publico, TB Trolleybus, and TR Aerial Tramway.
- Select Service Types (3):** Checkboxes for DO: Directly Operated, PT: Purchased Transportation, and DP: DO and PT combined (for Florida Standard Variables only).
- Select Aggregate Modes (for Florida Standard Variables only) (4):** Checkboxes for ST: Systemwide Total, FT: Fixed-route Total (all except DR), RT: Rail Total (AG, CC, CR, HR, IP, LR, MO), and NT: Non-rail Total (DR, FB, JT, MB, PB, TB, TR, VP, OR).
- Select Systems (5):** A list of systems including OR 7 Lane Transit District, WA 19 Intercity Transit, WA 20 Kitsap Transit, WA 21 Whatcom Transportation Authority, OR 25 Salem Area Mass Transit District, NV 9001 Regional Transportation Commission of Washoe County, CA 9012 San Joaquin Regional Transit District, and CA 9027 Fresno Area Express.
- Selected Systems (5):** A list of selected systems including WA 2 Spokane Transit Authority, OR 7 Lane Transit District, WA 19 Intercity Transit, WA 20 Kitsap Transit, WA 21 Whatcom Transportation Authority, OR 25 Salem Area Mass Transit District, and CA 9012 San Joaquin Regional Transit District.

Buttons at the bottom include "Forms", "Reports", "Tables >>", "Delete", and "Clear".

Figure A10. Process for selecting analysis years, modes, and service types.

Select the years to be used for the analysis, using the pull-down menus (1 in Figure A10) located above the list of groups. A 5-year period is recommended for a trend analysis, but any combination of consecutive years may be used.

- Full-year NTD data may not be available yet for the most recent years shown in the drop-down lists, as it takes the FTA some time to process the data after it is submitted. The INTDAS front page shows the most recent year for which full-year data are available.

If the peer group was created using agency-wide data, select “[All Individual Modes]” from the “Select Individual Modes” box (2). Otherwise, select the specific mode used to create the peer group.

The area in the upper-right (3) portion of the window specifies the service types to use in the analysis. The NTD distinguishes between service directly operated (DO) by an agency and service purchased from another provider (PT). For certain common performance measures and ratios—called “Florida Standard Variables” by FTIS—the software offers a third option, DP, which provides agency totals combining directly operated and purchased service.

- Most performance-measurement applications use performance measures based on ratios (e.g., passengers per revenue hour, cost per passenger, etc.). The NTD only collects and reports individual measures (e.g., passengers, revenue hours, operating costs). To make life easier for users, FTIS calculates a variety of common performance ratios as part of the

“Florida Standard Variables,” based on the raw NTD data. A list of the Florida Standard Variables is provided at the end of these instructions.

- The boxes to check in this section depend on the application, as some analyses may need to distinguish between directly operated and purchased service. If there is no need to distinguish by service type, and all performance measures to be used in the analysis are included in the Florida Standard Variables, then simply check DP. If there is a need to distinguish by service type, then check the DO and/or PT boxes, depending on the analysis needs. Otherwise, check both the DO and PT boxes; later on, one will need to manually add the DO and PT values in a spreadsheet to create an agency-wide value.

The center-right (4) portion of the window is used to specify how FTIS should aggregate values by type of mode. FTIS will always provide values for each mode specified in box (2), for each service type specified in area (3). One can optionally also obtain system-wide totals (ST), totals for all fixed-route modes except demand response (FT), rail-mode totals (RT), and non-rail-mode totals (NT). The abbreviations shown next to the aggregation options indicate which modes are included in the aggregation; the mode codes are the same as those used by the NTD and are also shown in box (2).

- Lists of common abbreviations used by FTIS are provided at the end of these instructions.

Click the “Tables >>” button (5) to proceed to the next screen.

Specify Performance Measures

FTIS offers several options for specifying the performance measures to be used in an analysis. These options appear on the screen illustrated in Figure A11.

In the upper-left corner ❶, NTD measures can be selected directly by scrolling through a list or by searching for text used in a measure's name. Measures are sorted by the NTD form they come from and the order in which they are entered on the form.

In the center-left section ❷, NTD measures can be selected from NTD forms. Click on a form name and then check the desired variable names from the form(s) they appear in. The list of forms provided in the "Using NTD Forms for Screening" section can serve as a guide for determining where to find a particular measure.

In the lower-left section ❸, pre-selected groups of Florida Standard Variables can be selected. Click the "see definitions" link above box ❹ to see which measures are included in each group, or refer to the list at the end of these instructions. Any user-saved groups of measures will also appear in this section. (The process for saving groups of measures is described later.)

- The "TCRP Project G-11 Variables" option selects all of the additional variables added to the FTIS database by the TCRP Project G-11. Most of these TCRP Project G-11 variables describe system characteristics and were used in the peer grouping step. They are not intended to be a recommended group of measures for comparing performance, but can provide useful descriptive information to help in-

terpret the results in a later step. The TCRP Project G-11 variables can also be selected individually using box ❺.

- Use the browser's "Back" button to return to this window after clicking the "see definitions" link, to avoid accidentally closing FTIS.

In the upper-right section, the Florida Standard Variables ❹ includes a set of commonly used measures derived from the NTD. Look here for pre-calculated performance ratios.

- If a desired performance ratio is not part of the Florida Standard Variables, select the components of the ratio using one of the other boxes and manually calculate the ratio later on in a spreadsheet.
- Use of the Florida Standard Variables' per-capita and per-square-mile ratios is not currently recommended, as this information is not yet reported consistently to the NTD by transit agencies.
- Standardized urbanized area population and size values from the Census Bureau can be used to calculate per-capita and per-square-mile ratios; these are available in box ❺.

The center-right section ❻ contains non-NTD measures added to FTIS by TCRP Project G-11. As mentioned above, most of these measures are used for peer grouping, but "urban area" and "urban area population" are useful for creating per-capita and per-square-mile ratios, and "mean wage rate" can be used later in the analysis to manually adjust cost data, if desired.

Select measures by clicking on them (or by checking the box in the form, if the "Forms" box was used). The selected

❶ Select Original NTD Variables by List:

Search:

- [F10].[Other Federal Funds Earned During Period]
- [F10].[Other Federal Funds Expended on Operations]
- [F10].[Other Federal Funds Expended on Capital]
- [F10].[Other Federal Funds Desc]
- [F10].[Total Federal Funds Earned During Period]
- [F10].[Total Federal Funds Expended on Operations]
- [F10].[Total Federal Funds Expended on Capital]
- [F10].[Federal Bond & Loan Payments Funds Expended on Oper
- [F10].[Federal Bond & Loan Payments Funds Expended on Capit
- [F10].[Total Funds Earned During Period (Summary)]
- [F10].[Total Funds Expended on Operations (Summary)]

❷ Select Original NTD Variables by NTD Forms:

B10 B20 B30 F10 F30 F40 F50 A10 A20
A30 S10 R10 R20 R30 FFA10 GPRA10 405-1 405-2

❸ Select Variable Groups:

- Florida General Performance Indicators
- Florida Effectiveness Measures
- Florida Efficiency Measures
- TCRP Project G-11 Variables
- Monthly Unlinked Passenger Trips
- Monthly Vehicle Revenue Miles
- Monthly Vehicle Revenue Hours

❹ Select Florida Standard Variables by List: ([see definitions](#))

Search:

- [FSV].[Revenue Hours]
- [FSV].[Revenue Hours Per Employee FTE]
- [FSV].[Revenue Hours Per Total Vehicles]
- [FSV].[Revenue Miles]
- [FSV].[Revenue Miles Between Failures]
- [FSV].[Revenue Miles Between Incidents]
- [FSV].[Revenue Miles Per Route Mile]
- [FSV].[Revenue Miles Per Total Vehicles]
- [FSV].[Revenue Miles Per Vehicle Mile]
- [FSV].[Route Miles]
- [FSV].[Route Miles Per Square Mile of Service Area]

❺ Select TCRP Project G-11 Variables:

- [TCRP-G11].[Census UZA]
- [TCRP-G11].[Primary UZA]
- [TCRP-G11].[Urban Area (in square miles)]
- [TCRP-G11].[Urban Area Population]

❻ Selected Variables: [9]

- [FGV].[Arebox Recovery (%)]
- [FSV].[Spare Ratio (%)]
- [FSV].[Passenger Trips]
- [FSV].[Passenger Miles]
- [FSV].[Total Operating Expense]
- [TCRP-G11].[Urban Area Population]
- [F10].[Total Funds Expended on Operations (Summary)]

<< Back

Figure A11. Performance measure selection screen.

measures will appear in the box in the lower-right corner of the screen ⑥.

It is recommended that one save one’s list of variables as a group, to save time the next time one uses FTIS, or in case one is automatically logged out from FTIS due to inactivity. To do so, click the “Save” button ⑦ above the “Selected Variables” section of the screen. To continue, click the “Tables” button ⑧ in the lower-right corner of the screen.

Data Retrieval

After Step 4a is completed, a new window (Figure A12) will open that contains a table of performance measure values for each of the selected agencies for each of the selected years. Each row of the table represents data for one year, one agency, one mode (or aggregation of modes), and one service type (or aggregation of service types). For example, row 1 in Figure A12 lists selected data for Intercity Transit’s DO demand-response (DR) service for the year 2003.

- Mode and service type abbreviations are the same as those shown on the previous screen. A reference list of these abbreviations is provided at the end of these instructions.

The darker blue columns in Figure A12 specify the year, agency name, location, mode, and service type associated with the performance data. The lighter blue columns provide the results for each of the measures selected in the previous window.

- Reported values for measures shown with an asterisk (*) in the top row, such as the “Total Funds Expended on Operations (Summary)” measure shown in the example on the previous page, are totals for the agency, even if a specific mode and service type is shown for the row. The NTD only collects system-wide data for those measures; therefore, mode- and service-specific data are not available.

Click the “Excel” button ① to open a window with the data in Excel format, and then select “Save As” from the “File” menu. Enter a name for the file and select the appropriate file type for the spreadsheet that will be used to analyze the data.

- Other buttons at the top of the page are used for creating cross-tables, simple charts, summations, regressions, and summary statistics. It is also possible to adjust cost data for inflation and sort the data in different ways. Most of these buttons are self-explanatory and duplicate functions available in a spreadsheet and so are not covered in these instructions.
- Another quick-summary option is the “Reports” button in either of the previous two windows. This button opens a new window that can produce quick-summary reports, by mode and service type, for a default set of performance measures for the peer group.

Organize Data in the Spreadsheet

The ultimate goal of this sub-step is to create a two-dimensional table for each performance measure, with (for

1

(*): Values are totals for the entire transit agency.

No.	Year	NTD ID	Company Name	Location	State	Mode	Service	Total Funds Expended on Operations (Summary) (*)	Farebox Recovery (%)	Passenger Miles	Passenger Trips	Revenue Hours	Revenue per Hour
1	2003	19	Intercity Transit	Olympia	WA	DR	DO	\$14,017,683	1.65	609,311	109,587	41,937	
2	2003	19	Intercity Transit	Olympia	WA	DR	DP	\$14,017,683	1.65	609,311	109,587	41,937	
3	2003	19	Intercity Transit	Olympia	WA	MB	DO	\$14,017,683	13.05	9,273,413	2,586,011	128,339	1
4	2003	19	Intercity Transit	Olympia	WA	MB	DP	\$14,017,683	13.05	9,273,413	2,586,011	128,339	1
5	2003	19	Intercity Transit	Olympia	WA	ST	DO	\$14,017,683	12.93	19,394,791	2,959,522	201,451	3
6	2003	19	Intercity Transit	Olympia	WA	ST	DP	\$14,017,683	12.93	19,394,791	2,959,522	201,451	3
7	2003	19	Intercity Transit	Olympia	WA	VP	DO	\$14,017,683	100.41	9,512,067	263,924	31,175	1
8	2003	19	Intercity Transit	Olympia	WA	VP	DP	\$14,017,683	100.41	9,512,067	263,924	31,175	1
9	2003	2	Spokane Transit Authority	Spokane	WA	DR	DO	\$35,453,570	3.03	2,134,361	288,434	95,167	1
10	2003	2	Spokane Transit Authority	Spokane	WA	DR	DP	\$35,453,570	2.27	3,727,667	454,503	159,421	2
11	2003	2	Spokane Transit Authority	Spokane	WA	DR	PT	\$35,453,570	0.85	1,593,306	166,069	64,254	1
12	2003	2	Spokane Transit Authority	Spokane	WA	MB	DO	\$35,453,570	19.55	31,058,934	7,504,713	351,239	4
13	2003	2	Spokane Transit Authority	Spokane	WA	MB	DP	\$35,453,570	19.55	31,058,934	7,504,713	351,239	4
14	2003	2	Spokane Transit Authority	Spokane	WA	ST	DO	\$35,453,570	17.16	35,649,672	7,895,573	456,740	6
15	2003	2	Spokane Transit Authority	Spokane	WA	ST	DP	\$35,453,570	15.83	37,242,978	8,061,642	520,994	7
16	2003	2	Spokane Transit Authority	Spokane	WA	ST	PT	\$35,453,570	0.85	1,593,306	166,069	64,254	1
17	2003	2	Spokane Transit Authority	Spokane	WA	VP	DO	\$35,453,570	87.75	2,456,377	102,426	10,334	

Figure A12. Example performance measure results screen.

example) the agencies as the rows and the years as the columns. Keeping the exported data on one spreadsheet tab, while creating new separate spreadsheet tabs for each performance measure, is a good way to organize the data. Start by sorting the exported data by mode code, agency name, and year. Next, follow the steps below to pick out the rows of interest from the exported data, and then copy the performance measure values from those rows to the corresponding spreadsheet tab.

