

Public **JOURNAL OF** Transportation

■ Joseph P. Schwieterman, Ph.D.
Lauren Ames Fisher

The S-Curve of Technological Adoption:
Mobile Communication Devices on Commuter Trains
in the Chicago Region, 2010–2015

■ Dr. Louis A. Merlin, AICP

Comparing Automated Shared Taxis and
Conventional Bus Transit for a Small City

■ Yanbo Ge
Parastoo Jabbari
Don MacKenzie
Jiarui Tao

Effects of a Public Real-Time Multi-Modal
Transportation Information Display
on Travel Behavior and Attitudes

■ Xin Li, Ph.D.
Yingling Fan, Ph.D.
John W. Shaw, P.E.
Yunlei Qi

A Fuzzy AHP Approach to Compare Transit System
Performance in US Urbanized Areas

■ Fatemeh Baratian-Ghorghi, Ph.D.
Hossein Ahmadianyazdi, M.Sc

Recommendation of a New Transit Performance Measure
in the National Transit Database

■ Eric Bruun
Marianne Vanderschuren

Assessment Methods from Around the World
Potentially Useful for Public Transport Projects

■ Emiel Cracco
Nicolas Dirix
Christopher P. Reinders Folmer

The Role of Specificity and Apologies
in Excuse Messages Following Train Delay

■ You-Lian Chu, Ph.D.
Yi Deng, Ph.D.
Rongfang (Rachel) Liu, Ph.D.

Impacts of New Light Rail Transit Service
on Riders' Residential Relocation Decisions

Joel Volinski, *Editor*
Lisa Ravenscroft, *Assistant to the Editor*

EDITORIAL BOARD

Robert B. Cervero, Ph.D.
University of California, Berkeley

William W. Millar
American Public Transportation Association

Chester E. Colby
E & J Consulting

Steven E. Polzin, Ph.D., P.E.
University of South Florida

Gordon Fielding, Ph.D.
University of California, Irvine

Lawrence Schulman
LS Associates

Jose A. Gómez-Ibáñez, Ph.D.
Harvard University

George Smerk, D.B.A.
Indiana University

Naomi W. Ledé, Ph.D.
Texas Transportation Institute

Vukan R. Vuchic, Ph.D., P.E.
University of Pennsylvania

The contents of this document reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation, University Research Institute Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

CONTENTS

The S-Curve of Technological Adoption: Mobile Communication Devices on Commuter Trains in the Chicago Region, 2010–2015 <i>Joseph P. Schwieterman, Ph.D., Lauren Ames Fisher</i>	1
Comparing Automated Shared Taxis and Conventional Bus Transit for a Small City <i>Dr. Louis A. Merlin, AICP</i>	19
Effects of a Public Real-Time Multi-Modal Transportation Information Display on Travel Behavior and Attitudes <i>Yanbo Ge, Parastoo Jabbari, Don MacKenzie, Jiarui Tao</i>	40
A Fuzzy AHP Approach to Compare Transit System Performance in US Urbanized Areas <i>Xin Li, Ph.D., Yingling Fan, Ph.D., John W. Shaw, P.E., Yunlei Qi</i>	66
Recommendation of a New Transit Performance Measure in the National Transit Database <i>Fatemeh Baratian-Ghorghi, Ph.D., Hossein Ahmadianyazdi, M.Sc.</i>	90
Assessment Methods from Around the World Potentially Useful for Public Transport Projects <i>Eric Bruun, Marianne Vanderschuren</i>	103
The Role of Specificity and Apologies in Excuse Messages Following Train Delay <i>Emiel Cracco, Nicolas Dirix, Christopher P. Reinders Folmer</i>	131
Impacts of New Light Rail Transit Service on Riders' Residential Relocation Decisions <i>You-Lian Chu, Ph.D., Yi Deng, Ph.D., Rongfang (Rachel) Liu, Ph.D.</i>	152

The S-Curve of Technological Adoption: Mobile Communication Devices on Commuter Trains in the Chicago Region, 2010–2015

Joseph P. Schwieterman, Ph.D.
DePaul University

Lauren Ames Fischer
Columbia University

Abstract

Urban transit riders' use of mobile communication devices has grown markedly in recent years. Studies evaluating the usage of these devices have generally focused on only one or two points in time, limiting their ability to describe long-range trends. To foster insights into this issue, this study evaluated data from 15,531 passenger observations collected on 156 commuter trains on the metropolitan commuter rail system of Chicago, Illinois, from 2010 through 2015. The data show that the rate of technological usage is following an S-shaped pattern among passengers. The share of passengers using mobile communication devices at observed points grew sharpest during the first three years, rising from 25.6% in 2010 to 47.8% in 2013, a compounding annual rate of 23.1%. Between 2013 and 2015, the share rose to 56.2%, an annualized rate of just 8.4%. Over the five-year period, the share of passengers conducting visually-oriented activities on their devices increased at a faster rate than usage as a whole, whereas the share of passengers engaged in audio-only tasks has dropped. Multiple regression analysis shows that the rate of device usage on trains is highest on outbound trips (traveling away from downtown) and positively related to the income associated with the route traveled, with differences of more than five percentage points between lines of varying levels of affluence.

Keywords: *Public transportation, technology, mobile communication devices*

Introduction

Many passengers on urban transit systems routinely use mobile communication devices (“mobile devices”) while making their journey. Mobile devices such as smartphones and tablet computers enable passengers to perform work or leisure activities, access travel information, and insulate themselves from the noise and distractions normally experienced during their trip. As devices have become smaller, less costly, and more sophisticated, the range of activities performed by passengers has grown sharply.

This study documented the changing prevalence of mobile device use among commuter rail riders in metropolitan Chicago to assess how changes in technology affect the way passengers allocate their time while using commuter rail. The study draws upon 15,531 unique passenger observations collected on 156 suburban Metra trains in the Chicago metropolitan area between 2010 and 2015. Although other studies have explored the use of mobile devices on transit systems, this study is unique in providing a time-series assessment for a five-year period, allowing the authors to gauge how the use of mobile devices has changed over time and to identify trends in technology use by travelers that may aid in the planning and management of urban transportation.

The results can help guide the discussion about how transit agencies can take advantage of a growing range of digital activities available to passengers as well as respond to the “digital divide” that persists among various income groups in the use of technology. Finally, the data can help transit agencies evaluate the need for enhancements that support technology use among passengers that have been unable or hesitant to extensively use mobile devices as part of their journey.

Background

As recently as the late 1980s, most electronic devices were too bulky and had a battery life that was too limited to be conveniently brought onto buses and trains (Farley 2005). Passengers generally had little or no ability to communicate with anyone except fellow travelers onboard. When communication technology first became available to travelers, it was available exclusively on intercity (longer-distance) services rather than on urban public transit. The first commercial cellular phone service in the United States was introduced in early 1969 on Metroliner trains between New York and Washington, DC (Farley 2005), and airline carriers first installed cellular phones in the early 1980s, which could be used by customers paying with a credit card. By the late 1990s, pay phones also had become available on many intercity rail-passenger routes in Europe and North America.

Technological innovation made spectacular advances by the late 1990s as cell phones and laptop computers became smaller, less expensive, and more sophisticated. These advances were followed, in the early 2000s by devices oriented toward multitasking, such as smartphones. More recently, tablets and e-readers have further enhanced the ease of personal entertainment and communication for passengers, a topic discussed in detail later in this paper.

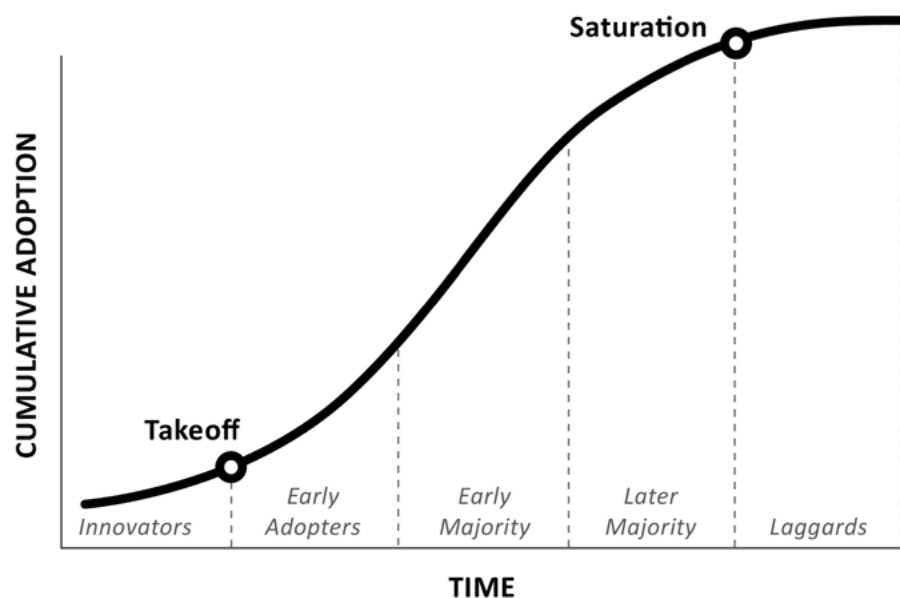
Some public transit operators have responded to these trends by investing in Wi-Fi systems on buses, and a smaller number have done so on commuter trains. The transit system in Albuquerque, New Mexico, for example, installed Wi-Fi on its trains in 2005, making it a pioneer in expanding internet access for passengers. In 2008, the transit operator in Boston (MBTA) launched a pilot program for the installation of Wi-Fi and has since made it available on nearly the entire system (MBTA 2013). In 2009, the Bay Area Rapid Transit (BART) system began providing wireless service on a portion of its system, and since then, commuter rail operators in Miami, Minneapolis-St. Paul, and New Jersey have also selectively installed Wi-Fi. Transportation systems outside the US also have offered these amenities to their passengers.

Nevertheless, the number of commuter rail systems that have installed Wi-Fi on their entire fleet remains relatively small. In Chicago, the suburban operator Metra did not offer Wi-Fi at the time data were collected for this study. The provider initially determined that offering the amenity on all its railcars was cost-prohibitive but adopted a plan for introducing Wi-Fi service on an experimental basis in selected cars on some of its trains in 2016 (Metra 2015).

Relevant Literature

The literature on the adoption of technology is shaped by Everett Rogers’ seminal work *Diffusion of Innovations*, which attempts to explain how, why, and at what rate new ideas and technology spread through cultures. A major contribution of Rogers’ work lies in his categorization of technology adopters based on their propensity to embrace new ideas, including the popular term “early adopter.” His work shows that rates of adoption tend to follow an S-shaped pattern that is affected by innovations, communication, social influences, and other factors (Rogers 1962). By the time the “late majority” adopts a technology, growth has sharply diminished (Figure 1).

FIGURE 1.
S-curve model of diffusion process (adopted from Rogers 1962).



Several decades would pass before mobile communication devices became a factor in decisions relating to urban travel. Gradually, however, as cell phones and other devices became prevalent, research drew upon Rogers' ideas to explore the complex interactions between technology and travel behavior (Andrew 2010; Golob and Regan 2001). Kim et al. (2008) examined how people use technology to plan for personal and business travel, while Yoshii and Sasaki (2010) studied how technology affects the need for non-business travel, focusing particularly on consumer behavior and online shopping. Much has changed, of course, since these studies were published.

Brakewood et al. (2014) use a stated preference survey of passengers on a pair of commuter rail lines in the Boston region to assess the degree to which passengers use mobile communication devices for mobile ticketing. The study concludes that among the approximately 76% of riders that used smartphones in 2012, 26% used them for mobile ticketing and 50% used them for making mobile purchases at other merchants. A related strand of research explored the varying rates of technology use by different socioeconomic groups while traveling (Jackson et al. 2008; Kim et al. 2008). A consulting firm, Ninth Decimal (2015), tracked the types of devices used by consumers when researching and purchasing tickets for transportation as well as the share of consumers relying on public Wi-Fi for these tasks.

Windmiller et al. (2014), through a survey of passengers on the light rail system in St. Louis, showed that passengers who use smartphones or access electronic messages, including emails and texts, tend to be more satisfied with their travel experience. These passengers also report greater feelings of personal security at transit stations and higher service quality when making connections. This study also demonstrated how rates of technology use are rising; whereas just 52% of passengers used their phone's text messaging capability in 2008, the number rose to 88% in 2012. A study by Tang and Thakuria (2010), focused on bus routes in Chicago, concluded that providing passengers access to real-time information on their smartphones fosters a slight increase in ridership compared to routes on which this was unavailable. The study isolates this effect while controlling for a wide variety of neighborhood-specific variables. Following a similar theme, Barbeau et al. (2010) explored how new mobile apps enhance rider experience by providing real-time information about their trips, including alerting customers when to disembark or enabling riders to signal the driver to stop. Guo et al. (2015) observed 1,739 passengers and surveyed 686 to evaluate usage patterns on the bus system in Vancouver, British Columbia. Their study used a predictive model to evaluate the factors (such as crowding and waiting conditions) that affect how riders spend their time.

Particularly relevant to the present study is the analysis of how mobile communication use affects the perceived cost of travel. Some of the most extensive research on this topic has been undertaken by Mokhtarian and Solomon, whose 2001 paper provides a framework for evaluating the value of time spent on buses and trains (Mokhtarian and Salomon, 2001). Their analysis, based on a study of residents in the San Francisco Bay Area, contested the notion that travel is purely a derived demand by providing evidence that many travelers take pleasure in the trip itself and benefit from the activities they conduct while making the trip. A later paper by Mokhtarian and Salomon (2014)

evaluated the utility that passengers gain from time spent traveling via train through Great Britain and showed that travelers derive significant benefit from the technology they bring with them. Lyons et al. (2012) employed similar methods to measure how rail passengers in Great Britain spent their travel time from 2004 and 2010. Not only did technology use dramatically increase for the three types of trips evaluated over this period—commuting, business, and leisure—but certain activities, such as listening to music, more than doubled in most categories considered.

Several other studies evaluated some of the psychological aspects of having the ability to use technology. Jain and Lyons (2008) were early contributors to this area of research by exploring how time spent traveling often is perceived as a gift that affords travelers transition time in such a way that it fills a need for “experiencing distance and the opportunity for gearing up to the destination’s demands” (83). Another benefit these authors identify is providing a “time out,” i.e., the opportunity to escape from obligations and provide “backstage” time to make a mental transition from one activity to another.

Ohmori and Harata (2008) showed that the amount of personal space available influences technology activity, and Holley et al. (2008) postulated that the value of travel time on trains is closely linked to passenger ability to perform tasks using devices. The interaction between technology use and traveler behavior also is explored by Simun (2000), Bull (2000), and Skanland (2011).

The final group of relevant research looks specifically at the effects of personal technology across modes on longer distance intercity trips. Hess and Spitz (2015), evaluating data from the Northwest and West Coast corridors, concluded that a desire to use technology has a positive effect on the propensity to travel by train and bus compared to driving for trips between 200 and 350 miles. Nevertheless, their research also shows that the importance passengers place on using mobile devices is small when compared to other factors, such as schedule convenience. Russell et al. (2011) observed bus and train riders in New Zealand, concluding that bus travelers are less apt to use mobile devices than train riders.

In summary, a rich body of literature explores the ramifications of the growing use of mobile communication devices on transit. What remains less clear is how the rate of growth in mobile device usage has changed in recent years, and the extent to which growth has followed the S-shaped pattern postulated by Rogers (1962). This study seeks to fill this gap in the literature by presenting findings from a five-year observational study on mobile device usage by commuters in one metropolitan region.

Chicago’s Commuter Rail System

The metropolitan Chicago region offers a rich environment for exploring the use of mobile communication devices in urban travel. It is home to one of the most extensive commuter rail systems in North America and consists primarily of routes that date back to the 1800s and radiate from the central city (known as “The Loop”). Most of these routes were operated by profitable railroads through the 1950s or 1960s until the

private automobile became affordable to a wider range of income groups. As services became increasingly unprofitable, Chicago-area residents approved a referendum in 1973 to create the Regional Transportation Authority to subsidize commuter rail, bus, and rail rapid-transit services in the metro area.

The region's commuter rail system, branded as Metra, currently has a 292,000 average weekday ridership over 11 main lines and 4 branch lines radiating out of downtown Chicago (APTA 2016). These lines depart from one of four downtown terminals to 237 outlying stations. Metra owns, maintains, and directly operates four of these main lines as well as three branches, and the others are operated under contract with private railroads. Metra's ridership has grown markedly in recent years; after modest declines in 2012, partially due to a fare increase, ridership grew 3% in 2013 and 12% in 2014. Ridership was also up slightly in 2015 (Metra 2016).

There is limited research on the factors affecting technology use among Chicago-area transit riders. Two studies looked specifically at factors affecting technology use among this group. Frei and Mahmassani (2011) administered a survey to Chicago Transit Authority passengers and determined that demographic issues and the quality of the environment significantly affect technology use. Older adults were found to be less likely to use devices than younger travelers, and females more likely than men. The other body of work relies on various passenger surveys commissioned by Metra. A 2011 survey showed that about half of all Metra riders used a smartphone during their trip, 27% carried a laptop, and 6% had a wireless card they could use while traveling (Metra 2013). Relatively few passengers (about 11%) report being willing to pay a fee for wi fi on board. A 2014 survey showed that 66% of passengers noted that the ability to "read/work/nap" was important to them (Metra 2015). The survey, however, did not include questions focusing specifically on mobile devices.

Methodology

The data presented below were part of *Technology in Travel*, a multi-year project that involved direct observation of passengers on various modes of transportation between 2010 and 2015. To measure the use of mobile communication devices among passengers, a data team passed through the aisles of buses and trains to record data on passenger technology device usage. Data collectors traveled as regular fare-paying passengers and collected data in real-time settings. To allow for greater consistency in comparisons among modes, this study considered only weekday trips.

Data collection began 10 minutes after leaving the station to ensure that passengers had time to get situated. Data were collected only on weekdays between 8:30 AM and 7:30 PM. The sample size differed by train, depending on the passenger load and time available for data collection. The data collection team adhered to a consistent protocol when responding to situations that created issues for data quality. For example, when two passengers are using the same device, only the passenger most closely situated to the device was counted as using the device, making our estimates conservative. Passengers who appeared to be below grade-school age (5th grade or younger)

were excluded from the observational count. Passengers using a set of earbuds or headphones plugged into an electronic device and who appeared to be sleeping were counted as using an audio device. Only when clear and unobstructed views were possible did the team record data of passengers sitting on upper levels of gallery cars. “Quiet cars” (in service only during peak periods), in which talking is banned, were not included in the sample. The number of riders observed on routes extending north, northwest, west, south, and southwest of the city was based on each region’s share of total ridership. The number of observations by region is depicted in Figure 2.

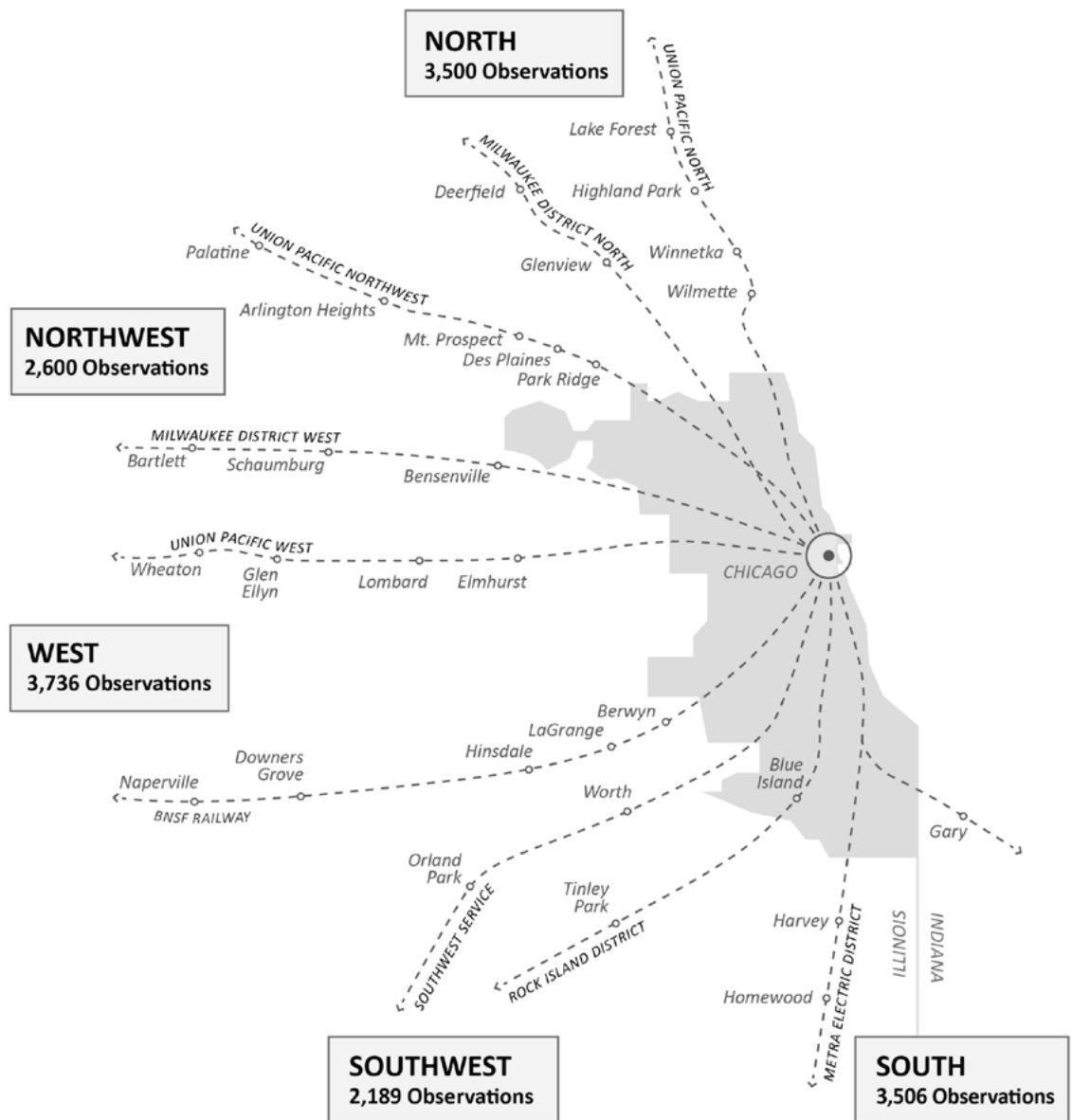


FIGURE 2. Passenger observations on commuter rail services in metropolitan Chicago, 2010–2015. Sample size by regional subcategory.

In addition to counting the number of passengers using an electronic device, data collectors also categorized each type of observed device use based on the type of activity being conducted by passengers (Table 1). Categories denoted whether a passenger was using a mobile device and, if so, whether the task was strictly audio (such as listening to music or talking on a cellular phone), visual, or audio-visual in orientation (these tasks typically involve using an LCD screen). Tablets and e-readers (including “phablets”—mobile devices that straddle the size of smartphones and tablets) were broken out into a separate category from other visually-oriented (LCD) devices starting in 2012.

TABLE 1.
Type of Activity Recorded

Activity	Description
No technology being used	Passenger not engaged with personal electronic device.
Audio activities	Passenger engaged in a task such as those involving cellphones or music players that can be used with earphones, speakers, or headsets, using only audio functions of mobile communication devices.
Visual or audiovisual activities on devices, not including iPads, Kindles, and other tablet use	Passenger engaged in visual or audiovisual features on laptop computers, Blackberries and other smartphones, DVD players, and iPods (includes any traveler looking at LCD screen for purpose of engaging in activity more substantial than placing a phone call).
Visual or audio-visual activities on iPads, Kindles, and other tablets or “phablets”	Same as Category 2 except focusing specifically on tablet and “phablet” use. Category introduced in 2012 to capture how small, lightweight devices affect traveler behavior.

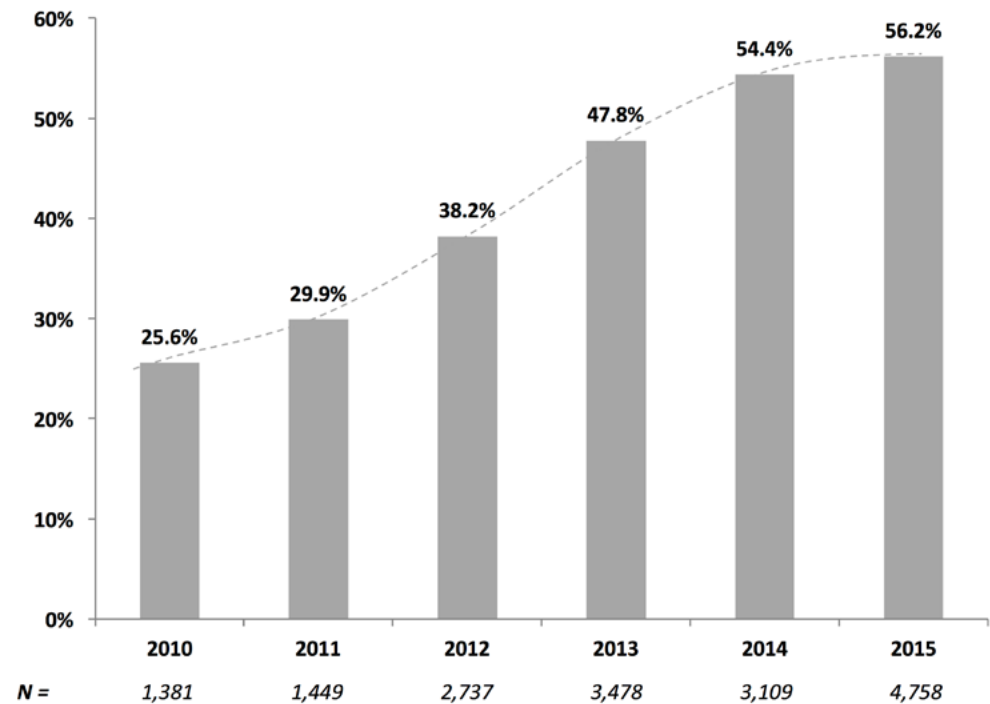
Data collection on passenger use of mobile devices began in 2010 and continued annually through 2015. For the first two years of the study (2010 and 2011), data were collected from December through March. In the following years (2012–2015), data were collected between January and May. The above issues may affect comparisons between the first two years of data collection and the last three years if it is assumed that technology use on transit varies considerably by month. The experience with this project suggests that this change in the months during which data collection occurred does not considerably change the conclusions presented herein.

As noted in Table 2, the number of passenger observations gradually grew from about 1,400 in 2010 and 2011, to 2,737 in 2012, 3,478 in 2013, and more than 4,000 in the final year. The sample was purposely expanded to allow for the analysis of sub-regional differences in usage (a topic considered in this paper by evaluating the effects of income differences along different lines). The number of trains sampled rose from 14 in 2010, to 18 in 2011, 24 in 2012, 36 in both 2013 and 2014, and 42 in 2015. The sample included at least 2,000 passengers on trains operating south, southwest, west, northwest, and north from downtown Chicago (Figure 2) and included both rush hour and non-rush hour. The number of observations was limited so that no single train accounted for more than 8% of all observations through 2013 and no more than 5% in subsequent years. Measures of onboard crowding were not considered due to the carrier’s ability to provide seats to the overwhelming majority of riders. (Some passengers nonetheless choose to stand rather to take empty seats when trains are nearly full, suggesting that this variable may be worthy of consideration in future analysis).

Descriptive Statistics

The data show a gradual increase in technology use between 2010 and 2015 as well as a shift toward more sophisticated technologies. Perhaps the most salient result is that the share of passengers using mobile communication devices gradually rose over the five-year project period despite changes in the economic performance of the regional economy. At randomly-selected points, 25.6% of passengers were engaged with technology in 2010, compared to 38.2% in 2012, 54.4% in early 2014, and 56.2% in 2015 (Figure 3). Since 2014, more than half of passengers have been engaged with personal devices at the observed points, compared to slightly over a quarter in 2010.

FIGURE 3.
Percent of commuter rail riders using mobile communication devices at observed points.



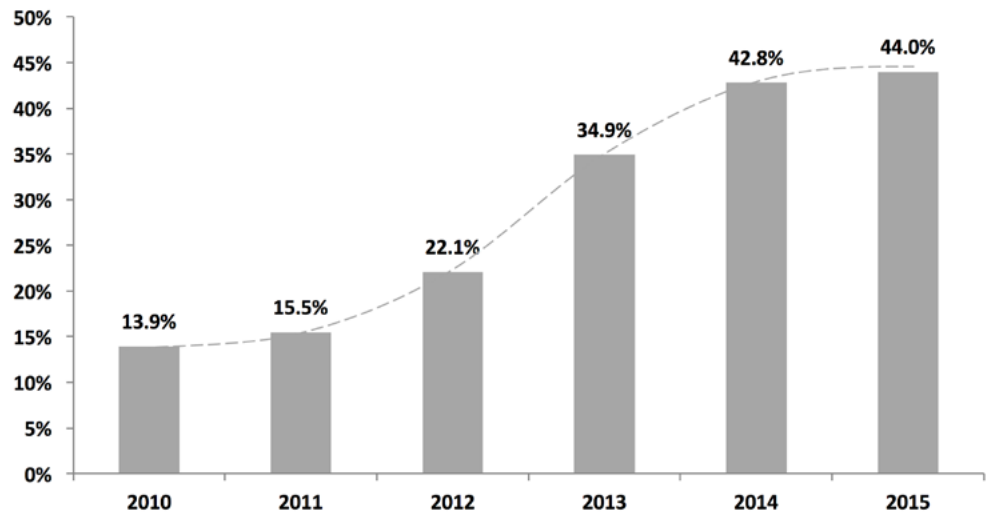
Although the methodology does not allow for measurement of the percentage of passengers who use technology at some point during the trip (which Metra estimates to be as high as 90%), the results suggest the intensity of technological engagement has risen steadily over the five-year period. The change in usage between each interval is statistically significant (99% confidence interval) with the compounding rate of growth from 2010–2013 being 23.1%, compared to just 8.4% from 2013–2015.

Observations also show that e-reader and tablet usage (tracked as a separate category) has been on the rise since 2012. The small sizes and quick boot-up times of these devices make them well-suited for space-confined environments and trips of shorter duration. Unlike laptops and notebook computers, tablets and e-readers (including “phablets”) can be stored easily in a briefcase or purse and take up much less room when in use. In 2012, the first year tablets were measured as a distinct category, just 4.9% of passengers were using tablets or e-readers. That share rose to 6.4% in 2013, 7.9% in 2014, and 9.3%

in 2015. As a result, by 2015, about 1 in 11 passengers was using a tablet or e-reader compared to just 1 in 20 in 2012. The share of all mobile devices users engaged with tables and e-readers rose from 12.8% in 2012 to 16.5% in 2015, suggesting that, although these devices are becoming more common, they remain far less prevalent than other types of devices.

As more travelers turn to sophisticated devices, such as tablets and e-readers, they increasingly diminish the time spent solely on audio-oriented functions, such as cellphone calls and listening to music. The share of Metra riders performing visually-oriented tasks at randomly-observed points rose from 13.9% in 2010 to 44.0% in 2015 (Figure 4).

FIGURE 4.
Percent of commuter rail riders engaged in visually-oriented tasks at observed points.



Both Figure 3 and Figure 4 align with the S-curve of technology adoption originally proposed by Rogers (1962). The dotted lines on these figures show how neatly the S-shaped curve fits to the collected data. Both figures shows accelerating growth in device use through 2013, followed by a deceleration, suggesting a flattening-out effect, perhaps due to the growing saturation of mobile devices in the general population. Figure 4 shows a slightly more pronounced S-shaped pattern than Figure 3, with use of visually-oriented activities growing at a compounding annual rate of 34.9% from 2010 to 2013, compared to just 12.3% between 2013 and 2015. This suggests that additional gains likely will continue to be made at a much slower rate in the coming years. The policy implications of this are discussed in the conclusion.

The data collected allow for a comparison of technology use by time and direction of travel. As noted in the Appendix, in the most recent year surveyed (2015), the use of technology was significantly higher on outbound trips than inbound trips, which provides evidence that technology use may be more pervasive (and beneficial) to travelers after their workday than before, although the authors cannot determine whether activities are work- or leisure-related. Technology use on outbound trips during the peak period exceeds that on inbound trips during the peak by an even

wider margin. Both results are statistically significant at a 0.05 level. Please refer to the Appendix for details.

Explanatory Model

A multiple regression model was developed to further investigate changing patterns of technology use on commuter trains. The analysis encompasses 148 trains surveyed between 2010 and 2015 for which at least 20 observations were recorded on each train. (Eight of the 156 trains had fewer than 20 observations and were omitted due to their small sample size.) The dependent variable is the proportion of passengers on a train using mobile devices, and the independent variables are listed in Table 2.

TABLE 2.
Independent Variables

Independent Variables	
PEAK	Trains observed between 7:00–9:30 AM and 3:30-7:00 PM Monday through Friday assigned a value of 1.
OUTBOUND	Trains traveling outward from central city assigned a value of 1.
PEAKFLOW	Trains servicing inbound commuters during morning peak period and outbound during evening peak assigned a value of 1.
YEARS	Number of years passed since first year of data collection (2010).
YEARS ²	Square of years passed since first year of data collection (2010). Used to capture nonlinear relationship.
INCOME	Average household income in ZIP codes of all suburban stops served on Metra line where data collected, derived from American Community Survey data from 2010. Income measured in \$1,000 annually.
Interaction Terms	
PEAK*YEARS	Number of years multiplied by peak travel designator to determine if peak period differential has risen or declined over time.
INCOME*YEARS	Number of years multiplied by income of a route to determine if effects of income have risen or declined over time.
OUTBOUND*YEARS	Number of years multiplied by outbound designator to determine if outbound differential has risen or declined over time.

Using a proportion as the dependent variable can lead to certain biases because values are bounded between 0 and 1. The authors, therefore, considered two sets of results: with and without a Tobit transformation of the dependent variable. The Tobit transformation involves a logarithmic transformation using the formula $\ln(p/(1-p))$. This adjustment results in a dependent variable that is normally distributed but generates coefficients that lack simple intuitive interpretations.

The vast majority of dependent values are clustered near the middle of the range, between 0.3 and 0.7, and not censored from above or below, and there are no values in the dataset at the extreme ends of the distribution (i.e., less than 0.1 and more than 0.9). The discussion below, therefore, focuses primarily on the standard model, with the Tobit results provided in the Appendix. This standard model also has the advantage of generating coefficients that have a simple and straightforward interpretation. The independent variables used in the model are provided in Table 3.

TABLE 3.
Results of Multiple
Regression Model

	Model 1: Time of Day, Direction of Travel, + Changes Over Time		Model 2: Time of Day, Direction of Travel + Changes Over Time (Nonlinear)		Model 3: Direction of Travel + Changes Over Time (Nonlinear), Route Income		Model 4: Direction of Travel, Changes Over Time (Nonlinear), Route Income, Income-Time Interaction	
	Coeff.	P Value	Coeff.	P Value	Coeff.	P Value	Coeff.	P Value
INTERCEPT	0.282**	(0.000)	0.243**	(.000)	0.123*	(0.023)	0.102	(0.278)
PEAK	0.021	(0.259)	0.025	(0.175)	0.022	(0.263)	0.021	(0.253)
OUTBOUND	0.042*	(0.026)	0.039*	(0.037)	0.020	(0.051)	0.035	(0.053)
YEAR	0.066**	(0.000)	0.123**	(0.000)	0.120**	(0.000)	0.129**	(0.003)
YEAR ²			-0.014*	(0.018)	-0.012*	(0.031)	-0.012*	(0.031)
INCOME					0.002**	(0.009)	0.002	(0.109)
INCOME*YEAR							-0.000	(0.793)
R ²	0.399		0.422		0.449		0.450	
ADJ. R ²	0.386		0.406		0.430		0.426	

Dependent variable: Proportion of travelers using mobile communication devices

Observations = 148 trains.

*Significant at 0.05.

**Significant at 0.01.

The dependent variables PEAK, OUTBOUND, and PEAKFLOW relate to the type and timing of service provided. The variable YEARS represents the number of years elapsed since the base year, 2010, and is intended to capture the average change in technology per year due to exogenous factors that have occurred over time, such as changing consumer habits and greater adoption of 3G and 4G services. A polynomial (squared) version of this variable, YEARS², captures the nonlinear relationship observed in Figures 3 and 4.

The INCOME variable is the mean of the median household income in the zip codes where suburban Metra stations are located for each line. Tabulating these values involved collecting data from 140 ZIP codes in which Metra stations are located. Values were tabulated by Metra line; the Union Pacific–North line, for example, is located in the ZIP code with the highest household income (\$99,600), followed by the BNSF line (\$82,580). The Metra Electric (\$60,360) and South Shore lines (\$56,000), both serving the south part of the region, have the lowest income values. Although INCOME is a simplified estimate, it serves as a useful measure of the affluence of the corridor in which the train operates. Income may be important due to the “digital divide” that exists in the adoption of technology as well as the higher opportunity cost affluent passengers face when spending time on the train. Three interaction terms are used to evaluate how the effects of these variables have changed over time.

The analysis considered four models using the proportion of technology use as the dependent variable and the independent variables discussed above. The models are discussed below, and the statistical results are presented in Table 4.

TABLE 4.
Rates of Technology Use by
Time and Direction of Travel

	Sample Size	Using Any Technology		Conducting Visual Task	
		Yes	No	Yes	No
Peak	2,925	56.4%	43.6%	45.7%	54.2%
Off-Peak	1,823	55.8%	44.2%	41.0%	59.0%
		X ₂ =0.1656		X ₂ =10.2156*	
Inbound	1,844	51.1%	48.9%	36.9%	63.1%
Outbound	2,904	59.5%	40.5%	48.3%	51.7%
		X ₂ =32.2167*		X ₂ =60.2585*	
Peak Inbound	1,207	49.0%	61.0%	36.4%	63.6%
Peak Outbound	1,718	62.7%	57.3%	52.1%	49.9%
		X ₂ =46.7688*		X ₂ =67.0287*	

* Denotes significance at 0.01 level.

Model 1: Basic Linear Model

This model evaluates how technology use has changed over time while controlling for the direction of travel (OUTBOUND), the time of travel (PEAK), and the number of years that passed since 2010 (YEAR). The results shows that the proportion of riders using mobile devices has risen an average of 6.6% annually (YEAR) and that this change is statistically significant. The OUTBOUND variable was also significant and indicates that device use is 4.2% higher on outbound trips than on inbound trips. The time of travel (PEAK) did not have a statistically significant impact on rates of mobile device use. The addition of the OUTBOUND*YEAR variable did not significantly improve the model’s predictive ability (it produced a coefficient that was positive but not statistically significant) and was dropped. Model 1 explains 39% (Adj. R² = 0.386) of observed variations in mobile device use.

Model 2: Nonlinear Rate of Growth in Technology Use

Model 2 is similar to Model 1 but includes both the YEAR and YEAR² variable to assess whether the use of devices changed in a nonlinear manner. The coefficient for YEAR remains positive and significant, whereas coefficient for YEAR² is negative and also statistically significant. This indicates that although device use increased, it did so at a declining rate, a finding consistent with the S-shaped pattern in Figures 2 and 3. The OUTBOUND variable remains statistically significant, although its coefficient value decreased slightly.

Model 3: Effects of Income

Model 3 considers how device usage is affected by the average income of the commuter line by adding the INCOME variable to Model 2. The INCOME variable is statistically significant and indicates that for each \$1,000 of average income on the line, there is an increase of 0.21% in device use among passengers riding that line. This suggests that the highest-income line has observed device use that is about 6% higher than the lowest

income line in our study. Inclusion of the INCOME variables results in the OUTBOUND variable being no longer significant, suggesting that the affluence of riders has a greater impact on device usage rates than does the direction of travel.

Model 4: The Changing Effects of Income over Time

The final model evaluates how the effects of income have changed over time by considering an interaction variable between income and years of time elapsed since 2010 (INCOME*YEARS). Although the INCOME coefficient remains roughly the same, it is no longer significant in this model; the interaction variable INCOME*TIME is also not significant. Based on these results, there is little evidence to suggest that the effects of income are falling significantly over time even as mobile device ownership expands. Nevertheless, the coefficient is negative, and it is possible that this effect could be definitively demonstrated with a larger sample. The addition of the interaction term lowered adjusted R² from 0.430 to 0.426.

A variety of other variables also were considered, including PEAKFLOW (indicating whether the train involved a traditional commuting schedule), but this variable was not significant when the OUTBOUND variable was included. The interaction effects between either OUTBOUND and PEAK with YEARS also were not statistically significant in any of the models.

Conclusions

Metra faces two distinct trends with regard to the use of mobile communication devices. First, passengers continue to shift toward more visually-intensive activities that are difficult or impossible to safely perform when using other forms of transportation, such as biking, driving, or walking. This suggests the intensity of activities being conducted is growing. At the same time, the overall rate of growth of device use by passengers has slowed. Mobile device usage is following the S-shaped pattern identified by Rogers (1962). Cultivating further gains in technology use may require efforts to cater to the “late majority” (those skeptical of innovation or with less financial liquidity) or “laggards” (those who often have an aversion to change). The findings indicate that rates of device use also are influenced by income, supporting claims that a digital divide persists on the basis of average income. Although the data show that traditional commuters traveling outbound are the heaviest users of mobile devices, the adoption of technology has progressed far enough along the S-curve that the differential in usage between the peak/off-peak periods is small.

The declining rates of growth in device use suggests that agencies should exercise caution in spending large sums to add amenities such as Wi-Fi to commuter trains, as the opportunities for expanding passenger use of technology may have fallen. Nevertheless, “tech-friendly” amenities may positively contribute to broader passenger satisfaction. Some of the most attractive opportunities may exist along routes serving lower-income riders, where technology use has not progressed as far along the S-curve as on higher-income lines. Although ubiquitous Wi-Fi service may be cost-prohibitive,

providing working electrical outlets and developing apps that improve the on-board experience may be reasonable short-term moves. Likewise, the prevalence of mobile device use among travelers may provide new opportunities for transit agencies to collect advertising revenue by partnering with “smart city” technology companies that provide real-time and geographically-specific advertising to passengers.

APPENDIX

The results from 2015 show that 56.4% of passengers on peak-period trips used technology, compared to 55.8% on off-peak trips (Table 4). This difference, however, is not statistically significant. An estimated 59.5% use technology on outbound trips, compared to 51.1% on inbound trips, a difference significant at a 0.01 level. A gap also was observed among inbound and outbound passengers with respect to visual tasks (see column b of Table 4). On outbound trips, 48.3% were engaged in visual tasks, compared to 36.9% on inbound trips. Furthermore, the share of technology use on outbound trips during the peak period exceeded that on inbound trips during the peak by 61.7% to 49.0%, a statistically significant difference. The gap was much smaller for off-peak trips.

A summary of the regression analysis with the Tobit transportation is shown in Table 5.

TABLE 5.
Results of Multiple Regression
Model with Tobit

	1. Time of day, Direction of Travel, + Changes Over Time		2. Time of day, Direction of travel + Changes Over Time (Nonlinear)		3. Direction of Travel + Changes Over Time (Nonlinear), Route Income		4. Direction of Travel, Changes Over Time (Nonlinear), Route Income, Income-Time Interaction	
	Coeff.	P Value	Coeff.	P Value	Coeff.	P Value	Coeff.	P Value
INTERCEPT	-0.961**	(0.000)	-1.12**	(0.000)	-1.63**	(0.000)	-1.83**	(0.000)
PEAK	0.021	(0.410)	0.089	(0.315)	0.070	(0.428)	0.023	(0.763)
OUTBOUND	0.074*	(0.024)	0.192*	(0.033)	0.179*	(0.045)	0.025	(0.053)
TIME	0.288**	(0.000)	0.524**	(0.000)	0.501**	(0.000)	0.626*	(0.002)
TIME ²			-0.0054	(0.052)	-0.048	(0.078)	-0.050	(0.054)
INCOME					0.003*	(0.023)	0.010	(0.060)
INCOME*TIME							-0.001	(0.466)
R ²	0.363		0.380		0.403		0.435	
ADJ. R ²	0.350		0.363		0.381		0.411	

Dependent variable: Proportion of travelers using electronic devices with TOBIT transformation

Observations= 148 trains

*Significant at 0.05.

**Significant at 0.01.

References

- American Public Transportation Association. 2016. "Public Transportation Ridership Report, Fourth Quarter 2015." March 2. <http://www.apta.com/resources/statistics/Pages/ridershipreport.aspx>, accessed May 27, 2016.
- Andreev, Pavel, Ilan Salomon, and Nava Pliskin. 2010. "Review: State of Teleactivities." *Transportation Research. Part C, Emerging Technologies*, 18: 3-20.
- Barbeau, S. J., M. A. Labrador, N. L. Georggi, P. L. Winters, and R. A. Perez. 2010. "The Travel Assistance Device: Utilizing GPS-enabled Mobile Phones to Aid Transit Riders with Special Needs." *IET Intelligent Transportation Systems*, 4(1): 12-23.
- Brakewood, Candace, et al. 2014. "Forecasting Mobile Ticketing Adoption on Commuter Rail." *Journal of Public Transportation* 17(1) 1-19.
- Bull, Michael. 2000. *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Oxford: Berg.
- Farley, Tom. 2005. "Mobile Telephone History." Privateline.com, http://www.privateline.com/wp-content/uploads/2016/01/TelenorPage_022-034.pdf.
- Frei, Charlotte, and Hani S. Mahmassani. 2011. "Private Time on Public Transit: Dimensions of Information and Telecommunication Use of Chicago Transit Riders." Presentation, Annual Transport Chicago Conference, Chicago, IL, June 3.
- Holley, David, Juliet Jain, and Glenn Lyons. 2008. "Understanding Business Travel Time Use and its Place in the Working Day." *Time & Society* 17(1): 27-46.
- Golob, Thomas F., and Amelia C. Regan. 2001. "Impacts of Information Technology on Personal Travel and Commercial Vehicle Operations: Research Challenges and Opportunities." *Transportation Research. Part C, Emerging Technologies*, 9(1): 87-121.
- Guo, Zhan, Alexandra Derian, and Jinhua Zhao. 2015. "Smart Devices and Travel Time Use by Bus Passengers in Vancouver, Canada." *International Journal of Sustainable Transportation*, 9(5): 335-347.
- Hess, S., and G. Spitz. 2016. "How Much Do Attitudes and Values Matter in Mode Choice?" Paper presented at 95th Annual Meeting of the Transportation Research Board, Washington, DC.
- Jackson, Linda A., Zhao Yong, Anthony Kolenic III, Hiram E. Fitzgerald, Rena Harold, and Alexander Von Eye. 2008. "Race, Gender, and Information Technology Use: The New Digital Divide." *CyberPsychology & Behavior*, 11(4): 437-442.
- Jain, Juliet, and Glen Lyons. 2008. "The Gift of Travel Time." *Journal of Transport Geography*, 16(2): 81-89.
- Kim, Dae-Young, Jungkun Park, and Alastair M. Morrison. 2008. "A Model of Traveler Acceptance of Mobile Technology." *International Journal of Tourism Research*, 10(5): 393- 407.

- Lyons, Glenn, Juliet Jain, and David Holley. 2007. "The Use of Travel Time by Rail Passengers in Great Britain." *Transportation Research Part A*, 41(1): 107–120.
- Metra. 2013. "2011 Regional Transportation Authority Customer Satisfaction Study." January. www.metrarail.com, accessed March 1, 2016.
- Metra. 2015. "Metra Will Test Wi-Fi on 10 Rail Cars Systemwide." September 25. https://metrarail.com/metra/en/home/utility_landing/newsroom/newsroom/metra-to-test-Wi-Fi-on-10-cars-systemwide.html, accessed February 16, 2011.
- Metropolitan Boston Transit Authority (MBTA). 2013. "MBTA Announces Wi-Fi Service Sponsorship Program." March 1. http://www.mbta.com/about_the_mbt/news_events/?id=26452, accessed February 11, 2016.
- Mokhtarian, Patricia L., and Ilan Salomon. 2001. "How Derived is the Demand for Travel? Some Conceptual and Measurement Considerations." *Transportation Research Part A*, 35(8): 695–719.
- Mokhtarian, Patricia L., Ilan Salomon, and Susan L Handy. 2006. "The Impacts of ICT On Leisure Activities and Travel: A Conceptual Exploration." *Transportation*, 33(3): 263–289.
- Neff, Jack. 2010. "Young Drivers Shift Gears: Is Digital Revolution Driving Decline in U.S. Car Culture?" *Advertising Age*. May 31. <http://adage.com/article/digital/digital-revolution-driving-decline-u-s-car-culture/144155/>.
- Ninth Decimal (formerly JiWire). "Measurement." <http://www.ninthdecimal.com/location-conversion-index-lci-measurement>, accessed February 11, 2016.
- Ohmori, Nobuaki, and Noboru Harata. 2009. "How Different are Activities While Commuting by Train? A Case in Tokyo." *Tijdschrift voor Economische en Sociale Geografie*, 34(1): 547-561.
- Pew Charitable Trust. 2016. "Who Uses Mobile Payments: Survey Findings on Consumer Opinions, Experiences."
- Rail Engineer Team. 2015. "Train Connectivity—The Future Beckons." *RailEngineer*. September 1.. <http://www.railengineer.uk/2015/09/11/train-connectivity-the-future-beckons/>, accessed February 2016.
- Rogers, Everett M. 1962. *Diffusion of Innovations*. London: The Free Press, Collier McMillan Publishers.
- Russell, Marie, Rachel Price, Louise Signal, James Stanley, Zachery Gerring, and Jacqueline Cumming. 2011. "What do Passengers do During Travel Time? Structured Observations on Buses and Trains." *Journal of Public Transportation*, 14(3): 123–146.
- Salomon, Ilan. 1998. "Technological Change and Social Forecasting: The Case of Telecommuting as a Travel Substitute." *Transportation Research Part C*, 6(1-2): 17–45.
- Sasaki, Kuniaki, and Kazuo Nishii. 2010. "Measurement of Intention to Travel: Considering the Effect of Telecommunications on Trips." *Transportation Research Part C: Emerging Technologies*, 18(1): 36-44.

- Schwieterman, Joseph P., and Lauren A. Fischer. 2011. "Variations in the Rates of Passengers Usage of Portable Technology on Intercity Buses, Trains, and Planes: Implications for Transportation Planning." *Journal of Transportation Law, Logistics and Policy*, 78(1): 435-444.
- Simun, Miriam. 2009. "My Music, My World: Using the MP3 Player to Shape Experience in London." *New Media & Society*, 11(6): 921-941.
- Skandland, M. 2011. "Use of MP3 Players as a Coping Resource." *Music and Arts in Action*, 3(2): 15-32.
- Susilo, Yusak O., Glenn Lyons, Juliet Jain, and Steve Atkins. 2012. "Rail Passengers' Time Use and Utility Assessment: 2010 Findings from Great Britain with Multivariate Analysis." *Proc. 91st Annual Meeting of the Transportation Research Board*, Washington, DC, January; forthcoming, *Transportation Research Record*.
- Tang, L., and Thakuriah, P.V. 2012. "Ridership effects of real-time bus information system: A Case Study in the City of Chicago." *Transportation Research Part C: Emerging Technologies*, 22: 146-161.
- Transportation Research Board. 2016. *Interregional Travel: A New Perspective for Policymaking, Special Report 320*. National Academy of Sciences.
- U.S. Public Interest Research Group Education Fund. 2013. "Our Changing Relationship with Driving and the Implications for America's Future." U.S. Public Interest Research Group. Spring 2013.
- Van der Waerden, Peter, Harry Timmermans, and Rob van Neerven. 2009. "Extent, Nature, and Covariates of Multitasking of Rail Passengers in an Urban Corridor: A Dutch Case Study." *Transportation Research Record*, 40(2110): 106-111.
- Windmiller, S., T. Hennessy, and K.E. Watkins. 2014. "Accessibility of Communication Technology and the Rider Experience Case Study of Saint Louis, Missouri, Metro." *Transportation Research Record*, 2415/Transit: 118-126.
- Yoshii, Toshio, and Kuniaki Sasaki. 2010. "Information/Communication Technologies and Travel Behaviour." *Transportation Research Part C: Emerging Technologies*, 18(1): 1-2.

