DEVELOPMENT OF A KEY PERFORMANCE INDICATOR SYSTEM TO BENCHMARK RELATIVE PARATRANSIT PERFORMANCE

Lindsey Morse, Corresponding Author
Railway and Transport Strategy Centre (RTSC), Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
Senior Research Associate; Project Manager, American Bus Benchmarking Group
Skempton Building, SW7 2AZ London
Tel: +44 (0)20 7594 6091; Fax: +44 (0)20 7594 6102; Email: l.morse@imperial.ac.uk

Mark Trompet
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
Associate Director, Head of Bus Benchmarking
Skempton Building, SW7 2AZ London
Tel: +44 (0)20 7594 1519; Fax: +44 (0)20 7594 6102; Email: m.trompet@imperial.ac.uk

Alexander Barron
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
Associate Director, Head of Metro Benchmarking
Skempton Building, SW7 2AZ London
Tel: +44 (0)20 7594 3974; Fax: +44 (0)20 7594 6102; Email: alexander.barron@imperial.ac.uk

Daniel J. Graham
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
Professor of Statistical Modelling
Skempton Building, SW7 2AZ London
Tel: +44 (0)20 7594 6088; Fax: +44 (0)20 7594 6102; Email: d.j.graham@imperial.ac.uk

Word count: 6,174 words text + 5 table/figures x 250 words (each) = 7,424 words

TRB Paper 17-01548
Accepted for publication in Transportation Research Record: Journal of the Transportation Research Board, No. 2650 (2017). DOI: 10.3141/2650-01.

November 15, 2016

ABSTRACT

The Americans with Disabilities Act of 1990 prohibits discrimination against people with disabilities. US transit agencies are therefore required to offer services to eligible customers that complement the mobility opportunities provided to the general public on fixed-route public transit. While these paratransit services are necessary and just, they represent a proportionally large cost to agencies: approximately eight times the cost per boarding compared to fixed-route bus service. To be able to identify opportunities for (cost) efficiencies, and to further improve the quality of paratransit services offered, the twenty agencies of the American Bus Benchmarking Group (ABBG) decided to benchmark their relative performance in paratransit management and operations. To ensure comparability of agencies' performance and hence ensure the usefulness of the benchmarking program, a key performance indicator system was developed and associated data items were defined in detail. The scope of this system went beyond the data already provided to the National Transit Database, both in amount and granularity of data collected, as well as the detail of definitions. This paper describes the challenges, respective solutions, and other lessons identified during four years of paratransit benchmarking development led by Imperial College London, the ABBG facilitators. The paper provides transit agencies and authorities as well as benchmarking practitioners and academics an opportunity to apply these lessons for the further benefit of paratransit services and their customers around the U.S.

Keywords: Paratransit, performance, benchmarking, demand-response transit, indicators, data quality, comparability

INTRODUCTION

The Americans with Disabilities Act of 1990 (ADA) prohibits discrimination against people with disabilities in employment, public accommodation, communications, governmental activities, and transportation. For US transit agencies this means that transportation services have to be offered to eligible customers that complement the mobility opportunities provided to the general public on fixed-route public transit. These paratransit services have an important social and economic function in society and, with an aging population, demand is generally increasing. Costs to provide paratransit service, as well as service quality needs, are high and therefore any cost efficiencies and quality improvements that can be identified are welcomed by transit agencies. The members of the American Bus Benchmarking Group (ABBG) therefore identified paratransit service as an important priority area for performance benchmarking and best practice sharing.

Founded in 2011, the ABBG is a group of public medium-sized urban bus operators that agreed to compare performance and share best practices with peers in similar cities throughout the United States. The ABBG is managed and facilitated by the Railway and Transport Strategy Centre (RTSC) at Imperial College London and is modeled on the International Bus Benchmarking Group (IBBG), a global group of 15 urban bus systems that was founded in 2004. The ABBG initially focused on performance benchmarking of fixed-route services, drawing upon lessons learned from the IBBG (1, 2, 3), but agreed from the start that paratransit benchmarking would also be a key group objective. Subsequently in 2012 the group added the development of paratransit performance benchmarking to the overall research efforts.

As of 2016 the ABBG is comprised of twenty agencies:

- Austin Capital Metro
- Buffalo NFTA
- Chicago Pace
- Cleveland GCRTA
- Des Moines DART
- Dayton RTA
- Eugene Lane Transit
- Flint MTA
- Fort Worth FWTA
- Hampton Roads Transit
- Orlando LYNX
- Nashville MTA
- Rhode Island RIPTA
- Rochester RTS
- San Bernardino Omnitrans
- San Joaquin RTD
- Spokane STA
- St. Petersburg PSTA
- Salt Lake City UTA
- Vancouver C-TRAN

All these agencies have contributed to the research in this paper, but performance data from Nashville MTA and Flint MTA were not yet available at the time of writing, as these organizations joined the ABBG in October 2015.

On average, the ABBG organizations provided one round-trip per week to their active customers in 2014; these trips represent a wide range of connections to medical care, education, jobs, and other activities and services that support the physical, mental, and economic health of the individuals. Across the group, the paratransit operating cost per boarding was $34 in 2014, compared to $4 per boarding for fixed-route service, highlighting the need for benchmarking to find cost efficiencies while maintaining service quality.

Building upon years of IBBG know-how and ABBG fixed-route benchmarking experience, a paratransit key performance indicator (KPI) system was developed from 2012-2016. To ensure comparability, the scope of this system went beyond the data already provided to the National Transit Database (NTD), both in amount and granularity of data collected, as well as the detail of definitions. In developing this

paratransit benchmarking system, seven key challenges in performance indicator development, data collection, and definitions were identified that needed to be overcome:

- *Scope:* The diversity of demand-response services and their respective funding programs and regulations make it a challenge to ensure that similar services are compared.
- *Operating models*: In-house, contracted, and taxi services may have different data availability, definitions, and performance, all of which are essential to understand the advantages and disadvantages of each model.
- *Boardings*: Consistently measuring passenger boardings was found to be challenging but important for benchmarking, even if considered secondary to origin-destination trips.
- *Trip classification*: Understanding and defining the full universe of trip types – from the initial customer request to the completed trip– was necessary to ensure data comparability.
- *Local contextual information:* Each agency operates in a specific context and is shaped by many local factors, which must be identified in order to understand differences in performance.
- *Policy differences:* Different agency policies, such as for on-time performance windows, required careful consideration on how to achieve comparability.
- *Defining performance:* The KPI system has to provide a concise, but balanced and holistic view of performance throughout the business, while identifying performance trade-off areas.