- The following steps assume that values for all service provided by an agency for a given mode are desired; however, a similar process can be followed if a peer-comparison application needs to distinguish between directly operated and purchased service.

If a measure is a Florida Standard Variable, find the data in the exported database as follows:

1. For an agency-wide comparison, copy values in rows with a mode of “ST” and a service of “DP.”
2. For a mode-specific comparison, copy values in rows with the corresponding mode code (e.g., “MB” for motorbus, “LR” for light rail) and a service of “DP.”

If a measure is not a Florida Standard Variable, the process is similar, but requires more steps:

1. For an agency-wide comparison, copy values in all rows for a given combination of agency and year.
2. For a mode-specific comparison, copy values in all rows with the corresponding mode code (e.g., “MB” for motorbus, “LR” for light rail).
3. After completing steps 1 and 2 for all combinations of years and agencies, the measure’s corresponding spreadsheet tab will contain a subset of the database, containing just the values for that measure, plus its identifying data (year, agency, etc.).
4. On a separate part of the tab, enter the agency names as a series of rows and enter the analysis years as a series of columns, as if they were row and column heads for a table.
5. Sum the values in the tab’s database that correspond to each combination of year and agency in the table.
 - For measures that only report system-wide values, do not sum the values. Instead, copy any of the (exactly the same) system-wide values for the combination of year and agency.
 - Advanced spreadsheet users can use a spreadsheet’s database or pivot table functions to achieve the same result without having to manually select the cells to be summed. Once selection criteria have been developed for one measure, they can be easily adapted for all other measures.

If the desired measure is a ratio of two other measures, and is not available as a Florida Standard Variable, follow the

above process for the ratio’s two component measures and then create a third spreadsheet tab to hold the summary table for the ratio. Complete this tab’s summary table by dividing the corresponding summary table value for the first performance measure by the corresponding summary table value for the second performance measure.

If desired, normalize cost data (see the Step 4a description in the body of the report) and perform any other desired supplemental calculations, such as computing a peer-group average for each year for each measure.

Step 4b: Analyze Performance

Data Checking

At this point, it is useful to create graphs for each measure to check for potential data problems such as unusually high or low values for a given agency’s performance measure for a given year or values that bounce up and down with no apparent trend. Examples of this process are described in the body of the report, as part of the Step 4b description.

Data Interpretation

The process of data interpretation is described in detail in the body of the report as part of the Step 4 description. FTIS can be used as part of the process to call up descriptive data about a particular agency that might explain particular performance. The Step 4b description includes an example of using FTIS data to explain why a particular peer agency appears to perform well in comparison to its peers in a number of areas, but not for one particular measure.

Results Presentation

The production of graphs and tables for presentation purposes will probably be done outside of FTIS. However, FTIS’ graphing and reporting features are valuable for quickly generating summary results for internal use. Results presentation is covered in more detail in the body of the report as part of the Step 4c description.

List of Common FTIS Abbreviations

Table A1 provides a list of common abbreviations that appear on FTIS screens and forms.

List of Florida Standard Variables

This list (Table A2) is current as of early 2010; consult the FTIS online list for the most up-to-date information and to identify the specific NTD measures used to calculate performance ratios.

Table A1. Common FTIS abbreviations.

Modes		Service Types	Data Aggregations
AG automated guideway	MB motorbus	DO directly operated	ST systemwide total
CC cable car	MO monorail	PT purchased transportation	FT fixed-route total (except DR)
CR commuter rail	OR other	DP both DO & PT	RT rail total
DR demand response	PB publico		NT non-rail total
FB ferryboat	TB trolleybus		
HR heavy/rapid rail	TR aerial tramway		
IP inclined plane	VP vanpool		
JT jitney	AR Alaska Railroad		
LR light rail			

Table A2. Florida Standard Variables.

General Performance Indicators	Effectiveness Measures	Efficiency Measures
SERVICE AREA DESCRIPTORS	SERVICE SUPPLY	COST EFFICIENCY
Service area population	Vehicle miles per capita	Operating expense per capita
Service area size		Operating expense per peak vehicle
	SERVICE CONSUMPTION	Op. expense per passenger trip
USAGE	Passenger trips per capita	Op. expense per passenger mile
Passenger trips	Passenger trips per revenue mile	Operating expense per revenue mile
Passenger miles	Passenger trips per revenue hour	Operating expense per revenue hour
	Average trip length	Maintenance exp. per revenue mile
SERVICE		Maintenance exp. per operating exp.
Vehicle miles	QUALITY OF SERVICE	
Revenue miles	Average speed	OPERATING RATIOS
Vehicle hours	Average headway	Farebox recovery
Revenue hours	Average age of fleet	Local revenue per operating expense
Route miles	Number of incidents	Operating revenue per op. expense
	Number of vehicle system failures	
EXPENSES	Revenue miles between failures	VEHICLE UTILIZATION
Total operating expense		Vehicle miles per peak vehicle
Total maintenance expense	AVAILABILITY	Vehicle hours per peak vehicle
Total capital expense	Revenue miles per route miles	Revenue miles per vehicle mile
	Weekday span of service	Revenue miles per vehicle
REVENUE	Route miles per square mile	Revenue hours per vehicle
Federal revenue		
State revenue		LABOR PRODUCTIVITY
Local revenue		Revenue hours per employee
		Passenger trips per employee
EMPLOYEES		
Total employees		ENERGY UTILIZATION
Transportation operating employees		Vehicle miles per gallon
Administrative employees		Vehicle miles per kilowatt-hour
VEHICLES		
Vehicles available for max. service		FARE
Vehicles operated in max. service		Average fare
Spare ratio		
ENERGY CONSUMPTION		
Total gallons consumed		
Total energy consumed		

APPENDIX B

Peer-Grouping Methodology Details

Introduction

This appendix presents the details of the peer-grouping methodology developed and tested by TCRP Project G-11. A summary version of the methodology appears in the body of the report as part of the description of Step 3 in Chapter 3.

The peer-grouping methodology was developed in conjunction with the project's broader benchmarking methodology. The general process used to develop the peer-grouping methodology was as follows:

- Prior to the start of the project, the oversight panel for TCRP Project G-11 specified their desired key characteristics for the peer-grouping methodology.
- The research team conducted outreach to the transit industry on the industry's desired aspects of a benchmarking methodology and reviewed the literature to determine what methodologies had been tried before.
- The research team developed initial concepts for the peer-grouping and benchmarking methodologies, conducted internal tests on the reasonableness of the results, and presented the concept to the panel for comment.
- Based on the panel's feedback, a second version of the benchmarking methodology was developed. No changes to the peer-grouping aspect of the methodology needed to be made at this point. A second outreach effort was conducted to obtain industry feedback on the reasonableness of the approach described in the methodology.
- The outreach feedback was incorporated into a third version of the methodology, which was implemented in spreadsheet form. Agencies were recruited to participate in a small-scale test of steps 2–4 of the benchmarking methodology, with the peer-grouping methodology being part of this test. The agencies provided a performance-measurement question to be answered, and the research team applied the methodology to form peer groups, identify appropriate performance measures, and present results. At each step of the process,

agency feedback was solicited on the reasonableness of the results. At the end of the testing effort, the feedback on the peer groups that were created was incorporated into a fourth version of the peer-grouping methodology.

- The peer-grouping methodology was incorporated into the FTIS software. Additional agencies were recruited for a large-scale test of the benchmarking methodology. This time, the agencies performed the work themselves (with the research team available to answer questions) and provided feedback. Their feedback on the peer groups was incorporated into the final version of the peer-grouping methodology presented here.

The remainder of this appendix describes the development of the peer-grouping methodology, including aspects of the methodology that were considered but discarded during the process, and provides the calculation details for the “likeness scores” used in the peer-grouping process.

Peer-Grouping Philosophy

Many of the overarching aspects of the peer-grouping methodology were determined by the project oversight panel at the start of the project. The panel's desired characteristics for the methodology were the following:

- Robustness – able to work with different transit modes, agency sizes, and operating environments.
- Practicality – relevant to and usable by transit agencies, state departments of transportation, and other interested stakeholders.
- Transparency – having an understandable process, with visible inputs, outputs, and intermediate results (i.e., not a black box).
- Uniformity – using readily available, uniformly defined and reported data.

- Innovation – going beyond traditional performance measures while avoiding previous peer-grouping approaches that have not been adopted by the industry.

The practicality and uniformity characteristics, in particular, drove the way the methodology was developed. The methodology is not intended to represent the best theoretical way that transit agency peer groups could be developed. During the course of the project, the research team identified peer-grouping factors that could be improved if better data were available, and some of the project’s recommendations in Chapter 5 reflect these data-definition and reporting needs. Instead, the methodology is designed to do the best possible job of meeting the industry’s needs, within the constraints of available data and tools.

During the course of developing the methodology, a few other desirable aspects were identified from the project’s industry outreach efforts and incorporated into the methodology:

- Adaptability – Not every user will share the same philosophy that underlies the methodology; therefore, users should be able to adapt the methodology for their own use and not be locked into a single approach to peer grouping.
- Accessibility – Easy-to-use tools for applying the methodology should be made readily available to the industry.
- Updateability – To make the methodology usable into the future, a process should be described for calculating any peer grouping factors not directly available from a national database.

Early in the project, the project oversight panel decided that rural and demand-response service was not to be a particular focus of the project. The final methodology can accommodate demand-response mode comparisons, although it has not been specifically tested on demand-response service (except as part of an agency’s overall service). The methodology should be adaptable to rural service; however, rural data from the NTD were not yet available at the time the project was completed to allow any testing to occur.

One aspect of the methodology that not all stakeholders agreed with is the philosophy that agencies that operate bus service but not rail service (with the exception of vintage trolleys or downtown streetcar circulators) should not be compared to agencies that operate rail service. Rail lines substitute for what would otherwise be an agency’s highest-demand, most productive bus routes. Therefore, the scale, function, and productivity of bus service in a city that also operates rail service would be expected to be different than in a comparably sized city that only operates bus service. However, two of the three largest bus-only operators in the United States have compared themselves in the past to rail-operating agencies and are comfortable doing so (the third has only compared itself to

bus-only operators). The TCRP Project G-11 methodology will not directly produce the peer groups those two large bus-only operators are used to seeing since it screens out rail operators as potential peers even when doing a motorbus-specific mode comparison. Nevertheless, it is possible for those two operators to take the peer-grouping spreadsheet that FTIS can export, remove the rail-related factors from the peer-grouping calculation, and re-calculate a likeness score. Agencies can make similar adjustments for other portions of the methodology they may not agree with, without having to abandon the entire methodology. Thus, the adaptability criterion is met.

Another aspect of the methodology that not all may agree with, as it has been incorporated into other peer-grouping methodologies in the past, is the deliberate exclusion of certain outcome measures as peer-grouping factors. The philosophy here is that outcomes should be the focus of a benchmarking effort: for example, if one looks solely at agencies with similar ridership, one is unlikely to find anything that could be used to improve one’s own ridership. On the other hand, if two agencies are similar in a number of input characteristics but have divergent outcomes (e.g., ridership, number of employees required for a given level of service, distance between mechanical failures), one is more likely to find something that can improve one’s own performance. At the same time, it is recognized that some peer-grouping factors included in the methodology, such as operating budget, vehicle miles operated, or amount of contracted service, could also be considered as outcomes for certain performance questions.

Finally, the methodology was not designed to be used as a means of ranking agencies to determine on a national basis the “best” agencies overall, or best at a particular aspect of service (although nothing prevents it from being applied that way). That approach has been tried before [e.g., Hartgen and Horner (*B-1*), Perk and Kamp (*B-2*)], but has not been widely accepted by the industry. Rather, peer-grouping and performance measurement is intended to serve as a starting point for an agency to ask questions and identify areas of possible improvement. That course—a true benchmarking process—holds much greater potential for long-term performance improvement.

Methodology Development

Initial and Outreach Versions

Description

The first two versions of the methodology used a three-step screening process to arrive at an initial peer group. In the first screening step, peers were screened out on the basis of population: an urban area population within $\pm 25\%$ for urban areas under 1 million population and $\pm 50\%$ for larger urban areas. The larger range for larger urban areas was to keep a reasonably large pool of potential peers available for subsequent steps

since there are fewer large urban areas. Urban area population was determined from the U.S. Census Bureau's American Community Survey.

Service area population would in theory be preferred to urban area population because it would allow agencies with similar market sizes to be compared. Unfortunately, although this variable is uniformly defined and collected for the NTD, it is not uniformly reported. (For example, a county-based system might report the entire county's population as its service area, although the actual population within the specified distance of transit service might be considerably less.) In addition, only one service area is reported to the NTD, even though the service areas of the various modes operated by an agency might be considerably different. As an alternative, the combination of urban area population and service type is used to identify agencies providing similar types of service within similar-sized urban areas.