About the Authors

JOSEPH P. SCHWIETERMAN, Ph.D. (jschwiet@depaul.edu) is a Professor in the School of Public Service and Director of the Chaddick Institute for Metropolitan Development at DePaul University in Chicago.

LAUREN AMES FISCHER is in the Graduate School of Architecture, Planning, and Preservation at Columbia University.

Comparing Automated Shared Taxis and Conventional Bus Transit for a Small City

Dr. Louis A. Merlin, AICP
Florida Atlantic University

Abstract

This study compared two hypothetical transit scenarios with the current bus transit system for serving the transit passengers of Ann Arbor, Michigan, for a typical fall weekday in 2013. One scenario consists of an automated taxi system that allows only one rider at a time, and the other consists of a similar automated taxi system that allows ridesharing for up to four passengers. The two automated taxi scenarios were modeled on simulated transit passenger travel demand data via agent-based models. All three scenarios were then compared for their level of service, cost, greenhouse gas emissions, and congestion impacts. The automated shared-ride taxi transit service could provide a higher level of service at lower cost and lower carbon emissions than the current bus system. An automated taxi service without ridesharing would provide high levels of service at lower cost, but with higher levels of carbon emissions than the current bus system. Ridesharing is essential to obtaining the full cost savings and environmental benefits for an automated taxi system. Both automated taxi systems would likely increase peak-hour congestion by increasing peak-hour vehicle kilometers traveled.

Keywords: *Automated vehicles, shared automated vehicles, shared mobility, greenhouse gas emissions, demand responsive transit*

What Does Vehicle Automation Mean for Transit?

With a background of steady technological progress towards vehicle automation, several researchers have speculated that highly-automated vehicles will lead to the advent of new vehicular travel modes. Taxis have long been available as a convenient mode of semi-transit, but have been limited in their share of the transportation market to high costs and difficulties in hailing. Automated taxis, in concept, solve both these problems—they are easily hailed from any mobile phone, and, without the costs of a human driver, the per-kilometer costs of automated taxi travel are presumably

lower than for conventional, human-driven taxis. In comparison with private vehicle ownership, automated taxis offer the potential of point-to-point travel on demand without having to worry about parking, maintenance, or insurance issues. In comparison with public transit, automated taxis offer much greater flexibility in routing and scheduling, including the possibility of demand-responsive service.

Technological innovation and entrepreneurship in the transportation sector have led to a variety of new transportation services captured broadly under the name “shared mobility” (Shared-Use Mobility Center 2016; Shaheen et al. 2015). Services such as Uber and Lyft (known as “ridesourcing”), car-share services, bike-share services, and ride-matching services are increasingly available in large cities and offer flexible versions of transportation on demand in urban areas. As car-share services have become more pervasive, new models of one-way car-sharing have arisen, such as Car2Go (Ciari and Bock 2014; Shaheen et al. 2015). Vehicle automation is likely to build upon these already-existing trends by allowing vehicles to come to passengers upon demand, thereby further opening up the markets for shared mobility services. Several companies, such as Uber and Ford, are investing heavily in developing shared, automated, on-demand mobility services (Boudette 2016; Chafkin 2016).

There are several reasons to suspect that the population of current transit riders may be among the first to widely adopt the use of a new automated taxi mode. First, transit riders have a demonstrated willingness to travel by modes other than their own private vehicle and a willingness to use shared forms of transport. Therefore, the modal switch to shared taxis would be a less dramatic behavioral switch for frequent transit riders than for those who currently rely predominantly on private vehicles. Second, automated taxis (and taxis in general) operate most efficiently and cost-competitively in urban environments in which travel demand is concentrated. In other words, automated taxis will be most prevalent and affordable in the same dense urban environments served by transit. Third, automated taxis are likely to be able to provide a higher quality of service than current fixed-route transit services by allowing fewer transfers, shorter wait times, and shorter access and egress approaches to boarding and alighting areas.

The question that leaps immediately to mind is how will the spread of automated taxis impact public transit? There are three possible answers to this question. The first is that automated taxis would increase current transit ridership by serving as effective first-mile/last-mile service to existing transit (Levine et al. 2013; Liang et al. 2016; Shepherd and Muir 2011). The second is that automated taxis and similar shared automated modes will compete directly with transit and gradually erode its market share (Martinez and Viegas 2016b). The third is that automated taxis and transit will be integrated into a system that will play off the benefits of each (Maheo et al. 2016). This study explored the second of these alternatives—that automated taxis might outcompete and replace public transit in certain areas—and investigates the implications of this potential outcome.

In particular, this study compared the performance of the current bus transit system in Ann Arbor, Michigan, with two alternative automated taxi systems. One automated taxi system does not permit ridesharing, serving one person or party at a time (a “single-rider” system); the other requires ridesharing, subject to certain passenger convenience

constraints (a “shared-ride” system). Both taxi systems could be considered a type of transit, in that they are required to serve all transit passengers within their service area and that vehicle movements are centrally coordinated across the taxi fleet. Agent-based models were used to simulate these alternative automated taxi systems under the premise that they would serve all Ann Arbor transit riders for a typical weekday in 2013 and then their performance was compared to the current transit system with respect to wait times, travel times, costs, carbon emissions, and congestion impacts. It was expected that a fleet of many small automated taxis would be able to serve Ann Arbor’s transit demand with shorter wait and in-vehicle travel times. On the other hand, a switch to automated taxis likely would also increase vehicle mileage externalities, such as carbon emissions and congestion impacts. The purpose of this simulation was to estimate the size of these impacts and assist in formulating policy initiatives to mitigate these potential impacts.

Recent Simulation Research Regarding New Automated Modes

Most of the current simulation and modeling research into automated taxis examined their competitiveness with private vehicle ownership (Burns et al. 2013; Fagnant and Kockelman 2014; Spieser et al. 2014; Fagnant et al. 2015). For example, Burns et al. (2013) found that a fleet of 18,000 automated taxis could serve most of the urban area around Ann Arbor with average wait times under 1 minute and costs of about 25 cents per kilometer (41 cents per mile). This is cost-competitive with current average private vehicle ownership costs of 37 cents per kilometer (59 cents per mile). Fagnant and Kockelman (2014) assumed that automated taxis would obtain a modest 1–2% of current private vehicle trips in Austin, Texas. Under such a system, they found that each automated taxi replaced about 8.4 privately-owned vehicles, whereas the system as a whole required 10% additional vehicle kilometers traveled (VKT) to serve the same travel demand. On the other hand, with dynamic rideshare,¹ the VKT increase could be as little as 1.5% over the present system (Fagnant and Kockelman 2014; Fagnant et al. 2015). The average wait time for their system was estimated at about 50 seconds, and 96.2% of travelers had a wait less than 5 minutes. Assuming a cost of \$70,000 per vehicle, including automation costs, they found that a shared automated vehicle (SAV) fleet in Austin would provide a 19% return on investment if the system operators charged 62 cents per kilometer (\$1 per trip mile) served. Bischoff and Maciejewski (2016) examined an automated taxi system serving 10% of private vehicle trips in Berlin and found that such a system would increase VKT by 14.6% with a 3-minute average wait time. In short, current models suggest that automated taxi systems could provide a high level of service with minimal wait times and competitive costs in moderately-dense urban environments.

A smaller body of research examined automated taxis as a first-mile/last-mile service feeding into existing transit (Levine et al. 2013; Liang et al. 2016; Shepherd and Muir 2011). Levine et al. (2013) found that private commute mode share decreases by

¹ Placing multiple passengers in the same vehicle in real time without pre-planning the trip.

between 7–29% with automated shuttles feeding into existing Chicago-area train stations. Shepherd and Muir (2011) found that transit mode share could increase 1–8% across various European cities with automated shuttles feeding into existing train stations. These studies document some possible growth in the transit market by using automated vehicles to feed into high capacity transit.

Martinez and Viegas (2016a) examined what would happen if two new forms of automated transit, automated taxis and advance-scheduled automated shuttles, served Lisbon, Portugal, replacing both private vehicle trips and current bus trips. They assumed that people within close proximity to the subway continued to use that mode. With this massive mode shift, the transportation system would see a 34% reduction in carbon dioxide emissions. Wait times would be limited to 5 minutes for automated taxis and 10 minutes for automated shuttles, and costs would be affordable, at about 32 cents per kilometer.

Maheo et al. (2016) redesigned the existing bus system of Canberra, Australia, with a system of human-driven taxis feeding into high-frequency bus lines. They found that for the same total costs, wait times could be cut in half. Their hub-and-spoke design offered a promising precursor for how automated taxis and large vehicle transit might work together as part of an efficient integrated system.

Reviewing the current research, automated taxis offer the possibility of a convenient, cost-competitive mode that could replace a large segment of private vehicle travel; however, automated taxis may replace or compete with current bus services as well. Many researchers have anticipated that automated taxis or automated shuttles could feed into current high-capacity transit systems, but few have studied whether and how transit systems themselves should be redesigned to best integrate with this proposed new mode.

A large caveat on this body of research is aggressive assumptions about market share and affordable automation cost. Several of these studies assumed a very large market share—in some cases, a 100% share of the current travel market—which makes it much easier to obtain economies of scale and create an efficient system. In the real world, any new mode would have to compete with all existing modes and could not obtain these efficiencies in the short run, or perhaps ever. Also, the cost assumptions regarding the marginal cost of automating a vehicle often are quite favorable without substantiation. For example, Burns et al. (2013) assumed that the cost to convert a regular vehicle to an automated one is a one-time marginal cost of just \$2,500 per vehicle, whereas Spieser et al. (2014) assumed a more reasonable but still speculative automation cost of \$15,000 per vehicle.

In comparison with these other simulations, this study assumed that all transit riders in a small city could be served by an automated taxi system, which is a more modest and arguably reasonable assumption than full market capture. It is even conceivable that a transit agency could decide to launch an automated taxi system to replace its current bus service. Indeed, some cities that are spread out and not well served by conventional transit have engaged Uber as a kind of transit substitute (Woodman 2016). The implications of such a switch are investigated in the rest of the paper.

The following section discusses the data used for transit passenger travel demand and the workings of the agent-based model in NetLogo. This is followed by a section that examines the three scenarios of current bus, automated taxi, and automated shared taxi from the perspective of wait times, travel times, costs, VKT, and carbon emissions. The section after that returns to considerations of what such an automated taxi system might mean for the future of public transit and transit policy. The Conclusion summarizes and reiterates the main points from the preceding sections.

Methodology: Agent-Based Models of Automated Taxi Fleets for Ann Arbor

This study compared three transit systems serving a typical weekday's ridership for the Ann Arbor Area Transit Authority (AAATA) system from 2013 with measurements for wait time, travel time, cost, vehicle miles traveled, and carbon dioxide emissions. Data for the performance of the actual AAATA bus system were from the National Transit Database (NTD) for 2013. The two automated taxi systems were simulated in NetLogo, with separate models for a single-ride system and a shared-ride system.

AAATA Background

The AAATA system is a small-city bus system without a rail component, but with a relatively high ridership due to a major university and downtown employment core and a large student population. In comparison with the other 472 bus systems reporting operations to NTD in 2013, Ann Arbor offered 8,916 vehicle revenue miles on a typical weekday, similar to the mean for all systems of 11,861 (although there is a large standard deviation across systems of 25,711) (Federal Transit Administration 2014). Average weekday vehicle occupancy, calculated as passenger miles divided by vehicle revenue miles, was 8.1 persons per vehicle, which was, again, comparable to the average for all bus systems in the US at 7.3 (4.3 standard deviation). AAATA trips were shorter than the average for US bus systems, with the average trip length being 5.0 kilometers for AAATA vs. 8.5 kilometers for the average US system, presumably because Ann Arbor is a relatively small town. The volume of passengers, however, was quite high, with 32.6 passenger trips per vehicle-hour, in comparison with a national average of just 22.4 (12.5 standard deviation). In other words, in comparison with the "average" bus system in the US, AAATA has typical occupancies, high volumes, and shorter passenger trip lengths.

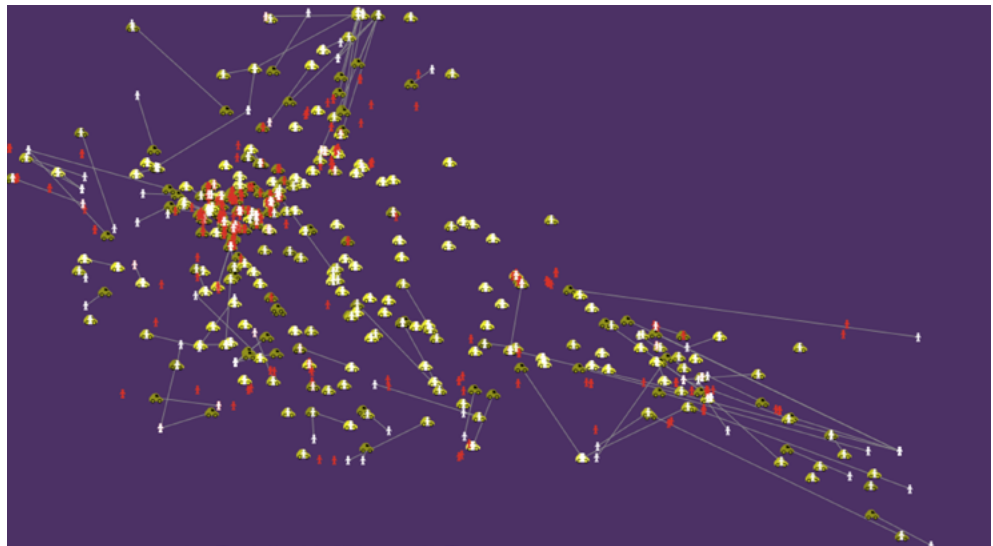
NetLogo

NetLogo is a graphical, agent-based modeling platform within which individual passengers and taxis are generated and follow prescribed sets of instructions or behaviors. Time itself is modeled, in this case, with one minute increments, and passengers and taxis move on a minute-by-minute basis. System-level patterns of travel emerge and are monitored both on the screen and through a number of aggregate metrics for passengers and vehicles. Within the NetLogo interface, the user can view

the movements of taxis and passengers in space as well as stop the simulation and monitor any of their internal variables (i.e., number of passengers for taxis, wait time for passengers) within the simulation in “real” time. That is, the user can watch the simulation unfold and observe the behavior of individual taxis and passengers to ensure that it comports with the intentions of the simulation logic.

Figure 1 shows a screenshot from NetLogo. Many small passengers are shown in white for passengers en route and in red for passengers who have completed their rides. Taxis are shown in yellow, but become darker as they are more fully occupied. White lines connect passengers with the taxis en route to picking them up.

FIGURE 1.
Screenshot of automated taxi
model in NetLogo



Generating and Serving Transit Passengers

True transit origin-to-destination trip data were not available from AAATA or the regional planning agency, the Southeast Michigan Council of Governments (SEMCOG), so transit trips were simulated based on other available data. A transit rider survey from 2010 by SEMCOG provided a sample of passenger boarding locations and alighting locations by time of day. Origins were sampled randomly from boarding locations, and destinations were sampled randomly from alighting locations, with separate samples for each of four time-of-day periods: morning, midday, afternoon, and evening (Nustats 2012). Since origins and destinations were each sampled at random without any correspondence, the sample did not represent true origin-destination pairs. Boarding data for the AAATA system from a weekday in 2013 were used to simulate the generation of new passengers for each 15-minute interval (Ann Arbor Area Transportation Authority 2014a). The average total passengers per day for the simulation were calibrated to the NTD total of 23,152 (Federal Transit Administration 2014).

For a single-rider taxi system, passengers hail the closest available (empty) taxi. Passengers who have been waiting longer hail taxis first. The hailed taxi goes directly

to the passenger and picks him/her up, then goes uninterrupted to that passenger's destination and drops him/her off. Passengers are assumed to have a boarding time of one minute and an alighting time of one minute, during which the taxi does not move.

For a shared-ride system, taxis en route divert to serve new passengers so long as no existing or committed passenger is inconvenienced by too great a degree. Specifically, for each passenger served by a shared-ride taxi, his/her wait time cannot exceed 10 minutes. In addition, for all existing, committed, and new passengers, their travel time may not exceed more than 10 minutes more than their direct travel time. Therefore, additional passengers are accommodated with a shared-ride only if existing passengers are still well-served. Also, each taxi has a maximum capacity of four passengers. If several taxis qualify to serve a new passenger, then the taxi that increases its VKT least is selected. This results in the smallest VKT increase for the fleet as a whole. To select the taxi that increases VKT least, a centralized system of taxi dispatch would be required. If each taxi was simply trying to maximize its own revenues, a different travel pattern outcome would result, presumably one with higher VKT.

Calibrating Taxi Speeds

The street network was not modeled, so all taxis traveled in a straight line to pick up and drop off their passengers. However, travel times from point to point were calibrated to real-world travel times to allow the queueing (waiting) aspects of the model as realistic as possible. The idea was that if travel times from each origin to each destination were realistic, then the time required for waiting and arriving at desired locations also would be representative.

Travel times were calibrated two ways. First, straight line travel distances were multiplied by a network factor to make them more closely resemble network travel distances. Based on a random sample of origins and destinations, a network inflation factor of 1.238 was obtained. Second, network travel speeds for peak and off-peak travel times were determined from a random sampling of origins and destinations run through Google directions. Together, these calibrations allowed for a realistic simulation of point-to-point travel distances and speeds.

Calculating Wait Times and Determining Fleet Size

A brief literature review was conducted of typical wait times for buses, with a focus on literature that related wait times to service frequency. Estimating bus wait times depends upon traveler behavior, including whether travelers arrive at a random time between buses or whether they optimize their arrival time to minimize their expected wait. Bowman's optimization model (Bowman and Turnquist 1981) has passengers optimize their wait time based upon service frequency and reliability and estimates a wait time of 2–3 minutes for 5-minute headways, 3–5 minutes for 10-minute headways, and 5–7 minutes for 20-minute headways. Fan and Machelme (2009) used empirical data on observed wait times and found that typical wait times can be reasonably estimated with a formula that is a function service frequency. This formula is wait time

= 2.28 minutes + 0.29 × service frequency. Based upon Fan’s formula, the expected wait time for the average AAATA rider was 9.5 minutes (see Table 1. Based upon an extrapolation of Bowman’s data, the expected wait time for the average AAATA rider was 6.2 minutes.

TABLE 1.
Estimated Wait Times for
AAATA System

Route	Route Name	Average Ridership	AM Peak Headways	Estimated Wait Times*	Estimated Wait Times**
1	Pontiac Dhu Varren	810	30	11.0	7.0
2	Plymouth	2,463	15	6.6	5.0
3	Huron River	1,337	30	11.0	7.0
4	Washtenaw	3,052	10	5.2	4.0
5	Packard	2,319	30	11.0	7.0
6	Ellsworth	2,248	30	11.0	7.0
7	S. Main East	1,241	30	11.0	7.0
8	Pauline	746	15	6.6	5.0
9	Jackson-Dexter	708	30	11.0	7.0
10	Ypsilanti-NE	506	60	19.7	10.0
11	Ypsilanti-South	243	60	19.7	10.0
12	Miller-Liberty	834	30	11.0	7.0
13	Newport	211	60	19.7	10.0
14	Geddes-E. Stadium	196	30	11.0	7.0
15	Scio Church-W. Stadium	253	30	11.0	7.0
16	Ann Arbor Saline Rd.	468	30	11.0	7.0
17	Amtrak Depot ST.	46	30	11.0	7.0
18	Miller-University	435	25	9.5	7.0
20	Ypsilanti-Grove-Ecorse	544	60	19.7	10.0
22	North/South Connector	772	30	11.0	7.0
33	EMU Shuttle	611	20	8.1	6.0
36	Wolverine Tower Shuttle	1,697	9	4.9	4.0
601	Pontiac-University	128	0	0.0	7.0
609	Dexter-University	142	30	11.0	7.0
Estimated per-passenger average wait time				9.5	6.2

Headway data from Ann Arbor Area Transportation Authority, 2014b

*Fan 2010

**Bowman and Turnquist 1980

In the NetLogo model, wait times are tracked for each individual passenger. The size of the taxi fleet was calibrated to ensure that the average wait time for the system was less than the smaller of the two expected wait times for the average AAATA rider, i.e., 6.2 minutes. For the single-rider system, fleet sizes between 500 and 1000 were tested, with the finding that a fleet of 800 vehicles performed well on wait times and optimally on costs. Then, for the shared-ride system, fleet sizes between 300 and 750 were tested, based on the assumption that the shared-ride system of the same fleet size would

almost certainly perform at least as well as the single-rider system with respect to wait times. The optimal cost system with satisfactory wait time performance for a shared-ride fleet was found to be 400 vehicles.

Calculating Costs

To compare the three systems on an equitable basis, only vehicle-related costs were accounted for, including vehicle capital costs, vehicle maintenance costs, insurance and taxes, and fuel and tire costs. Administrative costs and non-vehicle capital costs were excluded from all cost calculations.

For the bus transit system, vehicle operations costs, vehicles maintenance costs, and vehicle-related capital associated with existing transit lines were included. NTD provides costs on an annualized basis, so these costs were converted to per-kilometer costs to assign them to a typical weekday (Federal Transit Administration 2014). Costs per passenger-kilometer also were provided.

For the taxi systems, the costs estimates were based assuming a taxi fleet of model 2014 Prius Cs with a per-vehicle cost of approximately \$29,230. Vehicle automation costs were assumed to be \$50,000 per vehicle in addition to regular vehicle purchase costs. Kilometers-per-liter for this vehicle in urban areas is 22.2 (53 MPG), and the cost of a liter of gas in the Midwest was \$0.92 (\$3.47 per gallon) in 2013, according to the Energy Information Administration. The American Automobile Association (AAA) provided an estimate of maintenance cost of 3.2 cents per kilometer (5.2 cents per mile) and insurance and tax costs of \$3.98 per vehicle per day for small sedans (American Automobile Association 2013).

Vehicle depreciation was assumed to be based upon either time or distance, whichever usage was higher. This means that idle vehicles bear a depreciation cost even if not used. Vehicles were assumed to depreciate to their salvage value (10%) within 5 years or 400,000 kilometers, whichever comes first. Time-based depreciation is an important assumption because it drives the size of the minimum cost vehicle fleet; if vehicle fleet costs are assumed to vary only as a function of kilometers driven, then there is no performance disincentive for overly-large vehicle fleets.

Costs were reported as the sum total of all fleet vehicles costs per day, or total transit operator daily costs, as well as costs per passenger-vehicle kilometer.

Note that larger vehicle fleets have a number of ancillary benefits unrelated to costs. First, the larger the vehicle fleet, the smaller the wait time for passengers. Second, the larger the vehicle fleet, the lower the VKT because a nearby vehicle is more likely to be available for any given hailing passenger with a larger fleet. The only performance characteristic that declines with larger vehicle fleet size is system-level cost.

Calculating Vehicle Miles Traveled

The NTD provided the vehicle kilometers traveled for AAATA's bus fleet on a typical weekday. Note that vehicle kilometers traveled includes mileage traveled while out of

service. Peak hour VKT for AAATA was estimated based upon the number of vehicles operating during the AM peak hour.

For the NetLogo simulations, every kilometer for each vehicle was tracked within the simulation. This straight-line distance total was then multiplied by the network inflation factor to derive an estimated total for network-based taxi VKT. Vehicle kilometers also were tracked by hour of the day.

Calculating Carbon Emissions

For AAATA, the annual fuel use reported to NTD was converted into a figure for a typical weekday. Since the AAATA system uses biodiesel, carbon emissions were estimated to be 85% of that typical for diesel fuel (US Department of Energy 2015). In fact, the carbon savings of using biofuels is controversial, and if previously unfarmed land is converted to the production of biofuels, significant increases in carbon emissions are possible (Searchinger et al. 2008; Gnansounou et al. 2009). However, the precise impacts of biodiesel on greenhouse gas (GHG) emissions are beyond the scope of this study, so the simplifying assumption from the US DOE was employed here.

Carbon emissions for the automated taxi system were calculated based upon total kilometers miles traveled per day and the fuel efficiency of the Toyota Prius fleet, multiplied by the carbon intensity of gasoline fuel. The carbon emissions associated with burning 1 liter of fuel was assumed to be 2.35 kilograms per liter (19.64 pounds per gallon) (US Energy Information Administration 2015).

In addition to in-use emissions, certain components of life-cycle emissions related to vehicle manufacture and use were included as well. Vehicle manufacture, tire use, maintenance, and insurance related GHG emissions were derived on a per-vehicle-kilometer basis from Chester and Horvath (2009). The more vehicles are driven per day, the more of the vehicle's life-cycle of GHG emissions is consumed and accounted for.

Results: Performance Comparison of Conventional Bus Transit with Automated Taxi Fleets

The major results from the simulation and its comparison with the current Ann Arbor transit system are summarized in Table 2, including wait times, travel times, costs, VKT, and carbon emissions performance. Three systems were compared: the current Ann Arbor bus system, an automated single-rider taxi system, and an automated shared-ride taxi system.

TABLE 2.
Performance Comparison of
Transit System Alternatives

	Conventional Bus	Automated Taxi, Single-rider	Automated Taxi, Shared-ride
Service Performance			
Fleet size	65	800	400
Average wait time	6.2–9.5	5.6 (0.11)	5.9 (0.04)
95th percentile wait time	Unknown	16.1 (0.33)	13.9 (0.29)
Travel time	29.3*	9.1 (0.03)	16.1 (0.05)
Cost Performance			
System cost per day	\$ 92,323	\$ 73,777 (\$ 175)	\$ 33,196 (\$ 76)
% difference from bus system	na	-20.0%	-64.0%
% difference from single-rider system	na	na	-55.0%
Cost per passenger kilometer	\$ 0.794	\$ 0.601	\$0.271
Vehicle Miles Traveled Performance			
Vehicle kilometers traveled	15,636	185,412 (1,784)	82,626 (773)
x difference from bus system	na	11.9	5.3
% difference from single-rider system	na	na	-54.5%
Peak-hour vehicle kilometers traveled	1,359	20,729	8,412
x difference from bus system	na	15.3	6.2
Peak hour	na	9–10 AM	9–10 AM
Greenhouse Gas Performance			
Life-cycle carbon emissions per day (kg)	22,475	39,782 (383)	17,728 (166)
% difference from bus system	na	+77.0%	-21.1%

Mean values are from 50 simulations. Standard deviations indicated in parentheses.

*Total travel time including access and egress travel time from American Community Survey.

Wait and Travel Times

All of the automated taxi systems had wait time performance similar to the current Ann Arbor bus system, which was expected because fleet sizes were determined to match current average wait times. Both taxi systems had average wait times of around 6 minutes and 95th percentile wait times of around 15 minutes; the shared-ride taxi system had lower 95th percentile wait times because it could accommodate peak demand with greater flexibility via shared rides. Travel times for both automated taxi systems were shorter than the 29.3 minutes for the conventional bus system (US Census Bureau 2013), varying between a low of 9.1 minutes for the single-rider system and a higher 16.1 minutes for the shared-ride system. Passengers in a shared-ride system do not always travel directly to their destination and, thus, experience higher travel times than passengers in a single-rider system.

Daily System and Per-Kilometer Costs

The daily costs of automated taxi systems were lower than the current bus system, largely because of the removal of driver costs. The single-rider taxi system had 20.0%

lower vehicle-associated costs, whereas the shared-ride system had 64.0% lower vehicle-associated costs. The shared-ride system obtained significant efficiencies in this model and had costs 55.0% lower than the single-rider system. Note that parking costs were not included in this analysis.

Assuming that the marginal costs of vehicle automation per vehicle is a one-time cost of \$50,000/vehicle, the per-kilometer vehicular costs of an automated taxi system for the operator would be about 60.1 cents per kilometer for a single-rider system and about 27.1 cents per kilometer for a shared-ride system.

Vehicle Kilometers Traveled and Congestion Impacts

Automated taxi systems greatly multiply vehicle kilometers traveled relative to a transit system, especially for the peak period. For the single-rider taxi system, VKT were multiplied by 11.9 times over the course of a day and 15.3 times in a peak hour. The shared-ride system performed somewhat better, with a multiple of 5.3 for an entire day and of 6.2 for a peak hour. Link-specific congestion impacts were not obtainable from the current model because it did not incorporate the modeling of capacity and flows on individual network links. The shared-ride system reduced VKT over the single-rider system by 54.5%.

GHG Emissions

Although vehicle miles traveled were nearly 12 times greater for the single-rider automated taxi system, the carbon emissions were only 77.0% larger. This is due to the relative fuel efficiencies of the corresponding vehicle fleets—the taxi fleet obtained a fuel efficiency of 21.9 kilometers per liter (53 MPG) whereas the bus fleet obtained a fuel efficiency of 1.9 kilometers per liter (4.5 MPG).

Surprisingly, the shared-ride automated taxi system performed the best overall on carbon emissions, with 54.5% lower carbon emissions than the single-rider system and 21.1% lower life-cycle carbon emissions than the current bus system.

Sensitivity Analysis

To investigate the sensitivity of these results to specific assumptions, the impact of potential different vehicle types and different travel time diversion constraints were examined.

In addition to the Toyota Prius hybrid, other vehicles that were considered as potential fleet vehicles for the taxi system included other high-efficiency, compact, four-passenger vehicles from the 2013 vehicle cohort: 1) Honda Civic Hybrid, 2) Volkswagen Jetta Hybrid, and 3) Volkswagen Jetta Diesel. Analyzing alternative vehicles shed insight into how particular the results were to having a Toyota Prius vehicle fleet, especially for its GHG performance.

Daily system-level vehicle fleet costs varied little between vehicle types, as seen in Table 3. The least expensive costs were for the Honda Civic Hybrid, with daily costs of \$73,993, and the most expensive costs were for the Volkswagen Jetta Diesel, for which costs were \$80,764, a difference of 9.2%. Carbon emissions are much more sensitive to vehicle choice, as fuel efficiency is paramount to obtaining low GHG emission for such a high mileage system. Life-cycle carbon emissions per day were 16.0% greater for the Honda Civic Hybrid fleet, 36.3% higher for the Volkswagen Jetta Hybrid fleet, and 81.4% higher for the Volkswagen Jetta diesel fleet.

TABLE 3.
Sensitivity Analysis for Fleet Vehicle Types

	Toyota Prius	Honda Civic Hybrid	Volkswagen Jetta Hybrid	Volkswagen Jetta Diesel
Daily cost,* single-rider	\$ 74,241	\$ 73,993	\$ 76,634	\$ 80,764
Daily cost,* shared-ride	\$ 33,258	\$ 33,147	\$ 34,324	\$ 36,174
Daily life-cycle carbon emissions** (kg), single-rider	39,782.9	46,157.6	54,238.3	72,163.9
Daily life-cycle carbon emissions** (kg), shared-ride	17,728.7	20,569.5	24,170.6	32,158.9

*Daily costs are sum of costs across entire vehicle fleet for typical weekday.

**Daily Life-Cycle Carbon emissions are portion of life-cycle carbon emissions that can be attributed to one day of entire vehicle fleet's operations.

Another key assumption in the analysis was the diversion travel time constraint. Shared-ride passengers must be prepared for a detour as long as 10 minutes beyond their direct-to-destination travel time under current assumptions. Alternative assumptions of a maximum 5-minute detour and a maximum 15-minute detour were tested to examine the sensitivity of the results to this assumption. As expected, as the potential travel time delay increased, mean travel times increased, vehicle occupancies increased, and total fleet vehicle kilometers traveled decreased. In particular, a limitation to a 5-minute detour reduced average vehicle occupancy from 1.95 to 1.55 and increased daily VKT by 21.2%, whereas increasing the potential detour time to 15 minutes increased average vehicle occupancy from 1.95 to 2.39 and decreased daily VKT by 8.9%.

TABLE 4.
Sensitivity Analysis of Diversion Travel Time Constraint

	5-Minute Delay	10-Minute Delay	15-Minute Delay
Network VKT	100,328.4	82,805.8	75,400.0
Mean passenger travel time	13.7	16.1	18.2
Average vehicle occupancy*	1.55	1.95	2.39
Percent empty taxi distance	25.2%	20.1%	16.7%
Liters of gasoline used	7,165.7	5,914.2	5,385.3
In-use carbon emissions (kg)	16,863.8	13,918.5	12,673.7
Non-use carbon emissions (kg)	4,663.1	3,848.7	3,504.5
Total carbon emissions (kg)	21,526.9	17,767.2	16,178.1

*Average vehicle occupancy defined as total daily passenger minutes of movements divided by total daily vehicle minutes of movement.

Limitations

The shared-ride taxi system assumes that all passengers within the system must be willing to share rides. If such a system actually existed, there may be passengers who would be unwilling or uncomfortable with sharing rides or who would pay a premium to have a private travel experience (Zhang et al. 2015b). The willingness of passengers to share rides and their willingness to trade off the inconveniences of ridesharing for lower costs is both unknown and not well explored by current research.

The agent-based model was built on a simulated pattern of transit rider demand from bus stop to bus stop, not true origin-to-destination data. However, origin and destination locations were from actual bus boarding and alighting locations and were statistically representative of AAATA's 2013 transit ridership. Since travel was modeled from bus stop to bus stop rather than from true origin to true destination, the opportunities for ridesharing were increased by this data generation method. On the other hand, a taxi-based transit system could also make use of official stops to efficiently aggregate rides. Also, the model does not account for the possibility that a convenient taxi-based transit system might attract additional passengers. In addition, both paratransit agencies and transportation network companies have to deal with a substantial share of no-shows, a factor not accounted for in this model. No-shows would, of course, increase vehicle miles traveled, costs, and system environmental impacts.

The model does not account for the University of Michigan bus system and its ridership due to differences in data availability between the two systems; however, the two systems have much cross-over ridership. In addition, University of Michigan buses typically have higher occupancy and so would be more costly and inefficient to replace entirely with a taxi-only system than the AAATA system. However, it is interesting to note that the University of Michigan transit system is currently exploring modifying its system to a combination of high-capacity buses and demand-responsive vehicles to more efficiently meet the dispersed spatial and temporal pattern of transit demand on campus.

Many current models of automated taxi systems assume that taxis rebalance to areas of high demand when not in use (Fagnant and Kockelman 2014; Zhang et al. 2015a; Spieser et al. 2014). This is a likely behavior that reduces average passenger wait times, but at the cost of increasing total system VKT (Zhang et al. 2015a). Since the focus of this study was on the environmental performance of such automated taxi systems relative to the current bus system, no vehicle repositioning was included in the model. This increased average and 95th percentile passenger wait times, but decreased system VKT, energy use, and carbon emissions.

The US Environmental Protection Agency (EPA) reported fuel efficiencies of vehicles often are not obtainable in real-world conditions. *Consumer Reports* indicated that the Toyota Prius obtains closer to 15.9 kilometers per liter (37 MPG) rather than the 22.5 kilometers per liter (KPL) reported by the EPA (Consumer Reports 2013). Under these more conservative assumptions, automated taxi systems would have 43% higher

carbon emissions than current projections. Once again, it should be noted that the fuel efficiency of a taxi fleet is crucial to obtaining the possibility of improved environmental performance. However, it is likely that automated taxis would be operated in a manner that maximized fuel efficiency to reduce operational costs.

Discussion: Automated Taxis Will Reshape How Transit Functions

Automated taxi systems performed better on several metrics and worse on a few in comparison with the conventional bus system for serving the transit riders of Ann Arbor. According to the agent-based models implemented here, automated taxi systems would provide similar wait times, superior travel times, and lower daily and per-passenger-kilometer costs than the current conventional bus system. On the other hand, automated taxi fleets would greatly increase the system's VKT, especially during the peak period. Environmentally, the single-rider system would generate increased carbon emissions, but the shared-ride system would actually decrease carbon emissions relative to the current bus system, if EPA-reported fuel efficiencies are actually obtainable in operation.

The congestion impacts of a fleet of automated taxis are both uncertain and highly context-dependent. An increase in vehicle miles traveled does not necessarily lead to a major increase in congestion, unless a particular facility is operating at or near its capacity. An automated taxi fleet of 400 vehicles is small in comparison with the large number of private vehicles operating on Ann Arbor roadways, with more than 128,000 private vehicles used for commuting in Washtenaw County (US Census Bureau 2013). Nevertheless, only a few additional vehicles could greatly increase congestion during peak hours on near-capacity roadways. Further complicating matters, automated vehicle fleets could re-route themselves or otherwise coordinate in ways that might relieve congestion. Beyond the forecast for additional vehicle miles of travel, the location-specific congestion impacts are beyond the scope of this study.

The benefits of ridesharing, at least in concept, are substantial. With universal ridesharing, an automated taxi system could significantly reduce its costs and its environmental impacts, with both improving by slightly more than 50%. The energy efficiency of a shared-ride system is due to both its efficient vehicle fleet and the reduction in VKT due to high rates of ridesharing. The average occupancy over the course of a day for the shared-ride system was 1.95 passengers (although some of this occupancy was created by taking passengers on detours they would prefer not to take), whereas the average occupancy for the single-rider system was just 0.66. The advantages associated with ridesharing increase with urban intensity, because the opportunities for ridesharing occur in areas in which travel demand is concentrated. The cost of ridesharing, on the other hand, was an increase from about a 7-minute travel time, on average, for the single-rider system to a 16-minute travel time for the shared-ride system, plus the inconvenience of having to share a vehicle. The ride-sharing opportunity was increased because origins and destinations were concentrated at existing bus stops, but a future taxi-based transit system might also have designated stops to efficiently serve passengers.

Current transit services are substantially subsidized. These subsidies are justified, in part, on the grounds of providing a reasonable level of mobility to those who cannot afford to or otherwise cannot drive automobiles. But under an automated taxi or other demand-responsive system, many of these same populations could be served potentially more cost-effectively and with improved levels of service. The service provided by such systems could be an improvement over current bus transit by being more point-to-point, by reducing wait times, and by providing reduced in-vehicle travel times.

These results suggest that once shared automated taxi or ride hailing services become viable technologically and economically (i.e., a per-vehicle automation cost of less than \$50,000), the primary function of conventional, large-vehicle transit services will be on servicing large volumes of peak-hour demand and providing congestion relief. Although this study was for just one small city, the conclusions regarding the viability of automated taxis as a transit mode likely generalizes for many small and medium cities and suburbs in the United States in which densities are relatively low and current transit service is typically infrequent. Indeed, the city of Ann Arbor, with a large university campus and student population, has similar or better occupancy rates than many similar-size cities.

The regulatory and coordination implications of new automated taxi modes were beyond the scope of this study; however, it may be productive to start the discussion. Once automated taxi modes become prevalent, they may be required to provide universal service, including to those who are not technologically-savvy and/or have a disability. Even if automated taxi services are provided by the private sector, if these services become a substitute for current public transit, then they must be required to meet the transportation needs of the entire cross-section of the population.

Also, it may be desirable to limit the use of automated taxis or other ride-hailing services in certain times and places to maintain a viable ridership for high-capacity transit lines. Whereas automated taxis may be able to serve certain kinds of travel demand more flexibly than conventional transit, it would not make sense to have these systems competing directly with existing high-capacity transit lines. As seen in this study, the over-use of automated taxis has the potential to exacerbate peak-hour congestion; perhaps ridesharing could be required during peak hours, or perhaps automated taxis could be forbidden from using certain congested routes during peak periods. The policy questions surrounding this potential new mode are complex.

Conceptually, automated taxis and large-vehicle, high-capacity transit have the potential to work together in a coordinated fashion that takes advantage of the benefits of both. The possibility for such coordination is just beginning to be explored in the research literature (Maheo et al. 2016). Certainly, such coordination will involve new institutional relationships as well as new technical methods for managing passenger demand in real time.

As mentioned before, this simulation reflects how an automated taxi system might start in a small city. The first population to be served by such an automated taxi system likely would be those who currently use transit, because they have already demonstrated a willingness to use non-private modes. An automated taxi system theoretically could

serve this population with decreased costs and increased travel convenience. Once such a fleet was up and operating, it most likely would attract additional ridership, especially once its performance was demonstrated and its reliability proven. Those with destinations within the primary service area (under the assumption that such a system must serve a limited area to be efficient) might increasingly see such automated taxi systems as an attractive alternative to driving their own car and parking; some might even forego car ownership itself. Therefore, it is likely that once such an automated taxi transit system was established, it would grow its market share over time and eventually start to serve part of the market that currently relies upon privately-owned vehicles.

Conclusion

This study compared the performance of hypothetical automated taxi transit systems to the Ann Arbor Area Transportation Authority bus system with metrics for service performance, costs, vehicle kilometers traveled, and carbon emissions. A shared-ride automated taxi system performed favorably in comparison to the current bus transit system by providing comparable wait times, shorter travel times, significantly lower costs per day and per passenger kilometer, and somewhat lower carbon emissions. The bus transit system performed better on total vehicle kilometers traveled, which is significant because a switch from a bus system to an automated taxi system could generate new traffic congestion, with peak-hour vehicle kilometers traveled growing by a factor of 6 times. A single-rider automated taxi system performed better than the current transit system with regard to its travel times and costs, but worse with respect to carbon emissions and much worse with respect to peak-hour vehicle kilometers traveled.

These results suggest that a shared-ride automated taxi system could serve as a viable and efficient transit system for a small city, with a service areas similar to current transit service areas. Even if the cost of automation was as high as \$50,000 per vehicle, the costs of such a system would be competitive with current bus transit services. A shared-ride automated taxi system could initially be set up to serve current transit ridership, with the expectation that when its service is proven, the ridership likely would grow and shift mode choice for some people who currently rely upon private vehicles for travel.

The main drawback to a shared-ride automated taxi system would be the increase in traffic congestion resulting from a growth in system peak-hour VKT. Potential solutions to this growth in peak-hour congestion might include coordination of automated taxi systems with large-vehicle transit to efficiently manage heavy peak-hour flows. Other solutions could involve high occupancy requirements for such taxis or limitation of certain critical roadway segments to higher-capacity modes during peak hours (i.e., HOV lanes).

If automated taxi systems become widespread, the role of conventional, large-vehicle transit is likely to shift. The advantage of conventional large-vehicle transit relative to automated taxis is its ability to manage peak-hour flows and limit congestion (or provide an alternative to congested travel) during peak periods. The best transit systems

of the future, it would seem, would take advantage of both the flexibility of automated taxis and the capacity management capabilities of large-vehicle transit within a coordinated system.

Although it may be difficult to imagine the transition steps towards an automated taxi system, the social and economic benefits are potentially high; it is unknown when such a system might overcome the technical and regulatory barriers required for providing an automated vehicle ride-hailing service, but several companies such as Uber, Volvo, Ford, Google, and GM are actively working on developing such systems as part of their current research and development activities (Chafkin 2016; Stoll 2016; Boudette 2016; Grisold 2016). The research here indicates that the performance of a shared-ride automated taxi system could meet vital mobility, environmental, and economic needs. Nevertheless, such automated taxi systems will require forward-looking policy frameworks to ensure that the benefits to passengers are maximized while their potential negative externalities of increased VKT and greater congestion are mitigated.

References

- American Automobile Association. 2013. "Your Driving Costs."
- Ann Arbor Area Transportation Authority. 2014a. "Ann Arbor Area Transportation Authority Boarding Data." Ann Arbor, Michigan.
- Ann Arbor Area Transportation Authority. 2014b. "Ride Guide." Passenger Schedule. Ann Arbor, Michigan.
- Bischoff, Joschka, and Michal Maciejewski. 2016. "Simulation of City-wide Replacement of Private Cars with Autonomous Taxis in Berlin." *Procedia Computer Science*, 83: 237-244.
- Boudette, Neal E. 2016. "Ford Promises Fleets of Driverless Cars within Five Years." *New York Times*, August 16. http://www.nytimes.com/2016/08/17/business/ford-promises-fleets-of-driverless-cars-within-five-years.html?_r=2.
- Bowman, Larry A., and Mark A. Turnquist. 1981. "Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops." *Transportation Research Part A: General* 15 (6): 465–71. doi:10.1016/0191-2607(81)90114-X.
- Burns, Lawrence D., William C Jordan, and Bonnie A. Scarborough. 2013. "Transforming Personal Mobility." The Earth Institute, Columbia University, New York.
- Chafkin, Max. 2016. "Ubers' First Self-Driving Fleet Arrives in Pittsburgh This Month." *Bloomberg*, August 18. <http://www.bloomberg.com/news/features/2016-08-18/ubers-first-self-driving-fleet-arrives-in-pittsburgh-this-month-is06r7on>.
- Chester, Mikhail, and Arpad Horvath. 2009. "Life-Cycle Energy and Emissions Inventories for Motorcycles, Diesel Automobiles, School Buses, Electric Buses, Chicago Rail, and New York City Rail." Berkeley, CA. <https://escholarship.org/uc/item/6z37f2jr>.
- Ciari, Francesco, and Benno Bock. 2014. "Modeling Station-Based and Free-Floating Carsharing Demand : A Test Case Study For Berlin , Germany." *93rd Annual*

- Meeting of the Transportation Research Board, 13. Washington, DC. file:///C:/Users/louismerlin/Documents/Research/_In_Progress/Automated Vehicles as Transit/Ciari Bock 2014 - Modeling Free Floating and Station Based Carsharing.pdf.
- Consumer Reports. 2013. "Why You Might Not Be Getting the Efficiency Promised." *Consumer Reports*, August.
- Fagnant, Daniel J., and Kara M. Kockelman. 2014. "The Travel and Environmental Implications of Shared Autonomous Vehicles, Using Agent-Based Model Scenarios." *Transportation Research Part C: Emerging Technologies* 40 (March): 1-13. doi:10.1016/j.trc.2013.12.001.
- Fagnant, Daniel J, Kara M Kockelman, and Prateek Bansal. 2015. "Operations of Shared Autonomous Vehicle Fleet for Austin, Texas, Market." *Transportation Research Record*, 2536: 98-106.
- Fan, Wei (David), and Randy B. Machemehl. 2009. "Do Transit Users Just Wait for Buses or Wait with Strategies?" *Transportation Research Record*, 2111(1): 169–76. doi:10.3141/2111-19.
- Federal Transit Administration. 2014. "National Transit Database." Washington, DC. <http://www.transit.dot.gov/ntd>.
- Feigon, S., and C. Murphy. 2016. "Shared Mobility and the Transformation of Public Transit," TCRP Report 188. <http://nap.edu/23578>.
- Gnansounou, E., A. Dauriat, J. Villegas, and L. Panichelli. 2009. "Life Cycle Assessment of Biofuels: Energy and Greenhouse Gas Balances." *Bioresource Technology*, 100(21): 4919–30. doi:10.1016/j.biortech.2009.05.067.
- Grisold, Alison. 2016. "Self-Driving Taxis from General Motors and Lyft." *Quartz*, May 5. <http://qz.com/677394/self-driving-taxis-from-general-motors-and-lyft-will-hit-the-road-within-a-year/>.
- Levine, Jonathan, Moira Zellner, Yoram Shiftan, M. Arquero de Alarcon, and A. Diffenderfer. 2013. "Effects of Automated Transit, Pedestrian and Bicycling Facilities on Urban Travel Patterns." Ann Arbor, MI.
- Liang, Xiao, Gonçalo Homem de Almeida Correia, and Bart van Arem. 2016. "Optimizing the Service Area and Trip Selection of an Electric Automated Taxi System Used for the Last Mile of Train Trips." *Transportation Research Part E: Logistics and Transportation Review*, 93: 115–29. doi:10.1016/j.tre.2016.05.006.
- Maheo, A., P. Kilby, and P. Van Hentenryck. 2016. "Benders Decomposition for the Design of a Hub and Shuttle Public Transit System." <https://pascalvanhentenryck.engin.umich.edu/wp-content/uploads/sites/228/2016/01/BusPlus.pdf>.
- Martinez, Luis M., and Jose M Viegas. 2016a. "Concepts and Impacts of New Urban Shared Mobility Alternatives: An Agent-Based Simulation Model for the City of Lisbon, Portugal." *Transportation Research Board 95th Annual Meeting*, 16.

- Martinez, Luis M., and Jose M. Viegas. 2016b. "Concepts and Impacts of New Urban Shared Mobility Alternatives: An Agent-based Simulation Model for the City of Lisbon, Portugal." Transportation Research Board 95th Annual Meeting: 20.
- Nustats. 2012. "Southeast Michigan Council of Governments Regional On-Board Transit Survey."
- Searchinger, Timothy, Ralph Heimlich, Richard A. Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. 2008. "Use of US Croplands for Biofuels Increases Greenhouse Gases through Emissions from Land-Use Change." *Science*, 319 (5867): 1238-40.
- Shaheen, Susan, Nelson Chan, Apaar Bansal, and Adam Cohen. 2015. "Shared Mobility. Definitions, Industry Developments, and Early Understanding." Berkeley, CA.
- Shepherd, Simon, and Helen Muir. 2011. "CityMobil D2.3.2 Strategic Modelling Results." <http://www.citymobil-project.eu/>.
- Spieser, Kevin, Kyle Treleaven, Rick Zhang, Emilio Frazzoli, Daniel Morton, and Marco Pavone. 2014. "Toward a Systematic Approach to the Design and Evaluation of Automated Mobility-on-Demand Systems A Case Study in Singapore." Cambridge, MA. <http://hdl.handle.net/1721.1/82904>.
- Stoll, John D. 2016. "Volvo, Uber to Jointly Develop Autonomous Sport Utility Vehicles." *Wall Street Journal*, August 18. <http://www.wsj.com/articles/volvo-uber-to-jointly-develop-autonomous-sport-utility-vehicles-1471518577>.
- US Census Bureau. 2013. "American Community Survey (5-Year Estimates)."
- US Department of Energy. 2015. "Biodiesel." Energy Efficiency and Renewable Energy. <http://www.fueleconomy.gov/feg/biodiesel.shtml>.
- US Energy Information Administration. 2015. "Frequently Asked Questions." <http://www.eia.gov/tools/faqs/faq.cfm?id=307&t=11>.
- Woodman, Spencer. 2016. "Welcome to Uberville." *The Verge*, September. <http://www.theverge.com/2016/9/1/12735666/uber-altamonte-springs-fl-public-transportation-taxi-system>.
- Zhang, Wenwen, Subhrajit Guhathakurta, Jinqi Fang, and Ge Zhang. 2015a. "Exploring the Impact of Shared Autonomous Vehicles on Urban Parking Demand: An Agent-Based Simulation Approach." *Sustainable Cities and Society*. doi:10.1016/j.scs.2015.07.006.
- Zhang, W., S. Guhathakurta, J. Fang, and G. Zhang. 2015. "Exploring the Impact of Shared Autonomous Vehicles on Urban Parking Demand: An Agent-based Simulation Approach." *Sustainable Cities and Society*, 19: 34-45.

About the Author

LOUIS A. MERLIN, PH.D., AICP (lmerlin@fau.edu) is an Assistant Professor of Urban Planning at Florida Atlantic University. He holds a B.A. in Mathematics with Honors from Yale University and conducted National Science Foundation-funded graduate work towards master's degrees in Operations Research and City and Regional Planning from the Georgia Institute of Technology. He earned his doctoral degree in City and Regional Planning from the University of North Carolina, with research work funded by the Royster Society of Fellows. His current research work concerns how to better integrate transportation and land use planning through the application of accessibility-based planning tools, as well as the implications of automated vehicles for transportation systems.

