# Shared micromobility as a first- and last-mile transit solution? Insights from a novel dataset

*Zehui Yin[a], Greg Rybarczyk[b,c,d], Anran Zheng[f], Lin Su[f], Bingrong Sun[e], Xiang Yan[f]*

[a]Department of Management, University of Toronto Scarborough
[b]College of Innovation and Technology, The University of Michigan-Flint
[c]The Michigan Institute for Data Science (MIDAS), Ann Arbor, MI, USA
[d]The Centre for Urban Design and Mental Health, London, UK
[e]National Renewable Energy Laboratory, Golden, CO, USA
[f]Department of Civil and Coastal Engineering, University of Florida

## Abstract

The first- and last-mile (FM/LM) problem is a major deterrent to public transit use. With the rise of shared micromobility options such as shared e-scooters in recent years, there is a growing interest in understanding their potential to serve as a last-mile transit solution. However, empirical data regarding the integrated use of shared micromobility and public transit have been limited so far. As a result, much is unknown regarding the spatiotemporal patterns and characteristics of shared micromobility trips serving as an FM/LM connection to transit. This paper addresses these knowledge gaps by leveraging a novel dataset (i.e., the Spin post-ride survey dataset) that records thousands of transit-connecting shared e-scooter trips in Washington DC. Specifically, we used the dataset to reveal the spatiotemporal patterns of transit-connecting shared e-scooter trips in Washington DC, resulting in some major policy insights regarding the integral use of shared e-scooters and public transit. We further leveraged the dataset to validate if and to what extent a commonly applied buffer-zone approach can infer FM/LM micromobility trips accurately. Statistical tests showed that the actual FM/LM Spin e-scooter trips differ from inferred FM/LM Spin e-scooter trips in both spatial and temporal dimensions. This indicates that the common practice of inferring FM/LM micromobility trips with a buffer-zone approach can lead to inaccurate estimates of transit-connecting micromobility trips.

*Keywords:* micromobility, public transit, e-scooter, last-mile problem, spatiotemporal analysis

## 1. Introduction

A major issue in public transit is the first-mile and last-mile (FM/LM) problem, which refers to the difficulty of delivering transit services to the doorsteps of transit riders' trip origins or destinations. In many cases, the lack of convenient FM/LM travel options to reach transit stops is a major barrier for many individuals to use transit. As a result, studies have shown that strategies to improve FM/LM access to transit stops can enhance the coverage and reliability of transit systems, which in turn improves transport equity by providing better services for disadvantaged population groups and promotes environmental sustainability by improving transit's competitiveness with personal cars (Boarnet et al., 2017; Chen et al., 2021; Sun et al., 2021). In recent years, shared micromobility options such as shared e-scooters, which are especially convenient and efficient for serving short trips, have attracted growing attention as a potential FM/LM solution for bridging gaps in transit service areas.

There is indeed some preliminary evidence that demonstrates the potential for leveraging shared micromobility to enhance FM/LM transit connectivity. Accordingly to the 2021 State of the Industry report by the North American Bikeshare and Scootershare Association (NABSA), 63% of micromobility users have used shared micromobility to connect to transit, and 18% of all shared micromobility trips were for the purpose of connecting to transit (NABSA, 2022). These statistics were calculated by aggregating surveys mostly collected by shared micromobility operators. On the other hand, surveys conducted by public agencies and academic researchers have generally reported a lower percentage of travelers using shared micromobility to connect to transit, with values ranging from 4% to 39% (Ziedan et al., 2021). The statistic on the percentage of shared micromobility trips made to connect with transit is much harder to obtain because individual-level shared micromobility trip data available to researchers and policymakers rarely indicate if the trip is connecting to transit. As a result, much is unknown regarding the spatiotemporal patterns and characteristics of shared micromobility trips serving as an FM/LM feeder to transit.

As large-scale individual-level micromobility trip data, including precise latitude and longitude of trip origins and destinations, become increasingly available, researchers have sought to leverage such data to improve our understanding of how to integrate shared micromobility and public transit for FM/LM trips. The common practice is to assume that micromobility trips falling within the predefined buffer zone of a transit stop (the size of the buffer often ranges from 50 meters to 250 meters) are connecting to that stop (Ma et al., 2022; Oeschger et al., 2020; Yan et al., 2021). The problem with this approach is obvious: when a traveler starts or ends a shared micromobility trip close to a transit stop, the trip is not necessarily connected to the transit network. Some authors (e.g., Kong et al., 2020; Ma et al., 2018) have sought to improve the accuracy of inferred FM/LM trips by focusing on rail stations only and referencing the transit schedule (i.e., a FM/LM micromobility trip must be within 10 minutes of the bus/train arrival time at a stop). However, it is uncertain to what extent these mechanisms can improve the results. Another problem with this inference approach is that some FM/LM micromobility trips may occur outside of the predefined buffer zone, such as when a transit rider walks some distance to the nearest available shared micromobility device. Due to these issues, past findings obtained from inferred FM/LM micromobility trips can be unreliable or even misleading.