This paper describes these challenges and respective solutions identified during four years of paratransit benchmarking. To facilitate the discussion, the remainder of the paper is organized as follows: The literature review provides an overview of available academic research and government publications with regards to paratransit performance benchmarking, and identifies gaps that this research aims to address. The methodology section that follows describes the different stages of the ABBG paratransit performance benchmarking research. Thereafter each of the seven identified KPI development challenges and respective solutions will be discussed in detail. Finally, the paper will provide conclusions and offer opportunities for future work and research.

LITERATURE REVIEW

Most relevant documentation about ADA paratransit services are in the public realm, in the form of government-funded reports, and suitably concentrated in the U.S. In addition to these government reports, relevant academic peer-reviewed literature was examined. The most relevant publications are summarized below, but are also referenced throughout the paper where relevant.

There is a fairly large body of international work on demand-response transit (DRT), also known as flexible transport services, the history and status of which is summarized in Nelson et al. (4). Most publications primarily focus on public DRT and the effectiveness of specific strategies and technology aids, such as scheduling software or the use of zones. Research interest in this area was re-invigorated in 2014 by the TRB-sponsored International Paratransit Conference held in Monterey, CA, the first such gathering since 1997. The proceedings and significance of this conference are assessed by Mulley & Nelson (5). There have been few publications discussing DRT benchmarking and performance measurement, and these are described below.

Four studies were identified that have direct relevance to paratransit benchmarking, all of which relied on DRT systems data from the NTD, and applied statistical methods to explain relative performance or identify the best systems. Research by Dessouky et al. (2004) evaluated the impact of technology (e.g., CAD, AVL, automation) and management practices (e.g., financial incentives/penalties) on productivity by using linear regression to compare NTD data with survey responses from 62 large US transit agencies and 13 small California transit agencies (6). In a follow-up study (7), the authors further evaluated similar

strategies using NTD data and four paratransit-specific measures (booking window length, on-time window length, % cancelled, % no show) based on survey responses from 67 DRT agencies. Two other studies used data envelopment analysis (DEA) to measure the comparative efficiency of paratransit systems. Fu et al. (2007) looked at three inputs (total number of paratransit employees, total fuel expenses, and total number of vehicles used for paratransit services) and a single output measurement (revenue vehicle kilometers) to evaluate the efficiency of paratransit systems in 32 Canadian cities (8). Min & Lambert (2010) used NTD data for 75 US agencies, with a focus on explaining city characteristics that impact paratransit service efficiency (9).

Only one of these studies (7) used paratransit-specific measures and none looked at any output, or performance objectives, beyond efficiency (cost). These studies also did not consider the effects of the operating model or local fixed-route service characteristics.

Three government publications were identified that have the most direct relevance to paratransit benchmarking. The most comprehensive guide available for performance measures for DRT is TCRP Report 124: Guidebook for Measuring, Assessing, and Improving Performance of Demand-Response Transportation (10). It covers a wide range of performance measures, including those specific to ADA paratransit, and the data and comparability challenges posed by each. Although this guide is comprehensive in its consideration of measures and challenges, it does not provide a framework by which to consider the long list of measures, as is provided in this paper, nor does it fully test its concepts in application to the ADA paratransit industry. TCRP Synthesis 74 complements the Guidebook by documenting best practices in delivering paratransit services from 124 survey responses and 17 interviews (11). However, this study does not discuss performance measures. Finally, a 2014 Federal Transit Administration (FTA) report, Accessible Transit Services for All, summarizes survey results from 198 agencies (12). It covers ADA paratransit service delivery model characteristics as well as the use and effectiveness of specific strategies to improve schedule efficiency (e.g., staff training to select the most efficient trip placements), or to reduce no shows (e.g. customer incentive program). The report only covers a few performance measures, and only provides averages or ranges for a subset: cancellations, no shows, boardings per revenue vehicle hour, and cost per revenue hour. The report instead focuses on whether targets and financial dis/incentives are offered for customer (e.g., on-time performance) and safety measures. Although the report focuses on comparing cost efficiency across service delivery models, it does not cover cost effectiveness (cost per boarding and passenger mile) nor does it include models that use both in-house and contracted operations, which five ABBG agencies use. By combining relevant information and lessons from each of these three documents, useful input was created for the ABBG paratransit performance benchmarking research presented in this paper.

There has also been significant work done on public transport benchmarking more broadly, as documented in Trompet et al. (2). This publication found benchmarking to be useful and justifiable because there is sufficient variability in performance between operators due to factors within management control, rather than only those resulting from differences in external factors, thus allowing for operators to learn lessons from one another.

The National Transit Database (NTD) provided a starting point for ABBG's fixed-route benchmarking. In terms of ADA paratransit, the NTD has limited coverage and relevance, and thus TCRP Report 124 played a much larger role in providing a resource for inspiration and verification. NTD collects data for DRT, separated into directly operated and purchased non-taxi DRT and taxi-provided DRT. It collects the same data for DRT as for other modes, primarily consisting of financial and basic operational data (i.e., vehicle miles and hours, passenger trips). There are no data items specific to paratransit services (e.g., ADA eligibility approvals and denials, comparison of scheduled vs. completed trips, etc.). For ADA-specific

data, NTD only requires reporting of the proportion of trips and expenses attributable to ADA paratransit services.