Effects of a Public Real-Time Multi-Modal Transportation Information Display on Travel Behavior and Attitudes

Yanbo Ge, Parastoo Jabbari, Don MacKenzie
University of Washington

Jiarui Tao
University of California, Irvine

Abstract

This study reports on an experiment in downtown Seattle, Washington, to evaluate whether installing a public real-time multi-modal transportation information display screen in an office building lobby caused changes in building occupant self-reported awareness, attitudes, satisfaction, and usage of alternative transportation modes including transit, car-sharing, ride-sourcing, and bike-sharing services. Workers in the test building and two nearby control buildings were surveyed immediately before the screen was installed (N=550) and again six months later (N=455). Little evidence was found that exposure to the real-time display affected respondent travel choices, satisfaction, familiarity, or attitudes toward alternative modes. Although most respondents (70%) had noticed the screen and had generally positive reactions, two-thirds of this group never actually used it. These results, along with building occupant responses to open-ended questions, indicate limited benefits from this installation and suggest that site selection, screen placement, and marketing may help to maximize the effects of these types of displays on traveler satisfaction and mode shifting.

Keywords: *Real-time information, mode choice, commute trip reduction*

Introduction

Real-time information systems have emerged in recent years as a cost-effective way to make alternatives to driving more attractive, especially since traditional approaches such as expanding service areas, increasing frequency, and enhancing on-time

performance tend to be expensive. Alternative transportation services such as public transit, shared-use vehicle programs, and ride-sourcing platforms (also known as transportation network companies, or TNCs) provide mobility with higher sustainability and less environmental impact than privately-owned vehicles (Poudenx 2008). However, automobile use continues to predominate for multiple reasons, including autos' convenience, flexibility, and wide availability. Thanks to continuous progress in information and computing technology, real-time transportation information systems have emerged as a relatively low-cost approach to making alternative modes more attractive (Lyons and Urry 2005). Properly implemented, a real-time transportation information system is a travel demand management tool that presents current and potential travelers with dynamic, timely, and accurate information on alternative transportation services, such as vehicle arrival times, service availability updates, and service change notifications. Transit Cooperative Research Program (TCRP) Synthesis Report 104 provides a detailed review of the technical characteristics of real-time transit information systems and best practices for installation, largely in the context of signage at transit stops (Schweiger 2013). It also documents many of the benefits of such systems, including lower perceived wait times, greater feelings of safety and security, and better overall perceptions of service. Increasingly, the provision of real-time information is seen as essential to attracting passengers, increasing revenues, and projecting the image of a state-of-the-art transportation system (Dziekan 2004; Lyons and Harman 2002).

Many jurisdictions have invested in real-time transportation information systems based on their expected economic, social, and environmental benefits (Cham et al. 2006). However, compared with the large body of research on traveler information systems and driving behavior (Lappin and Bottom 2001), less work has evaluated traveler responses to real-time information about alternative modes. Although several studies have examined associations between ridership and real-time information among transit riders, few have convincingly addressed the causal effect of real-time transportation information displays on the choices of travelers. The key contribution of this study is to use a pre-test/post-test control group research design, analyzed by difference-in-differences, to identify the causal effects of installing a real-time information display on traveler attitudes, satisfaction, and choices.

Prior Work

Prior research has used stated preference, revealed preference, and simulation approaches to assess the effects of providing real-time transit information. Initial efforts used stated preference approaches and suggested the potential to increase transit usage by providing real-time transportation information (Abdel-Aty et al. 1996; Reed and Levine 1997; Abdel-Aty 2001). Others used simulations to evaluate the potential time savings and route choices of travelers provided with real-time information (Hickman and Wilson 1995). Once real-time information systems were deployed, several revealed preference studies assessed their real world impacts. For example, Zhang et al. (2008) used a before-and-after study to test the effectiveness of deploying ShuttleTrac on the

University of Maryland College Park campus. The ShuttleTrac system disseminates real-time bus arrival information via telephone, website, terminals at selected stops, and a large display at an activity center. This study found insignificant impacts on individual shuttle trip frequency, waiting anxiety, and feelings of security during the day. However, rider feelings of security after dark and their overall level of satisfaction increased with ShuttleTrac use. Recently, particular attention has been paid to evaluating effects using carefully-designed studies. In one noteworthy example, Brakewood et al. (2014) encouraged a randomly-selected subset of bus riders in Tampa, Florida, to use OneBusAway, a mobile application conveying transit arrival information. Their results provided strong evidence that the access to real-time information significantly improved the passenger experience of waiting for the bus, but found no effects on trip and transfer frequencies.

Real-time transit information may affect several outcomes, including mode choices, route choices, satisfaction, and perceptions of ease-of use, waiting time, and security. It reduces the uncertainty of accessing transportation services, so that travelers reduce their time wasted on waiting and the productivity lost to missed, delayed, or unavailable transportation service (Swanson et al. 1997). Smith et al. (1994) evaluated the effects of the application of advanced transport telematics in London, namely the Countdown project. The London Countdown system led to increased positive attitudes towards bus travel, the bus operator, and the local public transportation authority. For a ferry system along the Thames River in London, real-time information also enhanced the general impression of that particular travel option (Cassidy and White 1995). A conjoint analysis found that real-time information was expected to reduce the burden of waiting as the degree of certainty increased (Reed 1995). Thus, access to real-time information promotes feelings of reliability and convenience (Zito et al. 2011). When customer evaluations were conducted of bus status video monitor programs known as Transit Watch and Transit Tracker in Seattle, Washington, and Portland, Oregon, passengers felt less uncertainty and more in control after each implementation (Science Applications International Corporation 2003). Several studies have reported that real-time information affects wait times at transit stations in a positive way. The shortened wait time is associated with reduced disutility, less anxiety, and an increased feeling of personal security during the wait (Forsyth and Silcock 1985). In a before-and-after study, McCord et al. (2015) found that users of the Ohio State University's real-time bus information system reported more positive attitudes about the bus system's environmental and congestion benefits. Another recent study showed that real-time information via web-enabled and mobile devices caused modest increases in public transit ridership in New York City, particularly on heavily-traveled routes (Brakewood et al. 2015).

Earlier studies pertained primarily to information presented to travelers at transit stops, but in recent years mobile applications have emerged as a medium for providing real-time information directly to travelers. An example is the CTA Bus Tracker in the Chicago Transit Authority bus system. To investigate its impact on bus ridership, Tang and Thakuriah (2012) analyzed longitudinal data of route-level ridership. The incremental implementation of CTA Bus Tracker on different routes enabled their quasi-

experimental design. They estimated linear mixed models that indicated a significant, modest, time-varying increase in monthly average weekday ridership after the provision of Bus Tracker service than before. Watkins et al. (2011) found that both the actual and perceived wait times of transit passengers with access to real-time information (via the OneBusAway mobile app) were shorter than those of passengers without.

In general, prior research provides more support for the notion that real-time information improves attitudes and satisfaction with alternative modes, but less compelling evidence that it directly affects mode choices or trip frequency using the alternative modes. Notably, past work has focused primarily on real-time information provided through displays at transit stops or via mobile apps. The present work focused on both the travel behavior and perceptual effects of a different medium: real-time information provided through a display screen at a public location other than a transit stop, specifically the lobby of an office building.

Hypotheses

This study tested the following hypotheses about the effects on traveler perceptual and behavioral responses of a real-time transportation information display in a public location:

1. Individuals exposed to the real-time information display are more likely to agree that sufficient resources exist for transportation information.
2. Individuals exposed to the real-time information display report higher levels of familiarity with alternatives to personal car travel.
3. Individuals exposed to the real-time information display report more favorable attitudes toward the modes featured on the information display.
4. Individuals exposed to the real-time information display are more likely to choose alternative travel modes for their commutes in particular, and for travel in general.

Methodology

A field experiment was conducted to test the hypotheses that exposure to a real-time transportation information display affects travel behavior and perceptions of alternative modes. The experiment was based on a pre-test/post-test control group design and was analyzed using a difference-in-difference analysis (Card and Krueger 1993). First, workers in three office buildings were divided into a treatment group ("Building A") and a control group ("Building B" and "Building C"). Both groups participated in a web-based survey that measured travel behaviors, perceptions, and selected background variables. This pre-test survey was completed between late May and early June 2015. A real-time information display screen was installed in the treatment group building on June 15, 2015, and building occupants completed a post-test survey in December 2015.

Research Location and Transportation Context

The location for this experiment was selected based on several criteria. First, the location needed to have convenient access to alternative transportation modes so commuters would have viable alternatives to driving alone. Second, the treatment and control sites had to be physically close to one another so the difference-in-difference experimental design would be valid. Third, the sites needed to be large enough to provide a sufficient sample size. Finally, a property manager was needed who would be a willing partner in allowing the installation of the real-time display and helping to contact building occupants. These criteria led to selecting a site in downtown Seattle after reviewing several candidate sites suggested by the Seattle Department of Transportation.

This experiment was conducted in three buildings located in the 11-acre area of downtown Seattle known as the Metropolitan Tract. Managed by a single property management company, all three buildings are within 400–600 ft of one another, with similar access to transportation infrastructure and resources. Within a quarter mile of the buildings studied in this experiment are 167 different transit routes. In the half-mile circular area around the three buildings is access to ferry, water taxi, and the South Lake Union streetcar. Downtown Seattle is also well-served by TNCs (Uber and Lyft), carsharing (car2go and Zipcar), and traditional taxis. Due to the central location, excellent bike lanes, and convenient public transportation services, Walk Score has rated the area a walk score of 99, a transit score of 100, and a bike score of 64 to 74.

The plethora of viable alternatives to driving alone has led to high usage of alternative modes in downtown Seattle. According to the latest commuter survey (Commuter Seattle 2015) among downtown Seattle's estimated 228,000 employees, 31% of commuters drove alone to work, down from 35% in 2010 and 34% in 2012. Public transit was the most popular choice for downtown commuters (45%), followed by driving alone (31%), ridesharing (9%), walking (7%), teleworking (4%), and bicycling (3%).

Data Collection

Occupants (i.e., employees whose regular workplace is in the building) of the three office buildings were surveyed in May and June 2015. Subjects were recruited via emails sent by property managers to tenant companies, who forwarded the emails on to individual workers. As an incentive to complete the survey, respondents were entered into a drawing to receive one of two iPads valued at \$499 each. Out of a total of 2,575 occupants in the three buildings, 808 clicked through to the survey and 550 (21%) submitted usable responses. The second survey was conducted between December 7 and 21, 2015, approximately six months after the real-time information display was installed at the treatment site on June 15. Prior studies have suggested that a study period of six months should be sufficient to detect some longer-term responses to the availability of real-time information (Dziekan and Vermeulen 2006; Brakewood et al. 2014). Respondents again were offered the chance to win an iPad. In total, 709 of 2,579 occupants viewed the post-test survey, and 455 (18%) submitted valid responses. Also

identified were 137 respondents (5%) who completed both waves of the survey, which were analyzed separately.

The research team developed the survey instrument specifically for this project to elicit data on four measures of interest: (1) familiarity with, (2) attitudes toward, (3) satisfaction with, and (4) usage of alternative travel modes. Survey items included a question about commute mode to and from work for the past five days and asked respondents to complete a one-day, recall-based travel diary. The average time to complete the survey was 20 minutes. Full details of the survey instrument are reported by MacKenzie et al. (2016).

Experimental Intervention

The treatment in this study was the installation of a real-time multimodal transportation information display in the lobby of Building A on June 15, 2015. Like many other public real-time transportation information systems, the display used in this study incorporated countdown information for nearby transit stops. However, as shown in Figure 1, it also provided information on the quantity and location of available carsharing vehicles, the estimated arrival time of TNC vehicles, and the availability of nearby bikeshare bicycles, obtained from service providers' application programming interfaces (APIs). During the interface design stage, transit stations and stops were prioritized based on their proximity to Building A and their ability to serve the home ZIP codes of pre-test survey respondents. The content and design of the screen was updated in the initial few weeks after installation based on feedback from the building's property managers and Seattle Department of Transportation. A snapshot of the final version of the public display is illustrated in Figure 1.

The screen used in this experiment was a 65-inch, 1080p edge-lit LED LCD Planar display. It was installed along a wall near the main entrance, information desk, and elevators in the ground-floor lobby of Building A so most people could easily see the display upon entering and exiting and drivers who needed to use the garage elevator would also be exposed to the information. No displays were installed at Buildings B or C, and none of the three buildings was equipped with a real-time transportation information display before the study. For purposes of our analysis, it was assumed that people who worked in Buildings B or C would not go to Building A just to use the public display.

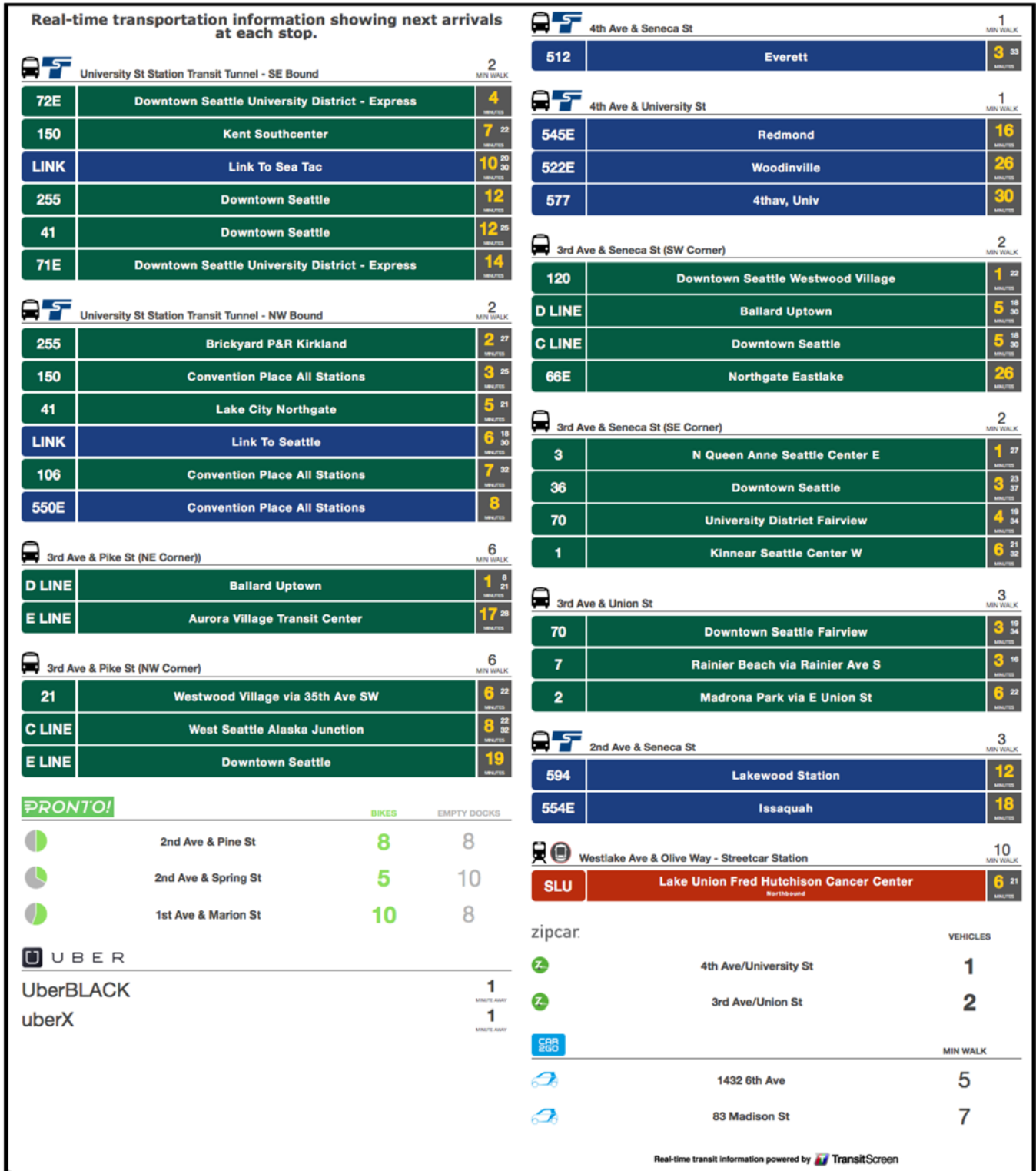


FIGURE 1. Screenshot of real-time multi-modal transportation information display

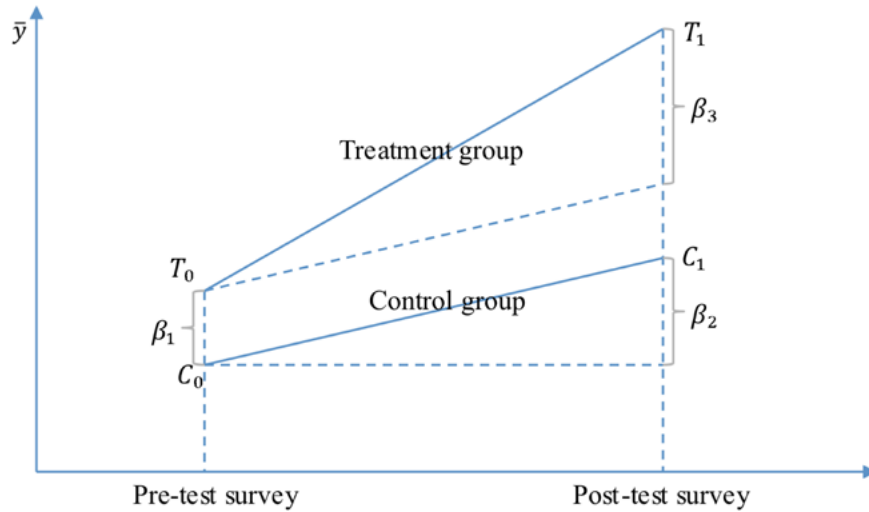
Data Analysis

This study used a difference-in-difference quasi-experimental design to control for time-varying factors and estimate the causal effects of introducing a real-time information display. In an experiment involving habitual behaviors such as travel, treatment effects may take time to materialize. Thus, we waited approximately six months after the installation of the display before conducting the posttest survey. However, simply comparing responses before and after the intervention does not provide a credible estimate of the causal effects, since many other factors (weather, gasoline prices, service quality, etc.) might also affect respondent choices and attitudes even if the screen had never been installed.

The intuition of the difference-in-difference design is simple (Card and Krueger 1993). There are two groups (treatment and control) and two time periods (before and after treatment), and the interests is in some outcome variable(s). The difference in outcomes for the control group is measured before and after the treatment, and the difference in the treatment group before and after treatment. It is then assumed that whatever difference is observed in the control group represents what would have been observed in the treatment group if the latter had not received the treatment. When this assumption is made, it can be concluded that the causal effect of the treatment is the difference between the two differences calculated previously: the “difference in differences.” This is illustrated graphically in Figure 2 for a generic outcome variable y . Note that there is no assumption that the treatment and control groups are exactly the same, only that their changes over time would have been the same if not for the treatment being administered.

	Control	Treatment	Difference	Treatment Effect
Before	C_0	T_0	$D_0 = T_0 - C_0$	$D_1 - D_0 = (T_1 - C_1) - (T_0 - C_0)$
After	C_1	T_1	$D_1 = T_1 - C_1$	

FIGURE 2.
Difference-in-difference estimator; β_3 is estimated average treatment effect



The difference-in-difference estimator was developed into a regression modeling framework. To represent the group assignment and time period, two dummy variables were created, as shown by equation (1) and equation (2).

$$g_i = \begin{cases} 0, & \text{if } i \text{ is in the control group} \\ 1, & \text{if } i \text{ is in the experimental group} \end{cases} \quad (1)$$

$$t_i = \begin{cases} 0, & \text{if the observation is in the pretest survey} \\ 1, & \text{if the observation is in the posttest survey} \end{cases} \quad (2)$$

For simple, continuous outcome variables, an ordinary-least-squares regression model typically would be used, where ε_i is a random disturbance that is assumed to be independent of the explanatory variables, as shown in equation (3).

$$y_i = \alpha + \beta_1 g_i + \beta_2 t_i + \beta_3 g_i t_i + \varepsilon_i \quad (3)$$

In the pre-test survey ($t = 0$), no treatment takes effect. The expected value of the dependent variable among the control group ($g_i = 0$) can be represented as equation (4):

$$E[y|g = 0, t = 0] = \alpha \quad (4)$$

The expected value in the treatment group ($g_i = 1$) can be represented as equation (5):

$$E[y|g = 1, t = 0] = \alpha + \beta_1 \quad (5)$$

Thus, β_1 represents the baseline difference between the two groups. In the post-test survey ($t = 1$), the treatment is applied only to the treatment group. The expected value of the outcome among the control group ($g_i = 0$) can be represented as equation (6).

$$E[y|g = 0, t = 1] = \alpha + \beta_2 \quad (6)$$

Among the treatment group, ($g_i = 1$), the expected value can be represented as equation (7).

$$E[y|g = 1, t = 1] = \alpha + \beta_1 + \beta_2 + \beta_3 \quad (7)$$

Thus, β_2 denotes the change over time in the control group, which is assumed to represent the change that would have occurred in the treatment group if it had not received the treatment. The coefficient β_3 captures the additional change in the treatment group beyond any initial differences with the control group and the change over time within the control group. Thus, β_3 is the estimate of the causal effect of the treatment on the outcome.

Since many variables were non-continuous or non-normally distributed, various generalized linear models were used, each using the basic specification above as its linear predictor (i.e., the “right-hand side”). For mode choices, which are discrete, a logistic regression model was used. For attitudinal measures, which were measured on a Likert-type ordinal scale, an ordered logistic regression model was used. For daily vehicle miles traveled (which is often exactly zero, a condition known as zero-inflation), a gamma hurdle model was used, which allows first modeling whether or not miles traveled is zero, and if it is nonzero, modeling its magnitude. In cases in which there were repeated observations from the same respondent, mixed-effects variants of these models were used to capture respondent-specific characteristics.

Results

Table 1 summarizes the number of workers and respondents from each building in each wave of the survey. About one-third of the post-test respondents were linked to responses in the pre-test survey, based on email addresses they provided. Therefore, two parallel sets of analyses were conducted. First, the full pre-test and post-test samples were considered as independent cross-sections. Second, the 137 respondents who could be positively identified as having completed both waves of the survey were analyzed as a panel data set. This section reports the effects of the real-time information display on traveler awareness, attitudes, and satisfaction toward various transportation modes and its effects on self-reported travel behavior and reviews respondent awareness, usage, attitudes, and comments regarding the real-time information display itself. More detailed results are reported by MacKenzie et al. (2016).

TABLE 1.
Building Occupants,
Respondents, and Response
Rates in Pre-Test and Post-
Test Surveys

	Pre-Test	Post-Test
Building A – Treatment		
Number of tenants (organizations)	29	33
Number of occupants (individuals)	1,192	1,110
Number of valid responses	267	175
Response rate	22.4%	15.8%
Repeat respondents	-	50
Building B – Control		
Number of tenants (organizations)	56	59
Number of occupants (individuals)	695	733
Number of valid responses	176	162
Response rate	25.3%	22.1%
Repeat respondents	-	54
Building C – Control		
Number of tenants (organizations)	38	40
Number of occupants (individuals)	688	736
Number of valid responses	107	118
Response rate	15.6%	16.0%
Repeat respondents	-	33
Total		
Number of tenants (organizations)	123	132
Number of occupants (individuals)	2,575	2,579
Number of submitted responses	567	466
Number of valid responses	550	455
Response rate	21.4%	17.6%
Repeat respondents	-	137
Number of viewers	808	709
Valid completion rate	68.1%	64.2%

Respondent Use and Evaluation of Real-Time Information Display

In the post-test survey of the treatment group, 175 valid responses were received, of which 124 (about 70%) reported knowing about the real-time information screen that had been installed in the lobby of their building. Among the 124 who knew about the screen, 84 did not use the information on the screen for their travel decisions and only 9 said they used the screen information daily (Figure 3). These results were similar when the sample was restricted to respondents who had commuted using one of the modes featured on the screen at least once in the week preceding the post-test survey. Among these 127 respondents, 88 (about 70%) knew about the screen, and 56 never used the information on it for their travel decisions.

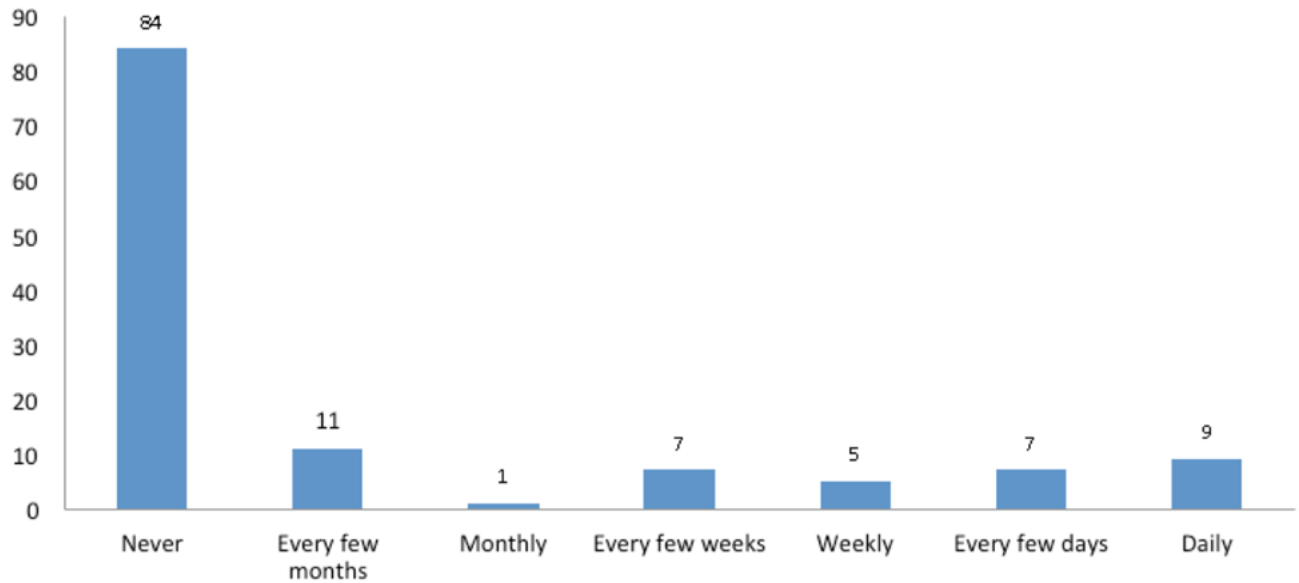
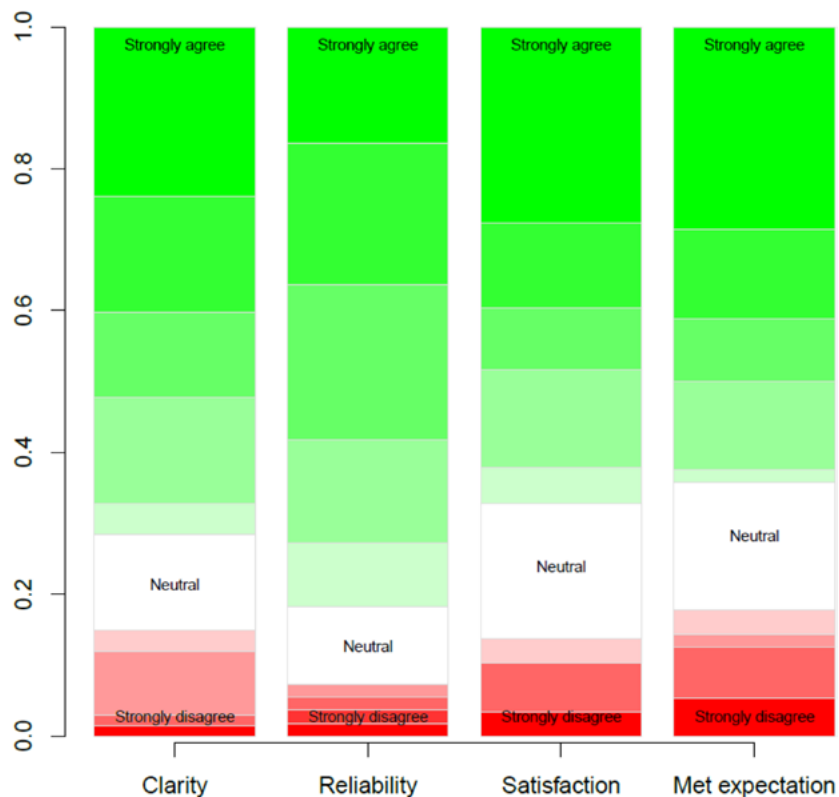


FIGURE 3. Screen usage frequency among treatment group respondents who knew about screen

Treatment group respondents who were aware of the screen were asked if the screen was easy to read and understand, if it displayed accurate and reliable travel information, if they were satisfied with it, and if it met their expectations. As shown in Figure 4, most thought the screen was easy to understand and reliable and met their expectations.

FIGURE 4. Perceptions of real-time display screen among treatment group respondents who were aware of its presence.



To gain a deeper understanding of people's perceptions of the screen, responses to an open-ended question about the screen and how it might be improved were reviewed. The full responses are provided in Appendix B of MacKenzie et al. (2016), and the following themes among the responses were noted:

- Numerous respondents noted that they prefer to use OneBusAway or similar smartphone apps to get the same information shown on the screen.
- Several comments implied that the respondent thought the display screen was showing schedule information, not real-time information.
- Several comments noted that the screen did not show route information for their transit routes.
- Several commented on the location of the screen—that it was difficult to see, in a corner, or too close to the building's security guard.

Effects of Real-Time Display on Awareness, Attitudes and Satisfaction

This section presents detailed results for the effects of the real-time information display on awareness, attitudes, and satisfaction with public transportation. Also presented are some key summary results for driving and other alternative modes (full results for these modes are reported by MacKenzie et al. [2016]). The results of the statistical analyses generally do not provide evidence that the real-time information display caused a change in satisfaction, attitudes, or awareness of any modes.

Table 2 summarizes the median ratings of perceptual indicators relating to various travel modes. A minimum score of 0 and a maximum score of 10 apply to each of these indicators. Across the treatment and control groups and both survey waves, respondents were very familiar with the local public transportation systems, moderately familiar with TNC services, and only slightly familiar with car-share and bike-share services. (Although "TNC" is used in this paper, "Hired car service [e.g., Uber, Lyft]" was used in the questionnaire to avoid confusing respondents; Table 2 reflects the language used in the questionnaire.) Respondents considered public transportation the most important among all travel options, followed by driving and walking. In terms of satisfaction, travel by walking received the highest evaluation, followed by public transportation, TNC service, and car-share service. For service quality factors such as convenience and reliability, TNC service had the highest ratings, even exceeding driving and public transportation.

TABLE 2. Perceptual Indicators by Group (0–10 scale)

Perception by Mode	Control Group Pre-Test N = 283	Control Group Post-Test N = 280	Treatment Group Pre-Test N = 267	Treatment Group Post-Test N = 175
	Median	Median	Median	Median
How familiar are you with the following travel options around the Seattle area? (0=not at all, 10=extremely familiar)				
Public transportation (e.g., buses, light rail)	9	9	8	9
Hired car service (e.g., Uber, Lyft)	5	6	5	7
Car-share services (e.g., Zipcar, car2go)	3	3	3	3
Bike-share service (e.g., Pronto)	1	1	1	1
How important are the following travel options for your daily travel? (0=not at all, 10=extremely important)				
Driving	6	6	7	7
Bicycling	0	0	0	0
Walking	7	6	5	7
Public transportation (e.g., buses, light rail)	10	10	10	10
Hired car service (e.g., Uber, Lyft)	0	1	1	1
Car-share services (e.g., Zipcar, car2go)	0	0	0	0
Bike-share service (e.g., Pronto)	0	0	0	0
Overall, how satisfied are you with the following travel options around the Seattle area? (0=extremely dissatisfied, 10=extremely satisfied)				
Driving	5	4	5	4
Bicycling	5	5	5	5
Walking	7	6	7	7
Public transportation (e.g., buses, light rail)	7	7	7	7
Hired car service (e.g., Uber, Lyft)	7	7	7	7
Car-share services (e.g., Zipcar, car2go)	6	5	5	5
Bike-share service (e.g., Pronto)	5	5	5	4
Convenience: X is convenient. (0=strongly disagree, 10=strongly agree)				
Driving	5	5	6	5.5
Riding public transportation	7	7	7	7
Using hired car services	8	8	8	9
Using car-share vehicle services	6	5	5	5
Using bike-share services	3.5	3	3	3.5
Reliability: X is reliable. (0=strongly disagree, 10=strongly agree)				
Driving	6	6	7	7
Riding public transportation	6	6	6	6
Using hired car services	7	8	7	8
Using car-share vehicle services	6	5	5	5
Using bike-share services	5	5	5	5
Sufficient information is available about X. (0=strongly disagree, 10=strongly agree)				
Driving	7	7	7	8
Using public transportation	7	7	7	8
Using hired car services	7	7	7	7

TABLE 2. (CONT'D.) Perceptual Indicators by Group (0–10 scale)

Perception by Mode	Control Group Pre-Test N = 283	Control Group Post-Test N = 280	Treatment Group Pre-Test N = 267	Treatment Group Post-Test N = 175
	Median	Median	Median	Median
Using car-share vehicle services	5	5	4	5
Using bike-share services	4	3	3	3
Expanding X is beneficial. (0=strongly disagree, 10=strongly agree)				
Public transportation services	10	10	10	10
Hired car services	6	7	6	7
Car-share vehicle services	7	7	6	7
Bike-share services	6	5	5	6
I prefer to X whenever possible. (0=strongly disagree, 10=strongly agree)				
Drive	3	3	5	5
Ride public transportation	7	7	5	6
Use hired car services	3	2	3	3
Use car-share vehicle services	2	1	2	1
Use bike-share services	1	0	0	0

Figures 5 through 9 summarize the distributions of reported familiarity, attitudes, and satisfaction with public transportation in the control and treatment groups, before and after the screen was installed. Overall, all groups were fairly similar in these metrics. Some small differences can be identified in the figures and are discussed here. Later in this section, whether these differences were statistically significant or if they could have occurred by chance are discussed. Figure 5 shows that both groups were very familiar with public transportation, with similar distributions before and after the screen was installed. Both groups also consider public transportation to be important to their daily travel (Figure 6), and its importance may have increased slightly between the pre-test and post-test. A large majority in both groups was satisfied with public transportation (Figure 7), but satisfaction appears to decrease slightly between the pre-test and post-test. In all groups, less than 20% disagreed with the idea that sufficient information was available about public transportation (Figure 8). Curiously, between the pre-test and post-test, the treatment group showed an increase in both the fraction strongly agreeing and the fraction disagreeing that sufficient information was available. A majority agreed that they preferred to ride public transportation whenever possible (Figure 9), and there may have been a small shift in the tendency of the treatment group to agree with this statement.

FIGURE 5.

Stated familiarity with public transportation for treatment and control groups before and after screen installation

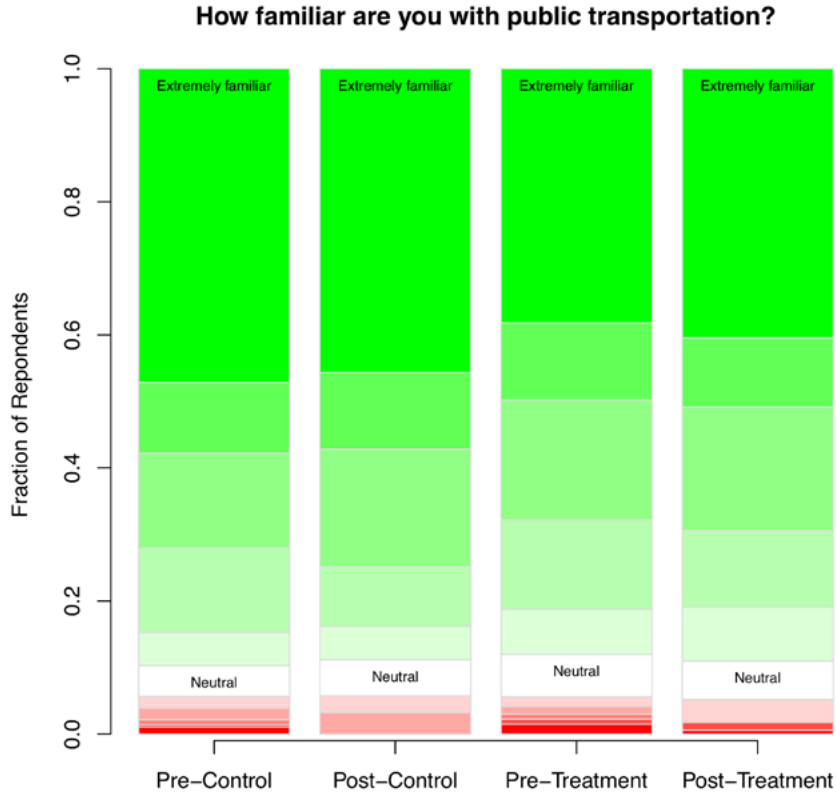


FIGURE 6.

Stated importance of public transportation for treatment and control groups before and after screen installation

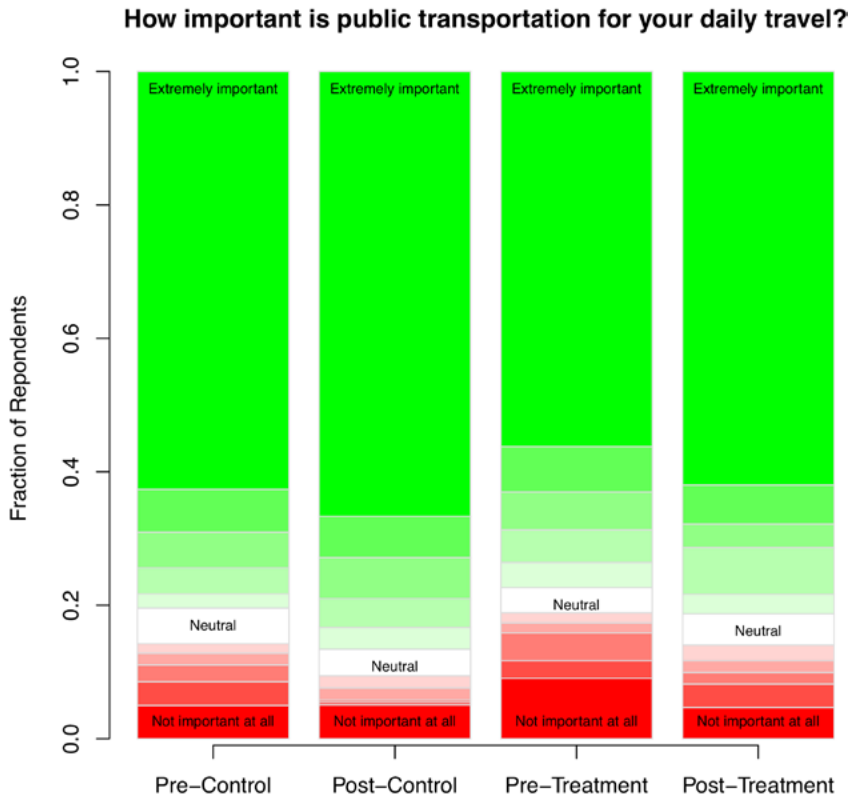


FIGURE 7.
Stated satisfaction with public transportation for treatment and control groups before and after screen installation

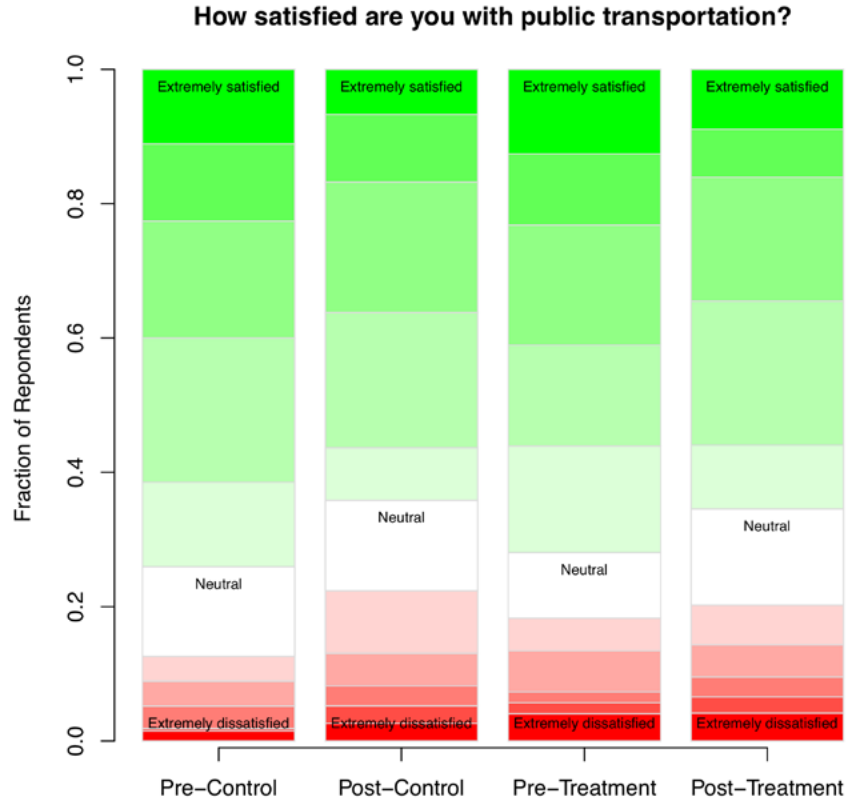


FIGURE 8.
Views on sufficiency of information about public transportation for treatment and control groups before and after screen installation

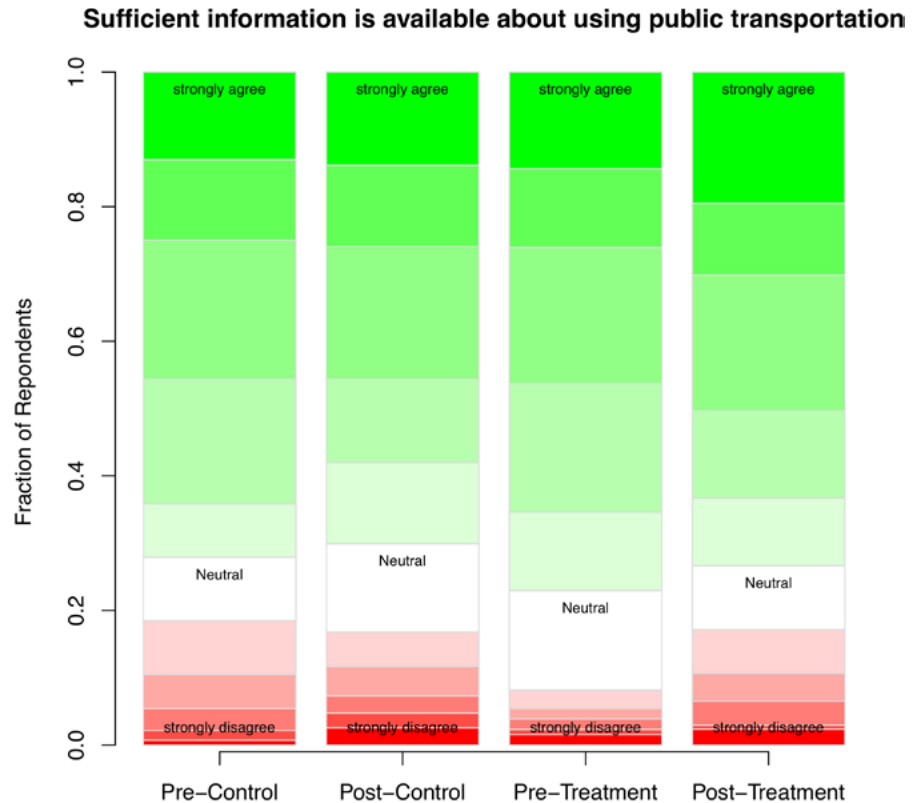
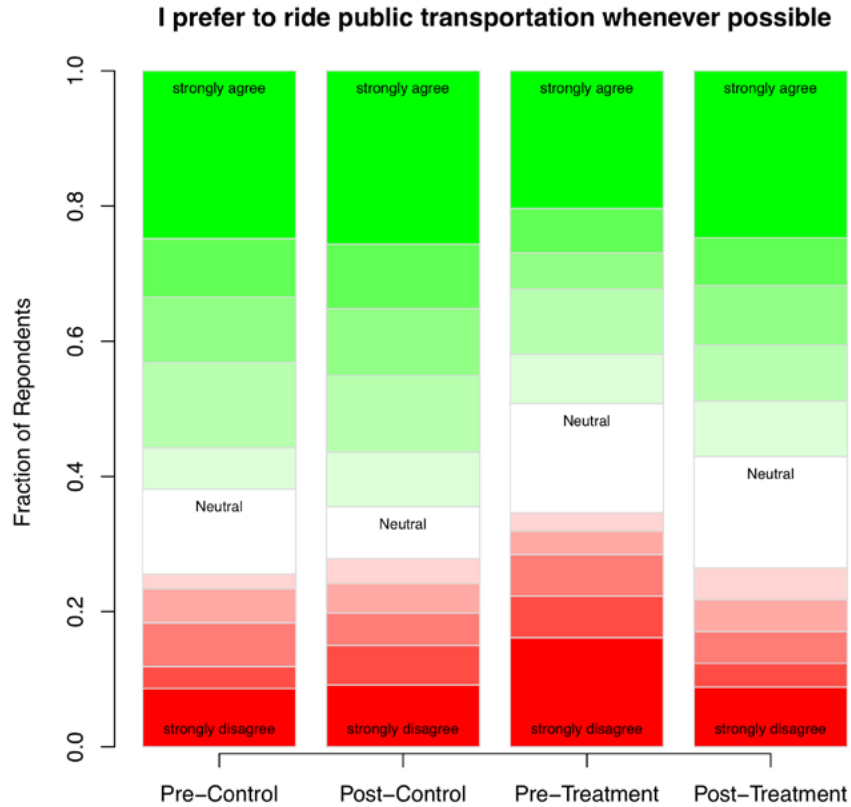


FIGURE 9.
Stated preferences for public transportation for treatment and control groups before and after screen installation



Ordered logistic regression was used to test whether the real-time display screen had a significant effect on satisfaction with or attitudes toward public transportation, using the model specification in equation (3). The estimated treatment effects and associated p -values from these analyses are summarized in Table 3. Each row corresponds to a single perceptual indicator. The first column is the estimated effect of the real-time display on that perceptual indicator, for the full sample of respondents. The second column contains the corresponding p -value for the causal effect estimate, based on a likelihood ratio test on the treatment-posttest interaction term. The third column contains the estimated effect of the real-time display on the perceptual indicator for the subset of 137 respondents who answered both waves of the survey. The fourth column contains the p -value of the estimate in column three, based on a likelihood ratio test. None of the estimated treatment effects related to public transportation were statistically significant at the $\alpha=0.05$ level. Table 3 also summarizes the estimated causal effects of the real-time display on familiarity, satisfaction, and attitudes toward driving and other alternative modes. Several of these estimates (noted in boldface) are statistically significant at conventional levels ($\alpha=0.05$). However, in an experiment such as this where multiple comparisons are being made, there is an increased risk of false positives. Since there were approximately 40 outcomes of interest and 2 modeling approaches (full-sample and repeat-respondents only), 80 comparisons in total were made. The large number of comparisons means more opportunities to make a type I error (a "false positive"). To mitigate this risk, a Bonferroni correction was applied, dividing the significance threshold by 80 (the number of comparisons). This reduces

the significance threshold from $\alpha=0.05$ to $\alpha=0.0006$. Once this was done, none of the effects in Table 3 appear to be significant. These results are consistent with a multivariate analysis of variance (MANOVA), which failed to reject the null hypothesis that perceptions are the same across all groups ($p=0.52$).

TABLE 3.
Estimated Treatment Effects for Real-Time Multi-Modal Display Screen on Familiarity, Attitudes, and Satisfaction with Various Modes, and Associated p-Values

		All Respondents		Repeat Respondents Only	
		Est. Treatment Effect	p	Est. Treatment Effect	p
Public transportation	Familiarity	0.083	0.722	-0.179	0.740
	Importance	0.041	0.872	1.130	0.169
	Satisfaction	0.089	0.701	0.282	0.565
	Convenience	-0.078	0.731	-0.278	0.572
	Reliability	0.055	0.810	-0.528	0.283
	Information sufficiency	0.080	0.729	-0.099	0.840
	Expansion is beneficial	0.271	0.314	0.419	0.512
	Prefer to use	0.357	0.117	0.732	0.173
TNC services	Familiarity	0.052	0.818	-0.219	0.662
	Importance	0.263	0.274	-0.108	0.831
	Satisfaction	0.145	0.635	-0.629	0.347
	Convenience	0.027	0.925	-0.725	0.256
	Reliability	0.001	0.996	-0.744	0.232
	Information sufficiency	0.104	0.703	-0.907	0.129
	Expansion is beneficial	0.342	0.220	0.437	0.480
	Prefer to use	0.457	0.095	1.261	0.042
Carsharing	Familiarity	0.049	0.828	-0.615	0.209
	Importance	0.428	0.103	0.593	0.191
	Satisfaction	0.549	0.168	0.358	0.677
	Convenience	0.362	0.331	0.762	0.351
	Reliability	0.717	0.071	0.849	0.301
	Information sufficiency	0.654	0.055	0.553	0.424
	Expansion is beneficial	0.696	0.046	0.597	0.383
	Prefer to use	0.429	0.215	1.292	0.114
Bikesharing	Familiarity	0.205	0.384	-0.360	0.497
	Importance	0.173	0.597	0.967	0.265
	Satisfaction	-0.571	0.250	-1.519	0.234
	Convenience	0.526	0.231	0.058	0.957
	Reliability	0.153	0.748	-0.181	0.871
	Information sufficiency	0.397	0.300	0.559	0.500
	Expansion is beneficial	0.752	0.047	0.101	0.917
	Prefer to use	0.828	0.040	3.355	0.002

TABLE 3. (CONT'D)

Estimated Treatment Effects for Real-Time Multi-Modal Display Screen on Familiarity, Attitudes, and Satisfaction with Various Modes, and Associated p-Values

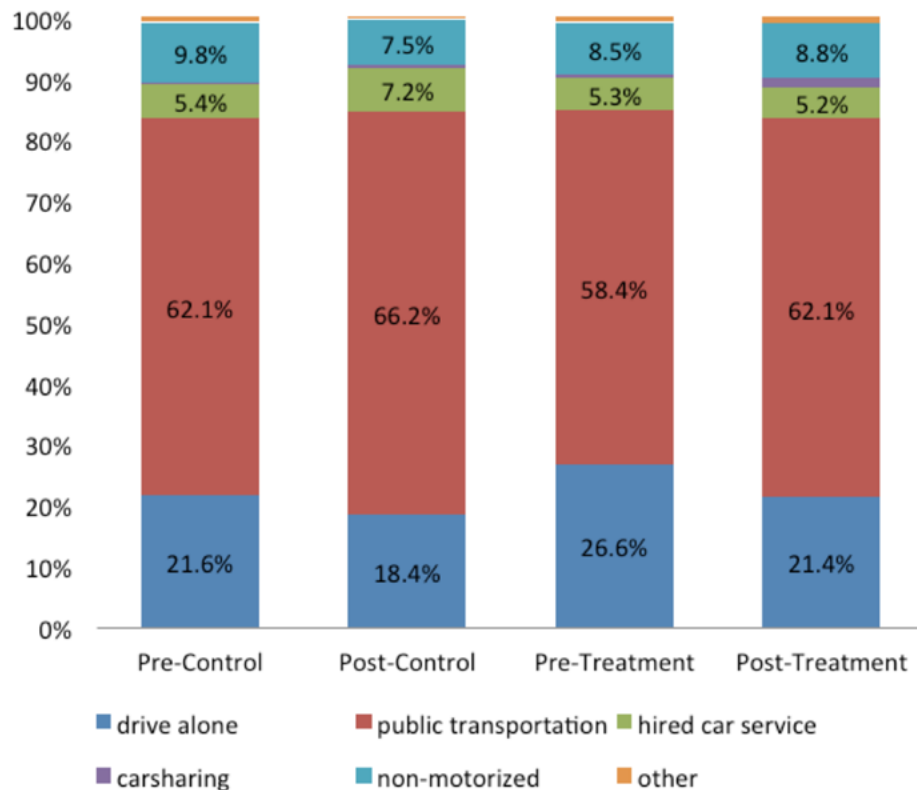
		All Respondents		Repeat Respondents Only	
		Est. Treatment Effect	p	Est. Treatment Effect	p
Driving	Satisfaction	-0.342	0.143	-0.340	0.484
	Convenience	-0.126	0.582	0.508	0.304
	Reliability	-0.134	0.561	0.504	0.320
	Information sufficiency	0.053	0.821	0.157	0.759
	Desire for another car	0.334	0.208	0.405	0.522
	Prefer to use	-0.298	0.195	-0.185	0.730

Effects of Real-time Display on Travel Behavior

The commute mode shares for the control group and the treatment group, before and after the installation of the real-time display, are shown in Figure 10. Public transportation was the top choice for most commute trips, followed by driving alone. Very few respondents used TNC services, car-share services, bike-share services, taxicab, private shuttle or bus, or other modes for commuting. Between the pre-test survey and the post-test, the percentage of respondents who reported driving alone as their commute mode decreased on all days for the treatment group, and four out of five days for the control group.