The current study aims to advance knowledge on shared micromobility and transit integration with a novel dataset (i.e., the Spin post-ride survey dataset) made available to us through a data-sharing agreement with a major micromobility company, Spin. This dataset contains a sample of shared e-scooter trips serving as FM/LM connections to transit in Washington DC, which is recorded by the Spin app and validated by Spin users. The dataset has two major advantages: first, it contains thousands of FM/LM e-scooter trips made by different individuals, which is a sample size hardly attainable for traditional surveys; second, since the FM/LM trips are reported by Spin users rather than estimated through an inference approach as described in the previous paragraph, the data have higher validity. Hence, the Spin post-ride survey dataset provides us with a unique opportunity to examine the spatiotemporal patterns and characteristics of transit-connecting shared e-scooter trips. In this paper, we leverage this unique dataset to address two important research objectives:

- Examine the spatiotemporal patterns of Spin e-scooter trips serving as FM/LM connection to transit in Washington DC;

- Validate if and to what extent the commonly applied buffer-zone approach can infer FM/LM micromobility trips accurately.

## 2. Literature Review

The integration of shared micromobility with public transit networks can potentially deliver multiple benefits: encouraging great use of the public transit system, expanding the catchment area of transit stops and transit network, improving mobility services for disadvantaged travelers, and reducing personal vehicle travel (Beale et al., 2022). Accordingly, there has been a growing research interest in examining the relationship between shared micromobility services and public transit.

A survey questionnaire is a common approach to studying the integration of micromobility with public transit (Guo and Zhang, 2021; Oeschger et al., 2020). Much of the existing research has focused on what shapes micromobility adoption and usage. For example, Qin et al. (2018) developed logit models based on survey data to identify key factors associated with the share of bike-and-ride users in Beijing. Cao et al. (2021) conducted a stated preference survey on e-scooter users in Singapore and estimated mixed logit models to identify factors influencing the choice of using shared e-scooters to serve short-distance transit trips. Some authors have also investigated whether shared micromobility complements or competes with public transit (Ziedan et al., 2021). Some studies found a complementary relationship, suggesting that shared micromobility can facilitate FM/LM connections to or from a transit stop (Yan et al., 2021; Ziedan et al., 2021). On the other hand, others suggested that shared micromobility may compete with transit and reduce public transit usage, especially when shared micromobility services are provided in areas well served by the public transit system (Fishman, 2016; Luo et al., 2021). For example, studies have shown that bikeshare substituted for transit usage in cities such as London and Washington DC (Fishman et al., 2015). Recent research further showed that shared e-scooter services have replaced some transit trips(Luo et al., 2021; Yan et al., 2021). Given the inconsistency in study findings and the presence of multiple confounding variables correlated with both micromobility and transit use, we are left with an inadequate comprehension of the true relationship between shared micromobility and public transit.

In recent years, the availability of shared micromobility trip data across cities has promoted a better understanding of shared micromobility and the factors that shape its use. (Caspi et al., 2020; Cheng et al., 2022; Hawa et al., 2021; Merlin et al., 2021; Tuli et al., 2021; Zhu et al., 2020). For example, Merlin et al. (2021) built a hurdle model to examine how e-scooter supply, socioeconomic factors, and built environment variables impact the e-scooter use. Some studies have focused on predicting spatiotemporal travel demand for micromobility using machine learning (Feng et al., 2022; Ham et al., 2021; Li et al., 2022;

Lin et al., 2018). Moreover, a growing number of studies have examined the spatial and temporal characteristics of micromobility services with a particular focus on their potential to serve as an FM/LM transit solution (Foissaud et al., 2022; Ma et al., 2018; Oeschger et al., 2020; Yang et al., 2019; Zhu et al., 2020). Studies have found that shared micromobility options are frequently used where public transportation is not well-developed (Schwinger et al., 2022). Saturday afternoons were found to be the peak of e-scooter usage in Louisville, Kentucky (Hosseinzadeh et al., 2021). Other studies have further shown that built environment, weather, transportation infrastructure, and socio-demographic characteristics shape the use of shared micromobility to connect with transit (Guo and He, 2020; Guo and Zhang, 2021; Hosseinzadeh et al., 2021).

However, empirical data on the integrated use of shared micromobility and public transit are difficult to obtain. Surveys can produce reliable data but often result in a small sample size, whereas the large-scale, individual-level trip data available to researchers do not indicate if a trip is connected to transit. Faced with these data challenges, some authors have applied a buffer-zone approach to infer FM/LM trips from shared micromobility trip data for further spatiotemporal analysis or modeling work (Cheng et al., 2022; Guo and He, 2020; Ma et al., 2022; Oeschger et al., 2020). The common practice is to assume all micromobility trips that fall within a predefined buffer zone of a transit stop to be transit-connecting trips. However, as discussed above, the validity of this approach is yet to be verified. If the inference approach is proven problematic, the spatiotemporal patterns and the causal relationships derived from the inferred FM/LM shared micromobility trips can be misleading.