In a 2012 report on ADA paratransit demand (13), the Government Accountability Office (GAO) reviewed the NTD's ADA-related data and "found that, according to GAO standards for data reliability, the data are not sufficiently reliable for the purpose of assessing changes in ADA paratransit demand and costs." In 2014, the Federal Transit Administration (FTA) proposed reporting a subset of ADA paratransit data (FTA–2014–0006) but withdrew the proposal in 2015 due to comments that such a requirement would impose a considerable burden on agencies because of the high level of integration between ADA and other DRT. This concern was certainly a consideration when defining the scope for paratransit benchmarking in the ABBG. However, ABBG agencies determined it was worthwhile to voluntarily commit to an additional effort of data collection, KPI development, and validation for paratransit, given its importance to society and the transit agencies themselves.

METHODOLOGY

Establishing a robust KPI system that uses a comparable dataset and obtaining the understanding of underlying differences in performance takes time. The KPI development process took approximately three years for both the IBBG and ABBG fixed-route KPI systems. The development of the paratransit KPI system greatly benefited from lessons learned from these projects and from existing literature. The ABBG fixed-route benchmarking also provided very important context and baseline from which to review and understand agencies' paratransit efforts. However, with limited NTD data available as a starting point, the initial ABBG paratransit KPI development process also took three years (2012 to 2014). The RTSC at Imperial College London worked with agency senior executives during these years to establish the initial set of KPIs and their respective data and definitions over multiple rounds of comprehensive indicator and data review. This necessary groundwork was followed by KPI development workshops in 2015 and 2016, where agency paratransit experts provided detailed feedback on content and feasibility and offered suggestions, leading to the current agreed-upon performance benchmarking system.

The full timeline of ABBG paratransit benchmarking development is as follows:

- Phase 1 (2011-12): Agency's own paratransit indicators and definitions were obtained for comparison and a survey on paratransit operations of the 12 initial agencies was conducted. Based on this research, agency leaders agreed on the initial paratransit KPI set and underlying data items to be collected during Phase 2.
- Phase 2-3 (2012-14): Initial results of paratransit KPI development were presented at the Phase 2 meeting, leading to lessons which were implemented and presented to agency leaders in Phase 3. The first paratransit KPI report with data from 17 agencies was issued in August 2014.
- Phase 4 (2014-15): Another KPI development survey was conducted in 2014, and the results were presented alongside 2013 KPI data at the Phase 4 Meeting. There, executives agreed more detailed work at a workshop of paratransit experts was needed to further develop paratransit KPIs.
- Phase 5 (2015-16): A paratransit KPI development workshop was held in May 2015. The workshop had 32 participants from 16 participating organizations, including 11 managers or directors of paratransit. The workshop results and revised 2014 KPIs were presented back to agency leaders at the Phase 5 ABBG Meeting in September 2015. It was then agreed to hold an annual expert paratransit workshop to allow for the exchange of best practices and continued KPI development.
- Phase 6 (2016-17): The second expert paratransit workshop was held in May 2016. The workshop had 27 participants from 15 member organizations, with nine managers or directors of paratransit

among other paratransit operations and data analysis staff. Currently, 2015 data is being collected based on refinements identified at this workshop.

As a result of this development and research, the 20 ABBG agencies are now providing paratransit data on an annual basis for approximately 150 clearly defined data items, some of which are further broken down by service delivery type (in-house, contracted, taxi). The data is a mix of performance-related data and background data items that provide necessary context and understanding. Up to 8 years of data, from 2007 to 2014, are currently available. Data collected during the earlier development years was revised based on definitional changes and resubmitted to ensure data trends are consistent.

As documented in Trompet et al., one of the main contributing factors to data quality is an open and honest sharing environment that is supported by a strict confidentiality agreement, which ABBG has adopted (2). As a result, the graphs in this paper have been anonymized where necessary. Nevertheless, the ABBG agreed that important lessons from its experience in paratransit benchmarking should be shared for the benefit of the wider public transit industry and policymakers. The following section of this paper therefore discusses the seven main challenges and respective solutions found during the ABBG paratransit performance benchmarking development process.

DISCUSSION

Challenge 1: Scope

As mentioned in the literature review, there are a wide range of demand-response services, not all of which are directly comparable. Through the ABBG paratransit benchmarking development, it was agreed to focus on a subset of these services, as shown in white in Figure 1.