FIGURE 10.

Commute mode shares in control and treatment groups, before and after installation of real-time display



A mixed-effect binary logistic regression model was estimated to test whether the installation of the screen had a significant effect on commute mode choices. In this model, the dependent variable was whether the traveler chose to drive alone or used some other mode for their commutes. A random intercept term was included to account for correlation in repeated choices made by the same individual, since each individual reported modes for 10 commute trips. The model produced an estimated regression coefficient of -0.096 for the treatment effect, but this effect was not statistically significant ($p=0.92$).

The analysis was repeated for only the 137 respondents who participated in both waves of the survey, and the results yielded an estimated regression coefficient of 1.88 ($p=0.0005$). This reflects the reported commute modes shown in Table 4—repeat respondents in the control group showed a 5.0 percentage point decrease in drive-alone commute trips, whereas those in the treatment group showed a 0.7 percentage point increase in drive-alone trips. This result suggests that the installation of the real-time display was associated with a significantly higher probability of driving alone. Considered in the context of the other results reported here, this may be a spurious correlation.

TABLE 4.
Commute Modes Reported
by 137 Respondents Who
Completed Both Survey
Waves

Group	Pre-Test / Post-Test	% Drive Alone
Control	Pre	17.4%
	Post	12.4%
Treatment	Pre	17.7%
	Post	18.4%

Comparing the pre-test and post-test surveys, average automobile miles traveled decreased slightly in the control group (from 11.6 miles to 10.8 miles) and more substantially in the treatment group (14.1 miles to 8.7 miles). However, upon analyzing these data using a gamma hurdle model, it was found that this difference was not significant at the 0.05 level. This was the case when both the full data set and the panel data including only the 137 respondents who responded to both waves of the survey were used.

Conclusions and Recommendations

A well-designed real-time multi-modal transportation information display can provide clear and reliable information and a satisfying experience for users. However, little evidence was found that the installation of a real-time multi-modal display screen in an office building lobby changed the building occupant travel choices, satisfaction, familiarity, or attitudes toward alternatives to private car travel over the course of a six-month study period. Based on the quantitative data collected in the survey as well as open-ended comments from respondents, the following recommendations for future installations of public real-time information displays are offered:

- **Target gaps in awareness and use.** Future investments in public information displays may be more effective if they target locations with lower usage, satisfaction, and/or awareness of alternative transportation modes. Even in the absence of the real-time information display, respondents in this study were very familiar with alternative modes, especially transit, and many reported using transit on a regular basis. A real-time information display might be more effective at shifting attitudes and behaviors if installed in a location with more “low-hanging fruit,” i.e., room to increase awareness and use of alternative modes.
- **Target gaps in information.** Many respondents in this study felt that adequate information about transit was already available from other sources. In particular, many mentioned their reliance on the OneBusAway smartphone app for obtaining real-time transit information. A real-time information display may have more to offer in locations in which real-time information is not available via smartphone apps or smartphone adoption is low or in areas with poor mobile data coverage.
- **Consider usability and location in installation.** Although most respondents were aware of the display screen, a majority never used it. Several respondents noted the physical location of the real-time display in this study was inconvenient, located out of the way and close to a security guard’s desk. Future installations should strive to locate the screen where it is easy and comfortable for travelers, including both building occupants and visitors, to view.
- **Consider marketing/public information at launch.** Some respondent comments revealed a lack of understanding of the screen’s purpose and the information it contained, indicating that they believed the screen contained schedule information, not real-time information. Although the display screen showed information on services other than transit, we did not detect changes in usage, satisfaction, or attitudes toward other services were not detected, and respondent open-ended comments suggested that they primarily viewed it as a source of transit information. Future installations might be more successful if the installation were accompanied by a marketing or public information campaign to ensure that potential users understand that the screen is displaying real-time information on multiple services.

In closing, some recommendations for future research in this area specifically and in the transportation field more broadly are presented. First, it may be worthwhile to evaluate the effects of real-time information displays that are responsive to the above site selection and installation recommendations. Second, future work may want to consider route choice as a behavioral outcome, since providing information in workplaces or other public locations may support choices between transit routes more effectively than providing the same information after someone has walked to a particular transit stop or station. Third, this work considered only building occupants whose regular workplace was in the study buildings, but visitors to the buildings may have different responses than occupants. Fourth, it may be worth evaluating impacts over a longer time horizons than the six months used in this study, especially since behaviors and

attitudes take time to evolve. Finally, other transportation researchers are urged to conduct more careful evaluations of interventions, using appropriate experimental or quasi-experimental research designs (Campbell and Stanley 1963). Sound evaluations should be planned in advance and should use control groups and, where possible, randomization. The use of control groups becomes particularly important in longer-term studies, in which time-varying confounders can undermine the validity of a simple before-and-after evaluation, with sometimes embarrassing results (e.g., Degraeuwe and Beusen 2013).

References

- Abdel-Aty, Mohamed A. 2001. "Using Ordered Probit Modeling to Study the Effect of ATIS on Transit Ridership." *Transportation Research Part C: Emerging Technologies* 9 (4): 265-77. doi:10.1016/S0968-090X(00)00037-1.
- Abdel-Aty, Mohamed, Ryuichi Kitamura, and Paul Jovanis. 1996. "Investigating Effect of Advanced Traveler Information on Commuter Tendency To Use Transit." *Transportation Research Record*, 1550 (January): 65-72. doi:10.3141/1550-09.
- Brakewood C., G. Macfarlane, and K. Watkins. 2015. "The Impact of Real-Time Information on Bus Ridership in New York City." *Transportation Research Part C: Emerging Technologies* 53:59-75.
- Brakewood, Candace, Sean Barbeau, and Kari Watkins. 2014. "An Experiment Evaluating the Impacts of Real-Time Transit Information on Bus Riders in Tampa, Florida." *Transportation Research Part A: Policy and Practice* 69 (November): 409-22. doi:10.1016/j.tra.2014.09.003.
- Campbell, D. T., and J. C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. New York: Houghton Mifflin.
- Cassidy, S., and P. R. White. 1995. "Use and Perceptions of a Real Time Passenger Information System." *Journal of Advanced Transportation* 29(1): 27-39. doi:10.1002/atr.5670290105.
- Cham, Laura, Georges Darido, David Jackson, Richard Laver, and Donald Schneck. 2006. "Real-Time Bus Arrival Information Systems Return-on-Investment Study." Federal Transit Administration. <http://trid.trb.org/view.aspx?id=793114>.
- Commute Seattle. 2015. "2014 Center City Commuter Mode Split Survey Results." <http://commuteseattle.com/2014-modesplit-survey/>.
- Degraeuwe, B., and B. Beusen. 2013. "Using On-Board Data Logging Devices to Study the Longer-Term Impact of an Eco-Driving Course." *Transportation Research Part D: Transport and Environment*, 19: 48-49.
- Dziekhan, K. 2004. "Customer Perceptions and Behavioural Responses to IT-Based Public Transport Information: Literature Review and What the Experts Say." *TRITA-INFRA*, No. 04-012. <http://trid.trb.org/view.aspx?id=840810>.

- Dziekanski, Katrin, and Arjan Vermeulen. 2006. "Psychological Effects of and Design Preferences for Real-Time Information Displays." *Journal of Public Transportation*, 9(1). <http://trid.trb.org/view.aspx?id=780950>.
- Forsyth, E., and D. T. Silcock. 1985. "Evaluation of Real-Time Information on the Arrival of Urban Public Transport Services." *Seminar on Public Transport Operations Research*. <http://trid.trb.org/view.aspx?id=238584>.
- Hickman, M., and N. H. M. Wilson. 1995. "Passenger Travel Time and Path Choice Implications of Real-Time Transit Information." *Transportation Research Part C: Emerging Technologies*, 3(4):211-226.
- Lappin, J., and J. Bottom. 2001. "Understanding and Predicting Traveler Response to Information: A Literature Review." Volpe National Transportation Systems Center. <http://trid.trb.org/view.aspx?id=697629>.
- Lyons, Glenn, and Reg Harman. 2002. "The UK Public Transport Industry and Provision of Multi-Modal Traveller Information." *International Journal of Transport Management*, 1(1): 1-13. doi:10.1016/S1471-4051(01)00002-7.
- Lyons, Glenn, and John Urry. 2005. "Travel Time Use in the Information Age." *Transportation Research Part A: Policy and Practice*, 39 (2-3): 257-76. doi:10.1016/j.tra.2004.09.004.
- MacKenzie, D., Y. Ge, P. Jabbari, and J. Tao. 2016. "Evaluation of Effects of Real Time-Multi-Modal Transportation Information Displays: Final Report." Prepared for Seattle Department of Transportation, February. <http://faculty.washington.edu/dwhm/publications/>.
- McCord, M., R. Mishalani, and M. Etefagh. 2015. "Effect of Real-Time Passenger Information Systems on Perceptions of Transit's Favorable Environmental and Traffic Reduction Roles." *Transportation Research Record*, 2538: 102-109.
- Poudenx, Pascal. 2008. "The Effect of Transportation Policies on Energy Consumption and Greenhouse Gas Emission from Urban Passenger Transportation." *Transportation Research Part A: Policy and Practice*, 42(6): 901-9. doi:10.1016/j.tra.2008.01.013.
- Reed, T. B. 1995. "Reduction in the Burden of Waiting for Public Transit due to Real-Time Schedule Information: A Conjoint Analysis Study." *Vehicle Navigation and Information Systems Conference, 1995*. Proceedings. In conjunction with the Pacific Rim TransTech Conference, 6th International VNIS, "A Ride into the Future": 83-89. doi:10.1109/VNIS.1995.518822.
- Reed, T. B., and J. C. Levine. 1997. "Changes in Traveler Stated Preference for Bus and Car Modes Due to Real-Time Schedule Information: A Conjoint Analysis." *Journal of Public Transportation*, 1(2). <http://trid.trb.org/view.aspx?id=481513>.
- Schweiger, C. L. 2013. "Use of Electronic Passenger Information Signage in Transit." TCRP Synthesis 104. Transportation Research Board. Washington, DC.

- Smith, R., S. Atkins, and R. Sheldon. 1994. "London Transport Buses: ATT in Action and the London Countdown Route 18 Project." *Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle-Highway Systems*. <http://trid.trb.org/view.aspx?id=424466>.
- Swanson, J., L. Ampt, and P. Jones. 1997. "Measuring Bus Passenger Preferences." *Traffic Engineering and Control*, 38(6): 330-36.
- Science Applications International Corporation. 2003. "Oregon Regional Intelligent Transportation Systems (ITS) Integration Program Final Phase III Report: Transit Tracker Information Displays." Available through National Technical Information Services.
- Tang, Lei, and Piyushimita (Vonu) Thakuriah. 2012. "Ridership Effects of Real-Time Bus Information System: A Case Study in the City of Chicago." *Transportation Research Part C: Emerging Technologies*, 22: 146-61. doi:10.1016/j.trc.2012.01.001.
- Tseng, Yin-Yen, Jasper Knockaert, and Erik T. Verhoef. 2013. "A Revealed-Preference Study of Behavioural Impacts of Real-Time Traffic Information." *Transportation Research Part C: Emerging Technologies*, 30: 196-209. doi:10.1016/j.trc.2011.11.006.
- Watkins, K., B. Ferris, A. Borning, G. S. Rutherford, and D. Layton. 2011. "Where Is My Bus? Impact of Mobile Real-Time Information on the Perceived and Actual Wait Time of Transit Riders." *Transportation Research Part A: Policy and Practice*, 45(8): 839-848.
- Zhang, Feng, Qing Shen, and Kelly Clifton. 2008. "Examination of Traveler Responses to Real-Time Information About Bus Arrivals Using Panel Data." *Transportation Research Record*, 2082: 107-15. doi:10.3141/2082-13.
- Zito, Pietro, Gianfranco Amato, Salvatore Amoroso, and Maria Berritella. 2011. "The Effect of Advanced Traveller Information Systems on Public Transport Demand and Its Uncertainty." *Transportmetrica*, 7(1): 31-43. doi:10.1080/18128600903244727.

About the Authors

YANBO GE (yanboge@uw.edu) is a Ph.D. student in Civil and Environmental Engineering at the University of Washington. She currently works as a research assistant in the Sustainable Transportation Lab and is interested in vehicle electrification, transit-oriented development, non-motorized transportation modes, and shared-use mobility services. She holds a B.S. in Transportation Engineering and an M.S. in Urban Planning, both from Tongji University, China

PARASTOO JABBARI (jabbari@uw.edu) is a master's student in Civil and Environmental Engineering at the University of Washington. She is a research assistant in the Sustainable Transportation Lab and is interested in energy and environmental aspects of transportation systems, particularly vehicle electrification, vehicle automation, and their relationships with public policy. She holds a bachelor's degree in Civil Engineering from Sharif University of Technology in Tehran, Iran.

DON MACKENZIE (*dwhm@uw.edu*) is an Assistant Professor of Civil and Environmental Engineering at the University of Washington. He leads the Sustainable Transportation Lab, which develops and evaluates technical and policy solutions for making our transportation system more economically-viable and environmentally-benign while providing access to opportunities for all. He holds a Ph.D. in Engineering Systems and an S.M. in Technology and Policy, both from the Massachusetts Institute of Technology, and a B.A.Sc. in Chemical and Biological Engineering from the University of British Columbia. He is a member of the Transportation Research Board's Standing Committee on Transportation Energy and chairs its Subcommittee on Energy and Demand Implications of Connected and Automated Vehicles.

JARUI TAO (*jiaruit@uci.edu*) is a Ph.D. student in Transportation Science at the University of California, Irvine. He was previously a master's student in Urban Design and Planning at the University of Washington and holds a bachelor's degree in City Management from Renmin University of China.

A Fuzzy AHP Approach to Compare Transit System Performance in US Urbanized Areas

Xin Li, Ph.D.

University of Wisconsin–Milwaukee

Yingling Fan, Ph.D.

University of Minnesota

John W. Shaw, P.E.

University of Wisconsin–Madison

Yunlei Qi

University of Minnesota

Abstract

Public transit systems in the United States often face multiple policy objectives. Typically, stakeholders desire frequent service on an extensive network, but funding and other resources are constrained, creating complicated relationships between service effectiveness goals and business efficiency goals. Using data from the National Transit Map (NTM), this study evaluated the general performance of transit systems across 294 Urbanized Areas (UZAs) in the US, which were stratified into six peer groups based on population. Transit efficiency and effectiveness were compared by developing a composite business efficiency index score and a composite service effectiveness index score for each urbanized area. The scores were generated using a fuzzy logic extension of the Analytical Hierarchy Process (AHP), which allows automated weighting of the measures. The NTM currently includes a limited set of performance measures, and each transit agency's data are associated with the largest urban area it serves; consequently, it is perhaps best-suited for identifying high-performing UZAs and less suitable for identifying the weakest performers. The analytical results suggest that a few UZAs (mainly densely-populated cities and university towns) are simultaneously able to achieve high scores on both business efficiency and service effectiveness. In most small- and medium-size conurbations, business efficiency appears to be a higher policy priority than service effectiveness.

Keywords: *Performance evaluation, US urban transit system, efficiency and effectiveness, fuzzy AHP*

Introduction

Public transit providers in the US are a unique category of organizations that blend some characteristics of private-sector businesses with some characteristics of public agencies. Managers and policymakers involved in transit typically need to strike a balance among at least three competing perspectives or sets of objectives:

- Perspective 1: Transit is a *business* that needs to meet customer expectations while using personnel, equipment, and financial resources efficiently.
- Perspective 2: Transit is a *social service* that provides essential mobility for non-drivers, people with disabilities, and other socially- and economically-disadvantaged people.
- Perspective 3: Transit is a *policy intervention* that might help relieve traffic congestion, reduce energy consumption, diminish pollutant emissions, promote compact urban form, and/or stimulate revitalization of distressed areas.

These objectives often conflict, so it is necessary to strike a balance between business efficiency and service effectiveness, which were defined by Fielding et al. (1985) and Chu and Fielding (1990):

Business efficiency is a statement about the achievements of an agency in transforming a set of inputs into a set of outputs. For example, can existing levels of transit ridership be obtained with less equipment or fewer operational resources? Can increased services generate more fare box revenue? Service effectiveness reflects the ability of transit operations to meet certain goals. Do the services attract a significant number of users or transit-dependent populations from a given level of input?

Often, there are perceived trade-offs between efficiency and effectiveness. For example, assuring that there is basic mobility for non-drivers throughout an urban area implies that a transit system will need to provide a considerable amount of service in areas with relatively low demand, which could drive down measures of operational and financial efficiency.

Performance measurement and benchmarking tools provide opportunities for transit system managers to clarify the links between policy decisions and system performance outcomes. For example, a carefully-chosen set of performance measures could assist in balancing (or re-balancing) conflicting objectives that affect both strategic long-range planning and tactical day-to-day decisionmaking. Moreover, performance evaluation can contribute to better understanding of the relationships between efficiency and effectiveness.

Transit Cooperative Research Program (TCRP) Report 141 (Ryus et al. 2010) suggests a performance evaluation process that begins by hand-selecting a peer group with similar characteristics and then computing cross-comparison metrics using publicly-available data sets or information obtained directly from the peer agencies. In this study, we compared transit performance at the urbanized area (UZA) level. As defined by the US Census, each UZA is a contiguous urban region with a population greater than 50,000.

For this analysis, UZAs from the 2010 Census were first stratified into six groups based on population, as shown in Table 1. The implicit assumption is that transit resources and objectives are somewhat comparable within these groups. Next, a Fuzzy Analytical Hierarchy Process (Fuzzy AHP) model was used to determine how similar certain transit performance measures are within each group and compare the efficiency performance and effectiveness performance by generating a business efficiency index score and a service effectiveness index score for each urbanized area, along with composite scores that include both efficiency and effectiveness measures for all UZAs within each of the six peer groups. The fuzzy-AHP approach includes an automated process to weigh individual performance criteria, which avoids subjectivity in the weighting and scoring process.

TABLE 1.
Peer Groups Used for This Analysis

Peer Group	Urbanized Area (UZA) Population	Number of Census UZAs	Total Population of Census UZAs	Number of UZAs in This Study	Total Population of NTM UZAs
A	> 2.5 million	16	93 million	16	93 million
B	1–2.5 million	26	43 million	26	43 million
C	500,000–1 million	38	27 million	36	26 million
D	250,000–500,000	69	23 million	62	21 million
E	100,000–250,000	149	23 million	94	15 million
F	50,000–100,000	199	14 million	60	5 million

Transit performance characteristics for each UZA were obtained from the National Transit Map (NTM), a data set recently developed by other researchers on behalf of the Federal Transit Administration (FTA). It is important to acknowledge an important characteristic of the NTM at the outset: transit operators that serve more than one UZA (or serve an UZA and adjoining rural areas) are mapped exclusively to the largest UZA they serve. This characteristic has the potential to overstate (to a degree) the amount of service identified as being provided in large UZAs and understate the amount of service provided in small UZAs. It is also important to note that many Census 2010 UZAs have very irregular boundaries, which do not necessarily coincide with municipal boundaries or transit service area boundaries.

Literature Review

Transit system performance evaluation is an important task for transit service providers seeking to respond to passenger demand trends, operational constraints, stakeholder concerns, and changing service needs. It allows system managers to achieve better economic performance assessment, organizational administration, transit planning, and financial management. Its importance has been increasingly realized by oversight authorities, transit agencies, and groups representing passengers, major employers, and other stakeholders. Thus, assessing transit system performance has been one of the most widely investigated areas of research within the transit industry.

Fielding et al. (1978) carefully described nine performance indicators for transit management, among which Revenue Vehicle Hours per Employee, Revenue Vehicle Hours per Vehicle and Operating Expense per Revenue Vehicle Hour were regarded as indicators of efficiency; Revenue Passengers per Service Area Population, Percent of Population Served, Total Passengers per Vehicle, and Revenue Passengers per Revenue Vehicle Hour were used as effectiveness indicators; and Operating Expenses per Total Passenger and Operating Expense per Revenue Passenger were chosen for overall performance evaluation. Several years later, Fielding et al. (1985) used FY 1980 Section 15 data to identify a set of performance indicators using factor analysis. The selected indicators were also used to evaluate the performance of fixed routes.

When the application of Data Envelopment Analysis (DEA) began to become more widely used for system efficiency assessment, Chu et al. (1990) and Karlaftis (2004) applied the approach to evaluating efficiency and effectiveness of transit systems. To document the state of transit performance evaluation in the US, Fielding (1992) examined three programs (Federal triennial review, California performance audits, and Los Angeles program) in terms of four components: dimensions for policy objectives, indicators, information systems, and incentives.

Some of the previously-published research on transit performance evaluation focuses on the service level. These studies include user perception- (customer satisfaction-) based approaches, efficiency indicator-based approaches, and approaches that combine both user opinions and efficiency indicators (Abreha 2007; Badami and Haider 2007; Sheth et al. 2007; Nathanail 2008; Tyrinopoulos and Antoniou 2008; Lao and Liu 2009; Eboli and Mazzulla 2011).

To monitor transit system performance and support public transport development, some transit associations and organizations have developed efficiency assessment indicators and related agency guidance. For example, the International Association of Public Transport (known by its French-language initials, UITP) has suggested a group of indicators to compare public transport system performance across cities and regions. The indicators suggested by UITP include the population of transit users; the service coverage; and the number of bus routes, stations, etc. (UITP 2011). The Transportation Research Board's *Transit Capacity and Quality of Service Manual* (TCQSM) (TRB 2003) includes guidelines for evaluating the performance of public transport systems using a three-level evaluation index system: station, route, and system. Other resources, such as the annual American Public Transportation Association (APTA) *Public Transportation Fact Book* (Neff and Dickens 2013) provide extensive statistical data about resource inputs and production outputs for each agency, but offer only limited interpretive advice.

With support from FTA, TCRP Report 88 was published, a guidebook for developing a transit performance-measurement system (Kittelsohn & Associates, Inc. et al. 2003). The guidebook offers technical assistance and suggestions on how to implement and use performance measurement on an ongoing basis at a transit agency. It suggests a wide range of performance measure categories, including availability, service delivery, community, travel time, safety and security, maintenance and construction, economic,

capacity, paratransit, and comfort. Measures based on these categories can be grouped into four general types: individual measures, ratios, and indices and level-of-service indicators. The report also includes 12 domestic and international case studies such as Livermore, CA; Denver, CO; and Sydney, Australia. TCRP Report 141 (Ryus et al. 2010) also focuses on transit performance evaluation and highlights the importance of the peer comparison approach. The report suggests that peer-grouping and performance measurement should serve as a starting point for a transit agency to ask questions about performance and identify improvement opportunities.

In addition to the abundant studies evaluating the general performance of transit systems, some scholars have concentrated specifically on investigating the relationship between efficiency performance and effectiveness performance, sometimes claiming that these two objectives are mutually exclusive (Griffis et al. 2004). Other studies suggest that treating efficiency and effectiveness as a dichotomy is unwarranted. For example, Talley and Anderson (1981) explored the relationship between transit efficiency and effectiveness from the perspective of the bus company itself and in terms of government objectives for transit service. They argued that from the point of view of the transit firm, the aim is to maximize ridership within allowable deficit limits as well as minimize the operating costs to maintain a certain service level. They also argue that the government or public policy objectives for transit can be defined as maximizing social well-being, economic development, and environmental quality. Therefore, attainment of these objectives is predicated on the transit agency being both effective *and* efficient, and they should not be treated as mutually exclusive. Similar findings were presented by Karlaftis in 2004 using Data Envelopment Analysis (DEA) to analyze the relationship between transit system efficiency and effectiveness. Using data from 256 US transit systems over a five-year period, he found that efficiency and effectiveness were positively related—that is, the most efficient systems were also the most effective.

Together, the two TCRP reports and various journal articles on transit performance evaluation make many valuable contributions toward developing transit performance-measurement systems, but there are major limitations:

- Both TCRP reports suggest a wide range of measures that impact transit system performance but leave many details unresolved, such as how to weigh dissimilar criteria.
- Most studies have focused on measuring performance at the operator or agency-level, usually concentrating on service level. There is lack of a comprehensive performance evaluation and comparison at other spatial scales, such as cities and urbanized regions. Importantly, many US urbanized areas have multiple transit service providers. For example, buses and commuter rail often are operated by different entities. Although numerous UZAs span more than one county or state, transit agencies in some of these UZAs are not authorized to cross jurisdictional boundaries. From the perspective of transit riders, the most important issue is the overall quality of transit service available in a city or urban region. Comparisons at the transit agency/provider level are less relevant to transit users, particularly

since many operators have made efforts to integrate their routes, schedules, and fare payment systems.

- Although Karlaftis (2004) concluded that efficiency and effectiveness are positively related using NTD data (1990–1994), the theoretical limitations of DEA models are a critical issue, and the analysis is based on a data set that is now more than 20 years old. Many empirical studies have clearly identified major limitations of applying DEA models in efficiency assessment. For example, DEA evaluates the relative efficiency of decisionmaking units but does not allow for ranking of the efficient units themselves (Charnes et al. 1978; Andersen and Christian 1993; Cook and Seiford 2009). Further, some statisticians and economists have stated that DEA will lead to the deviation of efficiency evaluation when the number of samples is small (Korostelev et al. 1995; Simar and Wilson 1998; Song et al. 2013).

In addition to the TCRP reports and scholarly works, the US mass media has shown interest in transit performance evaluation. For example, in 2011, *US News & World Report* ranked the “10 Best Cities for Public Transportation” as follows: Denver-Aurora, CO.; New York-Newark, NY-NJ-CT.; Los Angeles-Long Beach-Santa Ana, CA.; Boston, MA-NH-RI; Portland, OR; San Jose, CA; Salt Lake City, UT; San Diego, CA; Seattle, WA; and Honolulu, HI. Unfortunately, it provided relatively few details about the ranking methodology. In addition, there was a discrepancy between the headline (“Best Cities”) and the ranking method (metropolitan areas). In some cases, the rankings seem to violate the principle of peer-group comparison suggested by TCRP Report 141—some of the cities differ considerably in terms of population, transit system size, and other basic characteristics.

Scope of This Analysis

The main objective of this research was to assess the relative importance of various performance measures in distinguishing transit performance within the six city peer groups identified in Table 1 and how well individual agencies perform relative to peer-defined norms. The study focused on the following research tasks:

1. Implement an extended AHP with the integration of fuzzy logic to evaluate overall performance for each peer group at UZA level. In essence, this is a software-driven process that mathematically establishes weighting criteria for each performance measure (or objective), ranks the peer group members on each individual performance measure, and produces an overall weighted ranking. In contrast to most other multi-criteria ranking systems, Fuzzy AHP makes no *a priori* judgments about the importance of the individual performance measures. This removes some of the subjectivity associated with traditional methods for establishing weighting criteria, such as analyst judgment or small-group consensus. The proposed approach offers the advantage of avoiding the uncertainty and potential lack of repeatability that can affect the assignment of weighting criteria by individual people or expert panels.

2. Develop a specific comparison of efficiency and effectiveness by generating both a composite business efficiency index score and a composite service effectiveness index score for each urbanized area to explore the relationship between efficiency and effectiveness in all peer population groups.

Note that the general purpose of this study was simply to demonstrate how the fuzzy AHP methodology can be applied to transit analysis. Given the limitations of the input data set, it is important to state clearly that the authors did not intend this demonstration to be interpreted as a definitive statement that “transit system X is better than transit systems Y and Z.” Although a multi-city analysis was used to demonstrate the methodology, if appropriate data was available, a similar methodology could potentially be applied to an internal analysis of operational entities within a transit system—for example, treating collections of bus routes or commuter rail lines as the peer groups.

Data and Data Limitations

Data for this study came from the National Transit Map (NTM) developed at the University of Minnesota, which aggregates agency-level performance data from the 2013 National Transit Database (NTD) to the UZA level. In contrast to the agency-level information provided in some other US transit data sources, the NTM combines the data for all operators serving an UZA. For example, the Chicago UZA has four main transit operators (Chicago Transit Authority, Metra, Pace, and Northern Indiana Commuter Transportation District), and the NTM combines their data to facilitate comparison of the Chicago UZA’s performance with other American UZAs. A very important characteristic of the NTM data set is that to avoid double-counting in summary statistics, the NTM associates each transit operator with only one UZA, specifically the most-populous UZA served by that operator. Thus, services that link two or more UZAs (or an UZA and adjoining rural areas) appear in the NTM only as part of the more populous UZA (Fan 2015).

A total of 497 UZAs were defined by the 2010 US Census (486 in the US proper and 11 in Puerto Rico) (US Census Bureau 2013). Of these, 294 are free of missing values for the performance measures used in this study, as shown in Table 1. Many of UZAs that have missing data did not report statistical data to FTA (Fan 2015). A few urbanized areas were removed from the evaluation list due to incomplete data; for example, Atlantic City, NJ, is missing data for two of the performance measures. The locations of the cities that were analyzed are shown in Figure 1. The majority of UZAs that were not included in this study are in the smaller UZA groups (especially Peer Groups E & F).

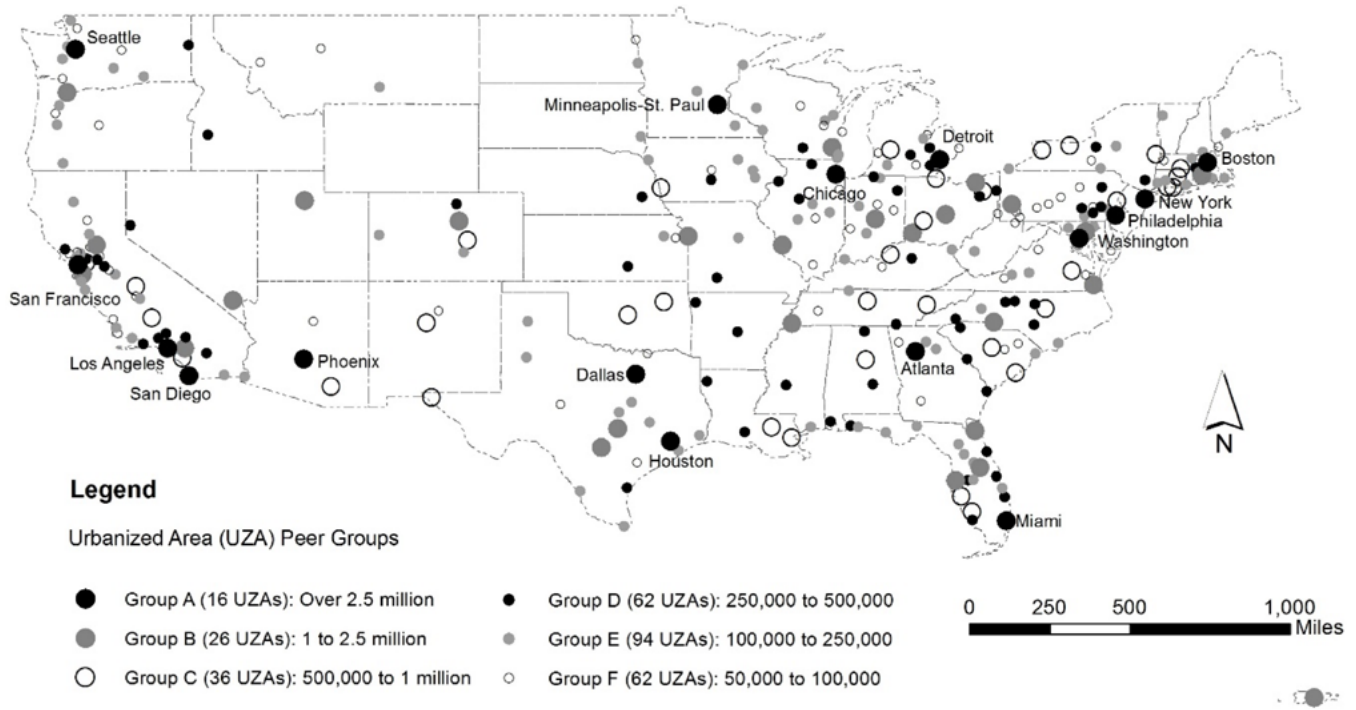


FIGURE 1. Geographic distribution of cities included in analysis

Since each transit operator’s services are attributed to the largest UZA it serves, some UZAs that are served by transit do not appear in the NTM listings. For example, Round Lake Beach, IL (a Peer Group D UZA) is served by Metra trains and Pace buses that would be reported as part of the Chicago data. Continuing this Chicago-area example, the commuter rail service provided by the Northern Indiana Commuter Transportation District (operator of the 90-mile South Shore Line from Chicago to South Bend, IN) is attributed to the Chicago UZA and not to South Bend (another Group D UZA) nor to any of the rural communities along the route. As a result, the South Bend area’s scores are based on its local bus systems and exclude the commuter rail service.

Since the NTM attributes cross-UZA services only to the largest UZA, the amount of transit service available to smaller communities is understated in some cases. This issue could potentially be addressed by pro-rating the NTM data to adjust for cross-UZA services. Due to the highly irregular boundaries of most Census-defined UZAs, this would require an in-depth analysis based on route-level (or stop-level) data, perhaps also factoring in service frequency. Since the primary purpose of this study was to probe the feasibility of using a fuzzy AHP analytical approach with a national transit data set, undertaking such an analysis was beyond the scope of the work described here, and the results should be interpreted accordingly.

Performance Measure Criteria

As shown in Table 2, five transit performance measures are currently included in the NTM database. Therefore, our demonstration analysis was based only on these metrics.

TABLE 2.
Service Characteristics Used
for This Analysis

Code	Characteristic	Normalization Divisor
VOMS	Vehicles operated in maximum service	Unlinked passenger trips
FRE	Fare revenues earned	Vehicle revenue miles
VRM	Vehicle revenue miles	UZA land area
UPT	Unlinked passenger trips	UZA population
TWM	Length of transit way system	UZA land area

All urbanized areas were scored based on the combined totals for the eight NTM transport modes (bus, bus rapid transit [BRT], commuter rail, heavy rail, light rail, street car, other rail, and other non-rail). The detailed definitions of the criteria are further defined by NTM project (Fan 2015), as follows:

- **Vehicles Operated in Maximum Service (VOMS)** – the revenue vehicle count taken during a transit agency’s maximum season of the year, on the day of the week that this maximum occurs; not taken on a day when a special event or other extreme set of circumstances would cause the resulting tally to represent a one-time event rather than a recurring maximum service requirement.
- **Fare Revenue Earned (FRE)** – all income received directly from passengers, paid either in cash or through pre-paid tickets, passes, etc.; includes donations from those passengers who donate money on the vehicle and reduced fares paid by passengers in a user-side subsidy arrangement.
- **Vehicle Revenue Miles (VRM)** – miles that vehicles travel while in revenue service; excludes miles that vehicles travel for deadhead services (leaving or returning to the garage or yard facility, changing routes, or when there is no expectation of carrying revenue passengers), operator training, maintenance testing, and school bus and charter services.
- **Unlinked Passenger Trips (UPT)** – number of passengers who board public transportation vehicles; passengers are counted each time they board vehicles no matter how many vehicles they use to travel from their origin to their destination.
- **Transit Way Mileage (TWM)** – length of transitway system; NTD’s transitway mileages include all fixed-route and fixed-schedule modes (i.e., track mileage for rail modes and lane/route miles for fixed-route and fixed-schedule non-rail modes).

To reduce scaling bias, all five criteria were normalized for each of the 294 UZAs. The normalization method was as follows:

- Fare Revenue Earned (FRE) was divided by Vehicle Revenue Miles (VRM) to obtain the fare income per revenue-mile operated.

- Vehicle Revenue Miles (VRM) was divided by the UZA's land area to provide an indicator of the operational intensity of the transit service.
- Unlinked Passenger Trips (UPT) was divided by Vehicles Operated in Maximum Service (VOMS) to provide a rough indicator of fleet utilization efficiency.
- Unlinked Passenger Trips (UPT) was divided by the UZA's population to serve as a rough indicator of the extent to which residents use transit.
- Transitway Mileage (TWM) was divided by the UZA's land area to provide an indicator of the area's transit network density.

Fuzzy AHP Evaluation Method

AHP is a biased multi-criteria decisionmaking process introduced by Saaty (1980) and subsequently used in several transportation system evaluation studies. For example, Yeh et al. (2000) employed a fuzzy multi-criteria analysis approach to evaluate the performance of urban public transport systems, and Hanaoka and Kunadhamraks (2009) used a fuzzy logic AHP to evaluate the logistics performance of intermodal freight transportation. Yu et al. (2011) developed a comprehensive AHP-based framework for ranking candidate location plans of multiple urban transit hubs. Li et al. (2015) proposed an enhanced fuzzy AHP approach to evaluate, monitor, and compare the development of public transportation systems towards transit metropolis status in different cities in China, in which two levels—policy and technical—were integrated into one framework.

AHP allows decisionmakers to decompose a complex problem into three hierarchical levels—goal, criteria, and alternatives. Three critical issues can arise when applying conventional AHP:

- How to handle criteria weighting when the judgment scale is very unbalanced.
- How to construct the pair-wise comparison matrix properly, given the potential for variability in the selection, judgment, and preferences of human decisionmakers.
- How to obtain a completely consistent pair-wise comparison matrix that satisfies the consistency-check rules.

To remedy these deficiencies in the structure of conventional AHP and limit the risk that judgment variations could result an unreliable analysis, an enhanced fuzzy AHP model was developed by Li et al. (2015). The advantages of the proposed fuzzy AHP structure lie in its ability to:

- Normalize the scales of different technical indicators.
- Construct a matrix of pair-wise comparisons using a fuzzy set.
- Optimize the weight of each criterion (using a non-linear programming model to maximize the judgment consistency).

The use of a fuzzy analytical model for weighting the technical criteria moves toward an evaluation method that is less dependent on human judgment and allows identification of transit system strengths and weaknesses with respect to each specific performance measure. To facilitate the model presentation, all definitions and notation used hereafter are summarized in Table 3.

TABLE 3.
Notation of Key Parameters
Used in Proposed Model

i	Index corresponding to criteria ($i = 1 \dots n$)
k	Index corresponding to cities to be evaluated ($k = 1 \dots m$)
x_{ik}	Indicator representing the selected city k being evaluated by criterion i
μ_{ik}	Fuzzy membership value corresponding to indicator x_{ik}
μ_i	Average fuzzy membership value for criterion i
$x_{i(min)}$	Minimal crisp value for criterion i
$x_{i(max)}$	Maximal crisp value for criterion i
s_i	Standard deviation of indicator values corresponding to criterion i
s_{min}	$\min\{s_i \mid i = 1, \dots, n\}$
s_{max}	$\max\{s_i \mid i = 1, \dots, n\}$
$A = (a_{ij})_{n \times n}$	Pair-wise comparison matrix
a_m	Comparison scale for the pair-wise comparison matrix
w_i	Weight for the policy criterion i
$Y = (y_{ij})_{n \times n}$	Consistency judgment matrix
$CIC(n)$	Consistency index coefficient
s_k	Synthesized evaluation score of city k

Details about the fuzzy AHP approach can be found in Li et al. (2015). The procedure can be summarized as follows.

Step 1: Fuzzy Scaling

Two types of indicators, i.e., “the-lower-the-better” and “the-higher-the-better,” are identified to normalize x_{ik} with their fuzzy sets, given by:

- For the lower-the-better indicators:

$$\mu_{ik} = \frac{[x_{i(max)} + x_{i(min)} - x_{ik}]}{[x_{i(max)} + x_{i(min)}]} \tag{1}$$

- For the higher-the-better indicators:

$$\mu_{ik} = \frac{x_{ik}}{[x_{i(max)} + x_{i(min)}]} \tag{2}$$

Step 2: Pair-Wise Comparisons

Using the standard deviation of indicators determines which criterion is more important and to what extent.

$$s_i = \sqrt{\sum_{k=1}^m (\mu_{ik} - \bar{\mu}_i)^2 / (m - 1)} \quad i = 1 \dots n \tag{3}$$

Then, a pair-wise comparison matrix $A = (a_{ij})_{n \times n}$ is created to measure the relative importance of criterion i over criterion j , given by:

$$a_{ij} = \frac{s_i - s_j}{s_{max} - s_{min}} \times (a_m - 1) + 1 \quad s_i \geq s_j \tag{4}$$

$$a_{ij} = \frac{1}{\left[\frac{s_j - s_i}{s_{max} - s_{min}} \times (a_m - 1) + 1 \right]} \quad s_i \leq s_j \tag{5}$$

where $a_m = \min\{9, \text{int}\left(\frac{s_{max}}{s_{min}} + 0.5\right)\}$ is a comparison scale (range from 1 to 9) for all criteria as recommended by Jin et al. (2004).

Step 3: Weight Determination

According to AHP analysis theory, a consistency-check of a pair-wise matrix is required to correctly reflect the importance of criterion j over criterion i . However, as mentioned in many previous studies (Bryson 1995; Jin et al. 2004; Saaty 1980; Sudhakar and Shrestha 2003; Yu 2002), it is usually difficult in practice to obtain a completely consistent pair-wise comparison matrix. Thus, this study proposed the following non-linear optimization model to estimate the weights $\{w_i | i = 1 \dots n\}$ from the inconsistent a_{ij} :

$$\min CIC(n) = \sum_{i=1}^n \sum_{j=1}^n \frac{|y_{ij} - a_{ij}|}{n^2} + \sum_{k=1}^m \sum_{l=1}^m \frac{|y_{ij} w_j - w_i|}{n^2} \tag{6}$$

$$y_{ii} = 1, i = 1 \dots n \tag{7}$$

$$\frac{1}{y_{ji}} = y_{ij} \in [a_{ij} - da_{ij}, a_{ij} + da_{ij}] (i = 1, \dots, n, j = i + 1, \dots, n) \tag{8}$$

$$w_i > 0, i = 1 \dots n \tag{9}$$

$$\sum_{i=1}^n w_i = 1 \tag{10}$$

In the above equations, $Y = (y_{ij})_{n \times n}$ is defined as the consistency judgment matrix, which is adjusted based on $A = (a_{ij})_{n \times n}$ during the minimizing process of the consistency index coefficient, denoted by $CIC(n)$. Based on extensive numerical

experiments this study has employed the convergence criterion $CIC(n) \leq 0.1$ to ensure that the resulting judgment matrix $Y = (y_{ij})_{n \times n}$ is consistent.

Step 4: Synthesis

After obtaining the weights for all criteria from the optimization model, the final ranking score of each city k will be synthesized by equation (11), which can be stated as:

$$S_k = \sum_{i=1}^n \mu_{ik} * w_i \quad (11)$$

The synthesis results will reflect the overall performance for all the areas with respect to the selected criteria.

Evaluation Process

Two types of evaluation were completed for this study. The first is an overall composite ranking for each of the six peer groups. The second compares efficiency and effectiveness metrics for each group. In the first objective, five normalized indicators—FRE/VRM, VRM/LandArea, UPT/VOMS, UPT/Population, and TWM/LandArea—were taken into consideration in the evaluation process. For the second analysis, two indices were generated: one represents *business efficiency*, including fare revenue per revenue mile (FRE/VRM) and operating vehicle utilization (UPT/VOMS); the second represents *service effectiveness*, including unlinked trips per capita (UPT/Population) and service coverage (VRM/LandArea).

When computing overall composite ratings, the proposed model estimates the weights for each of the five criteria independently for each peer group. For the purpose of comparing efficiency and effectiveness, the model is used to measure the relative importance between the two indicators for the “efficiency group” and the “effectiveness group.” After the weights were computed, the ranking scores were synthesized using equation (14) for both objectives.

Analysis Results

Overall Composite Ranking

Composite scores for each of the six peer groups (combining both business efficiency and operational effectiveness measures) are presented in Figure 2, Table 4, and Table 5. Figure 2 shows a complete ranking for Group A (although this group includes only 16 UZAs, it represents close to a third of the total US population). Table 4 provides complete results for Group B, including detailed results for all five metrics as well as the final synthesis and rankings. Due to space constraints, Table 5 for Groups C–F is abbreviated to include only the UZAs rankings in the top 10 and bottom 10 of each peer group.



FIGURE 2. Group A composite ranking

TABLE 4. Analytical Results for Group B

UZA	VOMS/ UPT	FRE/ VRM	VRM/ LandArea	UPT/POP	TWM/ LandArea	Synthesis	Ranking
Salt Lake City-West Valley City, UT	0.0683	0.0654	0.2091	0.1716	0.1795	0.6940	1
Portland, OR-WA	0.1048	0.1406	0.1215	0.2387	0.0454	0.6509	2
Baltimore, MD	0.0944	0.1151	0.1227	0.1941	0.0465	0.5727	3
Denver-Aurora, CO	0.0944	0.0956	0.1340	0.1692	0.0740	0.5673	4
Las Vegas-Henderson, NV	0.1033	0.1457	0.1045	0.1382	0.0434	0.5351	5
Pittsburgh, PA	0.0868	0.1414	0.0592	0.1494	0.0436	0.4803	6
San Jose, CA	0.0921	0.0798	0.1305	0.1051	0.0629	0.4704	7
Milwaukee, WI	0.1003	0.0971	0.0625	0.1272	0.0415	0.4285	8
Cleveland, OH	0.0974	0.0974	0.0505	0.1112	0.0348	0.3913	9
Sacramento, CA	0.0858	0.0898	0.0674	0.0742	0.0549	0.3720	10
San Antonio, TX	0.0914	0.0387	0.0848	0.1057	0.0476	0.3682	11
St. Louis, MO-IL	0.0971	0.0725	0.0598	0.0918	0.0348	0.3560	12
Austin, TX	0.0845	0.0518	0.0583	0.1058	0.0247	0.3249	13
Orlando, FL	0.0826	0.0545	0.0663	0.0782	0.0327	0.3142	14
Riverside-San Bernardino, CA	0.0804	0.0524	0.0649	0.0528	0.0637	0.3142	15
Providence, RI-MA	0.0816	0.0686	0.0465	0.0720	0.0412	0.3100	16
Charlotte, NC-SC	0.0816	0.0651	0.0416	0.0925	0.0220	0.3028	17

TABLE 4 (CONT'D). Analytical Results for Group B

UZA	VOMS/ UPT	FRE/ VRM	VRM/ LandArea	UPT/POP	TWM/ LandArea	Synthesis	Ranking
Columbus, OH	0.0818	0.0664	0.0447	0.0547	0.0375	0.2851	18
Cincinnati, OH-KY-IN	0.0666	0.1009	0.0331	0.0514	0.0279	0.2798	19
San Juan, PR	0.0230	0.0781	0.0574	0.1025	0.0090	0.2700	20
Tampa-St. Petersburg, FL	0.0784	0.0615	0.0396	0.0501	0.0309	0.2607	21
Virginia Beach, VA	0.0800	0.0521	0.0432	0.0517	0.0304	0.2574	22
Jacksonville, FL	0.0858	0.0432	0.0384	0.0471	0.0263	0.2408	23
Memphis, TN-MS-AR	0.0889	0.0636	0.0239	0.0390	0.0226	0.2381	24
Kansas City, MO-KS	0.0696	0.0449	0.0341	0.0449	0.0251	0.2185	25
Indianapolis, IN	0.0739	0.0533	0.0230	0.0285	0.0131	0.1918	26

TABLE 5.
Groups C-F Composite
Ranking

Top 10			Bottom 10	
Group	UZA	Rank	UZA	Rank
C	Urban Honolulu, HI	1	Baton Rouge, LA	27
	Buffalo, NY	2	Allentown, PA-NJ	28
	Albany-Schenectady, NY	3	Richmond, VA	29
	Rochester, NY	4	Columbia, SC	30
	Albuquerque, NM	5	Raleigh, NC	31
	New Orleans, LA	6	Birmingham, AL	32
	Fresno, CA	7	Toledo, OH-MI	33
	Tucson, AZ	8	Colorado Springs, CO	34
	Bakersfield, CA	9	Knoxville, TN	35
	Louisville/Jefferson County, KY-IN	10	Mission Viejo-San Clemente, CA	36
D	Durham, NC	1	Greenville, SC	53
	Anchorage, AK	2	Montgomery, AL	54
	Stockton, CA	3	Asheville, NC	55
	Madison, WI	4	Mobile, AL	56
	Ann Arbor, MI	5	Youngstown, OH-PA	57
	Spokane, WA	6	Wichita, KS	58
	Lansing, MI	7	Augusta-Richmond County, GA-SC	59
	Syracuse, NY	8	Jackson, MS	60
	Oxnard, CA	9	Huntsville, AL	61
	Flint, MI	10	Port St. Lucie, FL	62

TABLE 5. (CONT'D.)
Groups C–F Composite Ranking

		Top 10		Bottom 10	
Group	UZA	Rank	UZA	Rank	
E	Athens-Clarke County, GA	1	Elkhart, IN-MI	53	
	Seaside-Monterey, CA	2	Killeen, TX	54	
	Champaign, IL	3	Amarillo, TX	55	
	Gainesville, FL	4	Aberdeen-Bel Air North, MD	56	
	Santa Barbara, CA	5	Winter Haven, FL	57	
	Iowa City, IA	6	Leesburg-Eustis-Tavares, FL	58	
	Santa Cruz, CA	7	Texas City, TX	59	
	Eugene, OR	8	Gainesville, GA	60	
	Bellingham, WA	9	Fort Walton Beach-Navarre-Wright, FL	61	
	Portland, ME	10	Hickory, NC	62	
F	Ames, IA	1	Uniontown-Connellsville, PA	51	
	State College, PA	2	Salisbury, MD-DE	52	
	Ithaca, NY	3	Longview, WA-OR	53	
	Kahului, HI	4	Lodi, CA	54	
	Morgantown, WV	5	Kokomo, IN	55	
	Blacksburg, VA	6	Carbondale, IL	56	
	Davis, CA	7	Victoria, TX	57	
	Portsmouth, NH-ME	8	Sherman, TX	58	
	Harrisonburg, VA	9	San Angelo, TX	59	
	Lawrence, KS	10	Elizabethtown-Radcliff, KY	60	

In this methodology, when all members of a peer group all have similar outcomes on a performance measure, it will receive low weight in the ranking methodology. As the performance varies more widely, the weight assigned to that performance measure will increase. Thus, the algorithm puts the highest weight on the criteria that tend to more sharply distinguish higher-performing and lower-performing members of the peer group.