In the second screening step, peers were screened out on the basis of three factors:

1. Modes operated—for consistency in the mix of modes operated. NTD data were used.
2. Service area type (e.g., region-wide, suburban only, central city only)—for consistency in the types of routes operated and markets served. Agencies had to exactly match service types to continue as potential peers. This variable was developed by TCRP Project G-11.
3. Proximity of adjacent comparably sized or larger urban areas—to account for commuting differences between a stand-alone urban area (e.g., Boise) and two or more urban areas in close proximity (e.g., Raleigh and Durham). A threshold of 45 miles was used to determine if the target agency's urban area had another urban area in close proximity. Agencies had to match (i.e., either both have or do not have an adjacent urban area) to continue as potential peers. This variable was determined based on U.S. Census Bureau data for the geographic coordinates of the center of each urban area.

In the third screening step, a set of variables was used to further refine the peer group. These variables covered a number of factors that could differentiate one agency or region from another and that could account for observed differences in ridership or other outcomes. The variables were identified through a combination of the literature review, project oversight panel input, and project team brainstorming. A larger pool of variables was developed at this stage than was intended to be included in the final methodology in order to see which of several variables did the best job of distinguishing between regions and agencies.

The exact variables used in the screening depended on the type of application. Four types of peer-comparison applications

were identified, depending on the type of performance question being asked:

- Operations – questions relating to the service provided on the street, taken from the agency's viewpoint.
- Planning – questions in support of mid- to long-term planning efforts, often with policy and funding implications.
- Market focus – questions related to the service provided, taken from the viewpoint of the broad range of customers, including riders, non-riders, local jurisdictions, and policy-makers.
- Financial – questions related to the agency's financial performance.

The variables investigated are discussed below:

- **Agency Proximity.** This variable serves multiple functions. First, it serves as a proxy for other factors, such as climate, that are more difficult to quantify but tend to become more different the farther apart two agencies are. Second, agencies located within the same state are more likely to operate under similar legislative requirements and have similar funding options available to them. Finally, for benchmarking purposes, closer agencies are easier to visit, and stakeholders in the process are more likely to be familiar with nearby agencies and regions. Some past peer-grouping efforts grouped agencies by region of the country; however, that method is somewhat arbitrary and may not be useful for agencies located near the border of a region. Instead, proximity was based on distance, using the distance between the centers of the agencies' respective urban areas, as determined from U.S. Census Bureau data. This variable was used for all applications.
- **State Capital.** State capitals are typically associated with large employment centers that are frequently located in small to mid-sized communities. Because of the singular nature of capitals, a yes/no variable was used. This variable was developed by TCRP Project G-11 and used for all applications.
- **Percent College Students.** Urban areas with large college populations typically have more transit service and higher ridership since colleges provide natural activity centers on which to focus transit systems and often assist in funding transit service. However, the effect of a university on ridership is proportional to the size of the urban area (e.g., UC Berkeley influences travel patterns in the San Francisco Bay Area less than UC Davis does for Davis). The variable is derived from the U.S. Census Bureau's American Community Survey; data include community colleges as well as 4-year universities. It was used for all applications.
- **Population Growth Rate.** Transit agencies located in areas that are growing quickly often experience different challenges than those with more stable populations, including the need to expand service to keep pace with growth. Agencies

in regions that are shrinking face another set of challenges. This variable used the urban area's average growth rate between 2000 and 2006, using U.S. Census Bureau data. It was used for all applications.

- **Population Density.** This is a well-recognized factor in attracting transit ridership and increasing transit viability, and is readily calculated for urban areas using U.S. Census Bureau data. Because population density can be lowered by the existence of large open spaces within the urban area boundary, other density-related factors were also tested, as described below. Population density was used for all applications.
- **Census Block Density.** The number of census blocks per square mile within the urban area can be used as a proxy variable to measure network connectivity, and by extension, pedestrian access to transit. The variable was determined from U.S. Census Bureau data and was used for planning and market focus applications.
- **Percent Low-Income Population.** The proportion of an urban area's residents that are "low-income" (defined by the U.S. Census Bureau as members of a family with income less than 150% of the poverty threshold) affects transit ridership because those residents are more likely to use transit. Low-income statistics reflect both household size and configuration in determining poverty status and are therefore a more robust measure than either household income or automobile ownership, which are other factors known to influence ridership. This variable was used for all applications, based on U.S. Census Bureau American Community Survey data.
- **Population Dispersion.** While population density provides an overall measure of land-use intensity in an urban area, it does not reflect the homogeneity of land use. Urban areas with high-density cores and centers but low-density outlying areas may be more transit-friendly than those with population spread evenly throughout the area. Population dispersion is calculated by dividing an urban area's population density by its weighted density. Weighted density is calculated by multiplying each census block's population density by its proportion of total urban area population and summing across all census blocks. A dispersion value of 1 means that population is distributed evenly across all block groups, while a value closer to 0 means that residents are concentrated in specific areas. This measure was tested for planning and market focus applications using 2000 Census data.
- **Employment Dispersion.** This measure follows the same general principle and calculation methodology as population dispersion. It was tested for planning and market focus applications using 2000 Census data.
- **Transit-Supportive Area.** The amount of transit-supportive land-use found in an urban area plays a large role in transit operations. The *Transit Capacity and Quality of Service Manual* (TCQSM, B-3) uses the concept of "transit-supportive area" (areas capable of supporting at least hourly weekday transit service, based on population or employment density) to make apples-to-apples comparisons of different agencies' service coverage areas. While population density and population dispersion also address this issue to some extent, this measure can potentially do the work of both. Converting the TCQSM's household-based threshold to a population-based threshold suggested a value of 7.5 persons per acre as the minimum value to test. The measure was tested for planning and market focus applications using 2000 Census data.
- **Congestion Delay Per Capita.** Highway congestion has a large effect on bus operating conditions and may provide a greater incentive for persons to use all forms of transit for peak-period trips. Congestion is more likely to be an issue in larger urban areas. Data for this measure are available from the *Urban Mobility Report (B-4)* for larger urban areas. The measure was used for all applications, except financial, for agencies located in urban areas with at least 1 million population.
- **Freeway Lane-Miles Per Capita.** The extent of a region's freeway network may indicate the level of priority given to roadway investments compared to transit investments. It may also influence service design; for example, systems focused on large park-and-ride lots. In small and mid-sized urban areas, freeways may serve more intercity travel than intra-city travel and therefore have less of an influence on commuting patterns. The measure was based on *Urban Mobility Report* data and was used for all applications, except financial, for agencies located in urban areas with at least 1 million population.
- **Total Vehicle Miles Operated.** The total amount of service provided by an agency influences a number of transit service factors. This variable was used for operations and financial applications and was based on NTD data. For the other two applications, this variable was felt to be an outcome and was therefore not included.
- **Total Operating Budget.** *Total Operating Budget* influences many aspects of transit service. Structurally, operating budget is a measure of the scale of a transit agency's operations; agencies with similar budgets may face similar challenges. This variable was used for operations and financial applications and was based on NTD data. For the other two applications, this variable was felt to be an outcome and was therefore not included.
- **Mean Wage Rate.** Typical wages vary between regions. Higher wages will typically be associated with higher labor costs for transit agencies. This variable was used for operations and financial applications, and was based on Bureau of Labor Statistics data for metropolitan areas.

A "likeness score" approach was used for each variable (factor) included in an application. The factor likeness scores

were added together to form a (non-normalized) likeness score for an agency, with lower likeness score values indicating a greater degree of similarity. The factor likeness scores were, for the most part, calculated similarly to the scores used in the final methodology (described later). Whether or not to weight certain factors more heavily was considered at this stage, but not implemented, pending the results of more widespread methodology testing later in the project.

Factors Considered but Not Included in the Methodology

The following variables were considered but were dropped from further consideration after initial internal testing by the research team:

- **Median Household Income and Percent of Households Earning Less than \$35,000** (U.S. Census Bureau). Dropped because they do not reflect the size and composition of the household, unlike Percent Low-Income Population.
- **Automobiles Per Capita and Percent Zero-Car Households** (U.S. Census Bureau). The former was dropped because it had the lowest variation between urban areas of any of the tested measures. The latter also showed lower variation than most other measures. A lack of variation limits the ability of a variable to distinguish differences between regions or agencies. Poverty-related measures (e.g., Percent Low-Income Population) and density-related measures capture similar demographic characteristics.
- **Percent of Population Less Than 18 and Percent of Population 65 or Older** (U.S. Census Bureau). Dropped because of low variability between urban areas in the tests. While age may be a key consideration when making local service planning decisions, it does not provide much benefit when distinguishing between urban areas. The Agency Proximity variable can also help account for any regional differences that may exist.
- **Arterial Miles Per Capita and Freeway Miles Per Capita** [FHWA Highway Performance Monitoring System (HPMS)]. Dropped because of the work that would be involved each year deriving these measures from raw HPMS data to keep the database current. The *Urban Mobility Report* provides freeway miles data (B-4), albeit for a smaller set of urban areas.
- **Parking Cost** (Collier's Parking Cost Survey). Parking costs influence the decision to use transit. This measure was dropped because parking cost data are not available for most smaller urban areas. The Urban Area Population variable helps control for parking costs since larger urban areas will tend to have higher parking costs.
- **Sprawl Index** (Smart Growth America). More-sprawling regions are more difficult to serve with transit. SGA's

Sprawl Index provides a comprehensive, national source of data based on objective research. This measure was dropped because data are only available by county, making it difficult to assess the degree of sprawl for urban areas that span multiple counties or only a small portion of a single county.

- **USDA Plant Hardiness Zones and Annual Precipitation** (U.S. Department of Agriculture and National Oceanic and Atmospheric Administration). These are surrogates for climate. The former is based exclusively on average annual low temperature, which masks differences in summer extremes. The latter does not account for the distribution of precipitation throughout the year. The Agency Proximity variable helps control for climatic differences, as nearby agencies are more likely to have similar climates (although it is recognized that topography also plays a role).
- **Cost of Living Index** (ACCRA cost-of-living index and others). Cost-of-living differences between regions can influence agency costs. This variable was dropped because data are not available for all areas and require payment of a fee to obtain and distribute. Median wage was used as a surrogate for differences in costs between regions.
- **Park-and-Ride Spaces** (no standard source). This variable helps describe service structure, but was dropped due to a lack of a national data source.
- **Bicycle Friendly Community Rating** (League of American Bicyclists). Helps describe ease-of-access to transit since bicycle-friendly communities are typically also pedestrian-friendly. Dropped because ratings are only generated by request and are by jurisdiction, making them difficult to use for transit agencies serving multiple jurisdictions.

Small-Scale Testing

Description

During this stage of testing, the methodology was tested by 16 agencies: 10 transit agencies, 5 state departments of transportation (DOTs), and the Regional Transportation Authority in Chicago. The participating agencies were as follows:

- Transit agencies
 - Denver RTD (Denver, CO)
 - Utah Transit Authority (Salt Lake City, UT)
 - Santa Clara Valley Transportation Authority (San Jose, CA)
 - Lane Transit District (Eugene, OR)
 - Knoxville Area Transit (Knoxville, TN)
 - Triangle Transit Authority (Durham, NC)
 - Rochester Genesee RTA (Rochester, NY)
 - Greater Cleveland RTA (Cleveland, OH)
 - Greater Bridgeport Transit Authority (Bridgeport, CT)
 - Bay County Council on Aging (Panama City, FL)

- State DOTs
 - Florida DOT
 - Indiana DOT
 - Pennsylvania DOT
 - Texas DOT
 - Washington State DOT
- Other
 - RTA (Chicago, IL).

Each agency developed a performance measurement question or topic to be addressed, while the research team applied the methodology and presented the results to the agencies. Several feedback points with agency staff were built into the process to obtain feedback on particular steps of the peer-grouping and benchmarking methodologies and to make sure staff were comfortable with the results before continuing. These feedback points consisted of:

- Identifying a performance measurement topic of interest to the agency;
- Identifying an initial set of peers for the agency and an initial set of performance measures relating to the topic;
- Identifying a final set of peers and performance measures; and
- Discussing the performance results and the usefulness of the methodology.

Methodological Changes

As originally proposed, the FTIS software was going to be used for this round of testing. The methodology was programmed into FTIS, and a user interface was developed. This allowed more extensive testing of the initial methodology than had previously been possible. An initial research team observation was that the portion of the screening process that screened out potential peers based on modes operated, service area type, and proximity of comparably sized or larger urbanized areas did too good a job of screening and left too small a pool of potential peers.

Rather than continually update FTIS, the research team decided it would be faster and more cost-effective to implement a spreadsheet version of the methodology for the team's use for the small-scale testing, and then to update FTIS prior to beginning the large-scale tests, where agencies themselves would be applying the method. The original spreadsheet contained all of the data needed to use the peer-grouping methodology and, for each application type (operations, planning and market focus, and financial), produced summary lists of the 20 peers most similar to the target agency. There were five major versions of the spreadsheet, which was updated to add data for additional modes as the need to analyze them came up during the testing. In addition, each successive version also

corrected errors (mostly in the way agencies were assigned a service area type) in the non-NTD portions of the database.