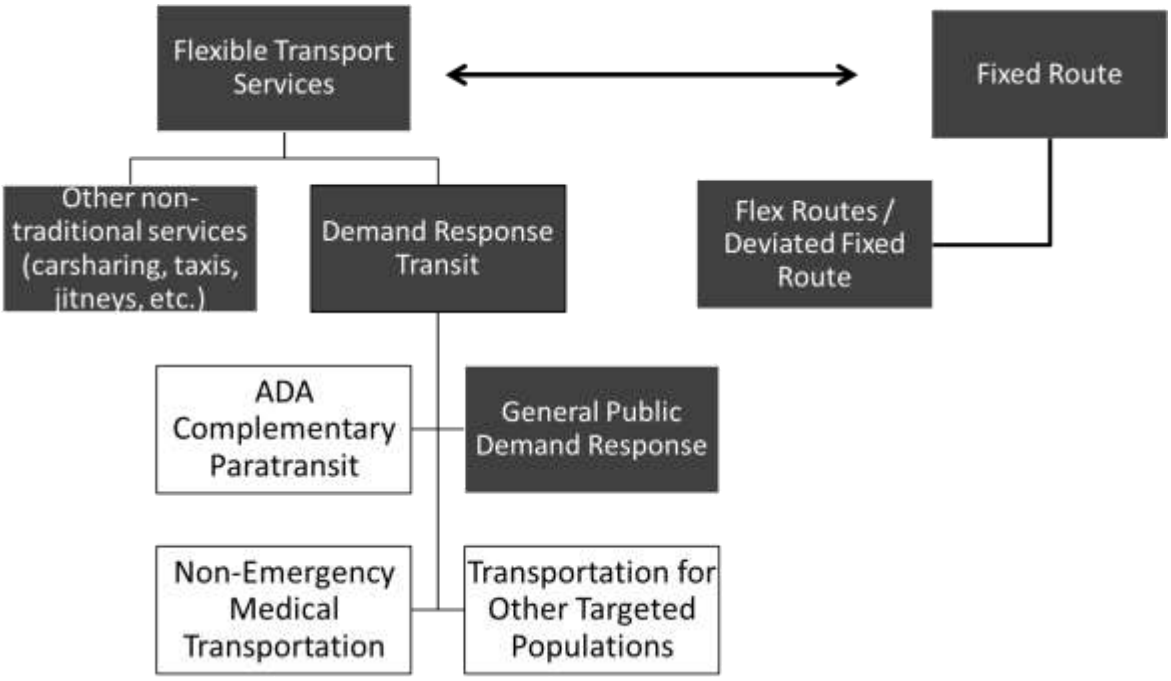

FIGURE 1 Diagram showing the types of services included in the ABBG paratransit dataset (in white).

The group has consistently agreed that the primary focus of the ABBG paratransit data is on 'complementary ADA paratransit services'. Unlike NTD's DRT data, the ABBG complementary paratransit

dataset excludes general public demand-response services, but includes other limited-eligibility services, which are explicitly excluded by NTD from the ADA data. Conceptually, it was agreed that general public DRT could be included if the service is operated in combination with ADA paratransit (e.g., same vehicles, drivers, maintenance, etc.) and the data cannot be separated easily, but currently no ABBG member has this setup.

Other limited-eligibility services include Medicaid's non-emergency medical transportation and non-medical transportation (offered under the Home and Community Based Services waiver) and state or county programs with eligibility standards that differ slightly from ADA (e.g., low-income). The inclusion of such services is driven by two main considerations: similarity and overlap in service characteristics, resources used, and customers served, and the difficulty in separating the data because of the high level of integration. In practice, only a few ABBG agencies include these services in the dataset. Many ABBG agencies include services for ADA-eligible customers that are purchased or funded by external entities.

Consistent with NTD, routes that generally have a fixed alignment or path but that either deviate a certain distance upon request for pick-up or drop-off or that have designated zones in which curbside pick-up can be requested are included in the fixed-route dataset rather than the paratransit dataset. This is true even in the case where deviations can only be requested by ADA paratransit customers.

The current scope has its data challenges, as some costs and functions are shared across services. For example, one ABBG agency shares vehicles and drivers between its deviated fixed-route and paratransit services. Other agencies share maintenance facilities with fixed-route services or have separate demand-response services, and all agencies with in-house paratransit operations share administration functions with fixed-route service. However, agencies are able to proportion the data between these services (e.g., with vehicle mileage, vehicle hours, or staff hours) as appropriate.

Challenge 2: Operating Models

Although several publications have focused on a comparison between the performance of in-house and contracted operations (e.g., 12), few have included or discussed split operation models. Through the paratransit benchmarking development process, it became clear that for split operation agencies, it is important to separate in-house and contracted data to understand data availability, ensure comparability of data, and to a lesser extent, understand differences in service provider performance. The ABBG has three basic operating models: in-house, contracted, and split (mix of the two), with taxi services as a third service provider that can exist in any of the three models and that also requires separate scrutiny (Figure 2).

Within both the contracted and split operating models, there are those organizations that own all or a portion of the vehicles used by the contractors, and there are also organizations that operate the reservation system, scheduling, and call center in-house, while others contract out some or all of these functions. Given these complexities, collecting data per each service provider separately is necessary. Fortunately, most ABBG agencies with split operations are able to provide the full data breakdown, but some are currently not able to provide safety or cost subcategory data for contractors.