Table 5 identifies the weighting criteria that the model assigned to each of the five performance measures for each of the six peer groups in the composite rankings. The table also highlights the factor that was most influential in driving the rankings of each peer group. For example, among the Peer Group C cities (500k–1M population), Route Coverage (transitway miles per square mile of land area) had the strongest effect on the rankings, whereas among the Group F cities (50k–100k population), the model assigned the greatest weight to Revenue per Vehicle Mile. This does not mean that transit managers in Group F urbanized areas are unconcerned about route coverage. Instead, it suggests that Group F communities might have relatively similar land use patterns or perhaps that decisionmakers in Group F communities have fairly similar viewpoints about how intensive the transit route coverage should be.

Although the NTM data show that the amount of revenue earned per vehicle-mile differs greatly in Group F cities, the reasons for these differences are not obvious.

Nevertheless, the rankings provide a starting point for probing possible explanations. For example, upon inspection, it is clear that the cities ranked near the top of the Group F list are mainly university towns such as Ames, IA (Iowa State University), State College, PA (Penn State University), and Ithaca, NY (Cornell University). Older industrial cities and cities with low/lenient fare policies tend to be toward the lower end of the list.

TABLE 6.
Weights Assigned by Model to Each Performance Criterion

Peer Group	Fleet Utilization	Revenue per Vehicle-Mile	Operational Intensity	Ridership Intensity	Route Coverage
	UPT/VOMS	FRE/VRM	VRM/Area	UPT/POP	TWM/Area
A	0.14	0.17	0.30	0.20	0.19
B	0.13	0.18	0.23	0.27	0.19
C	0.21	0.17	0.15	0.17	0.31
D	0.14	0.14	0.28	0.27	0.16
E	0.16	0.23	0.22	0.26	0.12
F	0.17	0.29	0.14	0.22	0.18

Although the methodology requires discretion regarding the selection of performance measures and the definition of peer groups, it does not depend on any *a priori* human judgement to set weighting criteria. As a result, it can serve as a starting point for developing correlations between transit operational characteristics and the resulting performance outputs. For example, there is a longstanding debate about the extent to which rail and bus rapid transit (BRT) serve as flagship services that help attract discretionary riders. Figure 3 explores this notion quantitatively by comparing the composite rankings of the Group A cities with the proportion of their transitway mileage that is comprised of BRT + rail. This graph suggests that there is some relationship between BRT + rail intensiveness and the overall performance of the system, but it is not the sole determinant of performance outcomes. More specifically, a linear relationship between BRT + rail intensiveness and composite score appears to exist for about a dozen of the 16 urbanized areas in Group A, but the three most rail-intensive conurbations (Boston, Chicago, Philadelphia) are clustered on the right side of the chart, with composite scores only near the middle of the pack. Dallas is also an outlier.

Proportion of Transit-Way Miles that are BRT or Rail vs Composite Fuzzy AHP Score

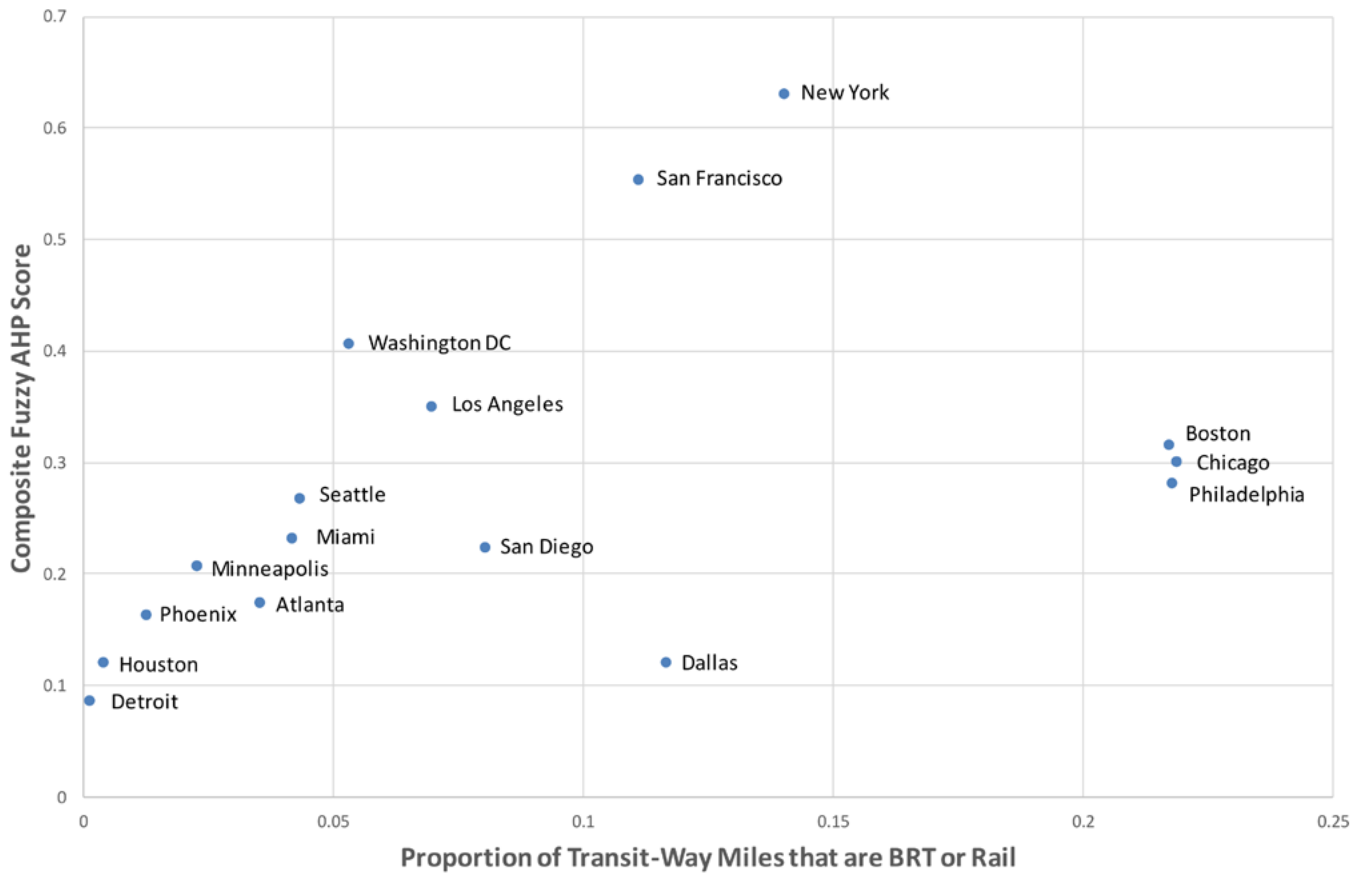


FIGURE 3. Proportion of transitway miles on BRT on rail vs. composite score for 16 Group A cities

The composite rankings potentially could be used to identify other types of performance outliers. For example, Salt Lake City, UT; Seaside- Monterey, CA; and Kahului, HI all scored near the top of their respective lists because they operate considerably more vehicle revenue-miles per square mile of land area than their peers. Since the NTM data set attributes all service to the largest UZA served by each agency, this potentially indicates that compared to their peers, these UZAs offer more service connecting to nearby rural or small urban areas. Conversely, communities that scored low due to low revenue per vehicle-mile are potential candidates for intensified marketing programs, or might simply have lower fares than their peers.

Comparison of “Efficiency” and “Effectiveness”

In addition to the overall performance analysis presented in the preceding section, the results of the fuzzy AHP process can be used to compare the extent to which there

are trade-offs between the business efficiency performance measures and operational effectiveness measures. As noted in the literature review, this has been debated in the academic literature, with some authors taking the view that higher efficiency comes at the price of lower effectiveness and other studies supporting the view that the two are somewhat independent. Our results suggest that both assertions are partially justified—it depends on which urban areas are included in the sample.

The efficiency and effectiveness scores for each of the 294 UZAs are plotted in Figure 4, with Business Efficiency on the horizontal axis and Operational Effectiveness on the vertical axis. Dots of various colors identify the members of each peer group, and regression lines are plotted for each of the six groups. The slope of these regression lines provides an indication of the tendency for the cities in each group to prioritize efficiency or effectiveness: a more vertical slope indicates an emphasis on effectiveness, and a slope closer to the horizontal indicates that greater emphasis has been put on efficiency.

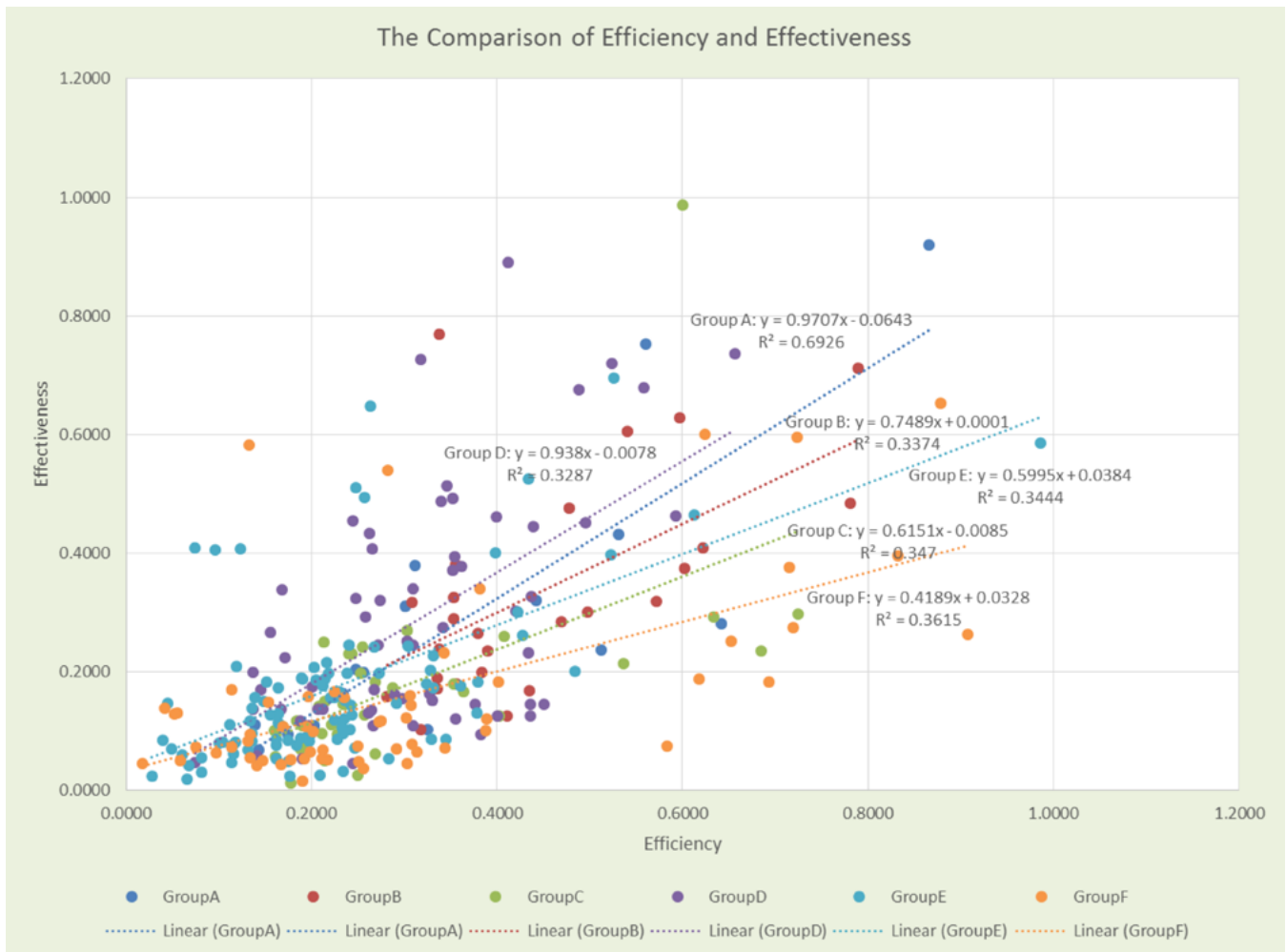


FIGURE 4. Scatter plot of business efficiency vs. operational effectiveness

The analysis suggests that the urbanized areas in Peer Groups A and D differ from the other four peer groups. Many of the UZAs in Group A score quite well on *both* efficiency and effectiveness measures, i.e., the data points tend to fall toward the upper right corner of the graph. Exogenous factors probably explain these results. For example, most Group A conurbations have relatively high population and employment densities, perhaps providing ridership sufficient to support extending the geographical extent of the systems well into the suburbs. Transit ridership in Group A cities is also influenced by traffic congestion (with the exception of Philadelphia and San Diego, automobile commuters in all Group A cities experience traffic delays exceeding 50 hours annually according to the 2015 TTI/Inrix mobility scorecard [Schrang et al. 2015]). In addition, several Group A cities have land use patterns that are relatively favorable to transit; New York, Philadelphia, Boston, and Chicago are examples of metro areas whose “main line” suburbs developed around commuter rail stations in the late 19th and early 20th centuries (Jackson 1985). Further research potentially could elaborate the relationships between land use patterns and the transit performance for these cities.

Group D includes a number of university towns. With a base of by student and university employee ridership, it appears that these communities are able to provide relatively comprehensive geographical coverage.

With the exception of some university towns and other unusual cases, the cities in Groups B, C, E, and F tend to have transit systems that appear to be more focused on efficiency than on operational effectiveness. This perhaps reflects the fiscal realities of these transit systems. When faced with budgetary challenges, transit planners often respond by cutting off-peak services (Nelson Nygard Consulting Associates 2009), trimming less-productive routes (Bizjak 2016), raising fares (Rutti 2016), and finding ways to use personnel and equipment as efficiently as the circumstances allow. Many of the cities in these groups also have low-density land use patterns and relatively uncongested traffic, and some lack supporting infrastructure such as sidewalks connecting transit stops to traffic generators—conditions that can make it difficult to attract discretionary riders or justify service expansions.

Conclusions and Opportunities

“Good performance” for transit systems is a stew with many ingredients. Focusing on a single performance measure (such as farebox recovery ratio) is problematic because transit systems serve a multitude of constituencies and policy objectives. Policymakers do not always agree about the relative importance of each objective, and experts might differ about how to weight the performance measures that correspond to the policy goals. This preliminary exploration suggests that the fuzzy AHP method could be useful in establishing an objective basis for comparing the performance of different systems, *without* the need to develop a consensus on which policy objectives are the highest priority. The methodology also provides tools that transit systems could use to compare themselves to peer groups based on readily-available data.

The analysis identified some strengths and limitations of the recently-developed NTM data set. Although the NTM currently contains only a limited set of transit performance measures, its spatial aggregation level (UZAs) is not readily-available in most other data sources. An important characteristic of the NTM is that each transit agency's performance data are associated only with the largest UZA it serves. Although this avoids double-counting when measures are summed on a national, regional, or statewide basis, when viewed from the perspective of individual communities, it tends to understate the amount of transit service provided in smaller UZAs. As a result, the NTM in its current form is perhaps best-suited to identifying well-performing UZAs and is less reliable as an indicator of poorly-performing UZAs.

The methodology presented in this paper could be enhanced by adding additional performance measures that were not available in the NTM data set. As the number of metrics increases, greater caution will be necessary to minimize the effects of correlations between the analytical metrics; for example, as the number of transitway miles increases, it is likely that the number of vehicle revenue miles also will increase. Further investigation with additional metrics also could help determine which performance metrics produce the most stable and consistent analytical results. For example, expansion of the methodology could help analysts avoid situations where the rank order is unduly influenced by the addition or omission of data for a small number of peer group members.

In future research, many other relationships could be explored using the composite scores. For example, university towns and urban areas with high land values (such as New York City, San Francisco, and Honolulu) tend to score well, perhaps because they have high relatively low car ownership rates compared to their peers. Similarly, there might be a relationship between system performance and the overall economic situation in each community; underlying factors could be explored by combining this ranking system with UZA-level economic data.

It might also be worthwhile to investigate applications of the fuzzy AHP methodology for comparisons of the performance of internal accounting cost centers within a transit system, such as individual rail lines or bus routes. Potentially, such investigations would assist agencies in identifying opportunities to strengthen overall system performance by pinpointing lines or routes that are not performing as well as their peers.

References

- Abreha, D. A. 2007. "Analysing Public Transport Performance using Efficiency Measures and Spatial Analysis: The Case Study of Addis Ababa, Ethiopia." Master's thesis, International Institute for Geo-information Science and Earth Observation, Enschede.
- Andersen, P., and N. P. Christian. 1993. "A Procedure for Ranking Efficient Units in Data Envelopment Analysis." *Management Science*, 39(10): 1261-1264.
- Badami, M. G., and M. Haider. 2007. "An Analysis of Public Bus Transit Performance in Indian Cities." *Transportation Research Part A: Policy and Practice*, 41: 961-981.

- Barnum, D. T., S. Tandon, and S. McNeil, 2008. "Comparing the Performance of Urban Transit Bus Routes after Adjusting for the Environment, Using Data Envelopment Analysis." Great Cities Institute, University of Illinois at Chicago.
- Bizjak, T. 2016. "Light-Rail Green Line May Shut in Sacramento to Save Money." *Sacramento Bee*, May 20. <http://www.sacbee.com/news/local/transportation/article78954437.html>.
- Charnes, A. C., W. W. Cooper, and E. L. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research*, 2: 429-444.
- Chu, X., G. J. Fielding, and B. W. Lamar. 1992. "Measuring Transit Performance Using Data Envelopment Analysis." *Transportation Research Part A: Policy and Practice*, 26(3): 223-230.
- Cook, W. D., and L. M. Seiford. 2009. "Data Envelopment Analysis (DEA)—Thirty Years On." *European Journal of Operational Research*, 192(1): 1-17.
- Eboli, L., and G. Mazzulla. 2011. "A Methodology for Evaluating Transit Service Quality Based on Subjective and Objective Measures from the Passenger's Point of View." *Transport Policy*, 18: 172-181.
- Fan, Y. 2015. "National Transit Map—United States: A Global Transit Innovations Data System." University of Minnesota.
- Fielding, G. J., R. E. Glauthier, and C. A. Lave. 1978. "Performance Indicators for Transit Management." *Transportation*, 7(4): 365-379.
- Fielding, G. J., T. T. Babitsky, and M. E. Brenner. 1985. "Performance Evaluation for Bus Transit." *Transportation Research Part A: General*, 19(1): 73-82.
- Fielding, G. J. 1992. "Transit Performance Evaluation in the US." *Transportation Research Part A: Policy and Practice*, 26(6): 483-491.
- Griffis, Stanley E., Martha Cooper, Thomas J. Goldsby, and David J. Closs. 2004/ "Performance Measurement: Measure Selection Based upon Firm Goals and Information Reporting Needs," *Journal of Business Logistics*, 25(2): 95-118.
- Hanaoka, P., and P. Kunadhamraks. 2009. "Multiple Criteria and Fuzzy Based Evaluation of Logistics Performance for Intermodal Transportation." *Journal of Advance Transportation*, 43(2): 123-153.
- Jackson, K. T. 1985. *Crabgrass Frontier: The Suburbanization of the United States*. London: Oxford University Press.
- Karlaftis, M. G. 2004. "A DEA Approach for Evaluating the Efficiency and Effectiveness of Urban Transit Systems." *European Journal of Operational Research*, 152(2): 354-364.
- Kittelsohn & Associates. Federal Transit Administration, Transit Cooperative Research Program, and Transit Development Corporation. 2003. *Transit Capacity and Quality of Service Manual, Vol. 100*. Transportation Research Board.
- Korostelev, A., L. Simar, and A. B. Tsybakov, 1995. "Efficient Estimation of Monotone Boundaries." *Annals of Statistics*, 23: 476-489.

- Kurtzleben, D. 2011. "10 Best Cities for Public Transportation." *US News & World Report*. <http://www.usnews.com/news/articles/2011/02/08/10-best-cities-for-public-transportation>.
- Lao, Y., and L. Liu, L. 2009. "Performance Evaluation of Bus Lines with Data Envelopment Analysis and Geographic Information Systems." *Computers Environment and Urban Systems*, 33: 247-255.
- Li, X., Y. Liu, D. Liu, and Z. Gao. 2015. "Multidimensional Assessment of Developing an Urban Public Transit Metropolis in China." *Journal of Urban Planning and Development*, 142(3): 04015021.
- Nathanail, E. 2008. "Measuring the Quality of Service for Passengers on the Hellenic Railways." *Transportation Research Part A*, 42: 48-66.
- Neff, J., and M. Dickens. 2013. *Public Transportation Fact Book*. American Public Transportation Association, Washington, DC.
- Nelson Nygard Consulting Associates. 2009. "Stranded at the Station: The Impact of the Financial Crisis in Public Transportation." <http://www.environmentalcouncil.org/mecReports/Stranded-at-the-station.pdf>.
- Rutti, R. 2016. "RTA Approves Fare Increases, Service Cuts to Help Erase \$7 Million Budget Gap." *Crain's Cleveland Business*, June 7. <http://www.crainscleveland.com/article/20160607/NEWS/160609845/rta-approves-fare-increases-service-cuts-to-help-erase-7-million>.
- Ryus, P., M. Connor, S. Corbett, A. Rodenstein, L. Wargeling, L. Ferreira, Y. Nakanishi, and K. Blume. 2003. "Guidebook for Developing A Transit Performance-Measurement System." TCRP Report 88.
- Ryus, P. 2010. "A Methodology for Performance Measurement and Peer Comparison in the Public Transportation Industry, Vol. 141." Transportation Research Board.
- Saaty, T. L. 1980. *The Analytic Hierarchy Process*. New York: McGraw-Hill.
- Schrank, D., W. Eseele, T. Lomax, and J. Bak, 2015. "Urban Mobility Scorecard." Texas A&M Transportation Institute and Inrix, Inc. <http://d2dtl5nnlpfr0r.cloudfront.net/tti.tamu.edu/documents/mobility-scorecard-2015.pdf>.
- Sheth, C., K. Triantis, and D. Teodorovic. 2007. "Performance Evaluation of Bus Routes: A Provider and Passenger Perspective." *Transportation Research Part E*, 43: 453-478.
- Simar, L., and P. Wilson. 1998. "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-Parametric Frontier Models." *Management Science*, 44: 49-61.
- Song, M. L., L. L. Zhang, W. Liu, and R. Fisher. 2013. "Bootstrap-DEA Analysis of BRICS' Energy Efficiency Based on Small Sample Data." *Applied Energy*, 112: 1049-1055.
- Talley, W. K., and P. P. Anderson. 1981. Effectiveness and Efficiency in Transit Performance: A Theoretical Perspective." *Transportation Research Part A: General*, 15(6): 431-436.

- Transportation Research Board of the National Academies. 2003. *Transit Capacity and Quality of Service Manual*. Washington, DC.
- Tyrinopoulos, Y., and C. Antoniou. 2008. "Public Transit User Satisfaction: Variability and Policy Implications." *Transport Policy*, 15: 260-272.
- US Census Bureau. 2013. "2010 Census Urban Area Facts." <http://www.census.gov/geo/reference/ua/uafacts.html>.
- UITP (International Association of Public Transport). 2010. "Report on Statistical Indicators of Public Transport Performance in Sub-Saharan Africa." <http://www.uitp.org>.
- Wang, Y. M., C. Parkan, and Y. Luo. 2008. A Linear Programming Method for Generating the Most Favorable Weights from a Pair Wise Comparison Matrix." *Computers and Operations Research*, 35 (12): 3918-3930.
- Yeh, C-H., H. P. Deng, and Y-H. Chang. 2000. "Fuzzy Multi-Criteria Analysis for Performance Evaluation of Bus Companies." *European Journal of Operational Research*, 126(3): 459-473.
- Yu, J., Y. Liu, G. L. Chang, W. J. Ma, and X. G. Yang. 2011. "Locating Urban Transit Hubs: A Multi-Criteria Model and Case Study in China." *ASCE Journal of Transportation Engineering*, 137(12): 944-952.

About the Authors

XIN LI, PH.D. (li44@uwm.edu) is a Post-Doctoral Researcher at the University of Wisconsin–Milwaukee.

YINGLING FAN, PH.D. (yingling@umn.edu) is Associate Professor in the Urban and Regional Planning Program of the Humphrey School of Public Affairs at the University of Minnesota.

JOHN W. SHAW, P.E. (jwshaw@wisc.edu) is a Researcher at the University of Wisconsin–Madison.

YUNLEI QI (yunleiqi@umn.edu) is a Ph.D. student in the Urban and Regional Planning Program of the Humphrey School of Public Affairs at the University of Minnesota.

Recommendation of a New Transit Performance Measure in the National Transit Database

*Fatemeh Baratian-Ghorghi**, Ph.D.
Parsons Corporation

Hossein Ahmadianyazdi, M.Sc
Iran University of Science and Technology

* Corresponding Author

ABSTRACT

The frequency of incidents negatively changes public perception with regard to the public safety in transit systems. The level of safety expressed as performance measures, ensures contractors and users that a quality safety level is maintained. Both the private and public sectors in the U.S. annually report accurate data to the National Transit Database (NTD) to be used in assessing the progress of the nation's public transportation systems. Although they must provide both the annual report and monthly reports including Safety and Security data, measures are only grouped into the indicators of efficiency and effectiveness of a system and there is a distinct lack of safety performance measures.

This paper makes a comparison between the safety indicator values in large and small areas, and finds the correlation between system effectiveness and safety measures. Comparison results provide evidence that not every effective system is safe. Finally, three safety indicators are suggested which enables transit agencies to find their system's weaknesses in terms of safety.

Keywords: *Safety, Performance Measure, Service Effectiveness, National Transit Database, Transit System*

INTRODUCTION:

The NTD, as the best national transit data source, can be used to compare service measures across the transit agencies in the U.S. All agencies benefiting from Urbanized Area Formula Program grants are required by FTA (1) to report certain statistical information to the NTD annually (2). However, organizations operating nine or fewer

vehicles in a non-fixed-guideway service are free from the obligation of reporting. The results show how Federal Transit Administration (FTA) funds should be allocated to the transit agencies. Currently, the NTD includes the information gathered from 930 separate agencies all over the nation. Information is available for revenues, wages, maintenance, incidents, employee, expenses, amount of service provided (e.g., vehicle miles, vehicle hours, and days of service), and amount of service consumed (e.g., unlinked trips and passenger miles). This information is readily available for agencies, planners, researchers, and others (3). Measures in the NTD database are grouped into two basic dimensions of performance, namely efficiency and effectiveness. Efficiency indicators define the relationship between operating expenses and the number of vehicle hours or vehicle miles (often referred to as “Produced Output Type”). Effectiveness indicators, however, are categorized into two groups: the first group measures the operating expenses against passenger miles and unlinked passenger trips (often referred to as “Consumed Output Type”) while the second group measures the unlinked passenger trips to the amount of service produced. As for the service effectiveness, the NTD has selected four of the many possible measures based on the availability of reported data. Still, there is a distinct lack of safety performance measures in the agency profiles reported to the NTD (5).

In this paper the appraisal of the transit system’s safety through the use of performance measures is examined. A set of indicators was collected from many past studies, and the ones possible to calculate based on the available data are employed for the measurement of transit safety performance. Starting with an initial set of 11 performance indicators, correlation tests are used to find and suggest the proper indicators to assess transit safety performance for every single mode.

LITERATURE REVIEW

Many studies have strived to propose methods to identify indicators which can properly measure transportation system performance (6-14). Different agencies may use and report different performance measures consistent with their objectives. However, not every researcher proposes using multiple indicators. Gordon et al. (7) identified and gathered a set of performance measures for transit systems and then suggested the bests to be used. Authors considered several criteria in the selection process such as distribution of values in the variable to be normal-like, the ease of gathering data, and the percentage of missing data during the selection process. The safety performance measures they suggested were: 1,000,000 vehicle miles traveled (VMT) per accident and accident and revenue vehicle hours per accident.

Texas Transportation Institute (TTI) recognized performance indicators to measure and monitor the performance of mobility throughout Texas and at a national level (8). Researchers summarized 20 case studies of the mobility management programs with the best examples of performance measures such as accidents per 100,000 miles of service, incidents per 1,000 passenger boarding, rate of serious injuries among transit riders per 100 million passenger miles traveled (PMT), and the average age of revenue vehicle fleet.

The United States Department of Transportation (USDOT) strategic plan, entitled “Transportation for a New Generation”, outlined the performance measures related to achieving strategic goals of safety (9). For example, the rate of transit-related fatalities per 100 million PMT was selected as the indicator of safety performance and the related goal was its reduction from 1.13 in 2009 to 1.03 in 2013. In addition to the passenger and vehicle safety an additional indicator was considered to measure the non-occupant (pedestrian and bicycle) safety as well (i.e. non-occupant fatalities per 100 million VMT).

Phase I of the Valley Metro/Regional Public Transportation Authority Efficiency and Effectiveness Study (10) listed the performance measures and related targets for transportation system performance factors. They used the safety incidents per 100,000 VMT as the only safety indicator of their transit system. Another study provided a summary of best existing practices in transit service planning and explained how to calculate safety performance measures (11). For example, no more than 6 accidents may occur per 100,000 miles operated. Fielding (1992) examined three programs: Federal Triennial Reviews, California Performance Audits and the Los Angeles Program. He used factor analytic techniques to reduce 48 transit performance indicators to a smaller set of measures, resulting in a model with 7 factors and 7 key variables, termed “markers” (12). The only safety indicator he used was the total vehicle miles between accidents.

Livermore Amador Valley Transportation Authority (LAVTA) (14) used three performance measures to assess their transit system, including traffic accidents per 100,000 VMT, passenger injury per 100,000 passenger boarding, percentage of preventive maintenance inspections completed within 10% of scheduled mileage.

The Transit Cooperative Research Program (TCRP) report 88 (3) provides a step-by-step process for developing a performance-measurement program. It describes the characteristics of an effective performance-measurement system and categorizes the performance measures. Data sources, data collection and management techniques that can be employed are described in this report. The program uses twelve case studies of successful performance-measurement programs and presents over 400 performance measures. Regarding safety, this report introduces five categories of safety and security goals as shown in Table 1.

Another interesting topic related to transit performance is the examination of the relationship between system performance indicators. Some studies have previously found the correlation between effectiveness and efficiency. Karlaftis and McCarthy (15) found that efficiency and effectiveness of a system are strongly related. However, their results disagreed with the findings of Chu et al. (16) who reported efficient systems are not necessarily effective. The reason behind this difference may be the different sets of data used in these studies.

What becomes quite apparent from the results of previous works is that there is a distinct lack of assessing transit safety performance and examining the relationship between safety indicators and the remaining attributes. Also, from the literature review, it was ascertained not every performance measure is attainable from NTD, and they are not necessarily related to safety, such as the average age of revenue vehicle fleet or the average annual miles of service per revenue vehicle (8). Thus, a single indicator,

or a smaller set of more reliable indicators, is needed to describe the transit safety performance.

Safety and Security Reporting Manual by the NTD explains how FTA funds are allocated to the transit agencies and what type of safety and security data are reported and are available to the public users (4). Also, the safety and security reporting forms are described and the process of completing and reporting incidents is clearly explained. Using this manual, the safety related data reported to the NTD was found.

TABLE 1.
Safety and Security
Transit Performance
Measures in TCRP

Category	Performance Measures	Examples
Vehicle and property damage	Accident Rate Number of Fires	<ul style="list-style-type: none"> • Vehicle accidents, Customer accidents • A measure of system safety
Passenger accidents	Passenger Safety	<ul style="list-style-type: none"> • Fatal accidents per passenger-miles/VMT • Injury accidents per passenger-miles/VMT • Property-damage-only accidents per passenger miles/ VMT • Response time • Incident/accident durations
Accident potential	Drug-related accident potential Bus operator accident potential Rail operator accident potential Maintenance-related accident potential	<ul style="list-style-type: none"> • Percent of positive drug/alcohol tests • Number of traffic tickets issued to operators • Percent of buses exceeding speed limit • Number of station overruns • Road calls • Fleet maintenance performance
Workplace safety	Employee Work Days Lost to Injury	<ul style="list-style-type: none"> • Number of work days lost to injury
Passenger security	The number of crimes committed on transit property The level of security provided Customer perceptions of the safety and security of the transit system	<ul style="list-style-type: none"> • Number of crimes (Crime rate) • Ratio of transit police officers to transit vehicles • Number (Percent) of vehicles with specified safety devices • Customer satisfaction • Incidents of vandalism

SAFETY PERFORMANCE MEASURE SELECTION

From the past studies reviewed, the safety indicators and their use in the assessment of transit systems was found. In selecting the key indicators for performance evaluation, particular attention was given to the availability and reliability of data in the NTD, necessary for calculating each indicator. As an example, the number of crimes was omitted from the database due to a high percentage of missing information. Data came from a total of 930 urban transit agency reports to the NTD in 2012. It must be noted that data for purchased services versus directly operated are not included in the data sets. Table 2 lists 11 performance measures selected for analysis from the past studies. Indicators are classified into four groups: Vehicle safety, Transit riders' safety and security, Employee and operator safety, and Non-riders safety.

TABLE 2.
Safety Transit
Performance Measures

Categories	Performance Measures
Vehicle safety	<ol style="list-style-type: none"> 1. Accidents per 100,000 miles of service 2. Accidents per 1,000 hours of service 3. Number of Incidents per 1000 Vehicle Miles 4. Number of Incidents per 1,000 passenger boarding 5. Number of Fires
Transit riders safety and security	<ol style="list-style-type: none"> 1. Number of transit-riders injuries per 100 million PMT 2. Number of transit-riders fatalities per 100 million PMT
Employee and operator safety	<ol style="list-style-type: none"> 1. Employees fatalities per 100 million VMT 2. Employees injuries per 100 million VMT
Non-riders safety	<ol style="list-style-type: none"> 1. Non-occupant fatalities per 100 million VMT 2. Non-occupant injuries per 100 million VMT

COMPARISON BETWEEN DIFFERENT MODES SAFETY

Prior to estimating the indicators, data were broken down by mode when more than one mode exists. One cannot assume the highest demanded agency has the best safety score (or less safety issues) because safety is directly related to the type of in-use modes. For example, a higher level of accidents is expected for an agency mostly accommodating passengers with bus compared to the other agency that uses heavy rail transit. It is quite clear that the probability of accidents with pedestrians, bicyclists and other vehicles is likely to be greater in the first mode. Furthermore, the amount of service produced and service consumed varies among modes which makes the safety ratios variable. In an attempt to check this assumption, a comparison was made between the safety indicator values of different modes. In doing so, it was necessary to classify modes of transit in different groups and then find the performance measures for the agencies operating the same mode. The “2012 National Transit Summaries and Trends” (NTST) prepared by the NTD (17) classifies the transit modes into seven categories as described in Table 3.

TABLE 3.
NTST Modal
Classifications (17)

Bus	Demand Response	Vanpool	Heavy Rail
Bus (MB) Commuter Bust (CB) Bus Rapid Transit (BRT)	Demand Response (DR) Demand Response Taxi (DT)	Vanpool (VP)	Heavy Rail (HR)
Light Rail	Commuter Rail	Other Modes	
Light Rail (LR) Streetcar Rail (SR) 2080 Hybrid Rail (YR) 9030 Hybrid Rail (YR)	Commuter Rail (CR) 0008 Hybrid Rail (YR) 6048 Hybrid Rail (YR)	Aerial Tramway (TR) Alaska Railroad (AR) Cable Car (CC) Ferry Boat (FB) Incline Plane (IP)	Jitney (JT) Monorail/Automated Guideway (MG) Publico (PB)

After splitting data, the percentages of agencies accommodating each category of modes were found to be 82 percent buses, 81 percent demand response (also called dial-a-ride or paratransit, provides services at the passenger’s request), 12 percent vanpool, 2 percent heavy rail, 4 percent light rail, and 11 percent other types of transit modes

(OM). It should be noted that none of the agencies have reported commuter rail data to the NTD, this data, however, are reported to the Federal Railroad Administration (FRA). Figure 1 indicates that different modes provide different levels of safety.

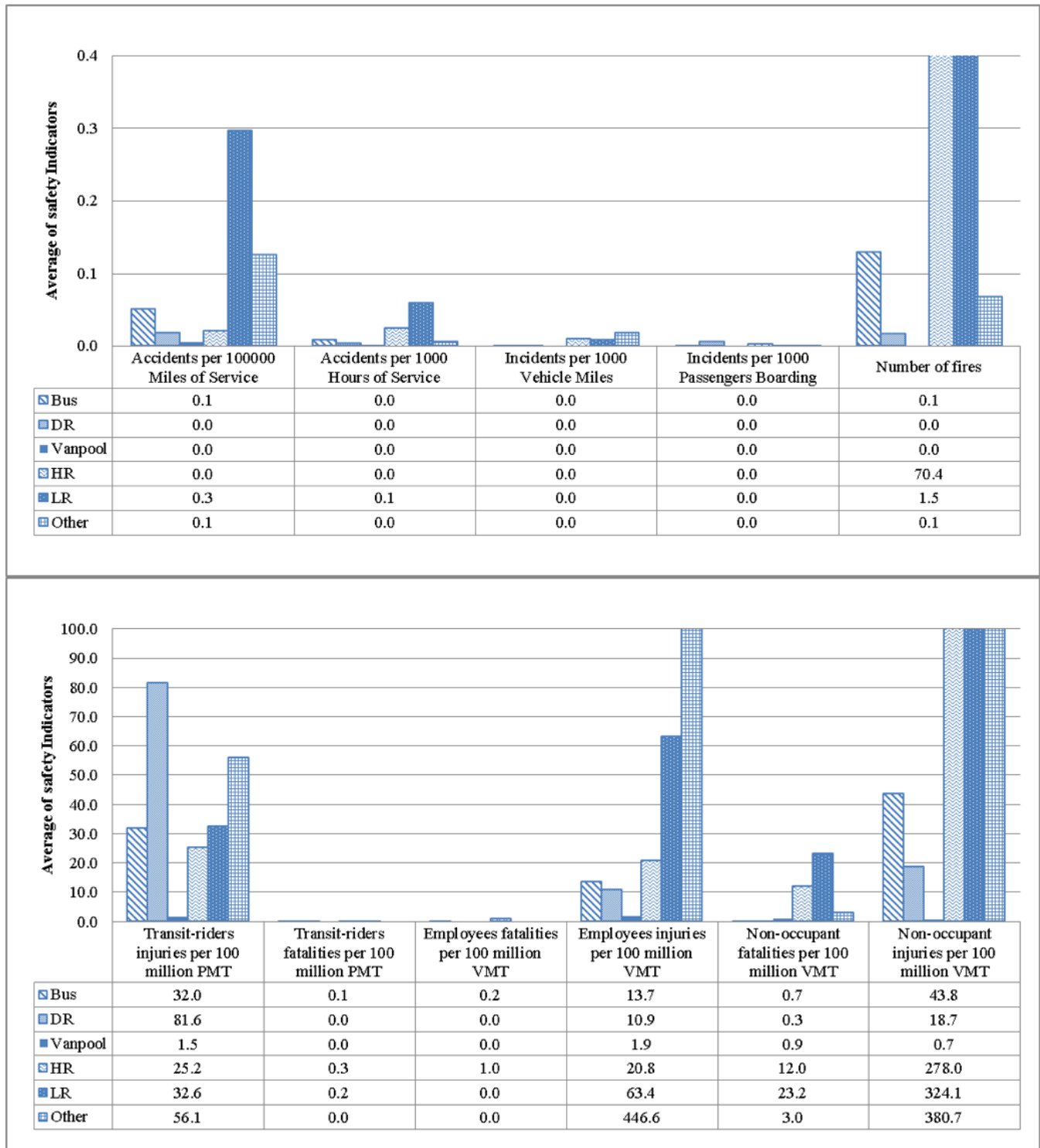


FIGURE 1. (a) Vehicle Safety for each Mode, (b) Riders, Non-riders, and Employees Safety for each Mode

Fig 1(a) shows the vehicle safety for each mode, and Fig 1(b) shows riders, non-riders, and employees' safety for each mode. The overall results show the variation between the safety indicators among transit modes. It is found from Fig.1 that vanpool is the safest mode. The Vanpool Safety Brochure (18) identifies several reasons for this including: vanpool drivers have several years of experience driving, vanpools operate in urban areas and on major highways with safer speeds, having 14 passengers in van keeps most drivers operating in a safer manner, and drivers have been trained before operating the vehicle. The findings reveal that mostly the victims in transit-related accidents are non-occupants. Numbers are rounded up to one decimal degree. Transit riders' fatalities only account for 3.8 percent of all reportable fatalities in 2012 (17). According to the Figure 1, the greatest values of non-occupant's safety indicators are for rail transit modes (i.e. Heavy and Light Rail Transit), meaning that these systems have the highest safety issues for non-riders. Also, the rail transit vehicles experience the highest rate of injuries (riders, employees and non-occupants). Meanwhile, accident rate values are the highest for light rail transit mode. The least safety levels regarding the transit rider's injuries are found in demand response mode. The reason can be the medical fragility and/or age of many of the passengers using this mode of transportation.

COMPARISON BETWEEN LARGE AND SMALL AREAS

One of the most revealing types of analysis is the comparison between safety performance measures of different agencies. However, transit systems work in areas with different sizes. The area size might have effects on the safety that a system provides. Thus, it is useful to find which indicators are dependent on the size of the area before making any suggestions.

In this section of the study, analysis was conducted to investigate any association between the transit system's safety and area size. Within the NTD, a numeric code is assigned to the Census-designated Urbanized Area (UZA) variable, which indicates where the system primarily operates. Numbers are assigned sequentially from the largest UZA (#1 = New York-Newark, NY-NJ-CT) to the smallest (#452 = Fond du Lac, WI), with exceptions made for UZAs in Puerto Rico (numbered from #500-#511), the special Virgin Islands UZA (only applies to FTA grants, numbered #600), and one UZA that was added after the initial UZA definitions were published (#453-Cumberland, MD-WV-PA) (3). Since the population of each area is not recorded in NTD, the comparison was made between the first half of the UZA codes, showing the larger areas, and the second half, indicating the small areas.

Data in Table 4 presents the average safety performance measures for large and small areas with regards to the transit mode. The p-values came from the results of a t-test, to compare the safety between the two areas. The sample size for each mode of transit in large and small areas are as follows: 270 and 380 for bus system; 241 and 392 for DR system; 49 and 23 for Vanpool mode and 27 and 9 for other modes. It should be noted that it was practically impossible to make a comparison between rail transit safety in large and small cities due to the lack of rail transit systems in small areas.

TABLE 4.
Large vs. Small Areas
Safety Indicator Values

Average values (Large area vs. Small area (p-value))				
Safety Indicators	Bus	DR	Vanpool	OM
Accidents per 100000 Miles of Service	0.05 vs. 0.05 (0.71)	0.03 vs. 0.01 (0.03)*	0.00 vs. 0.00 (0.79)	0.17 vs. 0.00 (0.20)
Accidents per 1000 Hours of Service	0.01 vs. 0.01 (0.97)	4.1E-03 vs. 1.8E-03 (0.03) *	0.00 vs. 0.00 (0.79)	6.2E-03 vs. 4.6E-05 (0.07) *
Incidents per 1000 VMT	0.00 vs. 0.00 (0.65)	0.00 vs. 0.00 (0.54)	0.00 vs. 0.00 (0.80)	0.02 vs. 0.01 (0.07) *
Incidents per 1000 Passengers Boarding	1.5E-03 vs. 1.4E-03 (0.08) *	0.01 vs. 0.00 (0.99)	0.00 vs. 0.00 (0.80)	1.6E-03 vs. 1.5E-04 (0.03) *
Number of fires	0.25 vs. 0.04 (0.00) *	0.01 vs. 0.02 (0.72)	0.00 vs. 0.00 (Note ²)	0.10 vs. 0.02 (0.20)
Transit riders injuries per 100 million PMT	22.31 vs. 42.15 (0.01) *	57.02 vs. 103.80 (0.23)	1.42 vs. 1.75 (0.87)	75.52 vs. 0.78 (0.04) *
Transit riders fatalities per 100 million PMT	0.09 vs. 0.11 (0.84)	0.04 vs. 0.00 (0.32)	0.00 vs. 0.00 (Note ²)	0.00 vs. 0.00 (Note ²)
Employees and operators fatalities per 100 million VMT	0.00 vs. 0.32 (0.40)	0.00 vs. 0.00 (Note ¹)	0.00 vs. 0.00 (Note ²)	0.00 vs. 0.00 (Note ²)
Employees and operators injuries per 100 million VMT	12.62 vs. 14.53 (0.67)	8.62 vs. 12.30 (0.65)	2.82 vs. 0.00 (0.11)	595.40 vs. 0.00 (0.17)
Non-occupant fatalities per 100 million VMT	0.88 vs. 0.54 (0.52)	0.22 vs. 0.36 (0.65)	0.11 vs. 2.70 (0.35)	4.02 vs. 0.00 (0.26)
Non-occupant injuries per 100 million VMT	45.45 vs. 42.70 (0.77)	27.12 vs. 13.49 (0.25)	1.03 vs. 0.00 (0.23)	506.52 vs. 3.25 (0.03) *

*statistically significant

¹The number of employees' fatality is zero

²The number of riders' fatality is zero

Table 4 reveals that there is a significant difference between the bus systems safety in large and small areas in terms of Incidents per 1000 Passengers Boarding, Number of fires, and Transit riders' injuries per 100 million PMT. Results indicate that in small areas bus fleets operate safer while transit riders are more likely to be injured. Similarly, it could be concluded from the DR and OM safety indicator values that higher vehicle safety is provided in small areas. Regarding the transit riders and non-riders, OM provides a safer environment in the small areas.

THE RELATIONSHIP BETWEEN SAFETY AND EFFECTIVENESS

In order to identify the appropriate indicators to be included within the NTD profiles, the relationship between the system effectiveness and safety performance measures was investigated.

Correlations between data items will be valuable for examining the relationship between two sets of data. Consistent results across the safety and effectiveness indicators would provide evidence that the more effective a system is, the more safety it provides. In order to examine this assumption, the correlation coefficient was employed. The equation for the correlation coefficient is as follows:

$$Correl(X, Y) = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$

Where, \bar{X} and \bar{Y} are the sample means of each pair of safety indicators and effectiveness indicators for each mode. The strength of an association is graded from zero to 1.00 and the direction of the relationship is expressed by a positive sign (+) if the relationship is direct, and by a negative sign (-) if the relationship is inverse. In order to find how strong a relationship is, three criteria are defined:

1. Correlations above 0.80 are regarded as high
2. Correlations between 0.50 and 0.80 are considered moderate
3. Correlations below 0.50 are regarded as low

Table 5 and Table 6 show the correlation coefficient for each pair of indicators.

TABLE 5.
Vehicle safety performance Measures vs. Service effectiveness

Service Effectiveness	Transit Mode	Accidents per 100000 Miles of Service	Accidents per 1000 Hours of Service	Incidents per 1000 VMT	Incidents per 1000 Passengers Boarding	Number of fires
unlinked passenger trip per vehicle revenue mile	Bus	0.11	0.05	0.15	-0.06	0.15
	DR	0.02	0.01	-0.03	-0.08	-0.06
	Vanpool	0.13	0.11	0.13	0.03	Note ¹
	HR	-0.25	-0.27	0.14	-0.76*	0.56
	LR	0.37	-0.02	0.29	-0.18	0.30
	OM	0.21	0.14	0.15	-0.09	0.00
unlinked passenger trip per vehicle revenue hour	Bus	0.14	0.09	0.18	-0.03	0.17
	DR	0.10	0.10	-0.03	-0.07	-0.05
	Vanpool	0.16	0.15	0.16	0.07	Note ¹
	HR	-0.15	0.03	0.34	-0.51	0.77*
	LR	-0.05	0.40	0.26	0.14	0.30
	OM	0.04	-0.01	0.03	-0.05	-0.02

¹Number of fires is zero for all agencies.

* Significant

The greater value of the safety indicator shows the less safety of the system. Thus, having a positive correlation means the higher effective systems are less safe. According to the Table 5, the number of incidents per 1000 VMT, as well as the number of fires in heavy rail systems is positively correlated with the effectiveness of the systems. Surprisingly, the more effective heavy rail systems are more likely to have fire incidents, and more incidents per VMT will occur.

TABLE 6.
Transit-riders,
Employees, and
Non-rider Safety and
Security Indicators
vs. Effectiveness
Indicators

Service Effectiveness	Mode	Transit Riders Fatalities per 100 million PMT	Transit Riders Injuries per 100 million PMT	Employees and operator fatalities per 100 million VMT	Employees and operator injuries per 100 million VMT	Non-occupant fatalities per 100 million VMT	Non-occupant injuries per 100 million VMT
unlinked passenger trip per vehicle revenue mile	Bus	-0.02	-0.05	0.02	0.07	0.09	0.05
	DR	-0.04	-0.03	Note ²	-0.01	-0.05	0.11
	Vanpool	Note ¹	0.18	Note ²	-0.05	-0.02	-0.06
	HR	-0.24	-0.28	-0.24	0.23	-0.12	-0.04
	LR	-0.23	0.01	Note ²	0.13	-0.24	0.35
	OM	Note ¹	-0.01	Note ²	0.18	-0.06	0.04
unlinked passenger trip per vehicle revenue hour	Bus	-0.02	-0.06	0.02	0.08	0.10	0.07
	DR	-0.02	-0.06	Note ²	-0.02	0.00	0.19
	Vanpool	Note ¹	0.20	Note ²	-0.05	0.05	-0.04
	HR	-0.25	-0.09	-0.10	0.06	-0.01	0.23
	LR	0.12	-0.16	Note ²	0.01	0.27	0.22
	OM	Note ¹	-0.04	Note ²	0.04	-0.03	0.00

¹Number of riders' fatality is zero

²Number of employees' fatality is zero

As no significant relationship was found between the safety and effectiveness indicators, one cannot conclude that the service effectiveness reflects the safety condition. Therefore, it is advantageous to ask the agencies report a set of safety indicators through NTD profiles.

SUGGESTING SAFETY PERFORMANCE INDICATORS

In order to select which indicators can represent the safety condition of the system, it is necessary to remove those highly dependent on the area size and effectiveness measures. In doing so, the results from Table 4 and Table 5 were taken into consideration.

From Table 4 it was found that:

- Considering 756 agencies operating DR, the value of “Accidents per 100000 Miles of Service” depends on the area size.
- Considering 786 agencies operating DR and OM the variable of “Accidents per 1000 Hours of Service” significantly depends on the area size.
- Variable of “incidents per 1000 Passengers Boarding”, the same as the previous indicator, depends on the area size. A total of 783 agencies all over the U.S. (operating bus and OM) provide different level of safety in terms of incidents per 1000 Passengers Boarding based on their area size.

- The indicator of “number of fires” and “transit riders’ injuries per 100 million PMT” have the same issue with area size.

From Table 5 it was found that:

- “Incidents per 1000 Passengers Boarding” as well as “the number of fires” in heavy rail systems are positively correlated with the effectiveness of the systems.
- “Number of transit Riders fatalities per 100 million PMT” and “Employees and operators fatalities per 100 million VMT” cannot be selected since a large number of systems did not experience riders or employees fatalities.
- The number of transit-riders injuries per 100 million PMT and the number of transit-riders fatalities per 100 million PMT cannot be calculated for one third of the agencies operating buses (323 out of 930) as they do not report PMT data to the NTD.
- Finally, the following indicators are suggested to be reported by transit agencies to the NTD:
 1. Incidents per 1000 VMT
 2. Employees’ and operators’ injuries per 100 million VMT
 3. Non-occupant fatalities per 100 million VMT

Injuries among Transit Riders per 100 million PMT

The last indicator showing the riders’ safety needs all agencies to provide the passenger mile traveled data to the NTD.