Implementing the peer-grouping methodology in a spreadsheet required some alterations to the original methodology. The main change was that all peer-grouping variables generated a likeness score instead of completely screening out agencies from further consideration. This had the positive side-effect of allowing the other 643 reporters in the NTD database to be assigned a likeness score and a likeness ranking relative to a given target agency. The original intent of using certain variables to screen out agencies from further consideration was retained by assigning a high factor likeness score for differences in those variables [rail-operator (yes/no), and rail-only operator (yes/no)].

Other changes to the methodology that were implemented prior to the small-scale testing, based on the FTIS testing, were:

- The variable on distance to the nearest comparably sized or larger urbanized area was dropped because testing showed it screened out too many potential peers.
- Agency proximity was weighted twice as heavily as before.
- Changes were made to the way the likeness score was calculated for the population growth rate variable.
- Agencies were defined as being rail operators if they operated more than 100,000 vehicle miles of rail service annually. This threshold was selected to distinguish between operators of vintage trolleys and downtown streetcar circulators and operators of full-scale light rail and commuter rail lines and systems.

Testing Feedback

The version of the methodology used for the small-scale applications identified four types of applications: operations, financial, planning, and market focus (the latter two applications shared the same peer-grouping variables). The performance topics selected by the participating agencies included three operations topics, three planning and market focus topics, and ten financial topics. (Some of the operations topics could also have been classified as financial topics and vice versa.)

The "financial" peer-grouping method resulted in the least number of requested changes to peers among the participating agencies. "Operations" performed reasonably well but generated more requested changes (particularly with Denver). When given a choice of using the "operations" and "financial" set of peers, Indiana DOT chose the "financial" set. The "financial" method included all of the peer-grouping variables used by the "operations" method, plus three others: vehicle-miles operated, total operating budget, and mean wage rate. The "planning" method tended to identify a large number of agencies within the same region, without regard to agency size. Service area population was added as an additional

screening variable but did not help much, regardless of the weight assigned to it (often because the reported service area population reflected the urban area population).

The range of average wage rates among agencies was not significant enough to cause that variable to influence the peer group selection (i.e., the list of peers might be shuffled a little, but the same peers would generally appear with or without the variable).

Two variables that were suggested as additional screening variables were the ratio of demand-responsive vehicles operated in maximum service to motorbuses operated in maximum service, and the ratio of purchased service to directly operated service (requested twice). Agencies that operate entirely demand-response service are accounted for in the methodology by the “service area type” screening variable, but not agencies that operate mostly demand-response service.

Ability to Obtain Local (non-NTD) Data

A significant peer-comparison challenge to overcome is the ability to obtain data not included in the NTD. The research team spent considerable time trying to track down such information. Particular issues include (a) the availability of staff at the target agency to find contacts at the peer agencies and request information from them, (b) the availability of staff at the peer agencies to track down the requested information, (c) the existence of the data, and (d) compatibility of measure definitions between agencies. We were, eventually, able to gather sufficient customer-satisfaction data from peers to be able to conduct a comparison. However, we were not able to gather absenteeism data or performance data specific to regional or express bus routes or to bus divisions (suburban vs. urban).

NTD Data Reliability and Detail

As expected, participants raised questions about the reliability of some of the NTD data. The most common problem that appeared was agency definitions of service area population. Some followed the FTA definition based on population within a certain distance of transit service, while others simply used the urbanized area population, regardless of whether they served the entire area. Agencies were also inconsistent from year to year in reporting population: for example, one agency reduced its service area population by 75% from 2005 to 2006, which caused obvious problems with “per capita” trends. Being able to compare performance on a per-capita basis can be very useful, but much more work appears to be needed to get agencies to report their service area population in a standard way.

Case study participants also commented about potential differences in how agencies reported passenger-mile and

vehicle-malfunction data and the general lack of detail of the maintenance data. The research team noticed problems at individual agencies (sometimes only in one year, sometimes every year) with some cost categories and in breaks/allowance time.

Large-Scale Testing

Description

During the final stage of testing, the methodology was tested by 22 agencies: 19 transit agencies, 2 state DOTs, and the Regional Transportation Authority in Chicago. The participating agencies are listed below. Agencies that also participated in the small-scale test are shown with an asterisk (*).

- Central Oklahoma Transportation and Parking Authority, Oklahoma City, OK
- Greater Bridgeport Transit Authority, Bridgeport, CT (*)
- Hillsborough Area Rapid Transit, Tampa, FL
- King County Metro, Seattle, WA
- Knoxville Area Transit, Knoxville, TN (*)
- MARTA, Atlanta, GA
- Metrolink, Los Angeles, CA
- North County Transit District, Oceanside, CA
- Oahu Transit Service, Honolulu, HI
- Orange County Transportation Authority, Orange, CA
- Pennsylvania DOT, Harrisburg, PA (*)
- Pinellas Suncoast Transit Authority, St. Petersburg, FL
- Regional Transportation Authority, Chicago, IL (*)
- Regional Transportation District, Denver, CO (*)
- San Joaquin Regional Transit District, Stockton, CA
- San Mateo County Transit, San Carlos, CA
- Sarasota County Area Transit, Sarasota, FL
- SEPTA, Philadelphia, PA
- StarMetro, Tallahassee, FL
- Texas DOT, Austin, TX (*)
- Utah Transit Authority, Salt Lake City, UT (*)
- Virginia Railway Express, Alexandria, VA

Each agency applied the methodology using instructions provided to them by the research team. The instructions provided background information on the purpose of the project and described the process for applying the methodology (including detailed instructions for using FTIS). At a minimum, agencies were instructed to provide feedback on the following questions:

1. Do you feel the peer group identified for your agency was reasonable? Were peer agencies identified that you feel are inappropriate (and if so, why)? Were agencies not identified that you feel should have been (and if so, why)?

2. Do you feel that the performance results were reasonable (i.e., reflect reality, to the best of your knowledge)? Were there any observed issues with the data (i.e., missing data, illogical trends, unexplainable results) that could affect the credibility of the results?
3. How easy was it to follow the instructions in this document and apply the software?

The research team was available throughout the process to answer any questions as they arose, but did not otherwise participate in the application of the methodology during the large-scale test.

Methodological Changes

Based on the results of the small-scale tests, only a single peer-grouping method was used for the large-scale test. This method was based on the “financial” peer-grouping application used in the small-scale test, with the following changes:

- Average wage rate was eliminated as a peer-grouping variable. There was not enough variation in the wage rate between regions to make a substantial difference in the peer-grouping results. The wage data were retained in FTIS to allow agencies to manually adjust costs based on wage rate differences if they so desired.
- The percentage of service that is demand-response was added as a peer-grouping variable for agency-wide and bus-mode comparisons. This variable helps distinguish agencies that mostly operate demand-response service from those that mostly operate fixed-route service.
- The percentage of service that is purchased was added as a peer-grouping variable for all types of comparisons. Agencies that purchase their service will typically have different organization and cost structures than those that directly operate service.
- Being a heavy rail operator (yes/no) was added as a third screening variable. A mismatch between the target agency and the peer agency for this variable resulted in a likeness score of 20 being assigned for this variable, effectively eliminating the potential peer from further consideration.
- The likeness score was reported as a normalized value by dividing the sum of the peer-grouping factors by the number of factors used (excluding the three rail-related screening factors). Guidance was provided that, in general, a total likeness score less than 0.50 was considered a very good match, a score of 0.50–0.74 was considered satisfactory, and a score of 0.75–0.99 indicated an agency that might be usable as a peer but with caution since there might be some significant differences that might need to be considered. A score of 1.00

or higher indicated that there were probably too many differences to make the agency a good peer.

Testing Feedback

Most of the feedback on the peer-grouping aspect of the methodology related to a problem with implementing the methodology in the FTIS software (since corrected), where the software assigned a likeness score for factors with missing data (indicating a very close match) rather than the intended value of 1,000 (to effectively drop the potential peer from further consideration). This problem resulted in inappropriate (mostly small) peers appearing in agencies’ lists of potential peers. At least nine test applications were affected by this issue. After fixing the issue, the lists of potential peers generally seemed appropriate.

As discussed previously in the Peer-Grouping Philosophy section, two large bus-only operators felt that the potential peers identified by the TCRP Project G-11 methodology did not match the ones they had used in previous efforts and were comfortable with. In both cases, the TCRP Project G-11 methodology identified a mix of smaller suburban operators from the same region and/or state, plus some larger national bus-only peers. In comparison, these agencies did not include suburban operators and included national peers that operated light rail systems. A third large bus-only operator only uses bus operators in its peer group and was comfortable with the results once the FTIS missing-data issue described above was addressed. These concerns were addressed by providing guidance on how to work with the peer-grouping data exported by FTIS to include rail operators as potential peers.

In some cases, agencies felt that a potential peer was inappropriate because one particular factor (e.g., urban area population, agency budget) was too big or small relative to the target agency. As described above, prior versions of the methodology considered setting absolute cutoffs (e.g., population within $\pm 25\%$ of the target agency), but found that these often reduced the pool of potential peers too much. Instead, agencies that are substantially different in one characteristic need to be quite similar in a number of other respects in order to end up with a total likeness score low enough to be considered as a potential peer. The concerns expressed by these agencies were addressed by providing guidance that agencies should identify limits for factors of concern to them prior to conducting the peer grouping and should then apply those criteria as part of a secondary screening process.

Other changes relating to the guidance related to the interpretation of likeness scores and how to address special cases. One special case involved a transit operator in Hawaii, where some additional spreadsheet work was needed to adjust the likeness scores to account for the long distances between the target agency and any potential peer.

Final Methodological Changes

The following changes were made to the methodology to address the feedback from the large-scale test:

- Distance was removed as a peer-grouping factor for mode-specific comparisons involving rail service. Rail operators tend to be widely spread apart outside the Northeast, and there is little expectation that peers will be located nearby. Removing distance as a factor for these comparisons allows the general guidance on interpreting likeness scores to be applied more consistently.
- The weights applied to different combinations of service types were adjusted in order to make it less likely that suburban operators would be matched as peers to central-city operators.
- More weight was applied to differences in service types between operators within the same urban area in order to compensate for the fact that these operators will be alike on the screening area factors that are based on urban area characteristics. This change was designed to make it less likely that suburban operators would be matched as peers to central-city operators within the same urban area.
- The definition of a rail operator was adjusted to count only those operating more than 150,000 vehicle miles annually since a downtown streetcar operator approached 100,000 vehicle miles in the 2007 NTD.

Likeness Score Calculation

Total Likeness Score

The heart of the peer-grouping methodology is the calculation of a total likeness score that indicates the degree of similarity between a target agency and a potential peer, based on a variety of factors that account for many of the differences between agencies and regions that can impact performance results.

A score of 0 indicates a perfect match between two agencies (and is unlikely to ever occur). Higher scores indicate greater levels of dissimilarity between two agencies. In general, a total likeness score under 0.50 indicates a good match, a score between 0.50 and 0.74 represents a satisfactory match, and a score between 0.75 and 0.99 represents potential peers that may be usable but for which care should be taken to investigate potential differences that may make them unsuitable. In some cases, peers with scores 1.00 may also be usable (with even greater caution) or, in a few cases, may be the only candidates available.

A total likeness score of about 70 or higher may indicate that a potential peer had missing data for one of the screening factors. In some cases, suitable peers may be found in this group by manually re-calculating the total likeness score in a

spreadsheet, removing that factor from consideration if the user determines that the factor is not essential for the performance questions being asked. Missing congestion-related factors, for example, may be more easily ignored than a missing total operating budget.

The total likeness score is calculated as follows:

$$\text{Total likeness score} = \frac{\text{Sum}(\text{screening factor scores}) + \text{Sum}(\text{peer-grouping factor scores})}{\text{Count}(\text{peer-grouping factors})}$$

Screening Factors

Three screening factors are used in the process. These are used to distinguish bus-only operators from rail operators and types of rail operators from each other.

- **Rail operator (yes/no).** A rail operator is defined as one operating 150,000 or more rail vehicle miles annually. A match on this factor produces a likeness score of 0; a mismatch produces a likeness score of 20. This factor is derived from the NTD.
- **Rail-only operator (yes/no).** A rail-only operator operates rail and has no bus service. A match on this factor produces a likeness score of 0; a mismatch produces a likeness score of 20. This factor is derived from the NTD.
- **Heavy-rail operator (yes/no).** A heavy-rail operator operates the heavy rail mode. A match on this factor produces a likeness score of 0; a mismatch produces a likeness score of 20. This factor is derived from the NTD.

Peer-Grouping Factors

Up to 14 peer-grouping factors are used in the process, depending on the type of analysis (rail-specific vs. bus-specific or agency-wide) and the target agency's urban area size (which determines whether the two *Urban Mobility Report* factors are included).

Most factor likeness scores are determined from the percentage difference between a potential peer's value for the factor and the target agency's value. A score of 0 indicates that the peer and target agency values are exactly alike, while a score of 1 indicates that one agency's value is twice the amount of the other. For example, if the target agency was in a region with an urbanized area population of 100,000, while the population of a potential peer agency's region was 150,000, the likeness score would be 0.5 because one population is 50% higher than the other.

For the factors that cannot be compared by percentage difference (e.g., state capital or agency proximity), the factor likeness scores are based on formulas that are designed to produce similar types of results—a score of 0 indicates identical

- **Distance.** Likeness scores are calculated as the distance between the two agencies' urban areas (in miles), divided by 500. The urban area centroid is derived from U.S. Census Bureau data.