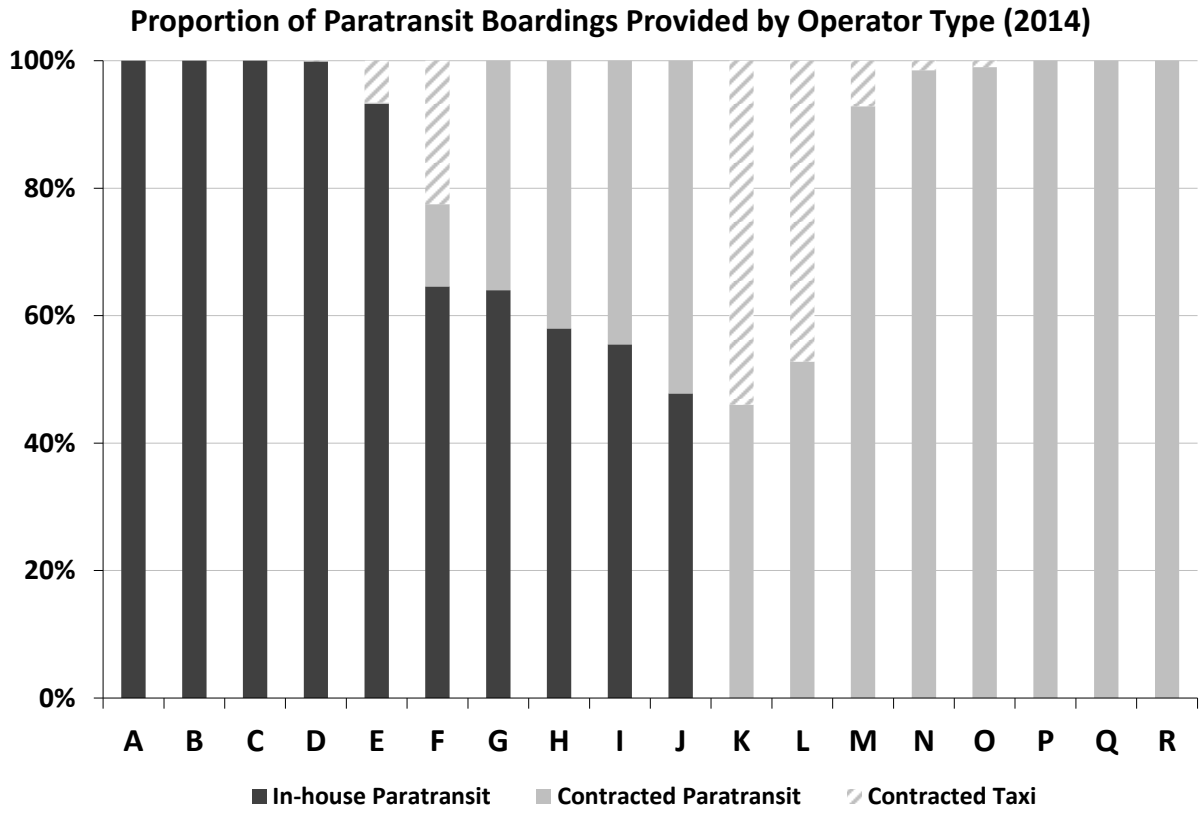

FIGURE 2 Graph showing the proportion of boardings provided by in-house (dark grey), contracted (light gray), and taxi (light grey hash).

Although taxi services are also a form of contracted or outsourced service, it was decided that it was valuable to split out taxi data separately, consistent with NTD. This is due to both data availability (to ensure data are normalized correctly, and numerators matched denominators) and the different role that taxis play in paratransit as a non-dedicated fleet. In terms of data availability, only two agencies are able to provide safety data for taxis and only one includes taxis in on-time performance. In addition, fleet data are only collected for dedicated in-house and contracted fleets. In terms of data comparability, taxi services only report revenue hours and miles, which represents the time and distance covered when a customer is in the vehicle. Revenue data for in-house and contracted service, on the other hand, includes time and distance between trips. Deadheading miles and hours, reflecting travel to/from the depot or other dispatch point, are also collected.

In addition to data challenges, use of contracted services for paratransit (especially taxis) also creates challenges of oversight, regulation, and market competition, although there are also benefits in terms of responsiveness and cost efficiency (14). This is an aspect of paratransit operations that is currently evolving, as several ABBG agencies have reported that local taxi companies have been forced out of business by the rise of transport network companies (TNCs) such as Uber.

Challenge 3: Boardings

Ensuring that agencies are reporting all relevant boardings (unlinked passenger trips) is a challenge identified and discussed in TCRP Report 124 but tested and expanded upon in ABBG. There are three possible categories of passengers that can board under ADA:

- The ADA paratransit customer, who has been certified as qualified to use the service by the agency, and pays the paratransit fare;
- The personal care attendant (PCA), who is designated or employed to assist the customer, the need for which should be documented on the ADA paratransit customer's profile, and who does not pay a fare; and
- Companions, including family or friends, with whom the ADA paratransit customer would like to travel, who generally pay the paratransit fare (there can be exceptions for young children).

The primary focus from external stakeholders is on ADA customer boardings, as they best reflect both the legal requirements and social objectives of the service. In addition, federal law requires that no fare is charged to PCAs, thus providing further incentive not to track their boardings. However, for agencies themselves, it is important to track all boardings in order to ensure sufficient capacity is available and to maximize scheduling efficiency, which also leads to cost efficiencies.

During the ABBG paratransit benchmarking process, it has been discovered that not all organizations are able to report actual boardings by PCAs and companions as they do not all have a consistent, accurate way to capture actual boardings. Some rely only on what is recorded in the reservation software, without verification of the number of people who actually board (as some of the non-ADA customers may ultimately not take the trip). Others rely on manual counts recorded in driver manifests.

In a comparison of ABBG and NTD data for 2014, several discrepancies were found: some agencies included PCAs and companions in the boardings reported to NTD while others did not, and some included non-ADA paratransit services while others did not. Furthermore, at least one agency is including service animals in its boardings. This is done in order to account for space taken onboard the vehicle.

Based on this challenge, ABBG agencies have now been asked to provide a breakdown of boardings by type of passenger (including 'other' for service animals) and to explain the counting methodology. This approach will lead to improved data for use as a key normalization factor for KPIs. As discussed in the next section, significant work has been done to improve the accuracy of 'completed trip' data. This allows for verification of the boardings data and provides a complementary normalization factor to boardings. Completed trips provide a better understanding of scheduling efficiency, while boardings indicate scheduling effectiveness.

Consistent with the scope discussion, the group has also agreed that boardings of paratransit-eligible customers on services that are included in the fixed-route KPIs, including flexible or deviated routes even if specific to ADA customers, are excluded from the paratransit dataset and are included in the fixed-route dataset.

Challenge 4: Trip Classification

Trips, or origin-destination pairs, are the primary determinant of route scheduling and efficiencies (rather than boardings) as they reflect the service provided rather than the number of passengers who may be traveling together (with the same origin and destination). . The ABBG research initially identified a couple key trip outcomes to measure, but found that to ensure comparability, all trip outcomes needed to be collected. Therefore a trip classification framework was developed to capture the reduction in trips from the initial request to the completed trip by cause (Figure 3). The majority of scheduled trips (88% on average) are successfully completed.