CONCLUSION

The research reported in this paper established a set of indicators that are particularly useful for the evaluation of transit systems safety performance. To accomplish this, past studies were reviewed to find out how a transit system’s safety performance is measured. The following criteria were used to collect a set of safety indicators from the available measures:

1. Variables collected by the agencies and reported to the NTD
2. Indicators that are easily understood by transit managers.

This study additionally strived to demonstrate whether or not there is a tangible perceptual differences among the variety of safety indicators in small and large areas. The higher values of safety performance measures were found for buses, DR and other modes (rather than rail transit) operating in small areas compared with large areas. Moreover, rail transit was found to have the high safety problems between all modes. Summing up the findings, it was concluded that the transit vehicles have higher safety issues in larger urbanized areas.

In order to suggest the safety indicators to be included within the NTD profiles, the relationship between the system effectiveness and safety performance measures was investigated. As no significant relationship was found between the safety and effectiveness indicators it was concluded that systems with high safety level do not necessarily perform effectively and vice versa.

The indicators best representing the system's safety performance were suggested as "Incidents per 1000 VMT", "Employees and operators injuries per 100 million VMT", and "Non-occupant fatalities per 100 million VMT".

Recently, some changes have been made in the Safety and Security Reporting Requirements that specifies how safety-related problems have to be reported in detail and highlights the importance of having safety measures (4). For example,

"Collisions involving transit vehicles that require towing away from the scene for a transit roadway vehicle or other non-transit roadway vehicle are automatically reportable" or "Rail transit vehicle collisions occurring at a grade crossing are automatically reportable." Addition of "Geographic location Longitude/Latitude" to the Basic Information Screen, also enables researchers to find the association between the characteristics of the service area and the level of safety.

Finally, it is important to remember this study does not show causation, but a relationship between service effectiveness and safety performance measures. This study does not ignore the importance of reporting safety data, but it suggests that the available and reliable data must be used to determine and present the most applicable and meaningful safety indicators of every system. Reporting more detailed information to the NTD makes it possible to consider new safety indicators. Therefore, it would be advantageous to keep exploring the relationship between the service performance measures to avoid presenting strongly correlated indicators together in a single report.

REFERENCES

1. Federal Transit Administration, National Transit Database, U.S. Department of Transportation, Washington, DC, 2000. <http://www.ntdprogram.com>.
2. Akoto, E.V. (2016). Exploring Diversified Performance Indicators for Evaluating Non-Urbanized Transit Program Outcomes, *Journal of Public Transportation*, 19(1):28-48.
3. Ryus, P. (2003). A Summary of TCRP Report 88: A Guidebook for Developing a Transit Performance-Measurement System. *TCRP Research Results Digest*, (56).
4. National Transit Database (2015). *Safety & Security Reporting Manual*. Office of Budget and Policy.
5. Federal Transit Administration. (2002). *Transit Profiles, 2000 Report Year*, Washington, DC.
6. Benjamin, J., Obeng, K. (1990). The effect of policy and background variables on total factor productivity for public transit. *Transportation Research* 24B (1), 1–14.

7. Fielding, G. J., Babitsky, T. T., & Brenner, M. E. (1985). Performance evaluation for bus transit. *Transportation Research Part A: General*, 19(1), 73-82.
8. Majumdar, S. R., Highsmith, M., Cherrington, L., & Weatherby, C. (2011). Performance measures for public transit mobility management (No. FHWA/TX-12/0-6633-1). Texas Transportation Institute, Texas A&M University System.
9. U.S. Department of Transportation. *Transportation for a New Generation. Strategic Plan for Fiscal Years 2012–16.*
10. Valley Metro/Regional Public Transportation Authority. (2007). *Efficiency and Effectiveness Study*. Project No. 064008S.
11. USF Center for Urban Transportation Research. (2009). *Best Practices in Transit Service Planning. Final Report.*
12. Fielding, G. J. (1992). Transit performance evaluation in the USA. *Transportation Research Part A: Policy and Practice*, 26(6), 483-491.
13. Anderson, M., & Khan, T. (2014). Performance Measures for the Analysis of Rural Public Transit in Alabama. *Journal of Public Transportation*, 17(4), 2.
14. Livermore Amador Valley Transit Authority. <http://www.wheelsbus.com/>
15. Karlaftis, M. G., & McCarthy, P. S. (1997). Subsidy and public transit performance: A factor analytic approach. *Transportation*, 24(3), 253-270.
16. Chu, X., Fielding, G. J., & Lamar, B. W. (1992). Measuring transit performance using data envelopment analysis. *Transportation Research Part A: Policy and Practice*, 26(3), 223-230.
17. National Transit Database. (2013). *2012 National Transit Summaries and Trends*. Office of Budget and Policy.
18. Web, Vanpool Safety Brochure, ACT Vanpool Council and ACT Government Affairs Specialist Jason Pavluchuk. Available at: http://actweb.org/wp-content/uploads/2014/12/ACT_Vanpool_safety_brochure.pdf

Assessment Methods from Around the World Potentially Useful for Public Transport Projects

*Eric Bruun, Consultant, Tuup, Ltd
Marianne Vanderschuren, University of Cape Town*

Abstract

The financial crisis in 2008/2009 has made many countries aware that public funds need to be managed better. A logical step is to identify methods that thoroughly assess the impacts of infrastructure and service investment. The authors argue that assessments should both identify and use performance indicators related to the 'triple bottom line', i.e. economic, social and environmental impacts.

This paper provides an overview of assessment methods used to evaluate public transport investments. Positive and negative aspects of various assessment tools are identified and discussed. Some developing world examples appear to be more elaborate and appropriate, than developed world examples, including examples from the United States (U.S.). Although the authors conclude that all methods/tools have challenges, they are of the opinion that a broad and inclusive assessment of public transport investment projects is a must and that the narrowly analyzed and ad-hoc investments witnessed around the world should be avoided.

Keywords: *Public transport, assessment, triple bottom line, project evaluation, multicriteria evaluation*

Background

For millennia, mankind has planned and implemented mega projects. The Ancient Egypt built pyramids, the Roman Empire was famous for its roads and, in Africa, Cecil Rhodes planned and partly implemented the 'Cape to Cairo Railway', that was meant to cross Africa and connect the south to the north.

In modern times, cities are large and have an ever growing number of people (and goods) that need be moved so that the urban economy keeps going. Traditionally, Cost-Benefit Analysis (CBA) was used to assess the implementation of these (public)

transportation systems. However, various authors have identified challenges around the use of CBA (Howarth, 2003; Vanderschuren et al., 2008; Salling and Leleur, 2011; Beukers et al., 2012).

Recognizing externalities related to the movement of people (and goods), the environment and social benefits of sustainable transportation (i.e. public transport and non-motorized transport) have long been identified – the triple bottom line (economy, environment and society) should be included (UITP, 2009). Public Transport (PT) increases mobility for all, improves economic productivity, improves quality of life, reduces car dependency and, consequently, the need for highway expansions (UITP, 2009). The success of PT, however, is dependent on the attractiveness of the service (Lo et al., 2010).

Assessment techniques are used extensively around the world. Henning et al. (2011) established a benchmarking tool for monitoring sustainable transport in New Zealand, a national level example. As will be elaborated upon later, there are also efforts to further develop or refine existing techniques from development banks, cooperative research organizations and NGOs. The literature also elaborates on the appraisal of user experiences and quality of service (Phillips and Guttenplan, 2003; Hawas et al., 2012; Olivková, 2015; Schiefelbusch, 2015; Godavarthy et al., 2015), the assessment of routes and networks (i.e. Banai, 2006), as well as the appraisal of maintenance investment (i.e. Paterson, 2015).

Although the many different methods tried over the years do aim to carry out a holistic appraisal, including all relevant key performance indicators, in practice it has been proven almost impossible to arrive at a totally reliable and widely accepted single value for PT investment projects (Cascajo and Monzon, 2014). Various countries, therefore, opt to use multiple methods, including a vast variation of indicators. This paper elaborates on PT project assessment methods used around the world and it reflects on the usefulness of these methods.

The Need for Ex Ante Assessment of Proposed Projects

Practice often displays a rush about which PT project alternatives to consider, as well as the selection of the preferred alternative. However, due diligence is required. PT projects are often amongst the most expensive public works projects done in a province, region or city. The results of these projects will be in place for decades and have a lasting impact on the daily lives of people, the economy and the environment. Therefore, the decision-makers who spend public money on these projects should be held accountable.

There are multiple considerations when assessing PT projects. The first is that there are always opportunity costs. In other words, the money can be used for something else that is also urgent, which becomes a serious issue, especially in poorer countries. Another consideration is that there is now a heightened need to look at the long-term effects of all investments. It is prudent to try to reduce greenhouse gases and energy dependency, and reduce all other negative impacts on non-renewable and natural

resources. There may also be unintended consequences. So it is also prudent to try to identify and consider these.

While a particular mode may sometimes seem like the obvious choice for a particular context, it needs to be confirmed through a rigorous assessment process. Thus, the purpose of this paper is to provide process guidance, based on practices in a sampling of nations. The next section outlines what the authors believe to be the attributes of a good assessment process and the reasons for their inclusion. The following section highlights project assessment methods used in several nations, including samples from both developed and developing nations with which the authors have personal experience, followed by a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis of the various assessment methods. The final section suggests some improvements.

Attributes of Good Assessment Methods

An overall assessment methodology for projects that are not strictly for private use or benefit, as is the case for almost all PT projects, should combine many attributes. What follows is one list, but not necessarily comprehensive:

- *Inclusive of multiple goals*; Assessments should not ignore any goals that are potentially addressed by a project. The importance of being inclusive cannot be overstated. The list needs to be clearly enumerated so that there are no misunderstandings about what is being considered. All too often analysts reach opposite conclusions about a project alternative because they do not have the same goals in mind.
- *Include multi-faceted goals*; It is essential that positive and negative impacts on the society and environment, at large, are compared to the expected costs and benefits of the project. National and regional goals need to be balanced with local requirements. Furthermore, the effects on a macroscopic, as well as a microscopic, scale need to be included. In order to be able to address multi-faceted goals, it might be necessary to create sub-goals that reflect how well a particular goal is met from the particular perspective of the users, the local community, the region, or the nation.
- *Weights the Relative Importance of the Multiple Criteria*; There is virtually never a situation when all criteria are equally important. Indeed, sometimes one goal is of overriding importance and any other goals that are achieved are incidental. Even more often, a close decision may turn on a particular project excelling in one area where others are weak. Because various project alternatives typically address a particular goal with different degrees of success, it is extremely helpful to be able to identify which goals are most important and which are less. Closely related is the ability to identify conflicting goals so that the trade-offs between them can be considered. An example is maximizing ridership versus minimizing operating subsidy.

- *Requires measurable objectives to monitor progress of project;* Most projects take several years from detailed design to construction. Completion may also take place in phases. Furthermore, it may take some years beyond completion before the full impacts are visible, particularly if some of the goals revolve around land development. Any plan submitted for public funding should specify key performance indicators that measures progress towards objectives that reflect the stated goals of a project. The measurement frequency depends on the goal such that actions can be taken. If insufficient progress is made or progress is made at a much higher than anticipated cost, then a review is conducted and the course of the project may be altered. Even if longer term objective outcomes, like densification, come long after project completion, the trends revealed by a carefully designed measurement program are invaluable to improve future project selection.
- *Uses data that is already available or can be collected;* In the richer, developed countries, data inputs are of central importance, for example, ridership estimates. These, in turn, require samplings of demographic data, traffic flow data, mode split data, origin-destination data, and so on. If not already on hand, an effort is expended to collect this information. This type of data collection, however, will often be unrealistic, if not impossible, in developing countries. In such cases, project objectives must be based on simpler data, such as spot check traffic and ridership counts, and margins of error for forecasts must be increased accordingly.
- *Computationally possible;* Methods that are fine, in principle, but cannot be applied in practice, should be rejected. As an example, the Net Present Value (NPV) method can almost always be used successfully by the private sector, since it concerns itself only with costs and benefits that accrue to itself. But NPV often runs into difficulty in public sector decisions. Some of the costs and benefits are incommensurable and an effort must, therefore, be made to turn them into the common basis of money. Yet, some goals like “improving air quality” can have highly contentious values. How to monetize the value of improved health, improved visibility, or improved liveability, for instance? In such cases, the NPV need not be discarded. Instead, it can be reported as a range of values and supplemented by additional indicators for objectives whose measures are not included in the NPV computation. However, caution is required when using the NPV, as larger projects will always look better, although the benefits per unit of investment might be lower than for smaller projects.
- *Comprehensible to decision-makers;* Even if it is technically and computationally possible to develop a sound decision from a “black box” process, it is highly unlikely that it would be accepted by politicians or board members of a public agency. It requires trust in the institution that performs the process. One decision-maker might believe the results based on the values and perceptions of persons working in the institution, while another quickly rejects them without further investigation, based strictly on opinions of past work. Any good assessment method must reach a balance between computational complexity

required for meaningful analysis and understandability by those outside the circle of analytical experts.

- *Defensible to stakeholders - must identify the incidences of costs and benefits;* The public should participate in project selection, since each potential project alternative must reflect community concerns. In particular, the distribution of benefits versus costs, must be identified. Nonetheless, it is not necessary that every single project provide equal benefits to all those who made financial contributions or who bear some of the costs.
- *Does not contain overly risky assumptions;* There are many unknowns about the future. Thus, using a single forecasted value for key design parameters is risky. For example, the rate of traffic growth (or reduction) depends a great deal on future fuel prices and on the locations where land development concentrates. It is, thus, a good idea to design project alternatives that are robust, meaning that the project viability does not depend on a particular projection. Furthermore, a robust project is such that capacity can be scaled up or down over a substantial range and still be feasible.
- *Is not biased towards the short term (except when the life of a project is short term);* Ever-increasing concern for selection of projects that promote sustainable development means that the longer term impacts must be included. It is clear that the “transportation-economic” method of using the monetary benefit-cost ratio or NPV, as historically used, makes any monetary or monetized costs or benefits virtually zero after about 30 years, regardless of the discount rate used. However, the best way to remove bias remains controversial as consensus has not been reached in either the practitioner or academic communities.
- *National government participation;* Financial contribution by the highest level of government is important, for two reasons. The first is that it provides a justification for placing a set of national goals and objectives into projects across the nation. Without it, regional projects may promote cross purposes. Second, national-level taxes raise far more revenue. Their availability is often a *sine quo non* to even initiate studies. Many regions will be net tax revenue losers for some years, but their turn will come when they initiate projects or their projects reach more expensive phases.

Highlights of Commonly Used Assessment Methods

Germany

One of the European nations that have had a mature and extensive assessment method is Germany. Specifically for PT investments, this country developed a benefit/cost method (B/C-ratio in 1976 (Bundesminister für Verkehr, 1996), that is still used today. The method includes costs and (dis)benefits from a government (national, provincial and local), operator and society perspective.

Federal Government has set financial limits for projects. Small projects are not funded. Furthermore, a province (Bundesland) or municipality requesting national funding, will have to generate a substantial percentage (50%) of the funds themselves.

The Federal Government has put strict guidelines together regarding all calculations, including the demand estimates. Criteria and attribute values are prescribed (and updated over time). This guarantees a fair comparison of large and small projects across the whole country. Furthermore, a government employee reviews investment applications at all stages. In many cases, the person requests additional alternatives to be reviewed. Due to the vast experience built up on a national government level, many project requests have been positively influenced during the funding application.

The Federal Government establishes a priority list on an ongoing basis. Projects that will receive funding are selected from the top of the list, depending on the budget available. Once a project has started, it is added to the multi-year funding plan, which guarantees funding until the project is completed. The method has two *per curial* aspects (although methodologically sound):

1. PT fares are excluded from the calculations, as they are a benefit to the operator but a cost to the user (neutral gains), and
2. The institution applying for funding has the liberty to self-fund aspects of a project, excluding the costs from the prescribed calculations. This improves the B/C-ratio of a project, moving it up the priority list. The first theoretical argument to allow this is that the local institution obviously believes in the project and is willing to invest more than the prescribed percentage. There must, therefore, be local benefits that are not captured in the current prescribed methodology. By allowing the financial exclusion of costs carried locally, this shortcoming of the method can be addressed. A second argument is that it allows a local government to select aspects of a project that are politically most palatable to fund, as opposed to simply asking the electorate to provide a matching fund to a larger, more diffuse project.

Switzerland

The National Government of Switzerland has used a social norm regarding PT investment for many decades. The policy is that all citizens should have access to PT, in particular a rail based connection (Nationalrat Schweiz, 1986). National Government, therefore, provides the infrastructure and operating subsidy (if required) for one train per direction at every location in the country. If a province or municipality wants to provide a higher frequency, they are responsible for the additional costs (required subsidy). The Municipality of Zurich, for example, extends this social norm thinking based on demand figures (see Table 1).

TABLE 1.
 Passenger Demand
 Thresholds to Justify
 Service Levels in
 Zurich

Criteria in Zürich	
Demand (Passengers per day)	Supply (Vehicles)
5 000	1/hour
10 000	1/30 minutes
15 000	1/20 minutes
20 000	1/15 minutes
Etc.	Etc.

Source: Verkehrsverbund Zürich, 1992

The Netherlands

Although clearly less ambitious than Switzerland, the Netherlands also has a social norm regarding PT services. Every settlement (small villages) is serviced by three buses a day (morning peak, around noon and evening peak). The National Government is willing to invest in other PT systems. Applications are accepted once a year and an auditing team assesses all applications and develops a priority list. Calculations in the Netherlands are not standardized, as is the case in Germany, making the prioritization process difficult. Relatively small and large projects are presented using various demand modeling tools, which make varying assumptions and use different project assessment theories. However, standardization suggestions in the early nineties and naughts were ignored.

England and Wales

The United Kingdom (U.K.) Central Government provides rigorous procedures for preparing CBA analyses for all transport projects involving over £20M (U.K. DfT, 2006a). Based on feedback, they modified the procedure to address other costs and benefits difficult to monetize, but they remain a side consideration (U.K. DfT, 2009). These procedures have limited and problematic application for PT projects. In 1986, PT services in the U.K. were privatized. Outside the London Metropolitan region, PT is unregulated and is supposed to run on a commercial basis. Operating subsidies are permissible for lower levels of government, for services, such as rural access and mobility for the disabled. Megaprojects, such as CrossRail, a new 113 km long network across London, including a 21 km twin tunnel, would be inconceivable without Central Government support. But, in general, capital support is rare away from London, other than the Network Rail (including some regional railway lines), which is a government owned entity.

Privatizing of rail services raised complex organizational issues. Sharing of rail infrastructure between operating franchises, via coordinated timetabling, regulates services. However, sharing between competitors caused serious problems in its earlier years (Watson, 2001) and can be expected to be problematic any time that a line capacity is approached.

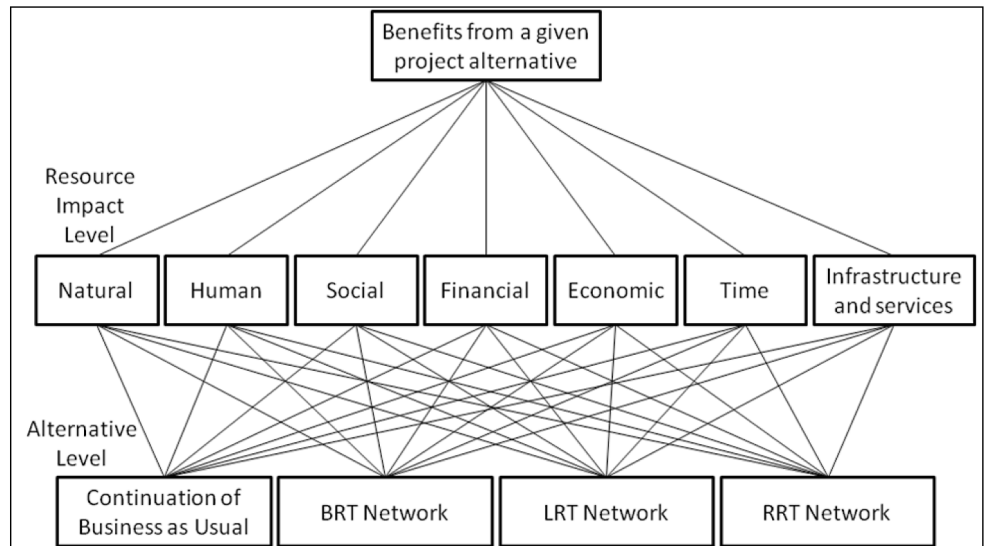
Should cities and regional planning bodies outside London (known as Passenger Transport Executives and Passenger Transport Authorities) decide that a major investment in other infrastructure projects is warranted, they are supposed to first try to raise any needed external financial investments through commercial sources, in a program known as the Private Finance Initiative, rather than ask the Central Government. This hurdle and other Governmental challenges to attempt creating integrated bus-rail services on competitive grounds have limited PT development. The U.K. government itself recognizes that there is serious public dissatisfaction with the investment decision process (U.K. DfT, 2006b; 2012).

Denmark

With the support of the Danish Government, researchers at the Technical University of Denmark, and their consortium partners, including on the Swedish side of Copenhagen/Malmö conurbation, have in the last few years, developed some project alternative assessment and selection tools that address many of the gaps and inadequacies of the historically used methods.

One of the new tools they have developed is Ratio Estimation of Non-Dominated Alternatives (REMBRANDT). It is analogous to the Analytic Hierarchy Process (AHP), a technique for including both quantitative and qualitative factors in the assessment using a similar scale for all. These include both positive impacts and negative impacts, for which the relative importance of each is established by decision-makers through a weighting procedure. Figure 1 shows four different projects (alternatives level) each of which has five different goals (impact level) which, together, form the total benefit from a project. Not shown is a similar hierarchy of costs or negative impacts from these projects. An individual is asked how important or preferred each goal is, relative to each of the other goals, on a scale from 1 to 9. They are further asked how effective each alternative is, relative to each of the other projects in helping to meet this goal, again on a scale of 1 to 9. In fact, in some cases, actual monetary costs can allow objective ranking on the scale. At the opposite extreme, such as how aesthetic an alternative is, the ranking is highly subjective.

FIGURE 1.
AHP Hierarchy
for Benefits from
Hypothetical Citywide
Transit Project



Several project alternatives are analyzed, for their relative effectiveness in maximizing benefits and minimizing costs, by experts. The selected project has the highest ratio of weighted benefits to weighted costs. Although highly intuitive and appealing, it has a few flaws. Some of the practical problems with, and theoretical objections about, AHP have been addressed through a different scoring system. The use of a geometric averaging calculation instead of a weighted linear combination determines the rankings of the various alternatives. Roberts and Goodwin (2002) found that relative weighting of the effectiveness of project alternatives towards meeting a certain goal or objective often tends to follow specific patterns. By using predetermined tables, much of the substantial, and often tedious, efforts needed to establish weights by interviewing individual experts, can be reduced or even eliminated.

The REMBRANDT tool is included in a package of assessment tools known as the SIMDEC (risk Simulation and Multi-criteria decision analysis in combination for DECision support approach). The monetary cost-benefit analysis results are presented in terms of a probability that a specific project alternative will achieve a B/C-ratio greater than 1, resulting in a relative ranking of alternatives. Each important cost and benefit element in the analysis is fitted with a probability distribution to address the uncertainties associated with major infrastructure projects (Salling and LeLeur, 2011; 2012). If an alternative does not have a sufficiently high probability of being feasible, then it can be discarded, even before anything else is considered.

At a decision conference, a software package was demonstrated and applied, showing how a particular individual's own weights of importance of goals varies from others and whether using their individual results alone, instead of the average, would make any difference in project selection outcome. This technique might satisfy public laws and regulations that require the use of CBA, as well as individuals and institutions that prefer Multi-Criteria Analysis (MCA).

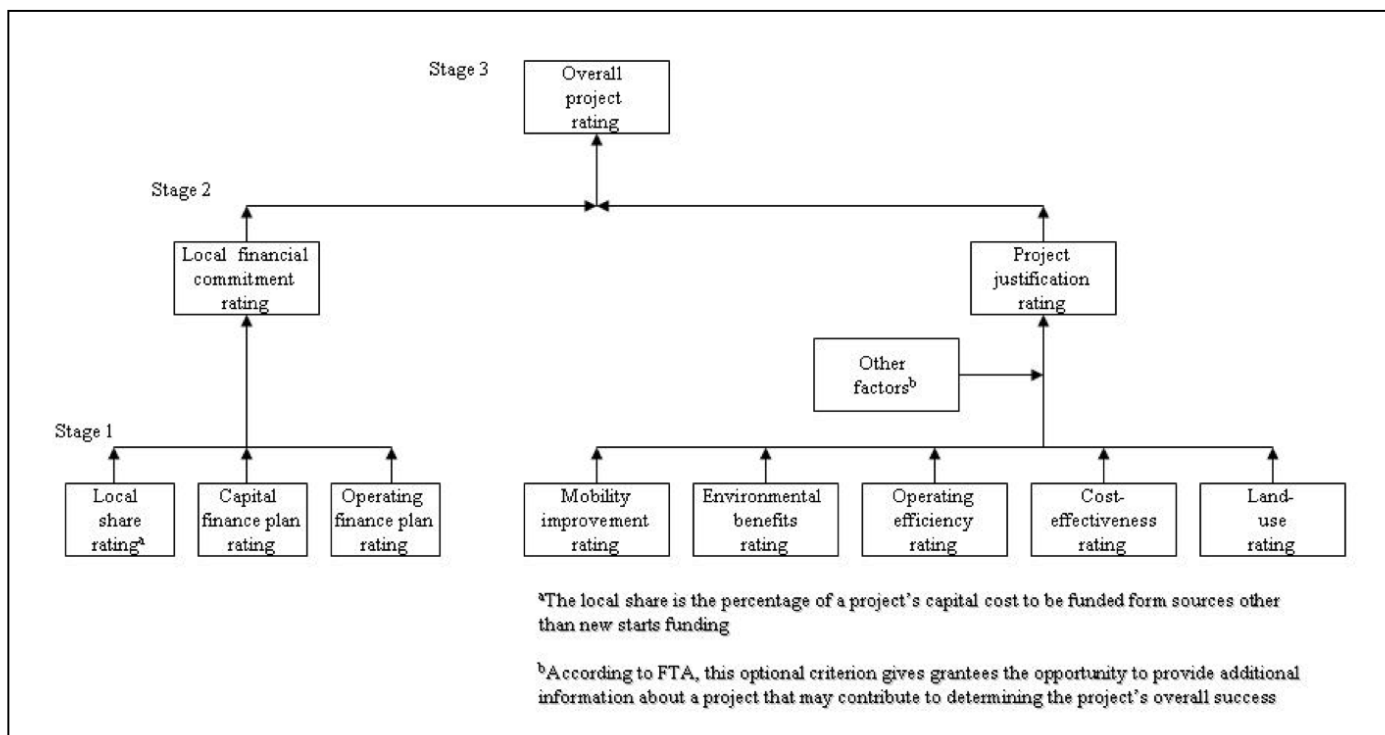
"It was applied to a real case that concerns an additional crossing between Denmark and Sweden, since the bridge/tunnel that opened in 2000 is already nearing capacity. For more details about this, part of the larger project is known as Oresund EcoMobility, see

Jensen et al. (2012). For a brief explanation of how the decision conference worked, in this case, see the related presentation (Oresund EcoMobility, 2011).

The research project entitled “SUSTAIN for National Transport Planning (2012-2016)” developed what is now known as the SUSTAIN Appraisal Framework. It provides yet another interesting case study showing the ongoing evolution, involving the Trans-Baltica Railway corridor, a recently approved high speed rail project linking Tallinn with Warsaw that has huge socio-economic implications (Jensen et al., 2013).

United States of America

The official U.S. Federal Government assessment procedure for major new projects, “New Starts”, used multiple criteria, as can be seen in Figure 2.



Source: U.S. General Accounting Office, 1999

FIGURE 2. The Federal Transit Administration “New Starts” Assessment and Rating Process

There were two basic categories of assessment:

- The first is whether the local/regional plan includes a credible means of covering its matching fraction of the capital costs, as well as the entire Operating and Maintenance (O&M) subsidy that may be required.
- The second category is about the cost versus benefit justification of the project. The ‘Mobility Improvement’ rating reflects time savings to existing users, improved travel for new users and for users of other modes. Environmental

Benefits rating includes air, water quality, noise reduction and other identifiable impacts. Land Use rating includes densification, farm land and wetland conservation/restoration and other identifiable impacts. Cost Effectiveness rating includes indicators such as: Investment cost per new passenger attracted and Construction cost per unit distance. Operating Efficiency rating excludes the capital element and focuses on indicators like O&M cost per passenger.

The enabling legislation to create this procedure dates from the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991. It took almost 8 years to finalize. It remains largely the same today, although some streamlined procedures were later enacted for “Small Starts,” in recognition that this highly involved and expensive process was incommensurate in effort with the impacts and cost of minor PT projects.

The assessment process and a recommendation rating are performed by Civil Service staff in the Federal Transit Administration. Although the procedure was certainly designed with the expectation that rail projects would dominate these submissions, it is mode neutral. The actual mode(s) used are decided by the applying entity. It is up to the regional or local authority requesting funding assistance to do its own pre-screening. In the end, one “Locally Preferred Alternative” is submitted. Because of the federal assessment procedure, regions were pressured into some conformity with other regions, with respect to their project justification. The competition for funding is intense and any criterion that might receive a poor review, relative to its peers, stimulates the search for project features that will offset it by outstanding performance in other criteria.

The actual details of the regulations and the authorized funding, for which the applicants compete, is set by multi-year legislation, but the actual level the Congress appropriates, in any given year, may be lower. The SAFETEA-LU Act covered 2005 until 2009. This was followed by 10 extensions continuing funding and programs at existing levels. Then the MAP 21 Act covered the years 2012 and 2013 with some significant changes (AASHTO 2014). A new Act passed with similar funding levels. Many projects at the State level were put on hold during the wait, to the detriment of the construction industry. Such regular delays and changes, and even cancellations to some of the authorized programs from the previous Act, also complicate the assessment process. In December 2015, the Fixing America’s Surface Transportation (FAST) Act passed. This is a 5-year authorization so it will lend some stability to both the highway and PT industries and to planning at lower levels of government. But there are few significant changes that affect PT assessment or decrease the intense competition for still-limited funding.

In principle, the official U.S. “New Starts” procedure has a lot of merit. Unfortunately, the application of the procedure, over time, has varied. Between 2001 and 2009 the Federal Transit Administration chose to focus only upon the one performance indicator, “net transportation system user benefit”, which is capital plus operating cost divided by time saved for both, private vehicles and PT. One year into the next presidential administration, intense lobbying by proposal submitters, restored the method developed under ISTEA (Bell, 2010). The lack of consistent use of an assessment procedure, despite it being legally mandated, reveals that a government might be free to ignore its own rulemaking. Furthermore, political clout by specific lawmakers often

steered “earmarked” money to favored districts and projects, clearly reducing the objectivity of the procedure. This is in sharp contrast with the long lasting, efficient method adopted and applied in Germany.

Canada

Canada is an exception amongst the richer, industrially developed countries in that, until about 2005, the Federal Government gave virtually no financial contributions towards PT projects, leaving it up to the provinces and regional/local governments (CUTA, 2011). The result has been a quite wide variation in the level of investment across the country, in general, and a significantly older fleet of vehicles than is seen in its U.S. neighbor.

On the other hand, Canadian cities tend to be more amenable to PT than U.S. cities. The sprawl has not been as severe, due to different tax policies and tighter land use controls, and despite continued population growth. Plus, cities are not as racially and economically segregated, reducing the perceptions of risk and stigma commonly attached to PT in the U.S. This has allowed Canada to operate with higher operating cost recovery factors, in general, than U.S. cities of the same size. In some instances, Toronto for example, under a parsimonious provincial government, some lower operating ratio lines were closed, reducing operating deficits but also service levels. Operating cost recovery rose to between 80% and 90%. Go Transit, the regional train service, also had a 90% cost recovery. These are phenomenally high numbers by North American (and even World) standards.

PT support remains sporadic and unpredictable for project planning purposes. In 2003, a substantial but one-time capital assistance of \$CAN 300M was promised from the Federal Government for development of the new Canada Line rail corridor in Vancouver. This was combined with provincial funding to accelerate the investment schedule.

Expansion in Calgary and Edmonton of their Light Rail Transit (LRT) systems was quite slow for decades. In 2010, the Green Trip incentive fund was created by the Province of Alberta. It initially provided \$CAN 2B that could be used for expansion anywhere in the province. It continues today with periodic application deadlines (Alberta Government, 2016). But Alberta remains a special case, due to its oil wealth and, consequently, is much less dependent on the Federal Government.

In Ontario, the same promise of forthcoming federal funding, as seen in Vancouver, gave new impetus to system investment. A new Bus Rapid Transit (BRT) system, Viva, began construction in 2003 and opened in 2005 in York (one of the member cities in the Greater Toronto Area regional planning body). In 2009, a further 37 kilometers of BRT began construction, as well as a 14 kilometer long LRT line (Ministry of Transportation (Ontario), 2009). Since then, an extension of a subway line through Vaughan to York is to be completed by the end of 2017, receiving only about one-third of capital financing from the Federal Government. A three-kilometer spur from a regional line to Pearson International Airport was funded on the condition that it cover its own expenses, thus, requiring a special branded service with a very high fare. It opened in June 2015 but demonstrated very low ridership so the fare was slashed by more than 50% within six

months (CBC News, 2016). A large expansion of the streetcar network within Toronto proper, and of the Go Transit regional rail systems, are also in the first stages of design but funding sources are not confirmed.

Montreal is where the consequences of the lack of adequate Federal support can, perhaps, best be seen. Competing uses of limited combined funds over the last decade included choosing between needed extensions of the Metro and additional regional rail lines to improve the limited area coverage, including Trudeau International Airport, versus replacement of antiquated rolling stock and equipment. Despite a new plan to inject \$CAN2.5B over 2015 to 2017, the deferred maintenance and replacement backlog is so huge that the life of some of the metro stock will be extended to more than 60 years (Progressive Railroading, 2014).

The lack of a guarantee of a certain level of funding from the Federal Government, for which the provinces and cities can compete if a formal process is followed, remains an obstacle to long-term investment planning in many urban regions across Canada. It is also clearly contributing to PT systems that are inadequate to cope with population growth. However, after ten years, in 2015 a newly elected national government promised to place a higher priority on environmental concerns, and the entire assessment and funding procedure is now undergoing major revision.

South Africa

South Africa has a policy and legislative environment that on a national, provincial and local level encourages the integration of all modes and, in fact, prioritizes the needs of PT and non-motorized transport above that of the private automobile. The Act governing land transport in South Africa is quoted as saying (Republic of South Africa, 2000):

“For the purposes of land transport planning and the provision of land transport infrastructure and facilities, public transport must be given higher priority than private transport. Public transport services, facilities and infrastructure must be so provided and developed as to integrate the different modes of land transport”.

Due to the lack of all mode inclusive guidelines, current practice relies on old assumptions and, for many engineers, it is business as usual. In some cases, the engineering firms suggest alternative practices to provincial and local government (the implementers of PT projects). An example is the assessment of improved services for the N1 corridor between Bellville and Cape Town. The project team identified nine different alternatives for PT provision ranging from ‘Do Nothing’, moderate upgrades of current bus and rail services, to the development of a Busway (Vanderschuren et al., 2008).

Traditionally, Cape Town used CBA for the assessment of transportation projects. The First Edition of the Guidelines for Conducting the Economic Evaluation¹ of Urban Transport Projects was issued in June 1992 after input from several stakeholders and practitioners. Since then, two reviews have taken place. In May 2002, the city adopted the Third Edition (CCT, 2002). This version allows for the possibility to assess road

¹ = Assessment

and PT infrastructure investments, as well as interchange facilities. Criteria included are: income distribution considerations, regional developmental benefits (economic developments) and environmental considerations (integration of Environmental Impact Assessment (EIA) requirements).

In order to have confidence regarding the suggested implementation suggestion, the project team decided to step away from the CBA and apply MCA techniques. Different MCA methods have been developed during the last 30 years to support decision-makers facing conflicting decision situations. An MCA aims to rank or score a finite number of decision options based on a set of assessment criteria. The number of MCA techniques, relevant to developing countries, has increased rapidly over the past several decades (for a recent review see Figueira et al., 2005). They provide practically limitless options for combining weights information with the assessment matrix to attain a result (Hajkowicz and Higgins, 2008). In the Busway example, two different MCA techniques, as well as two different weightings, were applied. The final results are provided in Table 2. As can be seen from the results, the final score can be negative. In the case of a negative score, the overall impact of the alternative will lead to reduced benefits. Implementation is, therefore, not recommended.

A final theory that has been explored in South Africa is the Sustainable Livelihood Approach (SLA). The combined contributions to neighborhood theory of Howard (1898), Perry (1929), Stein and Wright (1929), and Fisher (1984) have provided a rich basis for devising a set of criteria for the planning of sustainable neighborhoods: economic, social, technical and environmental sustainability (Choguill, 2008).

TABLE 2. Summary of the MCA Results for the N1 Busway in Cape Town

		Alternative											
		0	1	2	3	4	5	6a	6b	7	8a	8b	9
Method	Weight	Do nothing	Upgrade rail	Lane Balance to Freeay Lanes	Lane Balance & Rail Upgrading	BMT Lanes & Lane Balance	BMT Lanes, Rail Upgrade & Lane Balance	Busway & Lane Balance	Busway & Lane Balance (ticket R3)	Busway, Rail Upgrade & Lane Balance	Tram Train & Lane Balance	Tram Train & Lane Balance (ticket R3.5)	Road Pricing & Lane Balance
Weighted sum	Proposed	0.050	0.014	0.165	0.127	0.210	0.140	0.497	0.514	0.417	0.387	0.447	0.356
	Initial	0.039	-0.011	0.137	0.084	0.180	0.095	0.431	0.454	0.329	0.301	0.381	0.296
EVAMIX	Proposed	-3.469	-3.848	-1.485	-1.892	-1.125	-1.967	2.981	3.191	2.027	1.841	2.561	2.086
	Initial	-2.910	-3.457	-1.155	-1.737	-0.830	-1.850	2.757	3.037	1.542	1.316	2.276	1.811

Source: Vanderschuren et al., 2008

Scoones (1998) suggests that the SLA is useful as an umbrella to categorize criteria and attributes, because it acknowledges that, particularly in poor communities, people gain their livelihoods through multiple activities rather than one formal job. In contrast to the previous environment and development thinking aimed at sustainable development, Sustainable Livelihoods is a people-centered paradigm, which emphasizes people's inherent capacities and knowledge and is focused on community level actions (Chambers, 1986; UNDP and Wanmali, 1999). According to Chambers and Conway (1992) (cited in Scoones, 1998):

A livelihood comprises the capabilities, assets (including both material and social resources) and activities required for a means of living. A livelihood is sustainable when it can cope with and recover from stresses and shocks, maintain or enhance its capabilities and assets, while not undermining the natural resource base.

In South Africa, the SLA has been adapted by Barbour and Kane (2003), to include seven resources, which have been included in assessment tools in South Africa and Lesotho (Sah et al., 2008).

The resources applied are:

1. Natural resources, such as fuel consumption and land-use quality;
2. Human resources, such as access to basic services;
3. Social resources, such as neighborhood communication and emergency transport;
4. Financial resources, which includes all the costs related to a project;
5. Economic resources, such as job creation and entrepreneurship;
6. Infrastructure and services; and
7. Time, which includes free time available and independency of children to walk to school without adult supervision.

The move towards MCA and the SLA in South Africa, as demonstrated by various studies and authors has, unfortunately, not led to a change with regards to legal requirements.

Lesotho

The Kingdom of Lesotho, a land-locked country of about 30,500 km², with 2.2 million inhabitants, has recently adopted a SLA project assessment tool based, partly, following the South African example but taking the approach to new heights. Lesotho is mountainous over approximately two thirds of its area, where population densities are low. Economic development and survival is a challenge in these areas and, in winter, families are often divided for months as shepherds take their cattle to central points while leaving their wives and children behind. In a case of emergency, access to clinics and hospitals is often not available. Thus, improved rural public transport is important.

In the literature, the number and type of indicators (criteria), the difficulty of measuring indicators, as well as a lack of data, were identified as technical barriers (May et al., 2008). The question, therefore, arose on how to overcome these challenges in the Lesotho context. An in-depth analysis of government documents and local studies, as well as a review of data sources, led to the adoption of the seven resources identified previously in the South African case. In the final tool, a total of 81 attributes were included, clustered under the seven resources. The attributes were, in some cases, identical to the South African ones. However, there were others, such as soil erosion and access to agricultural land for urban dwellers that were specific to the Lesotho case. Furthermore, All 81 attributes were benchmarked for 8 different geographical area types. In this way, both rural and urban needs can be balanced.

The Lesotho Ministry of Public Works and Transport trusts that the newly developed tool will provide the balance between accessibility issues in rural areas and the need to promote economic development in urban areas. Lesotho Government, as well as the World Bank (sponsor of the project), scrutinized and accepted the tool. The six month practical trial period assisted in fine tuning local requirements.

Other international Assessment Approaches

The concern for improving assessment procedures is widespread, especially to better take into account longer term sustainability impacts. In 2008, The World Bank issued a policy paper that suggests that it will use a revised assessment method on future urban transport projects. It argues for evaluating such projects, not on a separate basis, but as to how they fit into a “core strategy” that supports broader development goals (Mitric, 2008). Once an initial plan is submitted for funding a particular project, it would be subjected to a set of criteria wider and more long term than simply the readily identifiable and direct economic development benefits. Some of the strategy elements include:

- All transport modes are within the strategy reach;
- Preference for PT modes as the main response to motorization, even when road investments are involved;
- Explicit concern for low-income and poor travelers and communities in both policy and investment dimensions;
- Emphasis on exclusive on-street space for PT trunk lines;
- Preference for public-private partnership (competition-for-the-market regulatory model), with service levels, fares and environmental parameters in public hands; and
- Focused institutional capacity to regulate the PT supply and manage traffic.

Note that this implies an MCA, but the World Bank did not specify that any particular MCA must be used. No doubt loan recipients will be able to submit and defend their own approach appropriate to their particular goals and the level of quantitative information that they have available.

The Asian Development Bank (ADB) has also made strides in developing a more comprehensive assessment policy that guarantees more consideration of the wider benefits of PT relative to auto-centric projects and whether there are supportive public policies like demand management and land use controls (ADB, 2009).

There are also NGOs offering advice on how to improve the design and selection of urban transport projects. For example, the Sustainable Urban Transport Project (SUTP) examines case study cities and distills some of the best practices into a recommendation for a detailed Sustainable Urban Mobility Plan planning cycle (SUTP, 2014). The Victoria Transport Policy Institute provides suggestions on a better overall approach to multi-modal transport project evaluation process based on accessibility instead of mobility (VTPI, 2016a), as well as a “Best Practices Guidebook” focused on a rigorous approach to determining PT benefits and costs in a CBA Analysis.

There are governmental organizations sponsoring research on improving the process as well. For example, the EVIDENCE project of the European Commission (EC) has multiple reports aimed at justifying the values used in CBA and other analyses. One, in particular, comes to the interesting conclusion that there is a list of measures like bike lanes, fare integration, better fare payment technology, real-time passenger information, and so on, that should be done anyway despite (so far) insufficient evidence to economically justify them on their own simply because they are likely to enhance the attractiveness of PT (Shergold and Parkhurst, 2016). Another EC-sponsored project, TIDE, offers a very helpful handbook on how to rigorously incorporate the new wave of transport innovations as possible project alternatives themselves and/or as elements in CBA, as well as an introduction to MCA methods (TIDE, 2013).

Space prohibits discussion of the numerous official methodologies based on CBA in use by governments at all levels. But of particular note is an official guideline from New Zealand, which includes a requirement entitled “Comparison of overseas best practice sustainable development impact assessment” (Abley, et al., 2010). The comparison, unfortunately, is limited to English speaking countries but can serve as a template for a wider comparison of nations that use CBA supplemented with additional factors and/or MCA. Finally, Holian and McLaughlin (2016) provide a thorough analysis of how the State of California CBA procedure compares with its peers and some local governments and suggests how it might be modified for multi-modal applications.

Highlights from the International Assessment Review

Every MCA assessment method appears to have advantages and disadvantages. The literature has not come to a conclusion with regards to a preferred method. There are two main schools of thought regarding MCA. The first unifies scores across alternatives, applies a weighting and sums the result per alternative. The second school of thought takes the comparison a step further. After the unification of scores, weighted alternatives are compared pair-wise. It is important to note that different assessment methods might lead to different conclusions. It is also important to note that there isn't even consensus that CBA should be replaced with MCA, especially as CBA methods are including more and more environmental and social criteria.

In the N1 Corridor project (Vanderschuren et al., 2008) it was decided to use two significantly different methods: The Weighted Sum method (appealing to the first school of thought) and the Evamix method (appealing to the second school of thought) (Vermeulen, 1986). The Evamix approach by Voogd (1982 and 1983), and described in Nijkamp et al., (1990) and Martel and Matarazzo (2005), treats data in the assessment matrix differently, depending on whether it is qualitative (ordinal) or quantitative (cardinal). This is an important contribution of Evamix to MCA. Evamix requires cardinal information on criteria weights. Evamix commences by identifying unique pairs of options. It then determines an ordinal and cardinal dominance score (Hajkowitz and Higgins, 2008). Besides applying two different MCA methods, two different weightings were applied, as well. In all calculations, with both methods, it was concluded that this strengthened the decision- making process.

Some countries, such as Switzerland and, to a lesser extent, the Netherlands, use social norms that guarantee a basic service level of PT for all. Other countries, such as Germany and the U.S., have a long history in using traditional assessment methods (CBA) that are based on a conversion of criteria into monetary terms. One very positive aspect of the process in these countries is the transparency provided. All proposers of new projects use the same set of criteria and values. At the same time, with the enhancements added over the years, the process provides enough flexibility so that circumstances, that are special to a particular location, can be explained and accommodated.

In the German case, it has proven to be extremely valuable to have continued funding guaranteed for a started project, as well as the scheduling of a project in a medium-term funding plan. Furthermore, the participation of government representatives during all planning stages has increased transparency and comparability of projects country wide.

In some cases, described weaknesses were country specific. In other cases, they were related to the method itself. Table 3 provides an overview of the comprehensiveness and, to some extent, advantages and disadvantages of different explored methods.

MCA and, subsequently, SLA were developed to address shortcomings in existing methods. The fact that disadvantages are reduced in more recently developed methods indicates that they are successful in addressing shortcomings. Although MCA have proven to be useful transport assessment tools, it can be concluded that the SLA is the most holistic and inclusive method described. Specifically, when decision-makers are investing vast amount of taxpayers' money in PT systems that have goals that address the needs of the most vulnerable in the society, decision-making needs to be sound and investment needs to reach as many as possible. This is even more important in developing countries, where financial resources are scarce.

TABLE 3.
Overview of Different
Basic Assessment
Methods

Method	Advantage	Disadvantage	Country of Application
Cost Benefit Analysis (CBA)	<ul style="list-style-type: none"> • Well structured • One dimension • Gives economic result • Sensitivity analysis possible 	<ul style="list-style-type: none"> • Problem converting non-monetary criteria • Many calculations to change everything into a currency • Monetary values for different countries vary severely • Discounts costs and benefits to future generations 	U.S. Germany England Wales Denmark
Multi-Criteria Analysis (MCA)	<ul style="list-style-type: none"> • Qualitative and quantitative dimensions included • Weighting of criteria creates transparency • Possible to handle vast amounts of information • Sensitivity analysis through change of weighting 	<ul style="list-style-type: none"> • No direct sensitivity analysis • Different methods might provide different answers 	South Africa Future World Bank Funded projects
Sustainable Livelihood Approach (SLA)	<ul style="list-style-type: none"> • Provides multiple goal approach • Qualitative and quantitative dimensions included • Weighting of criteria creates transparency • Possible to handle vast amounts of information • Sensitivity analysis through change of weighting 	<ul style="list-style-type: none"> • No direct sensitivity analysis 	Lesotho

The majority of assessment tools reviewed in this paper were computer based. Although a black-box approach should be avoided to guarantee transparency, user friendly systems are valuable to decision-makers. A reduction in processing time, as well as a better understanding by users that have not specialized in the field, is provided.

An ironic result of the elaborate procedure required to receive funds from the U.S. Federal Government is that achievement of environmental goals is, at best, delayed and, at worst, prevented. One reason is that many automobile-based master plans are ‘grandfathered’. They were approved many years ago and cause little delay once developers have financing arranged. By comparison, major PT projects require 3 to 5 years to reach groundbreaking. The answer is not to remove the environmental study requirement from PT studies, but perhaps to incentivize local governments into canceling master plans that do not adequately consider the environmental bottom line.

The lack of systematic and consistent support from the Federal Government of Canada has hampered the ability to assemble large sums by regional and provincial governments. It is reflected in both bus and rail rolling stock that remains in use for much longer than average for developed countries. This, in turn, raises operating and maintenance costs borne at the local and regional levels, particularly with the bus mode,

also reducing attractiveness for choice users. It is also reflected in the lateness, relative to their peers, in which large cities like Montreal and Toronto get rail connections to their major airports and the delay of metro extensions in very high demand corridors, like Toronto to York.

Transferring Best Features of Various Methods to Other Countries

No locality, region or nation can import a complete project assessment method, including criteria and attributes, from other places. There will always be specific issues that must be addressed and existing legal structures that must be respected. Nevertheless, it might well be possible to incorporate many positive features from other places. It is in this spirit that Table 4 has been constructed. It rates the various example nations and the World Bank, discussed in this paper, regarding how well they serve the attributes that the authors’ suggested characterized a good assessment method.