Service types are defined as follows:

1. Agency provides service only to non-urbanized areas.
2. Agency provides service to multiple urban areas (may also include non-urban areas) and is the primary service provider within at least one urban area central city.
3. Only agency operating within an urban area and has no non-urban service.
4. Agency is the primary service provider in the urban area's central city, where other agencies also provide service to portions of the urban area. Urban areas with multiple central cities (e.g., Tampa–St. Petersburg) may have more than one type 4 agency.
5. Agency provides service into an urban area's central city, but its primary service area does not include a central city.
6. Agency provides service within an urban area but does not provide service to a central city.
7. Only agency operating within an urban area and also provides non-urban service.
8. Other (e.g., special needs transportation service only, ferry-only, monorail-only, agency in Puerto Rico, agency provides funds to another NTD reporter that operates the service).

References

- B-1. Hartgen, David T. and Mark W. Horner. *Transportation Publication Report 163: Comparative Performance of Major U.S. Bus Transit Systems: 1988–1995*. University of North Carolina at Charlotte, May 1997.
 - B-2. Perk, Victoria and Nilgün Kamp. *Benchmark Rankings for Transit Systems in the United States*. National Center for Transit Research at the Center for Urban Transportation Research, University of South Florida, Tampa, Fla., December 2004.
 - B-3. Kittelson & Associates, Inc., KFH Group, Inc., Parsons Brinckerhoff Quade & Douglass, Inc., and Katherine Hunter-Zaworski, *TCRP Report 100: Transit Capacity and Quality of Service Manual*, 2nd ed., Transportation Research Board of the National Academies, Washington, D.C., 2003.
 - B-4. Schrank, David and Tim Lomax. *2007 Urban Mobility Report*. Texas Transportation Institute, Texas A&M University System, College Station, Tex., September 2007.
-

APPENDIX C

Task 10 Working Paper

Introduction

The purpose of Task 10, Interpret Results/Recommendations, was fourfold:

1. Present and interpret the results from the Task 8 and 9 test applications of the project's peer-grouping and performance-measurement methodology;
2. Revise and expand, if necessary, the list of potential applications for the methodology presented in Working Paper #2;
3. Provide recommendations for new standard performance measures (or modifications to existing measures) that would help support the methodology; and
4. Develop strategies for the adoption of the methodology by the transit industry.

The aspects of item #1 related to results presentation were provided in the three working papers developed for Tasks 8 and 9. In addition, some initial interpretation of the Task 8 results was provided in the appendix to Working Paper #4. Also, six of the case studies from Tasks 8 and 9 that illustrate different applications of the methodology have been re-worked using the final version of the methodology and are provided in Chapter 5 of the final report. The Chapter 5 case studies provide interpretations of the case study results and guidance on how the questions that were raised by the results could be explored further.

The results of item #2 are incorporated into Chapter 3 of the final report. The Task 8 and 9 testing uncovered new planning application examples related to (a) thinking to the future and what might happen when a region reached 200,000 population and its funding sources changed, and (b) comparing the performance of an agency without a dedicated local funding source to peer agencies with one. These examples have been added to the list of applications in Chapter 3. In addition, the lists of applications have been revised and reworded in response to panel comments over

the course of the project. Finally, the lists of standard performance measures that are applicable to benchmarking applications have been reorganized into descriptive and outcome categories, with multiple subcategories for each (e.g., cost-effectiveness, cost-efficiency, perceived service quality, delivered service quality).

The results of item #4 are provided in Chapter 6 of the final report. Recommendations are provided for transit agencies, state and regional transportation and funding agencies, standards development, the NTD, and future steps.

This working paper focuses on the remaining aspects of Task 10 not already documented in the final report. This working paper presents key examples from the Task 8 and 9 peer comparisons that serve to highlight lessons learned and methods for dealing with common challenges. It also provides recommendations on modifications to existing performance measures that would help support the peer-grouping and performance-measurement methodology.

The paper is organized around the eight-step benchmarking methodology described in detail in the final report. Typical questions and challenges associated with each step of the process are covered here, using specific examples from the Task 8 and 9 testing. Because of the project's time and resource limitations, the Task 8 and 9 testing covered only Steps 1–4 of the process. Guidance on conducting Steps 5–8 is provided in the final report.

Table C1 provides a summary of the questions and topics addressed in this paper, organized by the methodological step to which each question applies.

Summary of Case Study Results

Step 1: Understand Context

While relatively straightforward, understanding the context of the peer comparison is a key component of success. Specific lessons from the case studies included the need to

Table C1. Summary of working paper topics by methodological step.

Methodological Step	Topic	Example Agencies
<i>Step 1: Understand Context</i>	Carefully tailor the topic to avoid an ill-defined performance topic.	Utah Transit Authority (UTA), Washington DOT
	Understand the analysis timeframe.	Greater Bridgeport Transit Authority (GBTA)
<i>Step 2a: Performance Measure Selection</i>	Carefully review relevant NTD forms to understand content.	Santa Clara Valley Transit Authority (VTA), UTA
	Use descriptive measures to provide context to the analysis.	VTA, Hillsborough Area Regional Transit (HART)
	Use multiple measures to get a rounded perspective of a particular issue.	Denver Regional Transit District (RTD)
<i>Step 2b: Identify Secondary Screening Measures</i>	What are appropriate variables for secondary screening?	
<i>Step 2c: Identify Thresholds</i>	Setting thresholds to ensure that peers are relevant to performance question.	Laredo Transit, Knoxville Area Transit
	Setting thresholds to lend additional credibility to peers.	King County Metro
<i>Step 3a: Register for FTIS</i>	This step is self-explanatory. No additional detail is provided.	
<i>Step 3b: Form an Initial Peer Group</i>	Refine grouping methodology to remove specific variables.	GBTA, King County Metro, Oahu Transit
	Interpreting likeness scores and choosing a peer group.	North County Transit District
<i>Step 3c: Perform Secondary Screening</i>	Applying the thresholds identified in Step 2c.	King County Metro
<i>Step 4a: Gather Performance Data</i>	Exporting data to Excel for calculating non-standard performance measures.	
	Gathering non-NTD data.	GBTA, Rochester-Genesee RTA
<i>Step 4b: Analyze Performance</i>	Using Excel to create graphs and charts to display result.	
	Normalizing results to account for cost of living and inflation.	RTD
	Gather additional descriptive measures as needed.	

carefully formulate the peer comparison question and the need to understand the analysis timeframe, particularly if non-NTD data are being considered.

Determining the Topic

Determining the topic for the peer comparison effort requires careful consideration if the peer comparison is to prove beneficial to decision-makers. Without a well-defined topic it will be difficult to select an appropriate set of performance measures, select relevant descriptive measures, and determine if secondary screening of peers is required. Similarly, topics

that are too broad may yield results that do not directly address the intent of decision-makers.

Many case study participants had a tendency to select broad or ill-defined performance comparison topics. For instance, several agencies simply stated a desire to understand their “efficiency.” The resulting peer comparisons were often focused on high-level efficiency measures (e.g., cost per ride) across a range of areas, but did not provide depth in any particular area. That would not necessarily be a problem except that once the high-level results were identified, no additional work was performed to dig into the reasons for the results. Thus, the underlying reasons why a particular agency’s performance

was superior or inferior to its peers were not identified. To address this issue, the Altoona case study in the final report provides examples of digging deeper into the data to identify potential explanations for the high-level results.

In contrast, several agencies provided detailed performance questions that allowed the peer comparison to be more focused from the start. For instance, Utah Transit Authority, like many other agencies, was interested in understanding operating efficiency. Rather than attempting to cover all aspects of efficiency within a single peer comparison, however, UTA chose a more focused question to consider: “How efficient are my bus and rail operator schedules?” The methodology provided a well-rounded set of performance measures that addressed the topic to the extent possible through the NTD. (This aspect of the methodology is discussed further under Step 2a.) Other aspects of agency efficiency could be considered by follow-up peer comparisons in a similar manner.

Understand the Analysis Timeframe

At the outset of the peer comparison, it is important to understand the timeframe for which results will need to be available. This is particularly important if the target agency is considering the possibility of using non-NTD data for the peer comparison. The Florida Transit Information System software tool allows users to complete basic peer comparisons in a matter of hours when only NTD data and other standardized data provided by FTIS are required. However, the process for collecting non-NTD data is much more time-consuming.

The Greater Bridgeport Transit Authority (GBTA) case study provides an illustrative example. GBTA wished to analyze employee absenteeism among peer agencies, a topic requiring non-NTD data. Eight peer agencies were contacted to obtain information about absenteeism rates by job category. After 2 months, replies were received from five of the eight peers, only two of which were able to provide specific absenteeism data/rates by worker category. As a result, the peer comparison was not completed within the timeframe required by the research.

With additional time and resources to dedicate to the effort, it is likely that sufficient data for a peer comparison could have been obtained by requesting data from more than the top eight peers. However, the overall effort would certainly take months, compared to days for a peer comparison using NTD data. Thus, understanding up-front whether such a timeframe is viable is necessary to avoid wasted effort. At the same time, once the initial effort has been made, it may be easier in the future to obtain the same data from the same set of peers as part of peer-comparison activities in future years.

The state of the economy at the time of the research also influenced other agencies’ ability to respond to data requests. Due to the financial crisis and its impacts on tax revenue, most

transit agencies were facing funding shortfalls and agency staff did not have time to respond to outside data requests. Here again, having previously established relationships with peer agencies might make it easier to obtain data during such times. However, even an established benchmarking network such as TFLEX reported difficulty in getting its members to contribute data since the crisis started. Nevertheless, as pointed out in the final report, when funding is tight, it is more critical than ever for agencies to identify where they can improve their performance. Sharing data and practices with other agencies, while not necessarily providing an immediate benefit to an agency, helps establish relationships that can provide longer-term benefits.

Step 2a: Performance Measure Selection

The selection of an appropriate set of performance measures is one of the most important components of a successful peer comparison. There are hundreds of potential measures available through the NTD, and choosing a reasonably-sized set of measures (ideally fewer than ten) that provides the desired detail can be difficult. While definitive “cookbook-style” guidance, such as that provided for selecting a peer group, may be ideal from the guidebook user’s perspective, the case study experience clearly shows that there is no single correct set of measures for a particular question. Rather, each agency must custom-select performance measures that address their specific performance question and operating environment.

Case study participants that identified identical performance topics often ultimately chose considerably different performance measures. For instance, both the Greater Cleveland Regional Transit Authority (GCRTA) and the Lane Transit District (LTD) chose “What is a reasonable level of subsidy?” as their performance question. Table C2 shows the performance measures that each agency ended up selecting.

Table C2 shows that only one performance measure (*farebox recovery ratio*) was chosen by both agencies. In general, LTD chose to focus on per-capita measures of revenue sources (i.e., a funding perspective), whereas GCRTA focused on the proportion of overall operating expense from various sources (i.e., an operating perspective). In addition, LTD chose to include several descriptive measures (i.e., measures selected purely to provide context), such as *passenger trips per capita*, to understand how various funding strategies may impact other performance measures of interest.

While each agency chose to approach the question from a different perspective, both found the results of their respective peer comparisons beneficial. This highlights the need to approach the selection of performance measures for every peer comparison as a unique exercise. The following section summarizes some key issues to consider while selecting performance measures.

Table C2. Performance measures selected by GCRTA and LTD relating to subsidy level.

LTD Measures	GCRTA Measures	Overlap?
Farebox recovery ratio	Farebox recovery ratio	✓
Local operating funds per capita	Local revenue as percent of operating expense	
State operating funds per capita	State revenue as percent of operating expense	
Federal operating funds per capita	Federal revenue as a percent of operating expense	
Operating cost per revenue hour	Other directly generated revenue as percent of operating expense	
Operating subsidy per revenue hour		
Operating subsidy per capita		
Passenger trips per capita		
Revenue hours per capita		
Operating costs per capita		

Review Relevant NTD Forms

As described in the final report, most peer comparisons will rely on NTD data due to the lack of viable alternatives. Consequently, fully understanding the contents of the NTD is critical to understanding performance measure options. FTIS reports hundreds of NTD measures across 18 different NTD formatting forms, making it difficult for any individual to be familiar with everything that the NTD reports.

VTA and UTA peer comparisons (modified versions of which are included as case studies in the final report) provide good examples of the potential to mine NTD forms to develop nonstandard performance measures to answer specific peer-comparison questions. Research team staff worked closely with agency staff to review relevant NTD forms and develop a tailored set of performance measures. In both cases, reviewing the NTD forms allowed the groups to select measures that they were initially not aware of. Most of the selected performance measures are ratios, which provide greater comparability of results across agencies.

Table C3 summarizes the peer comparison questions and selected outcome measures for each agency. The variables listed in Table C3 show clearly how a variety of focused performance measures can be derived from NTD data by considering more than the most commonly used measures.

Use Descriptive Measures to Provide Context

In addition to the outcome measures that form the core of a peer comparison, it is often beneficial to collect data for descriptive measures as well. Descriptive measures are measures that do not directly address the performance question at

hand, but provide context as to why a particular result occurs. Many useful descriptive measures are included among the methodology's peer-grouping variables (e.g., *total operating budget*, *congestion per capita*), but other descriptive measures may also be valuable depending on the application. Chapter 4 of the final report provides lists of descriptive measures, arranged by topic.