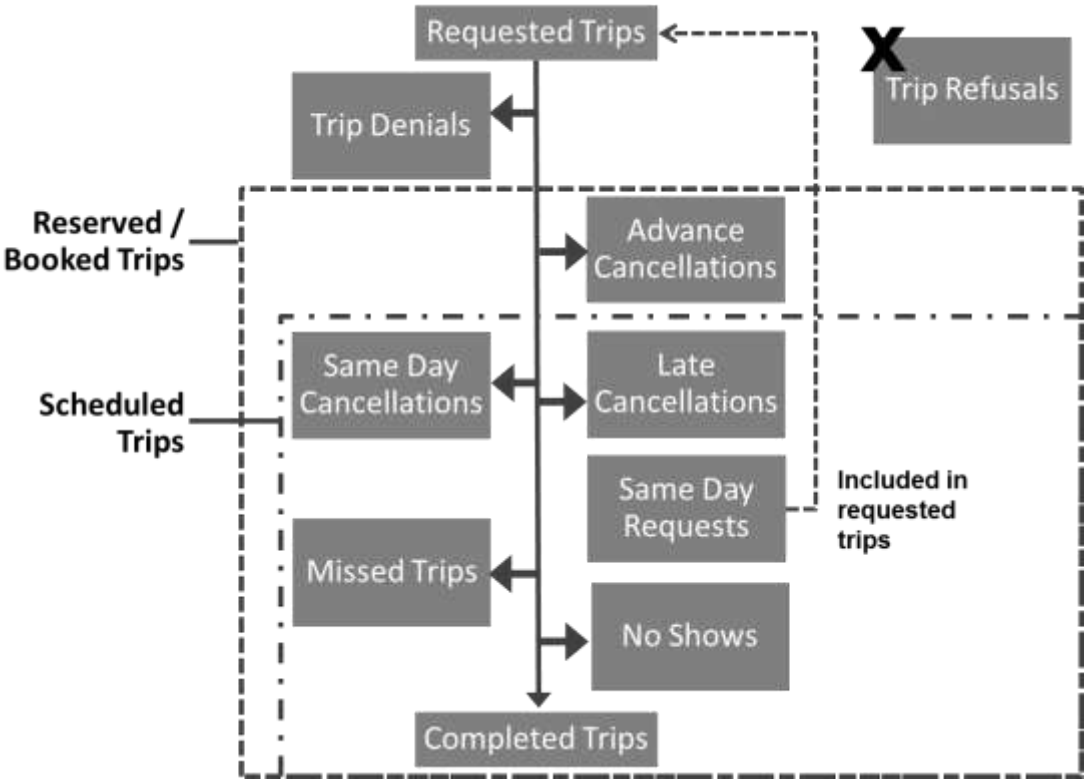

FIGURE 3 Diagram of trip classification from requested trip to completed trip.

All of these events are initiated by the customer except for trip denials and missed trips. Fortunately, trip denials are very rare (<0.1% of requested trips). Missed trips, which have the most direct and significant negative impact on the customer, are also infrequent (<0.2% of scheduled trips). The events with the most significant impact in this diagram are late cancellations, which are defined as a cancellation that occurs within each organization's late cancellation threshold (generally 1 to 2 hours), and no shows, where the customer does not show for the trip. These represent instances when both a driver and vehicle are deployed without any outcome (completed trip, boarding, etc.) and so have the most significant direct negative impact on operator performance. These two events each comprise on average 3% of scheduled trips. Same-day cancellations on average were 6%, but have less impact because operators have time to reschedule resources. One member reported that it purposely overbooks trips by 5% due to its 8-10% same-day cancellation rate, and makes use of late and same-day cancellations to accommodate same-day requests.

Challenge 5: Local Context

As summarized in Trompet et al., one of the necessary benchmarking conditions that contributes to improved quality of the dataset is the collection of background information and data to help understand differences in performance (2). For paratransit benchmarking, this information consists of agency policies, funding sources, community decisions and values, local demand, state and city programs, and the underlying fixed-route system.

Through this project, it has been found that to understand paratransit performance, it is essential to also understand fixed-route service characteristics and performance. Not only does the fixed-route service provide the basis for the extent of paratransit service that is required, but fixed-route characteristics are a central determinant of paratransit demand. Thus, expansions or cuts in fixed-route service impact paratransit (although many agencies continue to offer paratransit in areas that formerly were served by fixed-route service, at least for a year or two). In addition, if more paratransit-eligible customers can be accommodated on fixed-route service (or its flexible counterparts), then fewer such customers will require paratransit services. This is viewed as desirable because of the much higher cost efficiency of fixed-route service as well as the additional mobility opportunities the service provides for customers. The ways in which fixed-route services can better accommodate people with disabilities, such as travel training and improvements in access such as sidewalk enhancements, have been well documented (11, 15) as have strategies to integrate the two services, such as paratransit feeder routes to fixed-route hubs and flexible/deviated routes (15, 16). These are all strategies used by ABBG agencies, and as this project matures, it will be able to study the effectiveness of these strategies over time.

Challenge 6: Policy Differences

The unique characteristics of paratransit service, and the different ABBG agency policies, required careful discussion and consensus to identify both new KPIs and new approaches to known KPIs. Policy examples include late cancellation thresholds, eligibility certification processes, and call center functions. One example that demonstrates this particularly effectively is on-time performance policies.

For paratransit, the customer can request a pick-up time (at the origin) or a drop-off time (at the destination), which is primarily based on an appointment time. Measuring on-time drop-off performance is fairly straightforward, but pick-up performance required the ABBG to consider different agency policies and agree on what would be comparable.

Pick-up performance windows for each ABBG agency (Figure 4) differ in both length (10 to 30 minutes) and thresholds (from 20 minutes before to 30 minutes after). There are five different types of windows. The 0 in the graph represents the pick-up time scheduled. The thresholds represent how the window is communicated to customers. For example, if there was a scheduled pick-up time of 9:45am, the associated on-time window would be between 9:30-10am at agency C and between 9:45am and 10:15am at agency M. The main commonality among these windows is the length, which is 30 minutes or less. FTA considers pick-up windows longer than 30 minutes to be unreasonable because of the waiting time required. In addition, agencies indicated that early arrival of a vehicle (prior to the lower threshold) to pick up a customer did not count against on-time performance. The FTA advises agencies to track earlies and to ensure that customers are not pressured to depart earlier than desired, but considers including earlies in on-time performance to be appropriate. For the purposes of benchmarking, the group agreed that agencies with 30 minute performance windows have comparable on-time performance, regardless of thresholds communicated to customers.