TABLE 4. Summary of Criteria and Rating Attributes at the National Level

Category	Country	G	CH	NL	U.K.	Dk SUSTAIN	U.S.	CA	ZA	LS	World Bank
Costs	Project capital costs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Operational costs	Y	Y	Y	LDN ¹	Y	Y	N	Y	Y	Y
	Maintenance costs	Y	Y	Y	LDN ¹	Y	Y	N	Y	Y	Y
Economic	Job creation	N	N	Ps	N	N	N	N	Y	Y	Y
	Job security	N	N	Ps	Ra	N	N	N	Y	Y	N
	Entrepreneurship	N	N	UI	Y	N	N	N	Y	Y	Y
	Tourism	N	N	Ps	LDN	Ps	Ps	N	Ps	N	N
	Impact on logistics	N	N	UI	N	Y	Y	Ps	Ps	Ps	Y
	Financial wellbeing	N	N	UI	Ra	Y	N	N	Ps	Y	N
Social	Social security	N	N	UI	Ra	N	Y	N	Ps	Y	N
	Equity between regions/provinces	N	Y	L	N	Y	N	N	Y	Y	Y
	Densities and land use	Y	Y	Ps	N	N	Y	N	Ps	Y	Y
	Emergency services	N	N	Ps	LDN	N	N	N	Ps	Y	N
	Disaster management	N	N	Ps	LDN	N	Ps	N	Ps	Y	N
	Communication	N	N	UI	N	N	N	N	Ps	Y	Y
Legend:	Y = Yes N = No LDN = London Ps = Project specific Ra = Rural areas Uk = Unknown UI = Unlikely	E = Excellent VG = Vary Good G = Good F = Fair P = Poor VP = Very Poor L = Likely N/A = Not Applicable	¹ Contracting in London, otherwise socially necessary services only ² national required, local content not allowed ³ pertains to developing countries								

TABLE 4. (Cont'd) Summary of Criteria and Rating Attributes at the National Level

Category	Country	G	CH	NL	U.K.	Dk SUSTAIN	U.S.	CA	ZA	LS	World Bank
Human	Road safety	Y	N	L	Y	Y	Y	Y	Y	Y	Y
	Active transport and health	N	N	N	N	Y	N	N	N	N	N
	Access to basic services	Y	N	UI	Ra	N	N	N	Y	Y	Y
	Access to agriculture	N	N	UI	N	N	N	N	Y	Y	N
	Sense of freedom	N	N	UI	N	N	N	N	Ps	Y	N
	Transport related fear	N	N	UI	N	N	N	N	Ps	Y	Y
	Energy use	Y	N	L	N	Y	N	N	Y	Y	Y
	Air pollution	Y	N	L	LDN	Y	Y	Y	Y	Y	Y
	Noise pollution	Y	N	L	LDN	Y	Y	Y	N	N	Y
	Clean water provision	N	N	Ps	N	N	Y	N	UI	√	N
	Local materials availability	N	N	UI	Y	N	Y ²	Y ²	UI	√	N
	Soil erosion	N	N	UI	N	N	Y	N	UI	√	N
Time	Free time available	N	N	UI	N	N	N	N	Ps	√	N
	Child independence	N	N	UI	N	N	N	N	Ps	√	N
Assessment Method	Multiple goals	VG	N/A	F	P	E	G	VP	F	E	Y
	Relative weighting of goals	Via \$	N/A	L	P	E	F	N/A	F	G	N
	Multi-faceted	G	N/A	L	F	VG	G	Ps	G	G	G
	Measurable objectives (KPIs)	G	G	P	N/A	G	G	N/A	G	G	G
	Practicality of baseline data	G	G	P	N/A	G	G	N/A	G	G	P ³
	Practicality of computation	G	N/A	P	N/A	G	G	N/A	P	E	F ³
	Comprehensibility	G	E	L	N/A	VG	VG	N/A	G	G	G
	Defensibility	G	E	VP	F	E	G	F	G	E	G
	Valid assumptions	G	VG	P	P	VG	F	N/A	G	E	G
	Not biased against long-term	F	E	F	P	VG	F	P	G	E	G
National government participation	E	G	VP	P	E	G	P	P	VG	G	
Legend:	Y = Yes N = No LDN = London Ps = Project specific Ra = Rural areas Uk = Unknown UI = Unlikely	E = Excellent VG = Vary Good G = Good F = Fair P = Poor VP = Very Poor L = Likely N/A = Not Applicable	¹ Contracting in London, otherwise socially necessary services only ² national required, local content not allowed ³ pertains to developing countries								

Two observations are made, about the practical limits of formal assessment procedures, despite their apparent theoretical advantages in more equitable, consistent and inclusive results.

1. The formal assessment assumes cooperative government at all levels, each willing to participate in good faith and abide by the outcomes of the alternatives analyses and public participation (although it needs to be mentioned that public participation, in many instances, could benefit from refinement, including taking better advantage of social media and the internet). An example is the aforementioned SUTP's very logical planning cycle for a Sustainable Urban Mobility Plan. Step 7 is "Agree on clear responsibilities and allocate funding" (SUTP, 2014). In reality, there is a chronic lack of agreement in many jurisdictions.
2. There can be other laws or public policies adverse to laws that dictate a PT assessment procedure. An important feature of an effective procedure is that it mandates an array of alternatives be explored. An adverse law or public policy can take promising alternatives off of the table.

Some Suggestions for Further Improvements

There is a school of thought that the CBA should continue to be the dominant form of evaluation, albeit with continual refinements. The Organization for Economic Cooperation and Development (OECD) recently convened a workshop of CBA experts to discuss the future of transport assessment (Veryard, 2016). Amongst their conclusions, they thought it important to restate this quote (Dobes and Bennett, 2009):

"In MCA, objectives are generally weighted according to their priority for the decision maker (and to an extent their degree of overlap with other objectives). While guidelines are available to support MCA decision processes, economists are generally skeptical of its application due to issues of double counting, arbitrary implicit valuations, lack of viability thresholds, and "gaming" the system".

As for trying to evaluate several factors simultaneously (Dobes and Bennett, 2009):

"Several participants argued that in such instances, it is preferable to instead highlight these relevant impacts qualitatively alongside the CBA result, as is the practice in the U.K. Appraisal Summary, rather than adding a false sense of precision with the integrated result of the MCA approach".

The authors feel that similar accusations can be made about the CBA. This is especially so when forecasts are involved and future costs must be estimated. It is indeed an improvement to add qualitative relevant impacts alongside the CBA result. Thus, MCA should not be abandoned, but seen as valuable extension of CBA, as demonstrated in the Cape Town case, where all traditional CBA values were included in the assessment and in the new Danish SUSTAIN method where a probabilistic form of CBA is included.

The suggested MCA enhancement might come in the form of an AHP, a decision-making method that is already widely used in the marketing profession and in some corporate internal decisions involving major investments. It has shown promise, even for application in less developed countries, but there are many issues to overcome, such as lack of data and experts on needed topics (Khademi et al., 2014). The Danish (REMBRANDT based) method was the only one listed in our sampling that requires the use of anything similar to AHP. Its appeal is that it is intuitive and, thus, can be understood without deep mathematical capability. It addresses the problem of incommensurability of the various goals, as well as the issue of different importance of these goals.

Discounting also needs further refinement. The use of differential discounting between actual money expenditures and monetized values could have a profound impact on project selection. Martin Weitzman (1994) advocates use of discount rates that become smaller as the distance into the future becomes longer. It falls to 1% for 76 to 200 years into the future. Bruun (2013) discusses the possibility of using no, or a very low discount rate, for any monetized environmental costs and benefits while continuing to use the customary discount rate for actual monetary expenditures on civil works and equipment. This is based on the argument that sustainability, in the strictest definition, requires that environmental impacts to future generations be counted as equal to the current.

Finally, the type of delivery method that is appropriate for a particular project remains a contentious subject. This is unfortunate as the entire feasibility and affordability of a particular alternative may depend upon whether and how the private sector participates in the various parts of the lifecycle, from initial funding through the construction phase and daily operation. Tang and Lo (2010) provide an example of three “influence diagrams,” one each for the build, fund, and own decisions. For each of these, there are three possible agents, the railway operator, the government or a property developer. Conditional probabilities are then used in this interaction chain to estimate which agent is likeliest to give the best outcome for each of these decisions. In the spirit of trying to systematize and quantify important relationships between parties to help select the best possible candidate project alternatives, perhaps influence diagrams merit more widespread use.

Concluding Remark

This paper describes a wide range of assessment tools, and implementations from all over the world. Transport project assessment is specific to local policies and conditions. However, the authors are of the opinion that national governments and local authorities can learn from the experience of others. Furthermore, more inclusive assessment methods, such as MCA and SLA and packages that unify qualitative and quantitative factors, like AHP/REMBRANDT, are preferred over traditional CBA alone. For suggested criteria, see Table 4. In practice, analytic MCA and SLA methods would have to condense these criteria into a few key higher level groups.

It is worthwhile to follow current events as new methods actually combine the CBA with MCA, such as the Cape Town case and SUSTAIN. This overcomes many of the shortcomings of the more traditional and simpler methods. They show promise for adaption to a variety of circumstances, particularly in an era where participation at levels - public, professional and elected officials, is becoming easier and more efficient.

This paper does describe positive and negative aspects of various assessment tools. Although many, if not all, methods/tools have challenges, the authors are of the opinion that a broad and inclusive assessment of PT investment project is a must and that the ad-hoc investments witnessed around the world should be avoided.

References

- AASHTO. 2014. "Safetea LU versus Map 21 A Comparison. <http://map21.transportation.org/Pages/Safetealuvmap21.aspx>, accessed 19 September 2014.
- Abley, Steve, Paul Durdin and Malcolm Douglass. 2010. Integrated Transport Assessment Guidelines, Report 422, Land Transport New Zealand. <http://www.nzta.govt.nz/resources/research/reports/422>
- Alberta Government. 2016. GreenTRIP Green Transit Incentives Program: Procedures and Guidelines. <http://www.transportation.alberta.ca/Content/docType5409/Production/GreenTRIPProceduresandGuidelines.pdf>
- ADB (Asian Development Bank). 2009. Changing Course: A New Paradigm for Sustainable Urban Transport. <https://www.adb.org/sites/default/files/publication/27552/new-paradigm-transport.pdf>
- Banai R. 2006. Public Transportation Decision-Making: A Case Analysis of the Memphis Light Rail Corridor and Route Selection with Analytic Hierarchy Process, *Journal of Public Transportation*, Volume 9, Issue 2, pp. 1-24.
- Bell, John R. 2010. "LaHood Announces Adoption of Long-Sought 'New Starts' Criteria," *Passenger Transport*, January 18.
- Beukers, E., L. Bertolini and M. Te Brömmelstroet. 2012. Why Cost Benefit Analysis is Perceived as a Problematic Tool for Assessment of Transport Plans: A Process Perspective, *Transportation Research Part A*, Vol 46, pp 68–78.
- Bruun, E.C., 2013. *Better Public Transit Systems: Analyzing Investments and Performance*, 2nd Edition, Chapter 8, Oxon, U.K.: Routledge.
- Bundesminister für Verkehr. 1996. Anleitung für die Standartisierte Bewertung von Verkehrswegeinvestitionen des ÖPNV und Hinweise zum Rahmenantrag Schriftreihe des Bundesministers für Verkehr, Heft 51 (German).
- CBC News, 2016. Union Pearson Express fare slashed from \$27.50 to \$12: Airport train has struggled to make ridership goals since summer launch, Feb 23. Accessed 8 October 2016. <http://www.cbc.ca/news/canada/toronto/union-pearson-express-price-1.3459482>

- CUTA (Canadian Urban Transport Association). 2011. Federal Provincial and Territorial Public Transit Funding Programs in Canada.
- Dobes, L. and J. Bennett. 2009. Multi-Criteria Analysis: "Good Enough" for Government Work? *Agenda: A Journal of Policy Analysis and Reform*, Vol 16, Issue 3, pp. 7-29.
- Figueira, J., G. Salvatore and M. Ehrgott, Editors, 2005. *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, New York.
- Godavarthy R.P., J. Mattson, D. Peterson and J. Hough. 2015. Developing a Method for Assessing National Demand-Response Transit Level of Service, *Journal of Public Transportation*, Vol 18, Issue 4, pp. 1-15.
- Hajkowicz S. and A. Higgins, 2008. A Comparison of Multiple Criteria Analysis Techniques for Water Resource Management. *European Journal of Operational Research*, Vol 184, Issue 1, 1 January 2008 pp. 255-265.
- Hawas Y.E., B. Khan, N. Basu, 2012. Evaluating and Enhancing the Operational Performance of Public Bus Systems Using GIS-based Data Envelopment Analysis, *Journal of Public Transportation*, Vol 15, Issue 2, pp. 19-44.
- Holian, M. and R. McLaughlin. 2016. Benefit-Cost Analysis for Transportation Planning and Public Policy: Towards Multimodal Demand Modeling. MTI Report 12-4, Mineta Transportation Institute. San Jose, CA. USA. <http://transweb.sjsu.edu/PDFs/research/1203-benefit-cost%20analysis%20for%20transportation-planning-and-public-policy-toward-multimodal-demand-modeling.pdf>
- Howarth, R.B., 2003. Discounting and Sustainability: Towards Reconciliation, *International Journal of Sustainable Development*, Vol 6, issue 1, pp. 87-97.
- UITP (International Association for Public Transport). 2009. Assessing the Benefits of Public Transport, A UITP Position Paper, pp. 1-6.
- Jensen, A.V, I. Ambrasaite, K.B. Salling, M.B. Barfod and S. Leleur. 2012. "The EcoMobility Modelling Framework for Sustainable Transport Planning." Chapter 9 in *Rethinking Transport in the Øresund Region: Policies, Strategies and Behaviours*, edited by C-M. Carlsson, T. Emtairah, B. Gammelgaard, A.V. Jensen and Å. Thidell. Lund, Sweden: Lund University, pp. 149-164.
- Jensen, A.V., K.B. Salling, and S. Leleur. 2013. "The SUSTAIN Appraisal Framework: Flexible Decision Support for National Sustainable Transport Planning" extended version of paper presented at WCTR Conference, Rio de Janeiro, July.
- Khademi, N., K. Behnia, and R. Saedi, 2014. Using Analytic Hierarchy/Network Process (AHP/ANP) in Developing Countries: Shortcomings and Suggestions, *The Engineering Economist*, Vol 59, pp. 2-29.
- Lo, H.K., W.H.K. Lam, S.C. Wong and J.M.Y. Leung. 2010. Editorial: Advanced Methods for Public Transport Management, *Public Transport*, Vol 2, pp. 1-3, DOI 10.1007/s12469-010-0025-6.

- Martel J.M. and B. Matarazzo. 2005. Other Outranking Approaches. In: J. Figueira, G. Salvatore and M. Ehrgott, Editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, Springer, New York, pp. 197–262.
- Ministry of Transportation (Ontario). 2009. VIVA Rapid Transit Construction Underway: McGuinty Government and Partners Adding More Transit In York Region, December 14. http://www.metrolinx.com/Docs/News/NR_York_Rapidways_COPO_EN151209.pdf.
- Mitric, S., 2008. *Urban Transport for Development: Towards an Operationally-Oriented Strategy*, Washington DC: World Bank.
- Nationalrat Schweiz. 1986. Bundesbeschluss über den Leistungsauftrag 1987 an die schweizerischen Bundesbahnen und über Abgeltung ihrer gemeinschaftlichen Leistung, Art. 4, (S). Nijkamp, P., P. Rietveld and H. Voogd. 1990. Multicriteria Evaluation in Physical Planning, North Holland, Amsterdam, pp. 65–100.
- Olivková, I., 2015. Model for Measuring Passenger Satisfaction and Assessing Mass Transit Quality, *Journal of Public Transportation*, Vol 18, Issue 3, pp. 52-70.
- Paterson L. 2015. Evaluating the Regional Benefit/Cost Ratio for Transit State of Good Repair Investments, *Journal of Public Transportation*, Vol 18, Issue 3, pp. 15-28.
- Phillips R.G. and M. Guttenplan. 2003. A Review of Approaches for Assessing Multi-Modal Quality of Service, *Journal of Public Transportation*, Vol 6, Issue 4, pp. 69-87.
- Progressive Railroading. 2014. Montreal Transit Agency to Slow Phase-In of New Subway Cars. 30 October 2014.
- Roberts, R. and P. Goodwin. 2002. Weight Approximations in Multi-Attribute Decision Models. *Journal of Multi-Criteria Decision Analysis* Vol 11, no. 6, pp. 291-303.
- Sah, B.P., Choi, J. Sentihil, S. Babu B.M. and M. Vanderschuren (2008), Geographic Information Systems based Decision Support System for Transportation Planning in Maseru, Lesotho, Africa, 21st EAROPH World Congress and Mayors' Caucus, Oct. 21-24, 2008, Japan.
- Salling K.B. and S. Leleur. 2011. Transport Appraisal and Monte Carlo Simulation by use of the CBA-DK Model, *Transport Policy*, Vol 18, pp. 236–245.
- Salling, K.B. and S. Leleur. 2012. "Modelling of Transport Project Uncertainties: Feasibility Risk Assessment and Scenario Analysis" *European Journal of Transportation Infrastructure Research* Vol 12 no.1: 21-38. <http://www.ejtir.tbm.tudelft.nl>.
- Schiefelbusch, M. 2015. Analyzing and Assessing the Experience of Traveling by Public Transport, *Journal of Public Transportation*, Vol 18, Issue 4, pp. 46-72.
- Shergold, I. and G. Parkhurst. 2016. *The Economic Benefits of Sustainable Urban Mobility Measures Independent Review of Evidence: Report*, EVIDENCE Project, European Commission, Brussels. <http://tinyurl.com/m36e4un>.

- SUTP (Sustainable Urban Transport Project). 2014. SUTP Technical Document # 13: Urban Mobility Plans - National Approaches and Local Practices. http://www.sutp.org/files/contents/documents/resources/B_Technical-Documents/GIZ_SUTP_TD13_Urban-Mobility-Plans_EN.pdf.
- Tang S. and H.K. Lo. 2010. "Assessment of Public Private Partnership Models for Mass Rail Transit - an Influence Diagram Approach" *Public Transport* Vol 2: pp. 111–134, DOI 10.1007/s12469-010-0023-8.
- TIDE (Transport Innovation Deployment for Europe). 2013. Impact Assessment Handbook: Practitioners' Handbook for Cost Benefit and Impact Analysis of Innovative Urban Transport Measures, TIDE Project, European Commission, Brussels. http://www.tide-innovation.eu/en/upload/Results/T495_TIDE-Assessment-Handbook-Lite.pdf.
- U.K. DfT (Department of Transport). 2006a. Transport Analysis Guidance, Integrated Transport Economics and Appraisal, Department for Transport. <https://www.gov.uk/guidance/transport-analysis-guidance-webtag>, accessed 10 October 2016.
- U.K. DfT (Department of Transport). 2006b. Putting Passengers First: The Government's proposals for a modernised national framework for bus services. December. London. <http://webarchive.nationalarchives.gov.uk/20090806013054/http://www.dft.gov.uk/pgr/regional/buses/secputtingpassengersfirst/pdfputtingpassfirst.pdf>.
- U.K. DfT (Department of Transport). 2009. NATA Refresh: Appraisal for a Sustainable Transport System - New Approach to Transport Appraisal, http://bettertransport.org.uk/sites/default/files/dft_natarefreshApril_2009.pdf.
- U.K. DfT (Department for Transport). 2012. *Reforming our Railways: Putting the Customer First*, Presented to Parliament by the Secretary of State for Transport, March. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/4216/reforming-our-railways.pdf.
- U.S. General Accounting Office. 1999. FTA's Progress in Developing and Implementing a New Starts Evaluation Process. Report RCED–99–113. Washington, D.C.: U.S. Government Printing Office.
- Vanderschuren, M., A. Frieslaar and T. Lane. 2008. Assessment of the Improvement Strategies for the N1 Corridor between Bellville and Cape Town, Paper for the South African Transport Conference, Pretoria, July.
- Veryard, D. 2016. Quantifying the Socio-Economic Benefits of Transport: Roundtable Summary and Conclusions. Discussion Paper • 06, International Transport Forum, Paris, France. <http://www.itf-oecd.org/sites/default/files/docs/summary-roundtable-socio-economic-benefits-transport.pdf>.
- Voogd, H. 1982. Multi-Criterion Evaluation with Mixed Qualitative and Quantitative Data, *Environment and Planning Bulletin* 9, pp. 221–236.

Voogd, H. 1983. *Multi-Criteria Evaluation for Urban and Regional Planning*, Pion, London.

Watson, R. 2001. The Effect of Railway Privatization on Train Planning: a Case Study of the UK. *Transport Reviews*, Vol. 21, No. 2, pp. 181–193.

Weitzman, M.L. 1994. "On the Environmental Discount Rate," *Journal of Environment and Management*, Vol 26 No. 2, pp. 200-209.

The Role of Specificity and Apologies in Excuse Messages Following Train Delay

Emiel Cracco^{1*}

Nicolas Dirix¹

Christopher P. Reinders Folmer²

Abstract

An important issue in public transport is punctuality. Because delays are often caused by external factors, an efficient way to mitigate passengers' negative reactions is to point out these factors in an excuse. The current study investigated whether excuses following train delay can be optimized by making minor changes to their content. Specifically, we compared the effectiveness of specific and non-specific excuses. Furthermore, we investigated whether adding different types of an apology influenced the effectiveness of the excuse. The results indicated that specific excuses resulted in more forgiveness and a reduced intention to avoid public transport in the future. Further analyses showed that specific excuses were more acceptable and were therefore more successful at reducing perceived responsibility. The presence or absence of an apology did not influence excuse effectiveness. These findings suggest that minor adjustments to the communication strategy of public transport organizations can reduce passenger loss due to delays.

Keywords: *trust repair, public transport, delay, excuse, apology*

Introduction

When railway passengers purchase a ticket, they expect their train to arrive on time. However, these expectations are often violated. In Belgium, for instance, 11% of the trains in 2015 were either delayed by more than six minutes or cancelled (Infrabel, 2016). Research has shown that the inability to meet consumer expectations is perceived as a

¹Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

²Department of Private Law, Behavioural Approaches to Contract and Tort, Erasmus School of Law, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, the Netherlands

*Corresponding author. E-mail: emiel.cracco@ugent.be

Emiel Cracco and Nicolas Dirix contributed equally to this work.

violation of trust (Wang & Huff, 2007), which leads to complaints (Brown & Beltramini, 1989; Folkes, Koletsky, & Graham, 1987), negative word of mouth (Brown & Beltramini, 1989; Wang & Huff, 2007), and reduced repurchase intention (Folkes et al., 1987; Wang & Huff, 2007). In line with this evidence, recent polls indicate that the current situation of frequent delays has culminated in a general feeling of dissatisfaction among the Belgian railway passengers (Test-Aankoop, 2015). To avoid passenger loss, an important question is therefore how trust can be restored after a train delay.

Previous work has shown that the consequences of trust violation strongly depend on how the transgression is perceived. For example, studies have shown that the influence of inadequate service provision was reduced when consumers did not expect the event to reoccur (Folkes et al., 1987; Wang & Huff, 2007), evaluated the inconvenience as small (Brown & Beltramini, 1989), or felt that the supplier had no control over the situation (Brown & Beltramini, 1989; Folkes et al., 1987). As a result, an interesting approach to repair trust is to provide an excuse that lists the external factors contributing to the transgression (Kim, Dirks, & Cooper, 2009; Shaw, Wild, & Colquitt, 2003; Snyder & Higgins, 1988). Interestingly, train delays are often the result of factors beyond the control of the railway company. In Belgium, for example, 39.4% of the delays were caused by external factors such as cable theft (Infrabel, 2016). An efficient strategy to mitigate the reputational and financial costs associated with train delay is therefore to communicate these external factors to the passengers in a general announcement. In line with this suggestion, delay announcements typically consist of an excuse (e.g., *"The train with destination Antwerp is delayed by 15 minutes due to a technical failure"*) followed by an apology (e.g., *"Our apologies for the delay"*). However, excuses should be used with care as studies indicate that unconvincing explanations may have unintended harmful effects instead (Ferrin, Kim, Cooper, & Dirks, 2007; Kim et al., 2009). Therefore, the goal of the current study was to investigate how train delay announcements can be optimized by making minor changes to their content. Based on previous research, we will first investigate whether train delay announcements are more effective when they include a specific rather than an abstract excuse (Frey & Cobb, 2010; Greenberg, 1994; Shapiro et al., 1994). We will furthermore investigate whether apologies have a positive or a negative influence when they are combined with an excuse (Kim et al., 2009; Tomlinson, Dineen, & Lewicki, 2004). Finally, we will investigate whether the influence of an apology depends on whether it refers to the cause or to the consequence of the delay (Kim et al., 2009). As such, the present work aims to provide insight in how railway companies can augment their communication strategy to mitigate the negative impact of train delays.

Specificity of Excuses

Delay announcements often contain an excuse (e.g., *"... due to a technical failure ..."*). A widely used framework to interpret the influence of excuses is fairness theory (Folger & Cropanzano, 2001; Frey & Cobb, 2010; Gilliland et al., 2001; Shaw et al., 2003). Applied to train delays, this theory states that the degree to which passengers hold the railway company responsible depends on whether the delay could have been prevented (i.e.,

could counterfactual), on whether the delay should have been prevented (i.e., *should counterfactual*), and on whether they would have been better off without the delay (i.e. *would counterfactual*). Specifically, fairness theory postulates that the sense of injustice experienced by delayed railway passengers will depend on the number of questions to which the answer is affirmative (Shaw et al., 2003).

Because the goal of an excuse is to deflect responsibility to external factors, their effect can be situated at the level of could counterfactuals (Shaw et al., 2003). It is well documented that individuals automatically overestimate the internal and underestimate the external factors that contributed to a transgression (Kim et al., 2009). As a result, railway passengers are likely to initially assume that the delay could have been prevented. However, providing information on the contributing external factors may cause passengers to reconsider their initial internal attribution. It is assumed that such a shift from internal to external attribution will decrease the perceived responsibility of the railway company and will therefore reduce the negative consequences associated with the delay (Frey & Cobb, 2010; Gilliland et al., 2001; Shaw et al., 2003).

Importantly, research on excuses in interpersonal and organizational settings has identified that their effectiveness strongly depends on whether or not they are acceptable (Bies, Shapiro, & Cummings, 1988; Frey & Cobb, 2010; Mansour-Cole & Scott, 1998; Mellor, 1992; Shapiro, 1991; Shaw et al., 2003). Stated in terms of fairness theory, an excuse can be thought of as acceptable when it is legitimate and a legitimate excuse is more likely to reduce the perceived responsibility for a transgression. However, to successfully attribute a train delay to external factors it is especially relevant to know which factors constitute an acceptable excuse. Studies on this matter have revealed that acceptability is influenced both by the source and the content of the excuse. In particular, these studies indicated that an excuse should ideally be delivered in a sensitive way (Greenberg, 1993, 1994; Shapiro, Buttner, & Barry, 1994) by a trustworthy person (Greenberg, 1993) and contain detailed information regarding the external factors that contributed to the transgression (Frey & Cobb, 2010; Greenberg, 1994; Shapiro et al., 1994).

Whereas sensitivity and trustworthiness are difficult to manipulate in the short and formal messages that are used to communicate train delays, detailed excuses can be formed with minor effort. Therefore, the first goal of the current study was to investigate the influence of message specificity on the effectiveness of train delay excuses. Furthermore, to understand the process through which message specificity operates, we will additionally examine the role of two mediating variables. In line with previous work, we will first explore the mediating role of acceptability (Frey & Cobb, 2010). In addition, we will also explore the influence of perceived responsibility. In particular, we will test the assumption that acceptable excuses are more effective because they are more successful at reducing doubt concerning the responsibility of the railway company for the delay (Frey & Cobb, 2010; Gilliland et al., 2001; Shaw et al., 2003).

Combining Excuses with Apologies

Following the excuse, delay announcements frequently contain an apology (e.g., “... our apologies for the delay”). An apology can be defined as a statement in which the perpetrator admits responsibility and conveys regret for a trust violation. Although it entails an admission of guilt, it is assumed that such a message is successful because it communicates remorse and the intention to avoid a similar violation in the future (Ferrin et al., 2007; Kim et al., 2009). Indeed, the effectiveness of apologies is well established in organizational, interpersonal, and commercial settings (Bisel & Messersmith, 2012; Ferrin et al., 2007; Koehn, 2013; Ohbuchi, Kameda, & Agarie, 1989; Patel & Reinsch, 2003; Risen & Gilovich, 2007; Roschk & Kaiser, 2013). However, not much is known about the influence of apologies in the context of an excuse. According to an influential framework, excuses and apologies operate at a different level of trust repair (Kim et al., 2009). In particular, it is assumed that the success of an excuse depends on whether or not it is able to attribute the transgression to an external cause. Apologies, on the other hand, entail a confession of guilt and could therefore undermine the external shift intended by the excuse (Kim et al., 2009). In support of this idea, it has been shown that apologies are more effective when they are paired with an internal attribution compared with an external attribution (Tomlinson, Dineen, & Lewicki, 2004).

In light of the above, the second goal of the present study was to examine the influence of apologies on the effectiveness of train delay excuses. Because apologies imply a confession of guilt, it can be expected that apologizing for the delay has a negative impact on the legitimacy of the excuse. However, this may depend on the content of the apology. Specifically, an apology can refer to the cause of the delay (e.g., “our apologies for the incident”) or to the consequence of the delay (e.g., “our apologies for the inconvenience”). While an apology referring to the cause contains an internal attribution, an apology referring to the consequence does not. Instead, it brings across an empathic response. Consequently, it can be hypothesized that a cause apology may have a negative influence on the effectiveness of the excuse, whereas a consequence apology may have a positive influence.

To conclude, the current study aims to study how excuses following train delay can be optimized. First, we will study whether specific excuses are more effective than non-specific excuses. Second, we will examine if the addition of an apology has a positive or negative influence on excuse effectiveness and whether this depends on the type of apology.

Method

Participants

A sample of 128 Dutch-speaking Belgian participants took part in the study (82 females, 42 males, $M_{\text{age}} = 25.53$, $SD_{\text{age}} = 11.13$). The majority of the sample (88) was recruited online through social media advertisements and the remaining 40 participants were

first year psychology students of Ghent University who received partial course credit in exchange for their participation. Missing descriptive data were present for four participants. That is, three participants entered neither age nor gender information and one participant entered only age information. Participants indicated to make on average 15.49 train journeys per month (Table 2). Subjects were randomly assigned to the different conditions of the Excuse Type (specific or non-specific) x Apology Type (no apology, cause apology, consequence apology) design. Importantly, randomization was performed separately for credit students and social media volunteers to ensure that recruitment method was not confounded with the experimental manipulations.

Design and Procedure

Participants were asked to read and imagine a scenario (Supplementary Material). The scenario described a situation in which the subject was waiting for a train when it was announced that the train would be delayed by 20 minutes. The announcement started with an excuse. Importantly, the excuse could be either specific (“*due to a copper theft in the vicinity of Lokeren*”) or non-specific (“*due to problems on the railway track*”). Following the excuse there was either no apology, an apology referring to the cause of the delay (“*our apologies for this incident*”), or an apology referring to the consequence of the delay (“*our apologies for the inconvenience*”).¹ After participants had read the scenario, they were asked to answer a number of questions that assessed the effectiveness of the excuse. Finally, participants were asked how often they take the train and how often they are confronted with train delays.

Measures

All items were measured on a 7-point Likert scale going from “Strongly Disagree” to “Strongly Agree”. Example items are translated from Dutch to English. The full set of items is available in Supplementary Material.

Manipulation Check. The effectiveness of the manipulation was measured by assessing excuse specificity, excuse realism, and apology perception. The specificity of the excuse was measured with four items ($\alpha = 0.93$). An example item is: “*I think the explanation for the delay is specific*”. The realism of the excuse was measured with three items ($\alpha = 0.83$). An example item is: “*I find the cause of the delay unlikely*” (reverse scored). Finally, two questions were used to check whether the apology was perceived as an apology that referred to the cause or as an apology that referred to the consequence, namely “*I think the railway company apologized for the event that caused the delay*” (cause) and “*I think the railway company apologized for the inconvenience I experienced due to the delay*” (consequence). Note that these last two questions were only answered by participants for whom the announcement contained an apology ($N = 83$).

Mediator Variables. The acceptability of the excuse was evaluated with six items ($\alpha = 0.85$) based on previous research (Frey & Cobb, 2010; Riordan, Marlin, & Kellogg, 1983; Shapiro, 1991; Shapiro et al., 1994). An example item is: “*The explanation is acceptable*”.

The perceived responsibility of the railway company was measured with three items ($\alpha = 0.77$). An example item is: *"I think the railway company could have avoided this delay"*.

Outcome Variables. The effectiveness of the excuse was assessed by measuring forgiveness, revenge, avoidance, benevolence, and complaints. The degree of forgiveness was evaluated with five items ($\alpha = 0.89$) based on the study of Subkoviak et al. (1995). An example item is: *"After hearing this explanation, I could forgive the railway company for the delay"*. Items for revenge, avoidance, and benevolence were adapted from the Transgression-Related Interpersonal Motivations Inventory (McCullough, Root, & Cohen, 2006). The tendency to take revenge was measured with three items ($\alpha = 0.88$). An example item is: *"Despite hearing this explanation, I would hope that the railway company gets what it deserves"*. The degree to which participants wanted to avoid train travel as a result of the delay was measured with six items ($\alpha = 0.88$). An example item is: *"Despite this explanation, the delay would encourage me to no longer travel by train"*. Benevolence towards the railway company was measured by means of two items ($\alpha = 0.66$). An example item is: *"Even though I am disadvantaged, I would have goodwill for the railway company after hearing this explanation"*. Finally, the tendency to complain was assessed by four items ($\alpha = 0.69$). An example item is: *"Despite this explanation, the delay would encourage me to complain to the train conductor"*.

Control Variables. The frequency of train travel was measured by asking *"On average, how many times per month do you take the train?"* and the frequency of train delay was measured by asking *"On average, how many times per month is your train delayed?"*.

Analyses

All analyses were performed with SPSS for windows 21.0, except for the Bayesian analysis which was performed with JASP (Love et al., 2015) and the path analysis which was performed in R (R Development Core Team, 2013) with the lavaan package (Rosseel, 2012). P-values in the main analysis were corrected for testing multiple dependent variables according to Holm's procedure (Holm, 1979). The Bayes factors (BFs) obtained in the Bayesian analysis were interpreted according to Jeffreys (1961) with $1 < BF \leq 3$ indicating anecdotal evidence, $3 < BF \leq 10$ indicating substantial evidence, and $BF > 10$ indicating strong evidence. BFs measure the evidence for hypothesis A relative to hypothesis B. For example, a BF of 3 indicates that the data is 3 times more likely under hypothesis A than under hypothesis B.

One participant was excluded from all analyses because of excessive missing data (i.e., 21%). In addition, one participant was not included in the manipulation check analyses because no data was recorded for the relevant variables. For the other participants, we first performed Little's MCAR test to examine if there was a pattern in the missing data (< 7%). This revealed a normed χ^2 of 1.11. We therefore performed expectation-maximization to estimate the missing values at the item level.

Results

Manipulation Check

The manipulation check revealed that the specific excuse was perceived as more specific than the non-specific excuse, $t(124) = 12.39, p < 0.001$, but not as more realistic, $t(124) = 1.28, p = 0.203$ (Table 1). However, the cause apology was not perceived as referring more to the cause than the consequence apology, $t(80) = 1.61, p = 0.111$, and the consequence apology was not perceived as referring more to the consequence than the cause apology, $t(80) = 0.71, p = 0.483$ (Table 1).

TABLE 1.
Means and standard deviations of the manipulation check separately for the different conditions (columns) and dependent variables (rows)

	Non-Specific		Specific	
	Cause	Consequence	Cause	Consequence
Specificity	2.24 (0.81)	1.92 (0.81)	4.19 (1.21)	4.67 (1.14)
Realism	5.38 (1.18)	4.91 (1.03)	5.43 (1.19)	5.37 (1.27)
Cause	3.90 (1.97)	3.18 (1.62)	4.67 (1.74)	4.16 (1.95)
Consequence	5.40 (1.47)	5.50 (1.23)	5.43 (1.29)	5.74 (1.05)

Excuse Specificity and Type of Apology

To test the influence of Excuse Type and Apology Type on the evaluation of the excuse (Table 2), we first conducted separate ANCOVAs for each mediator and outcome variable with Excuse Type (non-specific or specific) and Apology Type (none, cause, or consequence) as factors and with Train Travel Frequency and Train Delay Frequency as covariates. As can be seen in Table 3, the main effect of Excuse Type was significant for acceptability, perceived responsibility, forgiveness, and avoidance. The main effect of Apology Type and the interaction between Excuse Type and Apology Type produced no significant effects. The results indicated that participants who received a specific excuse rated the excuse as more acceptable. Moreover, these participants were less likely to hold the railway company responsible for the delay, were more likely to forgive the railway company for the delay, and expressed a reduced intention to avoid train travel in the future. The presence of an apology did not have an influence on the outcome of the excuse.

TABLE 2.
Means and standard deviations of the Excuse Type x Apology Type analysis separately for the different conditions (columns) and dependent variables (rows)

	Non-Specific			Specific		
	Cause	Consequence	None	Cause	Consequence	None
Acceptability	4.80 (1.50)	4.22 (1.02)	4.32 (1.07)	5.45 (1.04)	5.31 (0.83)	5.41 (0.97)
Responsibility	3.17 (1.12)	3.70 (0.91)	3.86 (1.27)	2.70 (1.11)	2.65 (1.11)	2.58 (1.05)
Forgiveness	4.58 (1.50)	3.87 (1.22)	3.82 (1.28)	4.90 (1.34)	5.39 (1.04)	4.83 (1.35)
Benevolence	4.70 (1.32)	4.07 (1.24)	4.11 (1.56)	4.58 (1.40)	4.93 (1.09)	4.64 (1.47)
Avoidance	2.88 (1.40)	3.08 (1.25)	2.56 (1.22)	2.10 (1.22)	2.20 (1.19)	2.27 (1.15)
Revenge	2.25 (1.46)	2.21 (1.48)	2.10 (1.60)	1.54 (0.64)	1.59 (0.71)	2.26 (1.57)
Complain	3.01 (1.10)	3.16 (0.71)	3.09 (1.08)	2.73 (1.18)	2.91 (0.94)	2.97 (1.14)
Train Travel	13.20 (11.29)	19.00 (16.71)	14.86 (12.88)	15.57 (13.70)	16.85 (12.81)	13.45 (11.57)
Train Delay	7.45 (10.42)	9.38 (11.68)	5.09 (5.48)	5.71 (6.07)	5.55 (5.46)	6.55 (8.42)

TABLE 3.
Results of the Excuse Type x Apology Type ANCOVA

	Excuse		Apology		Excuse x Apology	
	F(1,119)	η_p^2	F(2,119)	η_p^2	F(2,119)	η_p^2
Acceptability	22.31***	0.16	1.42	0.02	1.03	0.02
Responsibility	21.16***	0.15	0.99	0.02	2.18	0.04
Forgiveness	15.33***	0.11	1.48	0.02	2.29	0.04
Benevolence	2.29	0.02	0.53	0.01	1.59	0.03
Avoidance	7.73*	0.06	0.40	0.01	0.33	0.01
Revenge	2.36	0.02	0.87	0.01	1.12	0.02
Complaints	1.04	0.01	0.36	0.01	0.02	0.00

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. P-values were corrected for testing multiple dependent variables according to Holm's procedure (Holm, 1979). Train Travel Frequency and Train Delay Frequency were included as covariates in the analysis.

To further evaluate the influence of Excuse Type and Apology Type, we additionally performed a JZS Bayes factor ANOVA (Rouder, Morey, Speckman, & Province, 2012) with default prior scales and with both Train Travel Frequency and Train Delay Frequency as nuisance variables (Table 4). We first compared a model with a single main effect for Excuse Type with the Null Model. This revealed strong evidence in favor of the Excuse Type Model for acceptability (BF = 4342.69), responsibility (BF = 4640.88), forgiveness (BF = 466.31), and avoidance (BF = 22.50). Anecdotal evidence for the Excuse Type Model was found for revenge (BF = 1.07). Anecdotal evidence for the Null Model was found for benevolence (BF = 0.72) and complaints (BF = 0.49). Next, we compared the Excuse Type Model with the Main Effect Model (i.e., *Excuse Type + Apology Type*)

and with the Interaction Model (i.e., *Excuse Type + Apology Type + Excuse Type x Apology Type*). This revealed substantial to strong evidence in favor of the Excuse Type Model for all dependent variables ($3.09 \leq BF \leq 70.00$). In summary, the Bayesian analysis confirmed that specific excuses were perceived as more acceptable and that these excuses reduced perceived responsibility, increased forgiveness, and reduced avoidance. Moreover, it also provided evidence against the hypothesis that there was a main effect of Apology Type or an interaction between Excuse Type and Apology Type.

TABLE 4.
Results of the Bayes factor ANOVA

	Excuse Type Model	Main Effect Model	Interaction Model
Acceptability	4342.69	1404.30	501.88
Responsibility	4640.88	1462.03	771.61
Forgiveness	466.31	145.46	97.77
Benevolence	0.72	0.09	0.06
Avoidance	22.50	3.17	0.57
Revenge	1.07	0.19	0.06
Complaints	0.49	0.05	0.01

Note. Each cell shows the Bayes factor for the Excuse Type Model, the Main Effect Model (*Excuse Type + Apology Type*), and the Interaction Model (*Excuse Type + Apology Type + Excuse Type x Apology Type*) relative to the null model with both Train Travel Frequency and Train Delay Frequency as nuisance variables.

Path Analysis

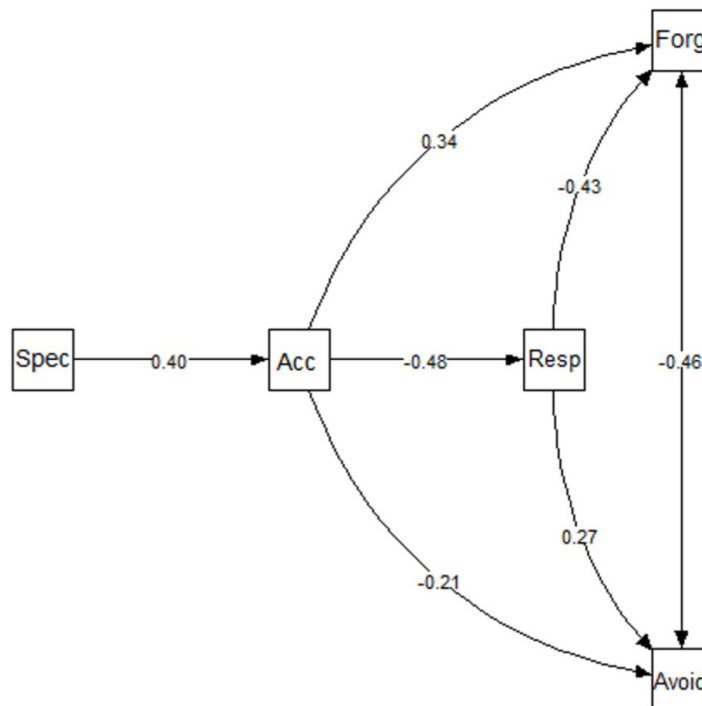
To understand the mechanism through which excuse specificity influenced forgiveness and avoidance, we furthermore performed a path analysis in which we tested the hypothesis that a specific excuse is more acceptable and that an acceptable excuse is more likely to reduce the perceived responsibility of the railway company. Two models were compared in the path analysis, namely a model in which the influence of acceptability on forgiveness and avoidance was completely explained by perceived responsibility (Model 1) and a model in which the influence of acceptability on forgiveness and avoidance was partially explained by perceived responsibility (Model 2).

Model fit was evaluated on the basis of the Comparative Fit Index (CFI) and the Standardized Root Mean Squared Residual (SRMR) because Monte Carlo simulation studies have shown that a cutoff criterion of $CFI \geq 0.96$ combined with a cutoff criterion of $SRMR \leq 0.09$ minimizes the risk of rejecting a true model or accepting a false model (Hu & Bentler, 1999). In addition, we included the Expected Cross Validation Index (ECVI) because this fit index takes into account model parsimony when assessing model fit. The ECVI is a comparative fit index that can be used to compare different models (Schermelleh-Engel, Moosbrugger, & Müller, 2003). The model with the lowest ECVI is preferred over the alternative models. Note that we chose not to report the Root Mean Square Error of Approximation (RMSEA) because it has been shown that RMSEA based

decisions inflate the risk of rejecting true models when the sample size and number of *dfs* is small, as is the case in the current study (Kenny, Kaniskan, & McCoach, 2014).

The evaluation of Model 1 revealed that it did not fit the data well, $\chi^2(5) = 28.83, p < 0.001$, CFI = 0.87, SRMR = 0.12. Model 2, on the other hand, did provide a good fit of the data, $\chi^2(3) = 9.34, p = 0.025$, CFI = 0.97, SRMR = 0.07, and proved to be a significant improvement over Model 1, $\chi^2_{diff}(2) = 19.49, p < 0.001, \Delta_{ECVI} = 0.12$ (Figure 1). This analysis thus indicates that specific excuses are more acceptable and that a reduction in perceived responsibility can partially explain why acceptable excuses are more effective in terms of forgiveness and avoidance.

FIGURE 1. visual representation of model 2. One-sided arrows reflect standardized regression coefficients. Double-sided arrows represent correlation coefficients. Legend: Spec = specificity, Acc = acceptability, Resp = perceived responsibility, Forg = forgiveness, Avoid = avoidance. Note that all reported coefficients were significant ($p \leq 0.020$)



Discussion

From an environmental perspective, an important societal challenge is to encourage individuals to make use of sustainable modes of transportation such as public rail transport. However, a weakness of rail transport is its punctuality. When passengers purchase a ticket, they expect their train to arrive on time. As a result, train delays have important reputational and financial costs (Brown & Beltramini, 1989; Folkes et al., 1987; Wang & Huff, 2007). One way to reduce these costs is to provide an excuse that points out the external forces contributing to the delay (Kim et al., 2009; Shaw et al., 2003; Snyder & Higgins, 1988). An efficient way to communicate such excuses to the passengers is to incorporate them in the message that announces the delay. However, excuses should be applied carefully because research suggests that unconvincing accounts can undermine their purpose (Ferrin et al., 2007; Kim et al., 2009). The

current study therefore investigated how train delay excuses are best implemented by examining whether minor changes to these excuses can influence how passengers experience the delay. First, we compared the effectiveness of specific and non-specific excuses for train delay. Second, we explored the influence of adding an apology to these excuses. Finally, we tested if the influence of the apology differed depending on whether it referred to the cause or the consequence of the delay.

With regard to excuse specificity, the results revealed that participants who received a specific excuse were more likely to forgive the railway company and less likely to look for an alternative mode of transportation in the future. Importantly, this suggests that excuse specificity may not only influence passengers' attitude towards the railway company but also the number of passengers that drop out. As such, providing specific excuses can be seen as a low-cost measure to mitigate the passenger loss caused by train delays. To understand the mechanism through which these effects arose, we focused on the role of acceptability (Bies et al., 1988; Frey & Cobb, 2010; Mansour-Cole & Scott, 1998; Mellor, 1992; Shapiro, 1991; Shaw et al., 2003) and perceived responsibility (Frey & Cobb, 2010; Gilliland et al., 2001; Shaw et al., 2003). This revealed that specific excuses were evaluated as more acceptable (Frey & Cobb, 2010) and that acceptable excuses produced favorable outcomes because they reduced doubt regarding the responsibility of the railway company (Frey & Cobb, 2010; Gilliland et al., 2001; Shaw et al., 2003). However, perceived responsibility could only partially explain the role of acceptability. This suggests that additional mechanisms contributed to the relation between excuse acceptability and excuse effectiveness. One possibility is that acceptable excuses also reduced the emotional impact of train delays. In support of this idea, Weiner and colleagues (1987) showed that excuses referring to an external cause reduce feelings of anger more effectively than excuses referring to an internal cause.

With regard to the addition of an apology, the results showed that concluding the delay announcement with an apology did not influence the effectiveness of the preceding excuse. This speaks against a large body of research showing that apologies are an effective way of repairing trust (Bisel & Messersmith, 2012; Ferrin et al., 2007; Koehn, 2013; Ohbuchi et al., 1989; Patel & Reinsch, 2003; Risen & Gilovich, 2007; Roschk & Kaiser, 2013). However, such research did not study apologies in the context of excuses. Because excuses achieve trust repair through external attribution and apologies through internal attribution, it has been suggested that they could potentially counteract one another (Kim et al., 2009; Tomlinson et al., 2004). The current study did not find evidence for this proposition. Specifically, neither an apology with high internal attribution (i.e., referring to the cause) nor an apology with low internal attribution (i.e., referring to the consequence) influenced the effectiveness of the excuse. However, the absence of a difference between these two types of apology should be interpreted with caution because the manipulation check suggested that participants did not perceive them differently. Restricting interpretation to the presence or absence of an apology, our results indicate that the addition of an apology had neither positive nor negative consequences. At least two explanations can be put forward for this finding. First, it could be argued that the study had insufficient statistical power to detect an influence of apologies on excuse effectiveness. However, this is unlikely considering that

a Bayesian analysis indicated substantial to strong evidence against an effect of apology type. Second, it is possible that the apology was not perceived as sincere in this context. Indeed, the scripted use of apologies in public transport communication possibly causes them to be perceived as less spontaneous and sincere. Given that these are important factors in determining the effectiveness of apologies (Koehn, 2013; Roschk & Kaiser, 2013; Tomlinson et al., 2004), it is possible that scripted apologies do not possess the necessary elements to produce an effect.

In short, the current study shows that the degree of message specificity but not the addition of an apology has an influence on the effectiveness of excuses following train delay. These findings show how minor adjustments to the communication strategy of public transport organizations can assist in preserving and restoring commuter trust in the face of delays. However, it is of note that the present work was restricted to situations where the railway company was not responsible for the delay. Although this situation is common (Infrabel, 2016), it is not always possible to attribute a delay to external factors. In such cases, it seems implausible that providing detailed information about the cause of a delay has favorable consequences. Importantly, this does not imply that public transport agencies should also attribute delays that could have been avoided to external factors. Apart from ethical considerations, research suggests that a denial of responsibility is associated with worse outcomes than an apology if later evidence refutes the denial (Kim, Ferrin, Cooper, & Dirks, 2004). It is therefore unlikely that such a strategy will be successful in the long run. Given these considerations, an interesting question is how delay announcements can be optimized in situations without an external cause. Because responsibility cannot easily be deflected in these situations, one possibility is that apologies gain importance in such events.

A further question that merits attention is the extent to which drop-out following train delay depends on the availability of transport alternatives. An interesting framework in this respect is the ASE model (De Vries, Dijkstra, & Kuhlman, 1988) which posits that the transition from intention to action depends on skills (e.g. driver license) and barriers (e.g. car ownership). From this perspective, a delayed passenger without a car is less likely to follow up on his/her intention of looking for alternative modes of transportation. Nevertheless, meta-analytic work suggests that intentions are a consistent predictor of subsequent behavior under a wide range of conditions (Webb & Sheeran, 2006). This suggests that obstacles reduce but do not eliminate the probability that individuals follow up on their intention. For example, individuals who do not own a car could avoid public transport by considering alternatives such as carpooling. Future research will be necessary to establish the influence of skills and barriers on avoidance of public transport following delay.

To conclude, the present study shows that railway passengers are less inclined to look for another mode of transportation following train delay when they are provided with a detailed rather than an abstract excuse. This finding offers railway companies an efficient and low-cost measure to reduce passenger loss associated with delay.

References

- Bies, R. J., Shapiro, D. L., & Cummings, L. L. (1988). Causal Accounts and Managing Organizational Conflicts: Is it Enough to Say It's Not My Fault? *Communication Research*, 15, 381–399. <http://doi.org/10.1177/009365088015004003>
- Bisel, R. S., & Messersmith, A. S. (2012). Organizational and Supervisory Apology Effectiveness: Apology Giving in Work Settings. *Business Communication Quarterly*, 75(4), 425–448. <http://doi.org/10.1177/1080569912461171>
- Brown, S. P., & Beltramini, R. F. (1989). Consumer Complaining and Word of Mouth Activities: Field Evidence. *Advances in Consumer Research*, 16, 9–16.
- De Vries, H., Dijkstra, M., & Kuhlman, P. (1988). Self-efficacy: The Third Factor Besides Attitude and Subjective Norm as a Predictor of Behavioural Intentions. *Health Education Research*, 3, 273–282.
- Ferrin, D. L., Kim, P. H., Cooper, C. D., & Dirks, K. T. (2007). Silence Speaks Volumes: The Effectiveness of Retinence in Comparison to Apology and Denial for Responding to Integrity- and Competence-Based Trust Violations. *Journal of Applied Psychology*, 92, 893–908. <http://doi.org/10.1037/0021-9010.92.4.893>
- Folger, R., & Cropanzano, R. (2001). Fairness Theory: Justice as Accountability. In J. Greenberg & R. Cropanzano (Eds.), *Advances in Organizational Justice* (pp. 1–55). Stanford, CA: Stanford University Press.
- Folkes, V. S., Koletsky, S., & Graham, J. L. (1987). A Field Study of Causal Inferences and Consumer Reaction: The View from the Airport. *Journal of Consumer Research*, 13, 534–539. <http://doi.org/10.1086/209086>
- Frey, F. M., & Cobb, A. T. (2010). What Matters in Social Accounts? The Roles of Account Specificity, Source Expertise, and Outcome Loss on Acceptance. *Journal of Applied Social Psychology*, 40(5), 1203–1234. <http://doi.org/10.1111/j.1559-1816.2010.00616.x>
- Gilliland, S. W., Groth, M., Baker, R. C., Dew, A. F., Polly, L. M., & Langdon, J. C. (2001). Improving Applicants' Reactions to Rejection Letters: An Application of Fairness Theory. *Personnel Psychology*, 54(3), 669–703. <http://doi.org/10.1111/j.1744-6570.2001.tb00227.x>
- Greenberg, J. (1993). Stealing in the Name of Justice: Information and Interpersonal Moderators of Theft Reactions to Underpayment Inequity. *Organizational Behavior and Human Decision Processes*, 54, 81–103. <http://doi.org/10.1006/obhd.1993.1004>
- Greenberg, J. (1994). Using Socially Fair Treatment to Promote Acceptance of a Work Site Smoking Ban. *Journal of Applied Psychology*, 79(2), 288–297. <http://doi.org/10.1037/0021-9010.79.2.288>
- Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.

- Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <http://doi.org/10.1080/10705519909540118>
- Infrabel. (2016). Stiptheidscijfers 2015 [Punctuality Statistics 2015]. Retrieved from <http://www.infrabel.be/nl/over-infrabel/stiptheid/rapporten/2015/1-0>
- Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford, UK: Oxford University Press.
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2014). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*. <http://doi.org/10.1177/0049124114543236>
- Kim, P. H., Dirks, K. T., & Cooper, C. D. (2009). The Repair of Trust: A Dynamic Bilateral Perspective and Multilevel Conceptualization. *Academy of Management Review*, 34(3), 401–422. <http://doi.org/10.5465/AMR.2009.40631887>
- Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the Shadow of Suspicion: the Effects of Apology Versus Denial for Repairing Competence- Versus Integrity-Based Trust Violations. *The Journal of Applied Psychology*, 89(1), 104–118. <http://doi.org/10.1037/0021-9010.89.1.104>
- Koehn, D. (2013). Why Saying “I’m Sorry” Isn’t Good Enough: The Ethics of Corporate Apologies. *Business Ethics Quarterly*, 23(2), 239–268. <http://doi.org/10.5840/beq201323216>
- Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., ... Wagenmakers, E.-J. (2015). JASP.
- Mansour-Cole, D. M., & Scott, S. G. (1998). Hearing it Through the Grapevine: The Influence of Source, Leader-Relations, and Legitimacy on Survivors’ Fairness Perceptions. *Personnel Psychology*, 51, 25–54. <http://doi.org/10.1111/j.1744-6570.1998.tb00715.x>
- McCullough, M. E., Root, L. M., & Cohen, A. D. (2006). Writing About the Benefits of an Interpersonal Transgression Facilitates Forgiveness. *Journal of Consulting and Clinical Psychology*, 74, 887–897. <http://doi.org/10.1037/0022-006X.74.5.887>
- Mellor, S. (1992). The Influence of Layoff Severity on Postlayoff Union Commitment Among Survivors: The Moderating Effect of the Perceived Legitimacy of a Layoff Account. *Personnel Psychology*, 45, 579–600. <http://doi.org/10.1111/j.1744-6570.1992.tb00861.x>
- Ohbuchi, K., Kameda, M., & Agarie, N. (1989). Apology as Aggression Control : Its Role in Mediating Appraisal of and Response to Harm. *Journal of Personality and Social Psychology*, 56(2), 219–227. <http://doi.org/10.1037/0022-3514.56.2.219>
- Patel, A., & Reinsch, L. (2003). Companies Can Apologize: Corporate Apologies and Legal Liability. *Business Communication Quarterly*, 66(1), 9–25. <http://doi.org/10.1177/108056990306600103>

- R Development Core Team. (2013). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria.
- Riordan, C. A., Marlin, N. A., & Kellogg, R. T. (1983). The Effectiveness of Accounts Following Transgression. *Social Psychology Quarterly*, 46(3), 213. <http://doi.org/10.2307/3033792>
- Risen, J. L., & Gilovich, T. (2007). Target and Observer Differences in the Acceptance of Questionable Apologies. *Journal of Personality and Social Psychology*, 92(3), 418–433. <http://doi.org/10.1037/0022-3514.92.3.418>
- Roschk, H., & Kaiser, S. (2013). The Nature of an Apology: An Experimental Study on How to Apologize After a Service Failure. *Marketing Letters*, 24(3), 293–309. <http://doi.org/10.1007/s11002-012-9218-x>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1–36.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes Factors for ANOVA Designs. *Journal of Mathematical Psychology*, 56(5), 356–374. <http://doi.org/10.1016/j.jmp.2012.08.001>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models : Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Shapiro, D. L. (1991). The Effects of Explanations on Negative Reactions to Deceit. *Administrative Science Quarterly*, 36, 614–630. <http://doi.org/10.2307/2393276>
- Shapiro, D. L., Buttner, E. H., & Barry, B. (1994). Explanations: What Factors Enhance Their Perceived Adequacy? *Organizational Behavior and Human Decision Processes*, 58, 346–368. <http://doi.org/10.1006/obhd.1994.1041>
- Shaw, J. C., Wild, E., & Colquitt, J. A. (2003). To Justify or Excuse?: A Meta-Analytic Review of the Effects of Explanations. *Journal of Applied Psychology*, 88(3), 444–458. <http://doi.org/10.1037/0021-9010.88.3.444>
- Snyder, C. R., & Higgins, R. L. (1988). Excuses: Their Effective Role in the Negotiation of Reality. *Psychological Bulletin*, 104(1), 23–35. <http://doi.org/10.1037/0033-2909.104.1.23>
- Subkoviak, M. J., Enright, R. D., Wu, C.-R., Gassin, E. A., Freedman, S., Olson, L. M., & Sarinopoulos, I. (1995). Measuring Interpersonal Forgiveness in Late Adolescence and Middle Adulthood. *Journal of Adolescence*, 18, 641–655.
- Test-Aankoop. (2015). Treinreizigers Ontevreden: Dienstverlening NMBS Moet Beter! [Railway Passengers Dissatisfied: NMBS Service Must Improve!]. Retrieved from http://www.test-aankoop.be/action/pers_informatie/persberichten/2015/nmbs
- Tomlinson, E. C., Dineen, B. R., & Lewicki, R. J. (2004). The Road to Reconciliation: Antecedents of Victim Willingness to Reconcile Following a Broken Promise. *Journal of Management*, 30(2), 165–187. <http://doi.org/10.1016/j.jm.2003.01.003>

Wang, S., & Huff, L. C. (2007). Explaining Buyers' Responses to Sellers' Violation of Trust. *European Journal of Marketing*, 41(9), 1033–1052. <http://doi.org/10.1108/03090560710773336>

Webb, T. L., & Sheeran, P. (2006). Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence. *Psychological Bulletin*, 132(2), 249–268. <http://doi.org/10.1037/0033-2909.132.2.249>

Weiner, B., Amirkhan, J., Folkes, V. S., & Verette, J. A. (1987). An Attributional Analysis of Excuse Giving: Studies of a Naive Theory of Emotion. *Journal of Personality and Social Psychology*, 52(2), 316–324. <http://doi.org/10.1037/0022-3514.52.2.316>

About the Authors

EMIEL CRACCO is a Ph.D. student at the Department of Experimental Psychology of Ghent University. His doctoral research focuses on action observation in social interactions that include multiple agents. However, he is also interested in the psychological aspects of public transport use and in emotion regulation in children and adolescents.

NICOLAS DIRIX is a PhD student at the Department of Experimental Psychology of Ghent University. He is currently affiliated with the “Language, Education, and Memory in Multilingualism and Academia” project (LEMMA), in which differences between first and second language processing and production are studied. In his part of the project, he studies eye movement data of bilingual students reading and studying in their first and second language.

CHRIS REINDERS FOLMER is a post-doctoral researcher at the research program Behavioural Approaches to Contract and Tort, affiliated with the Rotterdam Institute of Private Law and the Rotterdam Institute of Law and Economics. His background is in social and economical psychology (PhD 2008, Vrije Universiteit Amsterdam). His research integrates psychological, legal and economic perspectives to empirically test the assumptions that underlie legal practice and policy making, identify possible discrepancies, and develop alternatives informed by these perspectives. His research interests include legal psychology, trust, decision making, social dilemmas, ethics, and reputation.

Footnotes

Footnote 1

Note that the excuses and apologies were validated in a pilot study (N = 18) including 16 specific excuses and 7 non-specific excuses that were paired with an apology for the cause or an apology for the consequence. Based on the results from this study, we chose the excuses that differed maximally with respect to specificity, $t(17) = 9.56$, $p < 0.001$, and minimally with respect to realism, $t(17) = -0.93$, $p = 0.365$. The apologies were compared on the degree to which they referred to the cause or consequence

of the delay. This showed that the cause apology referred more to the cause than the consequence apology, $t(17) = 13.67, p < 0.001$, and that the consequence apology referred more to the consequence than the cause apology, $t(17) = -3.75, p = 0.002$.

Supplementary Material: Scenario

Scenario in Dutch

Gelieve je de volgende situatie aandachtig te lezen en zo levendig mogelijk in te beelden:

Het is vrijdagmiddag. Al je lessen zijn achter de rug en je bent gepakt en gezakt onderweg naar station Gent-Sint-Pieters om de trein te nemen naar Antwerpen. Je komt goed op tijd aan in het station en begeeft je vervolgens naar perron 1, waar je samen met enkele tientallen mede-passagiers wacht op het aankomen van de trein. Plots weerklinkt de volgende boodschap door de luidsprekers:

"Aandacht, spoor 1! **Vanwege een koperdiefstal ter hoogte van Lokeren [specifiek] / Vanwege problemen op het spoor [vaag]** heeft de trein van Gent-Sint-Pieters naar Antwerpen centraal van 16u05 een vertraging van ongeveer 20 minuten. **Onze verontschuldigingen voor het ongemak [gevolg] / voor dit voorval [oorzaak] / [geen]."**

Scenario in English

Please carefully read the following situation and try to imagine it as vividly as possible:

It is a Friday afternoon. Your classes are over and you are, packed and ready, on your way to the station of Ghent to take the train to Antwerp. You arrive well in time at the station and you go to platform 1 where you await the arrival of the train with a few dozen fellow travelers. Suddenly, the following message is played through the speakers:

"Attention, platform 1! **Due to a copper theft in the vicinity of Lokeren [specific] / due to problems on the railway track [vague]**, the 16:05 train from Ghent to Antwerp is delayed by approximately 20 minutes. **Our apologies for the inconvenience [consequence] / Our apologies for the incident [cause] / [none]."**

Supplementary Material: Questionnaires

All items were measured on a 7-point Likert scale (1 = "Strongly Disagree", 7 = "Strongly Agree").

Acceptability

Gelieve aan te geven in welke mate je het eens bent met de volgende stellingen.

1. De uitleg bevat onvoldoende informatie
2. De uitleg is adequaat
3. De uitleg is aanvaardbaar
4. Ik zou tevreden zijn met de uitleg
5. De uitleg is niet geloofwaardig
6. De uitleg is redelijk

Please indicate to what extent you agree with the following statements.

1. The explanation does not contain enough information
2. The explanation is adequate
3. The explanation is acceptable
4. I would be satisfied with the explanation
5. The explanation is not credible
6. The explanation is reasonable

Perceived Responsibility

Gelieve aan te geven in welke mate je het eens bent met de volgende stellingen.

1. Volgens mij kon de NMBS deze vertraging hebben vermeden
2. Volgens mij was de vertraging het gevolg van oncontroleerbare omstandigheden
3. Volgens mij lag de NMBS aan de oorzaak van de vertraging

Please indicate to what extent you agree with the following statements.

1. I think the railway company could have avoided this delay
2. I think the delay was caused by uncontrollable circumstances
3. I think the railway company was at the cause of the delay

Forgiveness

Gelieve aan te geven in welke mate de volgende stellingen voor jou van toepassing zijn.

1. Na het horen van deze uitleg zou ik de NMBS kunnen vergeven voor de vertraging
2. Ondanks het horen van deze uitleg zou ik boos zijn op de NMBS
3. Ondanks het horen van deze uitleg zou mijn vertrouwen in de NMBS geschonden zijn
4. Na het horen van deze uitleg zou ik begrip kunnen opbrengen voor de vertraging
5. Ondanks het horen van deze uitleg zou ik de NMBS deze situatie kwalijk nemen

Please indicate to what extent the following statements apply to you.

1. After hearing this explanation, I would be able to forgive the railway company for the delay
2. Despite hearing this explanation, I would be angry at the railway company
3. Despite hearing this explanation, my trust in the railway company would have been violated
4. After hearing this explanation, I would be understanding of the delay
5. Despite hearing this explanation, I would blame the railway company for this delay

Revenge

Gelieve aan te geven in welke mate de volgende stellingen voor jou van toepassing zijn.

1. Na het horen van deze uitleg zou ik de NMBS ongeluk toewensen.
2. Ondanks het horen van deze uitleg zou ik het de NMBS betaald willen zetten
3. Ondanks het horen van deze uitleg zou ik hopen dat de NMBS krijgt wat ze verdient

Please indicate to what extent the following statements apply to you.

1. After hearing this explanation, I would wish misfortune on to the railway company
2. Despite hearing this explanation, I would want to get even with the railway company
3. Despite hearing this explanation, I would hope that the railway company gets what it deserves

Avoidance

Ondanks de omgeroepen uitleg zou de vertraging me aanzetten...

1. ...tot het overwegen van andere vervoersalternatieven
2. ...om in het vervolg de trein te vermijden
3. ...om niet meer met de trein te reizen
4. ...om de NMBS te vermijden in de toekomst
5. ...om zoveel mogelijk afstand te houden van de NMBS
6. Ondanks het horen van deze uitleg zou ik geen vertrouwen meer hebben in de NMBS

Despite this explanation the delay would make me...

1. ... consider other transportation alternatives
2. ... avoid the train in the future
3. ... travel no longer by train

4. ... avoid the railway company in the future
5. ... take as much distance as possible from the railway company
6. Despite hearing this explanation, I would no longer trust the railway company

Benevolence

Gelieve aan te geven in welke mate de volgende stellingen voor jou van toepassing zijn.

1. Ook al ben ik benadeeld, zou ik na het horen van deze uitleg welwillend staan tegenover de NMBS
2. Na het horen van deze uitleg zou ik geen wrok koesteren tegenover de NMBS

Please indicate to what extent the following statements apply to you.

1. Even though I'm disadvantaged, I would have goodwill for the railway company after hearing this explanation
2. After hearing this explanation, I would not hold a grudge against the railway company

Complaints

Ondanks de omgeroepen uitleg zou de vertraging me aanzetten...

1. ...tot klagen mijn vrienden/familie
2. ...tot het schrijven van een klachtenbrief naar de ombudsdienst van de NMBS
3. ...tot klagen bij de conducteur
4. ...tot klagen bij mijn mede-passagiers

Despite this explanation the delay would encourage me...

1. ... to complain to my friends / family
2. ... to write a letter of complaint to the ombudsman of the railway company
3. ... to complain to the conductor
4. ... to complain to my fellow passengers

Manipulation check

Gelieve aan te geven in welke mate de volgende stellingen voor jou van toepassing zijn.

1. Volgens mij verontschuldigde de de NMBS zich voor hetgene dat de vertraging heeft veroorzaakt
2. Volgens mij verontschuldigde de NMBS zich voor de last die ik ondervond van de vertraging
3. Ik vind de oorzaak van de vertraging realistisch
4. Ik vind de oorzaak van de vertraging onwaarschijnlijk

5. Volgens mij treedt vertraging door deze oorzaak soms op in het echt
6. Ik vind de uitleg die voor de vertraging gegeven werd specifiek
7. Ik vind de uitleg die voor de vertraging gegeven werd vaag
8. Ik vind de uitleg die voor de vertraging gegeven werd gedetailleerd
9. Ik vind de uitleg die voor de vertraging gegeven werd uitgebreid

Please indicate to what extent the following statements apply to you.

1. I think the railway company apologized for the event that caused the delay
2. I think the railway company apologized for the inconvenience I experienced due to the delay
3. I think the cause of the delay is realistic
4. I find the cause of the delay unlikely
5. I believe that delays by this cause also occur in real life
6. I think the explanation for the delay was specific
7. I think the explanation for the delay was vague
8. I think the explanation for the delay was detailed
9. I think the explanation for the delay was extensive

Impacts of New Light Rail Transit Service on Riders' Residential Relocation Decisions

*You-Lian Chu, Ph.D.
Parsons Corporation*

*Yi Deng, Ph.D.
MTA New York City Transit*

*Rongfang (Rachel) Liu, Ph.D.
New Jersey Institute of Technology*

Abstract

Using the rider survey data collected from Hudson-Bergen Light Rail Transit System in New Jersey, this paper investigated the residential relocating decisions of the riders who have been riding the LRT for a period of time. Using the Heckman's sample selection model, the paper extends the current mobility literature by describing not only a rider's likelihood to move as a result of the new LRT service but also the movers' orientation toward their residence distances to LRT stations. Information on the socio-economic characteristics of the movers and their residence distances to LRT stations would help planners and developers identify areas where housing growth associated with specific characteristics of the riders will occur, and plan for these areas to provide affordable housing and amenities for relocating residents.

Key Words: *Light Rail Transit, Residential Moving Behavior, Sample Selection Model*

Introduction

Transportation, social, and economic impacts of light rail transit (LRT) service on the communities have been the subjects of many studies including accessibility, property value, dislocation, transit-focused development, auto ownership, and mode shares. However, very little work has been devoted to the examination of LRT's influences on residential moving behavior and the spatial redistribution of riders' residences. In

2008, New Jersey Transit conducted a platform survey along the northern segment of the Hudson-Bergen Light Rail Transit (HBLRT). The survey included the questionnaire about the extent to which the HBLRT service affected riders' moving decisions and their daily living. The survey results showed that 65% of the riders had actually moved in response to the new LRT service. For those riders who had moved, 46% and 24% of them consider, respectively, the LRT as a "very important" and "somewhat important" factor in their moving decision-making process. Based on the HBLRT survey data, this paper focuses on the residential relocating decisions of the riders who have been riding the HBLRT for a period of time. Two relocating decisions were analyzed, including whether or not a rider would move as a result of the new LRT service (discrete choice) and, having decided to move, how far a rider would live away from the nearest LRT stations (continuous choice).

The objective of this paper is to use an econometric model to jointly investigate the rider's discrete move choice and continuous distance choice. A major characteristic of such mixed discrete-continuous models is that the continuous dependent variables are censored because data on residence distance to the station usually exhibit certain number of observations clustered at zero (when no move occurs), with the rest of the observations being positive (when move occurs). In either case, the commonly used ordinary least squares (OLS) regression on the entire sample will produce biased parameter estimates because the residuals do not have a mean zero. Yet, estimation of OLS with a truncated sample created by dropping observations with zero values will also yield inconsistent estimates because of potential sample selection bias (Greene 2000). A typical model used to account for the censored nature of the data is the Tobit model (Tobin 1958) which analyzes censored data in a regression context. However, the Tobit model not only restricts the exact same variables affecting the rider's move and distance decisions but also restricts the relative effects of those variables to be equal in both decisions. To relax these restrictions, this paper adopted a Heckman-style two-stage sample selection model (Heckman 1979).

Literature Review

Very few studies have associated LRT service with respect to residential moving behavior. At most, Cao and Schoner (2013) investigated transportation impact of the Hiawatha LRT in the Twin Cities by exploring motivations for people moving into the LRT corridor and the socio-economic/demographic characteristics of the movers. They found that compared to non-movers (defined as residents who have lived in the Hiawatha corridor before the opening of LRT), Hiawatha movers are more likely to be well-educated, younger, employed, and renters. There are no significant differences in household size, income, share of female, and number of cars per driver between movers and non-movers in the Hiawatha corridor. Important relocating factors considered by movers include easy access to transit station, job accessibility, affordable and high quality living unit, and safe neighborhood.

Instead of focusing on LRT, a number of studies investigated the impact of a new rail service on residential location choice. Orchieng et al. (2002) conducted a survey analysis

of the residential relocations that followed the New Jersey Transit Midtown Direct rail improvement. The survey shows that improvement caused 8% of the riders to relocate their residence within five months of the start of the service. In addition, 15% of the respondents stated they would relocate if they could save between 31-45 minutes on their one-way commute. In the decision to choose residence, the most important attributes include accessibility to work, school, and services, neighborhood security, real estate values, and traffic congestion concerns. The above survey was also used by Holguin-Veras et al. (2002) to analyze the impacts of transit accessibility changes upon residential location choice. The result indicates that the decision to change residence is affected by overall accessibility (for all modes) rather than transit accessibility only. Riders also take into account the overall characteristics of the commute (including travel time, comfort, convenience, among other) while making residential choice decisions, as opposed to the sole consideration of travel time.

Using travel data from the San Francisco Bay Area, Cervero and Duncan (2008) employed a nested logit model to jointly estimate the traveler's decision to live near a train station (within ½ mile radius) and the decision to routinely take rail to work. The research reveals that station-area residents are most likely to be lower-income households, younger individuals, non-traditional households (traditional household is defined as two adults between the ages of 25 and 54 years with at least one child), Asian-Americans and Hispanics, and have lower levels of auto ownership. Also instrumental in the choice to live near transit is job accessibility via both transit and highway networks (Note: Many rail stations in the Bay Area have good highway access.) Lund (2006) conducted a survey of households who moved to transit-oriented developments (TODs) that were served by rail (including light, heavy, and commuter rail system in the San Francisco Bay Area, Los Angeles, and San Diego). Only about one-third of respondents reported access to transit as one of the most important reasons for choosing to live in a TOD. Other important reasons include cost of housing, quality of the living environment, and access to shops and services. Those who reported that their choice of residence location was motivated in part by access to transit were more likely to use transit than those who did not. Olaru, et al. (2011) also evaluated how households consider TOD characteristics in their residential location decisions with regard to new Mandurah railway line stations in Western Australia. Whereas "stage of life" (a continuous evolution of family structure) is found to influence household relocation decisions significantly, other important factors include affordable and safe locations, proximity to shops, services, and transit, and prospects for increased real estate prices.

From the review of literature, it appears that previous studies largely concentrated on the exploration of the characteristics of station-area residents and their main reasons for moving to the station area. Therefore, the primary contribution of this paper is to extend the current mobility literature by describing not only a rider's likelihood to move as a result of the new LRT service but also the movers' orientation toward their residence distances to LRT stations. Although recent studies (Holguin-Veras, et al. 2002; Lund 2006; Olaru 2011) claimed that access to transit was not always viewed as the most important factors for people to make relocation decisions, the HBLRT survey

showed that 65% of the riders had actually moved in response to the new LRT service. This result suggests that improved transit access can be a dominant relocation factor in certain conditions, particularly when the new LRT system can easily connect to existing transportation infrastructure (Note: The HBLRT provides connections to PATH subway, ferry service, bus stops, suburban commuter rail, and park and ride lots). This multi-modal connectivity would provide transit riders more options for more destinations in general and improves job accessibility to lower and midtown Manhattan and Newark in particular.

Model Description

In this paper, two mobility (move and distance) choices were analyzed using the Heckman's sample selection model (Heckman 1979). The model contains two latent variable equations: the move equation (first stage) and distance equation (second stage). The move equation uses a binary probit model to determine whether or not a rider would make a residential move as a result of the new LRT service. For those riders who have decided to move, the distance equation uses a linear regression to calculate how far a rider would live away from the nearest LRT station. To control for sample selectivity, the second stage regression appends the inverse Mills ratio (Greene 2000) calculated from the linear predictions of the probit model as an additional independent variable. A more detailed description of the model structure and its estimation method is included in Appendix 1.

Data Description

The primary rider data were drawn from New Jersey Transit's 2008 platform survey conducted at seven stations along the 5.5-mile northern segment of the Hudson-Bergen Light Rail Transit (HBLRT) between Tonnelle Avenue in Jersey City and the Hoboken Terminal in Hoboken. As shown in Figure 1, the seven stations include Tonnelle Avenue, Bergenline Avenue, Port Imperial, Lincoln Harbor, 9th Street, 2nd Street, and Hoboken Terminal. The survey technique was to distribute self-administered questionnaires at every station and ask passengers to return them by mail or to drop boxes at stations. The target population consisted of typical weekday boarding passengers traveling southbound from the selected stations during the time period 4:45 AM – 4:00 PM. Surveying in one direction is to reduce the likelihood of asking the same riders to fill in the questionnaire twice. (Note: Because questionnaires were distributed to southbound passengers only, the Hoboken Terminal would act as a collection station.) The sampling strategy adopted a census approach, in which an attempt was made to distribute a survey form to every boarding passenger. If the survey form was refused, it was put to the side not to be reused. The final count of the refused survey forms and field survey records confirm the total boarding volumes at each station and the total number of forms that were handed out.

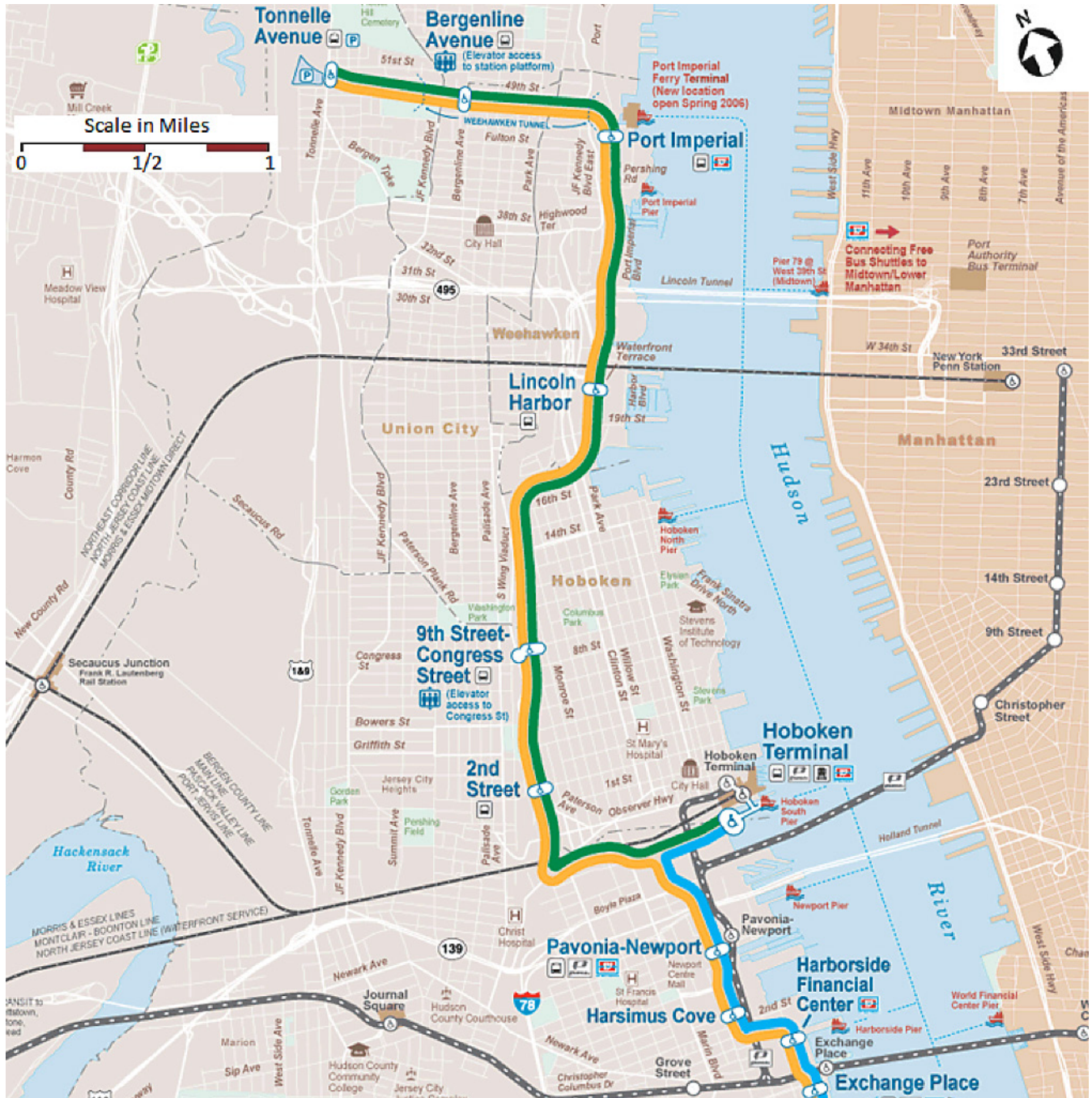


FIGURE 1. The Northern Segment of the HBLRT

The survey consisted of 39 questions covering the information on riders' travel patterns, customer satisfaction ratings, workplace locations, residential relocations, and socio-economic and demographic (SED) characteristics. To understand how the HBLRT service affects riders' moving decisions, one of the questions is that "If you moved within the last five years, how important was the availability of the HBLRT service in your decision to move?" Four categories are available for respondents to mark the importance of the HBLRT, including "Very important", "Somewhat important", "Somewhat unimportant", and "Not important at all". Survey results show that 46% of the movers considered LRT "very important" and 24% of the movers consider LRT "somewhat important". The above responding results also indirectly revealed that approximately 65% of the riders had made the decision to move after the HBLRT service began operation. Because riders who had made the decision to move also provided their new address information, geocoding of these addresses into a GIS-based database system allowed us to calculate the Manhattan distance, in miles, from a rider's residence to LRT station.

After the responses were reviewed and errors were corrected, the final data collection effort resulted in 1,023 usable survey forms, achieving a response rate of approximately 19%. The distribution of usable survey forms among the six stations surveyed matches the distributions of the total boarding volumes, implying a representative sample of passengers entering the northern segment of the HBLRT system. The 1,023 responses would achieve a system-wide confidence level of 90% with a $\pm 8.5\%$ margin of error for the survey results. A 8.5% margin of error means that if the survey was repeated, survey responses would lie $\pm 8.5\%$ of the initial survey responses 90 percent of the time, for each survey question. Or, we are 90% confident that the actual values are $\pm 8.5\%$ of the reported values.

Variable Specification

The Heckman's sample selection model contains two latent variable equations: the move and distance equations. The dependent variable of move equation is a binary variable that has the value of one if a rider's move was observed and zero otherwise. The dependent variable of distance equation is the amount of road network distance (in miles) from a rider's residence to the nearest LRT station. The distance variable would have the value of 0 if a rider did not make a move. For model estimation, the natural log of the distance was used to reduce the effects of the skewed nature of the distance variable as well as guaranteeing the nonnegativity of the distance predictions. To control for sample selection bias in the distance equation, it is suggested there should be at least one independent variable that appears in the move equation but not the distance equation (Wooldridge 2009).

The main category of the independent variables used in the analysis is individual (rider) and household socio-economic and demographic (SED) characteristics, including household income, age of individual rider, number of children less than 18 years old, race, ethnicity, automobile ownership, and tenure (own or rent). These variables were used to account for the taste variations in mobility choices between different

household groups, as well as capturing the effects of the life-cycle stage of a household and its orientation toward changing housing needs.

Whereas most of the SED variables are self-explanatory, household income was defined as a set of three income categories (from low, middle, to high) to encompass approximately 30, 40, and 30 percent respectively of all households. To allow for a nonlinear relation between income and dependent variables, each income category was specified as a dummy variable with value of one if a rider belongs to that category and 0 otherwise. Only low- and middle-income categories were entered into the model due to the collinearity problem in model estimation. Similarly, the age variable also was classified by three age categories: $\text{age} < 34$, $35 \leq \text{age} \leq 55$, and $\text{age} > 55$. Young-age ($\text{age} < 34$) category was used as the base category and, therefore, only middle- and old-age were entered into the model.

In addition to SED variables, access mode, workplace location, and length of time using LRT were also included in the models. The choice of these variables was based either on the findings from previous research, or on the statistical tests of the parameter estimates. Finally, constants were included to capture the mean effects of unobserved or unmeasured variables that affected rider relocation decisions, so that the random component of the model would be forced to have zero mean. Because rider information on relocation and transportation costs was not available, the constants would represent largely the costs and, therefore, their signs were expected to be negative.

Estimation Results

The parameter estimation results of the model are presented in Table 1. All the coefficient estimates for variables about which hypotheses were formulated have the expected signs. A positive coefficient for a variable in the move equation means that the likelihood of move increases with an increase in the value of that variable. Similarly, a positive coefficient for a variable in the distance equation means that the distance from a residential location to its nearest transit station increases with an increase in the value of that variable. Most individual coefficient estimates are significantly (at the 95% confidence or more) different from zero, implying that the selected variables can adequately explain variations in the move/distance data. Because the correlation coefficient ρ is statistically significant, the error term e_i in Equation 1 (in Appendix 1) is correlated with the error term ε_i in Equation 2. Therefore, unobserved factors driving riders' moving also have impacts on riders' distance choice, indicating that the effort to construct a Heckman sample selection model is worthwhile.

In the remainder of this section, model estimation results are discussed with emphasis on the effect of the independent variables on rider's mobility decisions. Based on the magnitude of the coefficient estimates, independent variables of age, income, walk access and renter in the move equation have relatively large effects on the rider's decision to move or stay. Similarly, independent variables of age, income, walk access, bus access and automobile ownership in the distance equation have relatively large effects on the rider's decision to determine the residence distance to LRT station.

As shown in Table 1, the constants are negative and highly significant, indicating diminishing utility for riders to make a residential relocation decision. The age variables show that younger adults exhibit a greater tendency than middle-aged and senior riders to move after the introduction of a new LRT service. This is possibly due to the fact that the life-cycle stage of the younger people experience more changes due to marriage and divorce. Furthermore, younger people experience more moves because of lower residential moving costs and more opportunities to get a better residence offer. Regarding residence distance to LRT station, the distance equation shows that senior riders are more likely than younger riders to live nearby LRT stations.

The value of income coefficients in the move equation is smaller with respect to high/middle-income riders and, therefore, increase in income would decrease the probability of a rider making a residential move as a result of the new LRT service. The magnitude of income coefficients decreases with increasing household income but at a decreasing rate, indicating a nonlinear effect of income on rider's move propensity. With regard to residence distance to LRT station, high-income riders are more likely than low-income riders to live closer to LRT stations once they have decided to make a residential relocation. This seems logical because of increases in the market value of residential properties located near transit stations. Note that the variable of middle-income was purposely not included in the distance equation in order to account for selection bias (Wooldridge 2009).

LRT riders who access stations by walking are more likely to move than riders who access by cars or buses. Furthermore, riders with bus access have a lower probability than riders with car access of making a move decision. This is possibly because that bus-access riders already have a direct bus connection to a LRT station and may lose bus availability and convenience after residential relocations. Regarding the residence distance to LRT stations, riders with walking or bus access show a greater propensity to live closer to LRT stations. In contrast, car-access riders are prone to live farther to LRT stations.

The presence of children has a negative effect on the rider's moving propensity and, therefore, as the number of children increases, the riders are less likely to move. Lower mobility rates are expected because the presence of children usually increases the cost of moving. In respect to residence distance to LRT station, a positive children coefficient indicates that the riders with more children are prone to live farther to LRT stations. Riders with a low car ownership level are prone to make a residential move as a result of a new LRT service. This is presumably because the riders with fewer cars are more likely captive to transit and, therefore, would like to move and live in housing within the walking distance of LRT station. For those riders who have decided to move, a lower car ownership level would be associated with riders living in housing near stations than in housing farther away.

Riders with rented housing are more likely to move than homeowners. For the renters higher mobility rates are expected, as the moving costs are lower. Regarding the residence distance to LRT stations, homeowners show a greater propensity than renters to live closer to the stations. This would suggest that the proposed housing developments near the stations should focus on sales rather than rental markets. Hispanic riders have the lower propensity for moves than other ethnic riders. However,

Hispanic movers are prone to like proximity to the stations than other ethnic movers. Similarly, white riders are less likely than nonwhite riders to move, but white movers are more likely than nonwhite movers to live near the stations.

Riders who work in New York City are prone to move than riders who work in New Jersey. For those riders who have decided to move, New York City workers are more likely than New Jersey workers to live further away from the stations. This result indicates that the LRT's regional connectivity might reduce total travel times of the New York City workers and, therefore, they could choose a new residence further away from the stations. Finally, length of time using the LRT service has a negative effect on the rider's moving propensity, implying a lower probability of relocating as the overall length of time reported by riders using LRT service increases. This result may suggest that the earlier the riders decide to move, the more likely they would view transit access (rather than housing cost or quality) as the top reason for moving. Once the riders have decided to move, the longer the riders use LRT service, the less likely they would live further away from LRT stations.

TABLE 1.
Move and Distance Model
Estimation Results

Model	Move		Distance	
Variable	Coefficient	t-statistic	Coefficient	t-statistic
Constant	-2.42	-7.15	-6.89	-6.93
High income	-0.97	-3.53	-1.57	-1.86
Middle income	-0.89	-4.48	--	--
Age > 55	-2.82	-8.31	-3.97	-3.54
35 < Age < 55	-0.77	-3.93	-1.10	-2.03
Children < 18 years	-0.46	-1.59	0.83	1.65
Automobile ownership	-0.29	-4.78	1.22	3.12
Walk access	0.75	4.64	-3.67	-4.61
Bus access	-0.46	-3.31	-2.36	-5.66
Renter	0.49	4.47	1.05	3.34
Hispanic/Latino	-0.39	-3.94	-1.01	-2.65
White	-0.21	-2.01	-0.69	-2.03
Work in New York City	0.33	2.89	0.68	1.81
Length of time using LRT service	-0.18	-4.24	-0.38	-2.65
σ_{ϵ}	--	--	0.91	2.94
ρ	--	--	0.69	3.62
Number of Observations	1,023			
Log likelihood at convergence	-1,899			

Conclusion and Discussion

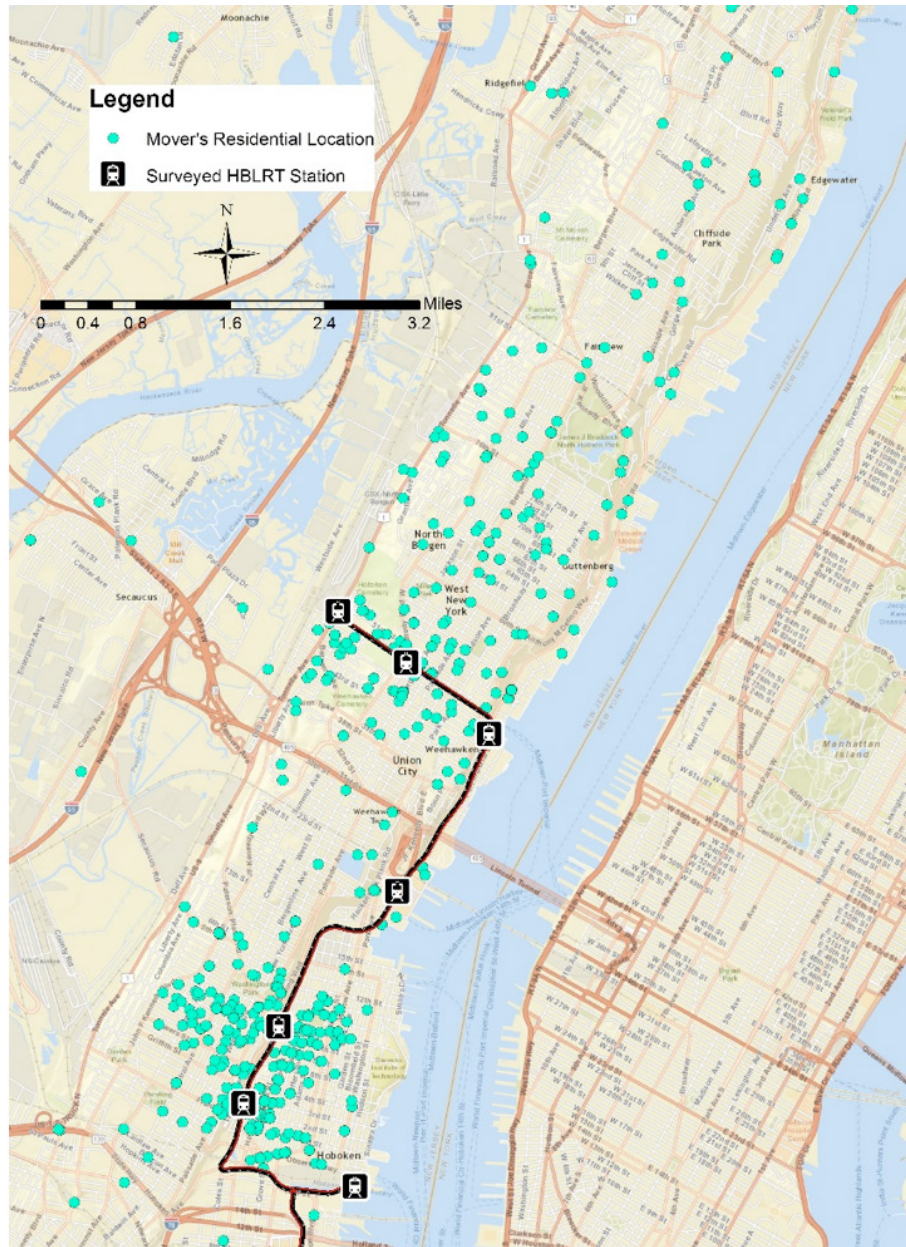
Using the HBLRT rider survey data, this paper conducted a cross-sectional analysis of the residential relocating decisions for riders who have been riding the HBLRT for a period of time. This transit investment opens new residential opportunities because the HBLRT provides the regional connectivity that makes it a viable transportation

option for relocated residents to enjoy easy access to a diverse array of destinations. Two relocating decisions were analyzed, including whether or not a rider would make a residential move as a result of the new LRT service (move choice) and, having decided to move, how far a rider would live away from the nearest LRT station (distance choice). Using socio-economic and demographic (SED) data from the survey, the Heckman sample selection model was used to explore the characteristics of the riders most likely to move and the movers' orientation toward their residence distances to LRT stations. Note that although the current model is largely employed to explain past riders' decisions, it can be used to forecast relocating decisions of the riders who are starting to use the HBLRT or a new LRT system if their SED data are available. This relocation forecasting would provide useful information to help planners/developers plan out housing and amenities for relocating residents, as explained later.

Regarding the effects of demographic characteristics on the move and distance choices, the model reveals that senior riders, Hispanic origin riders, white riders, or riders with more children are less likely to move as a result of the new LRT service. For those riders who have actually moved in response to the new service, senior riders, Hispanic origin riders, white riders, or riders with few children are prone to live near LRT stations. In respect to the effects of socio-economic characteristics, the model indicates that riders with high household income, home ownership, and high auto ownership level have the lower propensity for moves. For those riders who have decided to move, high-income riders, homeowners, or riders with low auto ownership level like proximity to LRT stations. The model also indicates that the riders who access stations by walking are more likely to move than riders who access by cars or buses. Walking-access riders show a greater propensity (than car- and bus-access riders) to live closer to LRT stations. Compared to riders who work in New Jersey, riders who work in New York City are prone to move and incline to live further away from the stations. Finally, the longer riders have used LRT services, the less likely they would move and more likely they would live near LRT stations.

By knowing the characteristics of riders most likely to move and the movers' residence distances to LRT stations, both planners and developers can identify areas where housing growth associated with specific characteristics of riders will occur, and plan for these areas to provide housing and amenities for relocating residents. More specifically, planners would understand the role and influence of LRT on residential moving behavior and the spatial redistribution of movers' residences. According to survey results, the movers' residential locations are shown as green circles in Figure 2. Although our model is not a residential location model, it can estimate how far a mover would live away from a LRT station. The model shows that 34% of the movers choose to live within 0.5 mile road network distance of a LRT station; 63% live within 1 mile of a station; 77% live within 2 miles of a station; 87% live within 5 miles of a station; and 13% live beyond 5 miles of a station. The average distance from movers' residences to LRT stations is 2.4 miles and the median is 0.7 mile. This information would allow planners to undertake planning efforts at specific locations by providing housing and amenities for relocating riders. Planning efforts may include the land use plan related to the type, location, density and intensity of development, and the transportation plan related to designation of a system of motorized and non-motorized travel that supports the land use plan.

FIGURE 2.
Distribution of Mover Residential Locations



On the one hand, survey results show that 56%, 24%, and 20% of the movers would access stations by walking, buses, and cars, respectively. On the other hand, the model indicates that the average distances between movers' home and stations are different along the mover's choice of access mode. The average distances for walking, bus, and car movers are 0.51, 2.53, and 7.59 miles, respectively. As described before, approximately 34% and 63% of the movers choose to live within 0.5 and 1 mile road distance of a LRT station. All of this information suggests that planners consider the integration of land use and transit development (e.g., developing transit-friendly neighborhoods to support transit infrastructure) within 0.7 or 0.8 mile road distance of a LRT station.

By knowing the characteristics of movers and their residence distances to LRT stations, developers would know the challenges and opportunities for developing specific housing types within various road distances of a LRT station to serve different income levels, family structure, and ownership. For those areas nearby the stations, for example, useful modeling information for the developers include:

- Whereas high/middle-income riders and homeowners are less likely to move, they are associated with those riders who would live closer to the stations provided they decide to move. In contrast, low-income riders and renters have a higher propensity to relocate, but are associated with those riders who would live farther to the stations. This implies that the greater share of new housing units being created near stations should be planned for sales (rather than rental) markets and for high- and middle-income households.
- Senior riders are less prone to relocate, but are associated with those riders who would like proximity to stations once they decide to change residence. In similar conditions, riders with more children and higher automobile ownership levels are less likely to move and more prone to live farther to stations. This result indicates that the new housing units (and their associated amenities) being created near stations should be designed for accommodating smaller household size, senior residents, and households with lower auto ownership level.

Therefore, instead of developing a diversity of housing types and price ranges in all areas, each area's development should be tailored to meet the housing demand associated with specific characteristics of the movers, including ages, races, income levels, ownership, and family status. Once knowing the characteristics of the movers in an area, planners and developers can develop policies and plans to offset the rising cost of land, keep housing affordable, and yet allow high quality housing to be built.

References

- Cao, J. and J. Schoner. 2013. *Transportation Impact of Transitways: A Case Study of Hiawatha Light Rail Transit in Minneapolis*. CTS Report 13-13, Center for Transportation Studies, University of Minnesota.
- Cervero, R. and M. Duncan. 2008. *Residential Self Selection and Rail Commuting: A Nested Logit Analysis*, Working Paper, University of California at Berkeley.
- Greene, W. H. 2000. *Econometrics Analysis*. Prentice-Hall, Upper Saddle River, NJ.
- Heckman, J. J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47: 153-161.
- Holguin-Veras, J., R. E. Paaswell, W. Aviles, V. Ochieng, and C. McKnight. 2002. *Multivariate Analysis of the Relationship between Transit Accessibility and Residential Location Choice*. Working paper No. 11, University Transportation Research Center, New York, February.

Lund, H. 2006. Reasons for Living in a Transit-Oriented Development, and Associated Transit Use. *Journal of the American Planning Association*, Vol. 72, 357-366.

Ochieng, V., H. S. Levinson, R. E. Paaswell, J. H.-Veras, and C. McKnight. 2002. *The Residential Relocation Impacts of Midtown Direct A Descriptive Analysis*. Working paper No. 10, University Transportation Research Center, New York.

Olaru, D., B. Smith, and J.H.E. Taplin. 2011. Residential Location and Transit-Oriented Development in a New Rail Corridor. *Transportation Research*, 45A: 219-237.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26: 24-36.

Wooldridge, J. M. *Econometrics*. India Edit. Michigan: Cengage Learning, 2009, pp. 459-461.

Appendix 1. Model Structure

The Heckman's sample selection model contains two latent variable equations: the move equation (decision to move or stay) and distance equation (residence distance to the nearest LRT station). They are defined as follows:

$$d_i^* = \alpha' z_i + e_i, \quad i = 1, 2, \dots, N \tag{1}$$

$$d_i = 1, \text{ if } d_i^* > 0; \quad d_i = 0, \text{ if } d_i^* \leq 0$$

$$y_i^* = \beta' x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad n < N \tag{2}$$

$$y_i = y_i^*, \text{ if } d_i = 1; \quad y_i = 0, \text{ if } d_i = 0$$

where d_i^* and y_i^* are the latent dependent variables for individual rider i ; N and n are the numbers of observations in the full and selected samples, respectively (Note: The selected sample includes uncensored observations ($d_i = 1$) only); z_i and x_i are vectors of observed explanatory variables; α and β are vectors of parameters to be estimated; and e_i and ε_i are the error terms assumed to be correlated through a correlation coefficient ρ , independently of (z_i, x_i) , and bivariate normally distributed, with zero mean and unknown covariance matrix.

$$\begin{bmatrix} e_i \\ \varepsilon_i \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix} \right) \tag{3}$$

where σ_ε^2 is normalized to 1 for identification purposes. The latent variable d_i^* in Equation 1 represents the rider i 's propensity to make a residential move as a result of the new LRT service, and it is continuous and unobserved. Instead, we observe the binary realization d_i , which takes the value $d_i = 1$ when $d_i^* > 0$ (move occurs), and $d_i = 0$ when $d_i^* \leq 0$ (no move occurs). The latent variable y_i^* in Equation 2 contains the Manhattan distance information (i.e., the shortest road distance (in mile) from residence to the nearest LRT station) of those riders for which the realization variable $d_i = 1$, that is $y_i = y_i^*$ when $y_i^* > 0$, otherwise their information is unobservable ($y_i = 0$).

The standard procedure to estimation of the model is the Heckman two-step estimator. The resultant estimates are consistent but not asymptotically efficient under the normality assumption. More efficient estimates can be obtained using the full information maximum likelihood (FIML) approach (Greene 2000). The log-likelihood for the full sample of observations is given as follows:

$$\ln L = \sum_{d_i=0} \ln \Phi(-\alpha' z_i) + \sum_{d_i=1} \left[-\ln \sigma_\varepsilon + \ln \phi\left(\frac{y_i - \beta' x_i}{\sigma_\varepsilon}\right) + \ln \Phi\left(\frac{\alpha' z_i + \rho \sigma_\varepsilon^{-1} (y_i - \beta' x_i)}{\sqrt{1 - \rho^2}}\right) \right] \quad (4)$$

Maximization of this function produces simultaneous estimation of the parameters of both the move and distance equations (i.e., α , β , ρ , and σ_ε). Compared with the Heckman procedure, the FIML estimator is computationally intensive and difficult to numerically find the maximum values. Hence, good starting values already close to the true parameter values become very important. In this study, the final values of the Heckman procedure were used as the starting values for the FIML procedure.

About the Authors

YOU-LIAN CHU (you-lian.chu@parsons.com) is currently a senior supervising transportation planner/engineer with Parsons in New York City. He received a Master of Architecture from Yale University, a Master of City and Regional Planning from Harvard University, and his Ph.D. in Transportation from the University of Pennsylvania.

YI DENG (yi.deng2@nyct.com) is a principal transportation planner in the Division of Operational Planning at MTA New York City Transit. She received her Ph.D. in civil engineering from New Jersey Institute of Technology and a Master degree in Transportation from Tongji University. She is a current member of the TRB Committee on Automated Transit Systems.

RONGFANG (RACHEL) LIU (RLiu@njit.edu) is a professor in the Department of Civil and Environmental Engineering, New Jersey Institute of Technology (NJIT). Her research and consultation interests are in the areas of multimodal and intermodal transportation planning. She has published extensively in various journals and conference proceedings and her new book is *Automated Transit: Planning, Operation, and Applications* (Wiley, 2016), the first in-depth, comprehensive analysis on automated and/or driverless transit.



Center for Urban Transportation Research

College of Engineering
University of South Florida
4202 E. Fowler Avenue, CUT100
Tampa, FL 33620-5375
(813) 974-3120, fax (813) 974-5168