For instance, as part of the VTA peer comparison for light-rail maintenance shown in Table C-3, VTA also used *miles of track*, *number of elevators*, and *number of escalators* as descriptive measures to provide information on factors that may drive non-vehicle maintenance costs. In general, identifying useful descriptive measures during Step 2 of the benchmarking process will reduce the need to collect supplemental data later in the process. Similarly, Hillsborough Area Regional Transit used such descriptive variables as the *total number of vehicles operated in peak service* to supplement their peer comparison on system-wide efficiency.

Use Multiple Overlapping Measures to Provide Perspective

Some agencies may rely primarily on a single performance measure to address a specific question (e.g., *cost per trip* as the measure of cost-effectiveness). While such an approach may make sense for a benchmarking exercise across a wide range of topics, the ease with which FTIS allows users to summarize NTD data means that data for multiple measures can be gathered with little additional effort.

During the Task 8 case studies, the Denver Regional Transit District (RTD) examined cost-effectiveness as its performance

Table C3. Outcome measures used in the VTA and UTA case studies.

VTA	UTA
<i>How cost-effective are VTA's vehicle and non-vehicle light-rail maintenance program?</i>	<i>How efficient are UTA's bus and rail operator schedules?</i>
Maintenance expenditures as percent of operating expense	Operating cost per passenger mile
Actual car miles per malfunction	Operating cost per passenger hour
Maintenance labor as percent of total maintenance cost	Revenue hours as percent of vehicle hours
Maintenance costs per actual car mile	Salaries/wages/benefits as percent of operating expenses
Vehicle materials and supplies cost per actual car mile	Operating wages as percent of operating expenses
Vehicle maintenance labor cost per actual car mile	Vehicle revenue hours per operating FTE
Non-vehicle maintenance costs per station	Passenger trips per operating FTE
Non-vehicle materials and supplies cost per station	Non-operating time as percent of total operating time
Non-vehicle maintenance labor per station	Breaks and allowances as percent of total operating time
	Premium hours as percent of operating hours

topic. Rather than using a single measure of cost-effectiveness, several measures were used, including *cost per trip*, *cost per revenue hour*, *revenue per trip*, *trips per revenue mile*, and *trips per revenue hour*. By examining the same issue from multiple angles, the peer comparison was able to provide more insight into the differences between RTD and its peers. For instance, RTD rated at the peer group average for *trips per revenue hour*, but lower than average for *trips per revenue mile*, as a result of RTD providing more long-distance, higher-speed service than the peer group as a whole.

Step 2b: Identify Secondary Screening Measures

As described in the final report, the selection of a peer group is a vital part of the benchmarking process to produce relevant results and establish credibility with stakeholders. While the peer-grouping methodology developed by this research and incorporated into FTIS is designed to produce a reasonable peer group for most situations, secondary screening may be needed in some circumstances, either to answer a performance question that requires a specific type of agency or to eliminate agencies deemed “too different” from the target agency. Table C4 lists each of the peer-grouping variables and provides general guidance on their appropriateness as secondary screening measures.

Note that secondary screening measures should be determined prior to identifying a preliminary peer group through FTIS in order to avoid the appearance of subjectivity (i.e., choosing a secondary screening measure to exclude a specific agency). Other potential secondary screening measures are also described in the final report.

Step 2c: Identify Thresholds

As described in Step 2b, agencies may identify secondary screening measures and thresholds to ensure the relevancy of the peer group to the question at hand and to ensure peer group credibility.

Setting Thresholds to Ensure Relevancy

The most common reason to conduct secondary screening is to ensure that peer agencies are relevant to the question at hand. For instance, the Texas DOT conducted a peer comparison for Laredo Transit to better understand the transit agency's funding options after Laredo's population reaches 200,000, at which time it will no longer qualify for state funding. The peer comparison focused on the mix of funding sources for peer agencies that operated in urban areas larger than 200,000, and thus it required a minimum population of 200,000 for any peer agency. Similarly, Knoxville Area Transit was specifically

Table C4. Potential applications of peer-grouping variables for secondary screening.

Peer-Grouping Variables	Potential Secondary Screening Applications
Urban area population	Commonly used for secondary screening, either because peers must fall into a specific population category to be relevant (e.g., same FTA funding eligibility) or because vastly different urban area populations may hinder the credibility of a given peer. Larger population tolerances are acceptable for larger urban areas because they will naturally have fewer peers. Typical population tolerances may range from 25% to 50%. Note that in most cases, the methodology will naturally select peers within this range.
Total annual vehicle miles operated	May be appropriate for secondary screening for operations- and finance-related applications, where the scale of a peer agency's operations is particularly important.
Annual operating budget	May be appropriate for secondary screening for operations- and finance-related applications, where the scale of a peer agency's operations is particularly important.
Population density	Used when typical regional land-use patterns are important to the comparison.
Service area type	May be appropriate for secondary screening where all peer agencies must operate comparable service. For instance, an agency that runs all service in a region may wish to only compare itself to agencies that do the same. In most cases, the methodology will select peers with identical service types, but will also include some agencies with similar service types (e.g., service that extends outside the urbanized area).
State capital (yes/no)	Not typically used for secondary screening but can be used when evaluating a marginal candidate peer (i.e., one with a total likeness score >0.74).
Percent college students	Not typically used for secondary screening, although an agency operating in an area with a high student population may set a minimum percentage for peer agencies. This may be particularly applicable for funding-related questions since systems serving large universities may receive funding from the university and/or be more likely to have free or reduced fares on at least some routes. For smaller college towns dominated by the presence of a university, the methodology will tend to select other college towns.
Population growth rate	Used when regional growth (or shrinkage) and an agency's response to the growth is important to the comparison.
Percent low-income population	Not typically used for secondary screening but can be used when evaluating a marginal candidate peer (i.e., one with a total likeness score >0.74).
Annual roadway delay (hours) per traveler	Not typically used for secondary screening but can be used when evaluating a marginal candidate peer (i.e., one with a total likeness score >0.74).
Freeway lane miles (thousands) per capita	Not typically used for secondary screening but can be used when evaluating a marginal candidate peer (i.e., one with a total likeness score >0.74).
Percent service demand-responsive	May be used for secondary screening, particularly if an agency dedicates an unusually large portion of its budget to demand-responsive service and wishes to have a peer group that does the same.
Percent service purchased	May be used for secondary screening, particularly for finance and operations-related comparisons. Not only may the amount of purchased service have a significant impact on operations, but some NTD measures (e.g., operating employee FTEs) are not reported for purchased service, limiting the usefulness of agencies with purchased service for certain comparisons.
Distance	May be used for secondary screening when having relatively nearby peers will aid in stakeholder acceptance of the process due to being familiar with the peers.

interested in various types of dedicated local funding sources and therefore only selected agencies with dedicated local funding sources as peers.

Setting Thresholds to Ensure Peer Group Credibility

The peer grouping methodology developed by the research team is specifically designed to produce a reasonable peer group with no secondary screening, and the results of the case studies indicate that this is typically the case. However, there may be some instances when an agency feels that a threshold should be set for a particular variable to ensure the credibility of the resulting peer group. This is most likely to be necessary when an agency's uniqueness limits the number of close-fitting peers.

For instance, King County Metro is the largest bus-only operator in the country, limiting the number of good-fitting peers available. Because of this, the TCRP Project G-11 methodology returned several much smaller transit agencies within the same urban area that King County Metro would not consider as peers. Although the final methodology was adjusted to address this situation, agencies that are among the largest in their class may still find it appropriate to set a minimum threshold for peers based on vehicle miles operated.

Step 3a: Register for FTIS

This step is self-explanatory, and no additional detail is provided.

Step 3b: Form an Initial Peer Group

In most cases, forming a peer group is simple, requiring only a straightforward application of the peer-grouping tool provided by FTIS. However, in some instances, agencies will wish to refine the peer grouping methodology to better fit their needs and/or may need to select a peer group from agencies whose likeness scores do not indicate close fits.

Refine Grouping Methodology

By exporting the FTIS peer grouping results table into an Excel spreadsheet, agencies are able to refine the methodology to provide a peer group that meets their individual needs. Most commonly, this will involve removing a specific peer grouping variable for one of three reasons:

1. The agency does not feel that the variable is relevant for establishing its peer group. For instance, the peer grouping methodology assigns a high weight to agencies that operate rail service in order to avoid selecting a bus-only agency as a peer for a rail-operating agency and vice versa. Two large bus-only agencies, King County Metro and Orange County Transit Authority, however, expressed

no problem with including rail agencies in their peer group, while a third (PACE in suburban Chicago) preferred a bus-only group. Distance is another peer-grouping factor that may make sense for agencies in isolated locations such as Hawaii or Alaska to eliminate.

2. The agency does not wish to exclude agencies with missing data for a particular measure from being in its peer group because the factor is not essential for the performance question being asked. For instance, GBTA indicated in the case studies that they would prefer not to exclude agencies with missing congestion data from the peer grouping.
3. A potential peer is operated by multiple NTD reporters (e.g., the Trinity Railway Express commuter rail line, which is jointly operated by the transit agencies in Dallas and Ft. Worth) and the operators' data need to be combined.

Table C5 shows an example peer grouping refinement for Oahu Transit Service to eliminate the distance factor. Through the screening, several potential peers are available for Oahu with total likeness scores less than 1.0, whereas none were available previously. Overall, the screening had the effect of replacing several west coast agencies with agencies located elsewhere in the country. Note, however, that five of the top six peers identified through the screened method are still located in warmer areas of the country (California, Texas, Nevada, and Florida) due to sharing other demographic characteristics with Honolulu.

Interpreting Likeness Scores

Agencies with unique characteristics will often have few potential peer agencies with total likeness scores that meet the ideal thresholds described in the final report. Many agencies may have difficulty identifying a full peer group with likeness scores less than 0.75, and some may even have difficulty finding peers with likeness scores less than 1.0. The final report provides several potential reasons why agencies may be unable to find a large peer group, several of which were encountered in the case studies.

For instance, North County Transit District (NCTD) in Oceanside, California, is unique among suburban bus operators in that it also operates commuter rail service (the "Coaster" train) and a diesel light-rail line. As a result, only three agencies received likeness scores less than 0.75, as shown in Table C6. Moreover, the top-ranked agency (Caltrain) runs primarily commuter rail service with only minimal bus service, making it a poor choice for an agency-wide comparison.

Despite this, NCTD was still able to form a peer group with which they felt comfortable and which provided useful results. However, moving ahead with a peer group with higher likeness scores requires that analysts pay greater attention to the performance results to understand if performance differences are likely caused primarily by fundamental differences between

Table C5. Example peer-grouping refinement for Oahu Transit Service.

<u>Peer Group Using Distance Factor</u>			<u>Peer Group Without Distance Factor</u>		
Rank	Agency	Likeness Score	Rank	Agency	Likeness Score
1	Las Vegas RTC	1.24	1	Capital Metro (Austin, TX)	0.75
2	Capital Metro (Austin, TX)	1.31	2	Central Florida RTA (Orlando, FL)	0.75
3	City of Phoenix Transit	1.33	3	Las Vegas RTC	0.85
4	Kansas City Area Transportation Authority	1.44	4	Kansas City Area Transportation Authority	0.86
5	Omnitrans (San Bernardino, CA)	1.45	5	City of Phoenix Transit	0.92
6	Central Florida RTA (Orlando, FL)	1.48	6	VIA (San Antonio, TX)	0.96
7	VIA (San Antonio, TX)	1.50	7	Capital District Transit (Albany, NY)	0.97
8	Riverside Transit (Riverside, CA)	1.52	8	Rhode Island Transit Authority	1.01
9	King County Metro (Seattle, WA)	1.62	9	Milwaukee County Transit	1.03
10	San Mateo County Transit	1.62	10	Omnitrans (San Bernardino, CA)	1.11

agencies. An alternative approach that could have been taken for the NCTD case study would have been to compare each of its modes separately, instead of doing an agency-wide comparison.

Step 3c: Perform Secondary Screening

As described above, secondary screening may be beneficial under several circumstances. Typically, secondary screening is relatively straightforward using the FTIS software once the thresholds have been identified in Step 2c.

Continuing with the King County Metro example, Table C7 shows an example of how a secondary screening could have been conducted for the top ten King County Metro peers. This

screening eliminates transit agencies with operating budgets of less than one-third of King County Metro's budget, as well as agencies with total likeness scores exceeding 1.0. Although having a budget only one-third of King County Metro's might seem like a low threshold to meet, there are very few bus-only operators in the country that meet that criterion, and AC Transit was an agency that King County thought was an appropriate peer.

Table C7 shows that two agencies (Pierce County and Snohomish County) would be eliminated based on operating budget, leaving four potential bus-only peers with likeness scores less than 1.0. While four peers is at the low end of the preferred size for a peer group, the secondary screening served to make the peer group more credible and relevant to King County. (Note that under the final version

Table C6. North County Transit District peer group.