For drop-off performance, the ABBG agreed that any arrival prior to the appointment time is considered on-time, with no early threshold. This may be revisited in the future, as drop-off on-time performance is increasingly attracting attention from FTA and external stakeholders and there is a concern that

customers should not be dropped off too early to appointments, especially in cases where the office is not yet open.

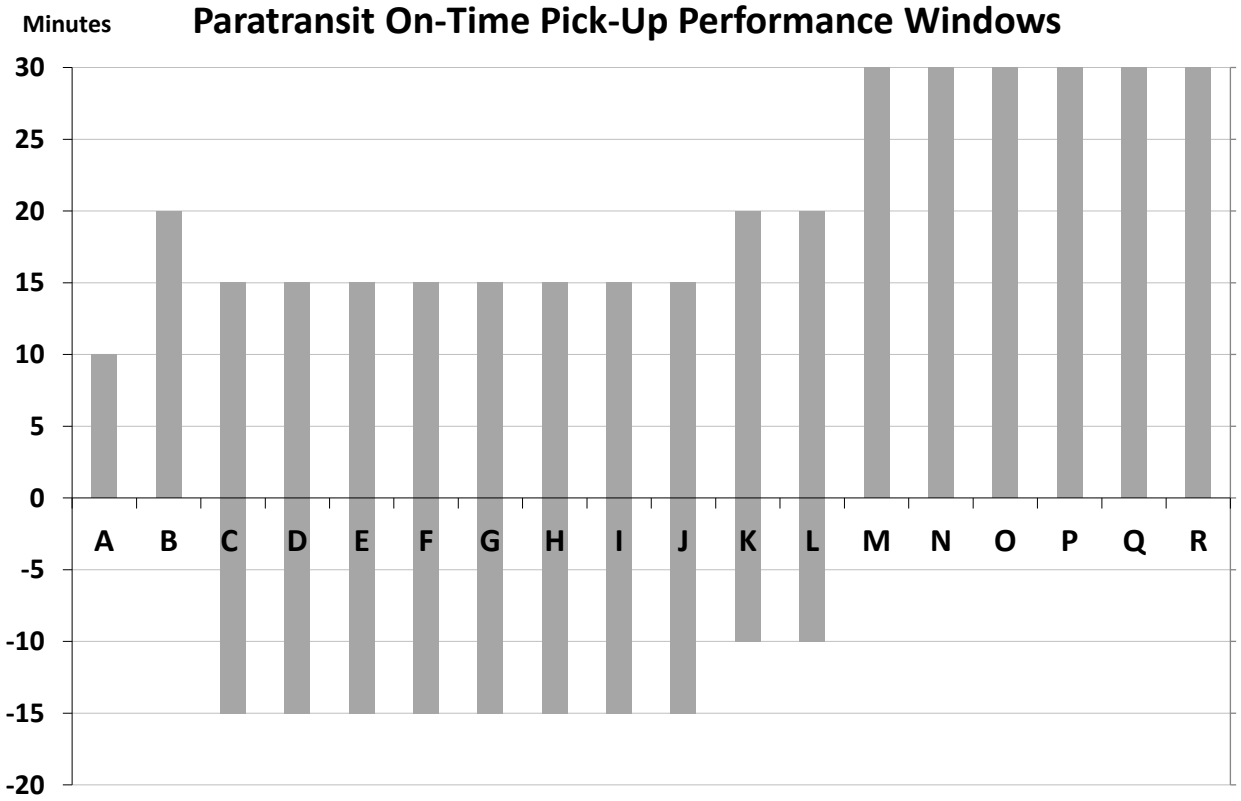

FIGURE 4 Graph of on-time pick-up performance window lengths and thresholds.

Challenge 7: Defining Performance

As previously mentioned, paratransit services are a civil right and perform an essential function in society. Transit agencies must balance the value of this service, and its quality, with cost and efficiency, while providing the minimum level of performance required. The framework for the ABBG paratransit KPI system is the same used for ABBG's fixed-route service and the IBBG. It was developed from the four success dimensions identified in the Harvard Business School Balanced Scorecard model: Financial, Customer, Learning & Growth (later Organizational Capacity) and Business Processes (17). The RTSC at Imperial College adapted this model to transit through identifying relevant and appropriate KPIs for each success dimension, and adding two additional dimensions that are integral to transit performance: Safety and Environment (1). The key principle of the balanced scorecard approach is that, to be successful, any organization must succeed not just in one area but across the whole of the business. To do this, a balance must be struck between the different areas – for example, balancing cost with safety, punctuality with productivity, etc.

Previous experience with fixed-route benchmarking and TCRP Report 124 provided a good starting point for identifying and defining relevant KPIs for paratransit benchmarking. As a result of four years of development, the unique performance needs of paratransit are now captured in the current ABBG paratransit KPI system (Table 1). Trade-offs in performance across this system are complex, and often involve trade-offs with fixed-route performance. For example, increased demand is generally always viewed as a positive objective in fixed-route service and one that will actually improve performance

across other metrics, including cost. In paratransit, reduced demand can be positive if the reduction is due to customers shifting to fixed-route service (and thus to a lower cost service), but can also be negative since it could represent a reduction in mobility and may lead to lower vehicle loads. Integration with other limited-eligibility services, expansion of the service area beyond the ADA requirements, and other agency and community decisions are also essential to consider, as concluded previously.