Rank	Agency	Likeness Score	Used as Peer?
1	Caltrain (San Francisco, CA)	0.58	No
2	Sound Transit (Seattle, WA)	0.65	Yes
3	Fort Worth Transportation Authority	0.72	Yes
4	Sacramento RTD	0.86	Yes
5	Santa Clara VTA (San Jose, CA)	0.99	Yes
6	Utah Transit Authority (Salt Lake City, UT)	1.04	Yes
7	Bi-State Development Agency (St. Louis, MO)	1.05	No
8	Metro Transit (Minneapolis, MN)	1.06	No
9	Memphis Area Transit Authority	1.07	No
10	TriMet (Portland, OR)	1.15	No

Table C7. Example secondary screening process.

Agency	Total Likeness Score	Annual Vehicle Miles Operated (000,000s)	% King County Operating Budget	Retained as Peer?
King County Metro	0.00	53.4	100.0%	N/A
VIA Metropolitan Transit	0.76	26.8	50.2%	Yes
Alameda-Contra Costa Transit	0.76	21.2	39.7%	Yes
Snohomish County Transit	0.83	12.0	22.5%	
Pierce County Transit	0.89	14.2	26.6%	
Orange County Transit	0.92	32.2	60.3%	Yes
Pace	0.96	46.6	87.3%	Yes
City of Detroit	1.05	16.5	30.9%	
Capital Metro (Austin)	1.05	17.6	33.0%	
Milwaukee County Transit	1.06	22.0	41.2%	
San Mateo County	1.06	10.0	18.7%	

of the methodology, Snohomish County and Pierce County would receive scores of 1.26 and 1.60, respectively, eliminating them from consideration prior to the secondary screening process.)

Step 4a: Gather Performance Data

FTIS makes the collection of performance data a straightforward task for peer comparisons that rely on NTD data. However, gathering data may be the most challenging portion of a peer comparison where non-NTD data are used.

Export Data to Excel to Calculate Non-Standard Variables

As described earlier, most outcome measures take the form of ratios (e.g., *cost per boarding*). While FTIS provides many of these variables directly through its Florida Standard Variables list, users will often need to calculate some of these measures manually. Because FTIS allows users to export performance data directly into an Excel spreadsheet, these calculations are fairly straightforward. This method was used to calculate performance measures associated with numerous case studies and to develop the updated performance measure results used in the Chapter 5 case studies.

Collecting Non-NTD Data

Most peer comparisons will rely on NTD data because of the difficulty with collecting non-NTD data. However, agencies frequently encounter situations where NTD data do not address

a particular performance question. Through the case studies, two agencies chose to perform a peer comparison using non-NTD data: GBTA and RGRTA.

For the GBTA case study, eight peer agency comparisons were conducted to obtain information about absenteeism rates by job category. However, after 2 months only two had provided usable data, and the effort was aborted. For the RGRTA case study, seven peer comparisons were conducted to obtain information on customer service and satisfaction using eight non-NTD performance measures. Of the seven peer agencies, four were able to provide data, and the peer comparison was able to proceed.

In general, these experiences suggest several recommendations for using non-NTD data:

- Agencies wishing to use non-NTD data for peer comparisons will need to invest significant time and resources to be successful,
- Agencies may wish to select unusually large peer groups, knowing that a large percentage of peers will be unable to provide information, and
- Forming benchmarking networks may be the most effective means to gather non-NTD data over the long term.

Step 4b: Analyze Performance

Use Charts and Graphs to Visually Enhance Results

The case studies showed repeatedly the value of using charts and graphs (primarily generated in Excel) to display the peer comparison results. Such graphics serve to enhance

understanding of the results and increase the likely impact that a peer comparison may have on stakeholders or decision-makers.

There is no one “best” way to display the peer comparison results. The following graphics display several examples from the case studies that participants found useful (see Figure C1).

Normalize Results to Account for Inflation and Cost of Living

Normalizing performance results to account for inflation and/or cost of living can be an important component of increasing the usability of results. The process for doing so is described in detail in the final report under the case study for the Denver RTD.

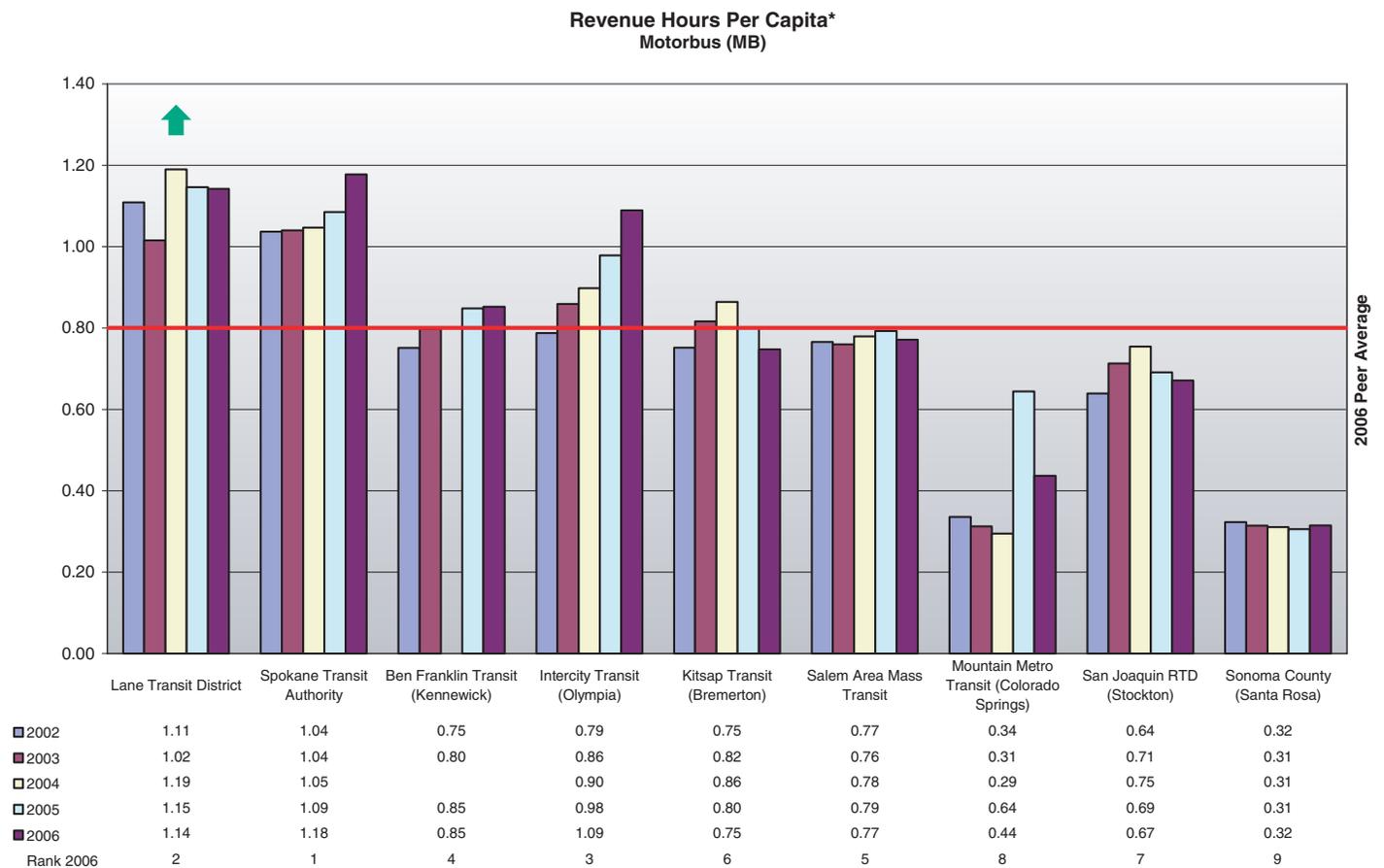
Gather Additional Descriptive Measures as Needed

The performance analysis will undoubtedly raise questions of why a particular agency does better or worse on a

given measure. In many cases, direct agency follow-up is required to understand the precise reasons for any result. However, additional descriptive measures may also help the analyst better understand the factors underlying the analysis results. For instance, an analysis that identifies a disparity in the number of vehicle malfunctions or vehicle maintenance costs between agencies may wish to also collect data on fleet age (if it has not already been collected) to assess whether this may contribute to the observed difference. The Altoona case study in the final report illustrates this process.

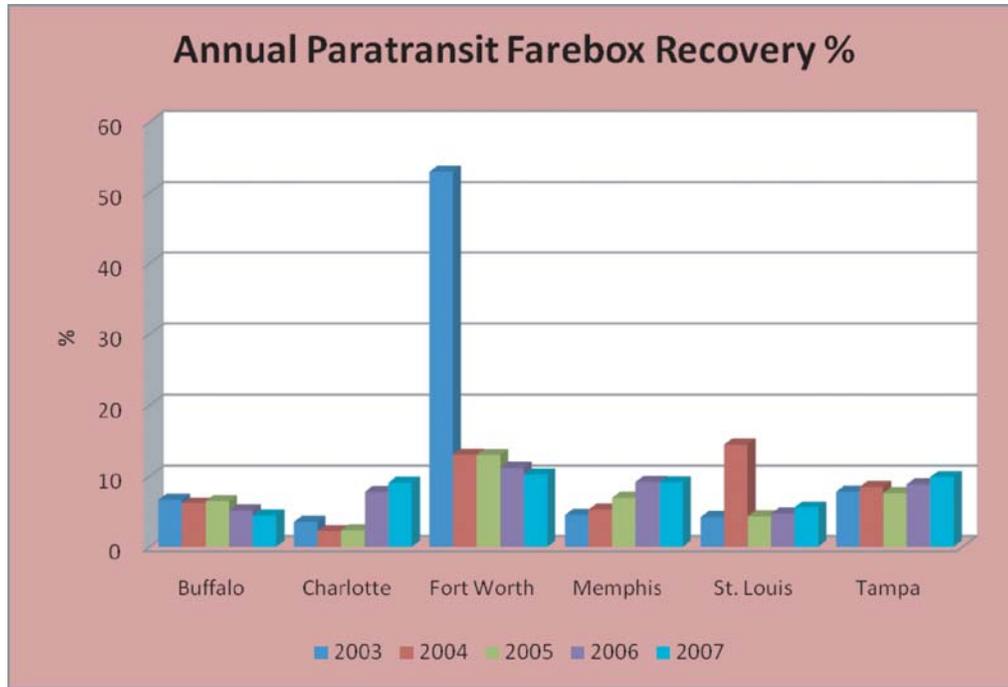
Performance Measure Recommendations

The case studies found that the NTD can be used to derive a variety of performance measures that are useful for peer-comparison applications. However, the case studies also found that many in the industry still have doubts about the accuracy of NTD data. Although data quality was generally not an issue

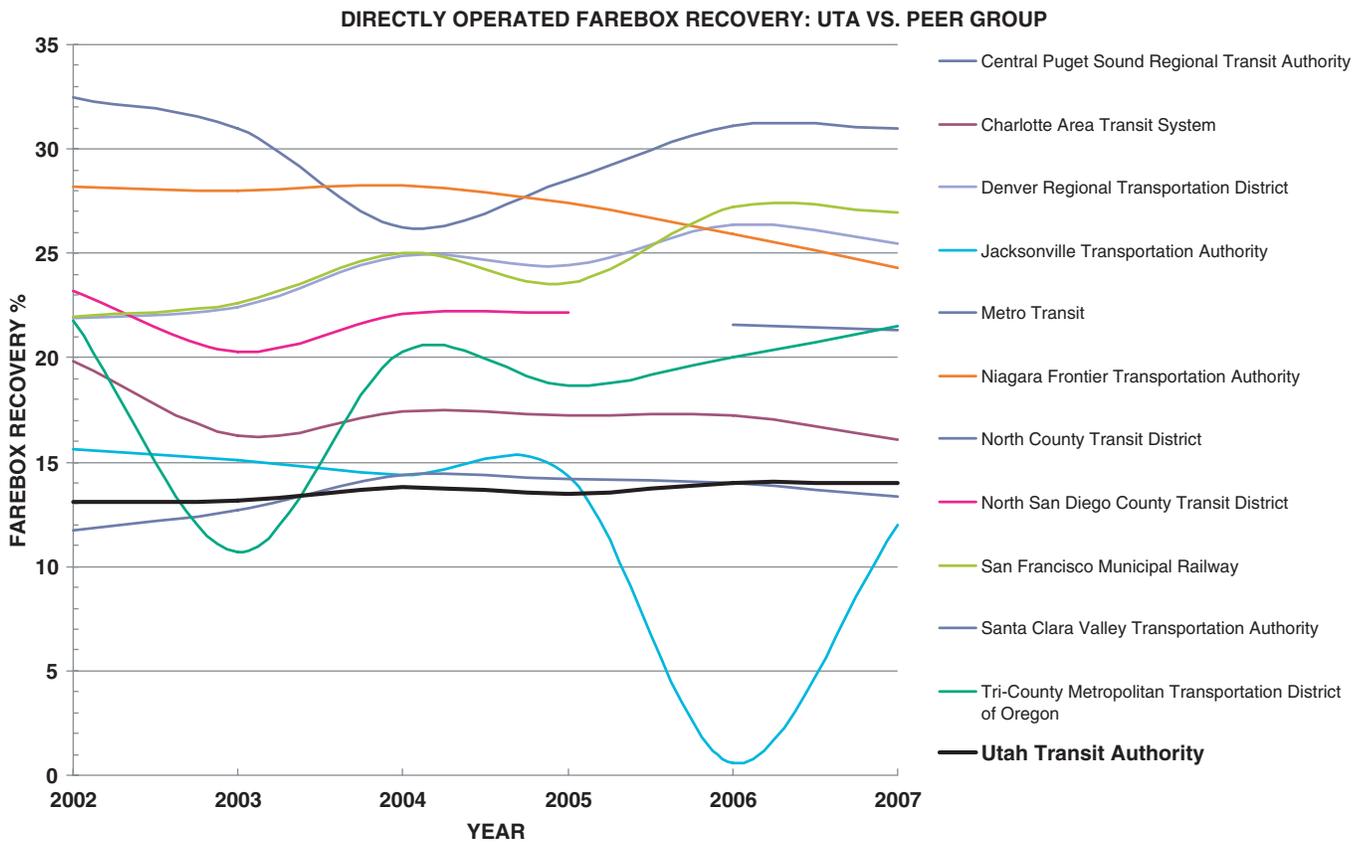


(a) Lane Transit District (Eugene, Oregon)

Figure C1. Examples of performance comparison graphics.

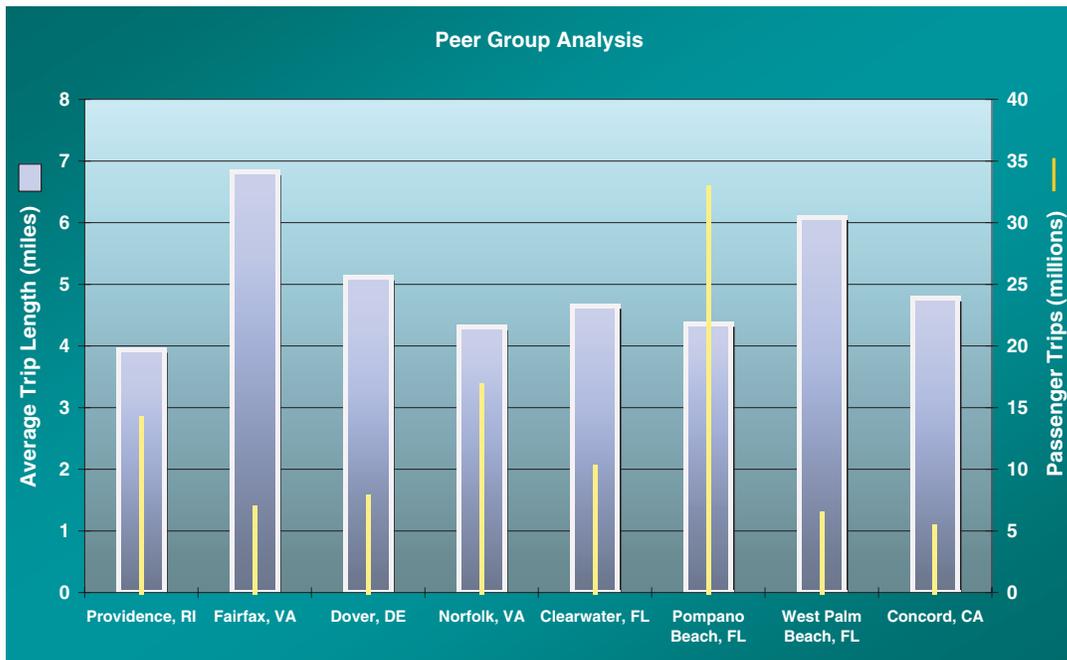


(b) Hillsborough Area Regional Transit (Tampa, Florida)



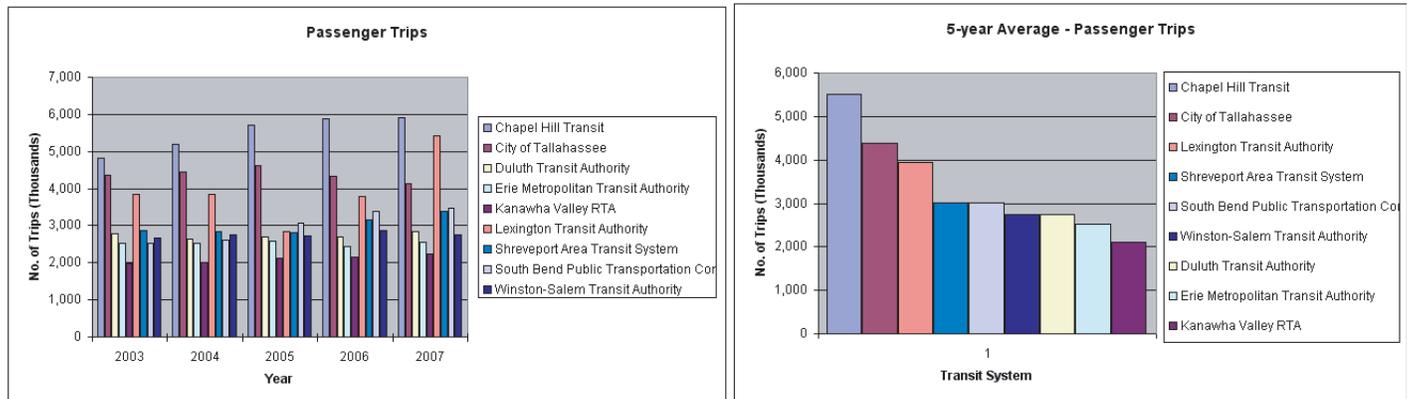
(c) Utah Transit Authority (Salt Lake City, Utah)

Figure C1. (Continued).



(d) Pinellas Suncoast Transit Authority (St. Petersburg, Florida)

	2003	2004	2005	2006	2007	5-year Average
Chapel Hill Transit	4,833,800	5,192,656	5,711,073	5,874,247	5,900,478	5,502,451
City of Tallahassee	4,372,762	4,459,371	4,612,725	4,333,213	4,136,790	4,382,972
Duluth Transit Authority	2,771,990	2,647,051	2,709,249	2,692,931	2,854,467	2,735,138
Erie Metropolitan Transit Authority	2,519,362	2,518,792	2,588,887	2,446,426	2,550,449	2,524,783
Kanawha Valley RTA	1,992,080	2,015,020	2,125,308	2,162,056	2,247,866	2,108,466
Lexington Transit Authority	3,835,065	3,846,412	2,838,102	3,794,115	5,422,578	3,947,254
Shreveport Area Transit System	2,861,687	2,827,977	2,814,415	3,145,320	3,381,851	3,006,250
South Bend Public Transportation Corp.	2,517,207	2,606,622	3,072,722	3,393,888	3,480,510	3,014,190
Winston-Salem Transit Authority	2,661,456	2,742,328	2,727,650	2,861,769	2,767,647	2,752,170



(e) Star Metro (Tallahassee, FL)

Figure C1. (Continued).

that came up during the case studies, there were occasions when isolated bad data were spotted, and there were particular measures observed where it appears that NTD reporters are not yet following the FTA’s guidance on how to calculate certain measures.

Because of the volume of data already reported to the NTD and because there is not yet widespread acceptance of the quality of NTD data or the value of NTD reporting, no new

NTD measures are recommended. Rather, the NTD-related recommendations focus on better reporting of certain existing NTD measures that would be valuable for peer comparisons. As NTD data quality improves, and as benchmarking becomes more of a standard transit industry practice, the time may come in the future when additional standard measures can be added. These are addressed in the second half of the recommendations.

NTD Recommendations

Service Area Size and Population

There are two significant issues with these NTD measures. First, the NTD defines service area as the area within three-quarters of a mile of bus routes and rail stations. However, many agencies report the entire urban area or transit district population and area. Second, only a single value is reported per agency, while the service areas of different modes operated by an agency may vary considerably.

Outcome measures based on per-capita ratios can be valuable comparison tools for comparing relative transit investments and productivity but require good service-population data to be able to make the comparison. This project's benchmarking methodology substitutes the combination of urban area population and agency service type as proxies for the number of people served. While these factors are useful in combination as a first-cut tool for identifying potential issues, it is readily acknowledged that they are not ideal and that service-area specific data would be preferable if the data could be relied upon.

Tracking regional population is not a normal transit agency function. MPOs, on the other hand, have the data and tools to readily perform these calculations. MPOs might be used in the future as the source of reliable service area and population data, rather than relying on transit agencies to supply these data.

Vehicle System Failures

There are few NTD-derivable measures that directly address the reliability of the service experienced by passengers. *Miles between vehicle system failures* is one such measure. However, the case studies found considerable inconsistency in how the NTD's *number of vehicle system failures* variable was reported, resulting in a lack of confidence in any measure based on that variable.

Track Miles versus Directional Route Miles

For commuter rail systems, the amount of single tracking impacts the amount of service that can be provided and, potentially, the reliability of that service. Comparing *track miles* and *directional route miles* should provide this information (1 mile of single track = 1 track mile and 2 directional route miles, while 1 mile of double track = 2 track miles and 2 directional route miles). However, all peer agencies in the Tri-Rail case study reported the same number of track miles as directional route miles, after rounding, indicating that *track miles* are not being reported correctly. (Some of the agencies were known to be mostly single-track operations and *track*

miles should also include sidings and non-revenue track, such as tracks in rail yards.)

Transit Industry Recommendations

The benchmarking literature indicates that a customer focus is important; however, very few NTD measures address service quality or customer satisfaction outcomes. At the same time, many agencies collect some form of service-quality data, for example by tracking complaints, conducting customer-satisfaction surveys, and measuring reliability and passenger loads. These data could be of great use in conducting peer comparisons related to service quality. The difficulty is that agency definitions for these measures are inconsistent. Rather than trying to force agencies to change their existing measures, it may be easier to encourage the development and storage of lowest-common-denominator measures that can be used to calculate related measures given a particular definition. For example:

- Measuring and storing *minutes early/late at time points* (something that is possible with automatic vehicle systems technology) can be used to calculate such reliability-related measures as *on-time performance* (using any desired definition of "on-time"), *excess wait time* (extra minutes passengers had to wait past the scheduled time, also accounting for early departures if desired), and *headway adherence*. These measures could be calculated for any desired time point along a line (e.g., start, middle, close to the end, end) and could be aggregated to a system average for a chosen location (e.g., the end of a line).
- For measures relating to passenger load, data on *number of passengers at the maximum load point* that was linked to specific fleet data (*number of seats*, already collected for the NTD, and *standing area*, not currently collected) would allow calculation of *load factor* and *area per standing passenger* measures. Data on passenger loads by route segment is already collected as part of the process for estimating passenger miles.
- *Number of complaints* is easy to track but for comparison purposes needs to be normalized by both the *number of comments received* and *total passenger boardings* since some agencies make it easier than others to submit comments and the number of people using transit service will obviously vary from agency to agency.
- *TCRP Report 47 (C-1)* provides recommended standardized questions for customer-satisfaction surveys. While every survey needs to be customized to the needs of the agency performing it, having industry agreement on at least a few core questions that could be asked using the same language and reporting scale would facilitate comparing customer satisfaction results between agencies.

Other case studies found desires for employee absenteeism data and more-detailed maintenance data, and there are any number of performance questions that could be conceived that would require specialized data. Rather than trying to encourage more NTD reporting, it may be more fruitful to encourage more widespread industry adoption of practices and definitions that will lead to greater collection and availability of standardized transit data.

The industry-related recommendations in the final report encourage (a) developing standard definitions for key non-NTD performance measures, (b) establishing performance measurement and benchmarking as standard practices through their standards program, and (c) establishing a confidential

clearinghouse for non-NTD data. The first recommendation addresses data-definition needs, the second would lead to more widespread internal agency data collection for a variety of needs, and the third would allow agencies to voluntarily and confidentially share their standardized non-NTD data with other agencies, for the betterment of the public transportation industry.

Reference

- C-1. MORPACE International, Inc., and Cambridge Systematics, Inc., *TCRP Report 47: A Handbook for Measuring Customer Satisfaction and Service Quality*. TRB, National Research Council, Washington, D.C., 1999.
-

Abbreviations and acronyms used without definitions in TRB publications:

AAAE	American Association of Airport Executives
AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
ACI-NA	Airports Council International-North America
ACRP	Airport Cooperative Research Program
ADA	Americans with Disabilities Act
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
ATA	Air Transport Association
ATA	American Trucking Associations
CTAA	Community Transportation Association of America
CTBSSP	Commercial Truck and Bus Safety Synthesis Program
DHS	Department of Homeland Security
DOE	Department of Energy
EPA	Environmental Protection Agency
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
HMCRP	Hazardous Materials Cooperative Research Program
IEEE	Institute of Electrical and Electronics Engineers
ISTEA	Intermodal Surface Transportation Efficiency Act of 1991
ITE	Institute of Transportation Engineers
NASA	National Aeronautics and Space Administration
NASAO	National Association of State Aviation Officials
NCFRP	National Cooperative Freight Research Program
NCHRP	National Cooperative Highway Research Program
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
PHMSA	Pipeline and Hazardous Materials Safety Administration
RITA	Research and Innovative Technology Administration
SAE	Society of Automotive Engineers
SAFETEA-LU	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (2005)
TCRP	Transit Cooperative Research Program
TEA-21	Transportation Equity Act for the 21st Century (1998)
TRB	Transportation Research Board
TSA	Transportation Security Administration
U.S.DOT	United States Department of Transportation