TABLE 1 ABBG Paratransit KPI System

Success Dimensions	**Key Performance Indicators**
Growth	Ridership: Passenger Boardings/Trips *(5-year % change)* Service Levels: Vehicle Revenue Miles and Hours *(5-year % change)* Ridership Relative to Agency Total Operating Cost Relative to Agency Total Eligibility *(% of applications denied, % approved conditionally and unconditionally)*
Customer	On-Time Pick-up Performance On-Time Drop-off Performance % of Late Cancellations % of No Shows Average Hold Time
Safety	Number of Vehicle Collisions per Vehicle Mile and Hour *(preventable and non-preventable)* Number of Passenger Injuries per Boarding and Passenger Mile
Financial	Total Operating Cost per Passenger, Passenger Mile, and Completed Trip Total Operating Cost per Vehicle Mile and Hour *(service operation, maintenance, taxi cost, administration)* Average Fare per Passenger Operating Cost Recovery
Internal Processes	Peak Fleet Utilization Actual Productivity *(boardings and completed trips per vehicle mile and hour)* Fleet Reliability *(miles between road calls due to technical faults)*
Environmental	Fuel Consumption *(per vehicle mile / passenger mile)* CO2 Emissions *(per vehicle mile / passenger mile)*

CONCLUSION

In conclusion, this paper presents a range of seven challenges, from defining scope to agreeing upon definitions and accounting for contextual differences, and the corresponding solutions identified over four years of paratransit benchmarking development. By undergoing this research, the resulting dataset and KPIs are significantly more comparable and more useful to ABBG agencies. The outcomes and lessons provided in this paper are applicable to transit agencies offering similar services throughout the U.S. and world and when applied offer an opportunity to benefit paratransit services and their customers. The ABBG will continue its research in paratransit benchmarking into the future, and will share further results and details in future papers, including KPI development and definitions and with an increasing focus on practices that lead to successful performance across all the KPI dimensions for both fixed route and paratransit.

Looking forward, it is clear that the role of technology and the rise of shared mobility, such as TNCs, have the potential to dramatically change the paratransit service model, resulting in new challenges and need for additional solutions. This will likely occur both through the integration of technology and the

substitution or addition of delivery models (18). In addition, per the discussion of scope in this paper, there may increasingly be opportunities for integration with other transportation services and customers. It will thus be important to continue to develop benchmarking to accommodate these changes and allow for opportunities for improvement to be identified and shared throughout the industry.

ACKNOWLEDGEMENTS

The authors thank the member agencies of the American Bus Benchmarking Group for their willingness to share its experiences with the development of paratransit benchmarking with the wider transport industry.

REFERENCES

1. Randall, E. R., B. J. Condry, and M. Trompet. International Bus System Benchmarking: Performance Measurement Development, Challenges and Lessons Learned. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
2. Trompet, M., Anderson, R. J., and D. J. Graham. Variability in Comparable Performance of Urban Bus Operations. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2111*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 177-184.
3. Trompet, M. and D. J. Graham. A Balanced Approach to Normalizing Bus Operational Data for Performance Benchmarking Purposes. Presented at 91st Transportation Research Board Annual Meeting, Washington, D.C., 2012.
4. Nelson, J., Wright, S., Masson, B., Ambrosino, G. and A. Naniopoulos. Recent developments in Flexible Transport Services. *Research in Transportation Economics*, Vol. 29, No. 1, 2010, pp. 243–248.
5. Mulley, C. and J. Nelson. Shaping the New Future of Paratransit: An Agenda for Research and Practice. In *Transportation Research Record: Journal of the Transportation Research Board*, *in press,* Transportation Research Board of the National Academies, Washington, D.C., 2016.
6. Dessouky, M., Palmer, K., and T. Abdelmaguid. Impacts of management practices and advanced technologies on demand responsive transit systems. *Transportation Research, Part A: Policy and Practice*, Vol. 38, No. 7, 2004, pp. 495-509.
7. Palmer, K., Dessouky, M. M., and Z. Zhou. Factors Influencing Productivity and Operating Cost of Demand Responsive Transit. *Transportation Research, Part A: Policy and Practice*, Vol. 42, No. 3, 2008, pp. 503-523.
8. Fu, L., Yang, J., and J. Casello. Quantifying Technical Efficiency of Paratransit Systems by Data Envelopment Analysis Method. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 2034*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 115-122.
9. Min, Hokey and Lambert, Thomas E. Lambert. Benchmarking and Evaluating the Comparative Efficiency of Urban Paratransit Systems in the United States: A Data Envelopment Analysis Approach. *Journal of Transportation Management,* Vol. 21, No. 2, 2010, pp. 48-62.
10. The Transit Cooperative Research Program (TCRP). *Report 124: : Guidebook for Measuring, Assessing, and Improving Performance of Demand-Response Transportation*, Transportation Research Board, National Academy Press, Washington, D.C, 2008.
11. Chia, D. “TCRP Synthesis 74: Policies and practices for effectively and efficiently meeting ADA paratransit demand. A synthesis of transit practice.” Transportation Research Board – Transit Cooperative Research Program, Washington, D.C., 2008.
12. Golden, M., Chia, D., Ellis, B. and R. Thatcher. *Accessible Transit Services for All.* FTA Report No. 0081. FTA, U.S. Department of Transportation, 2014.
13. *ADA Paratransit Services: Demand has Increased, but Little is Known about Compliance.* Publication GAO-13-17. U.S. Government Accountability Office, Washington, D.C., 2012.
14. Ellis, E. H. “TCRP Synthesis 119: Use of Taxis in Public Transportation for People with Disabilities and Older Adults.” Transportation Research Board – Transit Cooperative

Research Program, Washington, D.C., 2016.

15. Thatcher, R., and C. Ferris. “TCRP Report 163: Strategy Guide to Enable and Promote the Use of Fixed-Route Transit by People with Disabilities.” Transportation Research Board – Transit Cooperative Research Program, Washington, D.C., 2013.
16. Weiner, R. “TCRP Synthesis 76: Integration of Paratransit and Fixed-Route Transit Services: A synthesis of transit practice.” Transportation Research Board – Transit Cooperative Research Program, Washington, D.C., 2008.
17. Kaplan, Robert S., and David P. Norton. The Balanced Scorecard – Measures that Drive Performance. *Harvard Business Review*, Vol. 70, No. 1, January–February 1992, pp. 71–79. (Reprint #92105.)
18. Feigon, S., C. Murphy. “TCRP Report 188: Shared Mobility and the Transformation of Public Transit.” Transportation Research Board – Transit Cooperative Research Program, Washington, D.C., 2016 (pre-publication draft).