Our study will address these issues by leveraging a novel dataset (i.e., Spin post-ride survey data) to 1) shed light on the spatiotemporal patterns of FM/LM transit-connecting shared micromobility trips in Washington DC and 2) validate the commonly used buffer-based approach to infer FM/LM micromobility trips. We will analyze all transit-connecting trips for the first research objective; for the second research objective, we will focus on rail-connecting trips as existing studies have mostly applied the buffer-zone approach to only infer FM/LM micromobility trips connecting to rail or metro stations (Oeschger et al., 2020).[1]

## 3. Study area and Data

In this section, we will first describe the study context, focusing on the shared micromobility and transit systems in Washington DC. We will then introduce the datasets used for this study, i.e., the Spin post-ride survey data from May 2021 to September 2021, and the Spin trip data in June 2021 and June 2022.

### *3.1. Study area*

Washington D.C. (DC) is the capital of the United States. It is a highly urbanized city with an extensive transit system, consisting of Metrorail, Metrobus, and MetroAccess

[1]Moreover, Since bus stops in the U.S. are often in close proximity to other destinations such as shops and restaurants, the inaccuracies associated with inferring FM/LM micromobility trips around bus stops using the buffer-based approach can be much larger.

paratransit services (see Figure 1). In 2021, Metrorail included six rail lines and 91 rail stations, and Metrobus had about 270 bus routes that serve over 11,000 bus stops. Metrorail had a distance-based fare scheme that ranges from \$2.25 to \$6 in peak hours and from \$2 to \$3.85 in off-peak hours; the regular fare for a one-way bus trip is \$2.[2] In addition, the DC region has one of the largest station-based bikesharing systems (named Capital Bikeshare) in the United States.

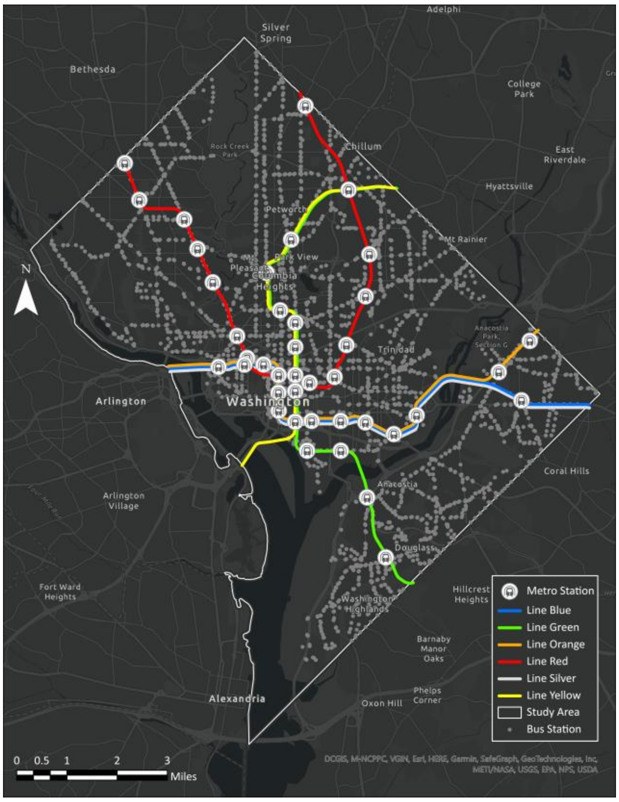


Figure 1: Washington D.C.

Dockless micromobility was first deployed onto DC streets in 2017 as a pilot program. Since 2021, over 10,000 shared fleet devices (including both dockless e-scooters and e-bikes) operated by six companies had been permitted in a geofenced operating area. The District Department of Transportation has developed detailed terms and conditions for fleet management, parking, data reporting, payment options, and low-income customer programs. Even though transit and micromobility coordination is emphasized in the DC region, there have not been formal efforts to promote the integration of privately-operated dockless micromobility services with the transit system.

[2]See https://www.wmata.com/fares/basic.cfm.

### 3.2. Spin post-ride survey data

The Spin post-ride survey data were collected by Spin, a micromobility company, from a randomly selected group of its users between May 2021 and September 2021. A three-question short survey was pushed to users' Spin app after they completed a Spin trip. These questions asked people which travel mode was replaced by the Spin e-scooter, whether the trip was connected to public transit, and why they used an e-scooter for the trip. Individuals' responses were matched with their Spin trip records to form a complete post-ride survey dataset. Of useful to the current study is the following information: a trip's start and end time, its start and end location (latitude/longitude), trip distance, and whether the trip was identified by users as connected to transit. The survey was randomly pushed to users through the Spin app for 33,530 completed trips, receiving 9,785 partial or complete responses. A total of 7,148 trips received a response regarding the question of if the Spin trip was connected to public transit, with 1,916 trips responding yes and 5,232 responding no. Out of these 1,916 trips which reported to be *transit-connected*, we further identified 294 trips as *rail-connected*, with an assumption that any transit-connecting trip starting from or ending at a location within 400 feet (122 meters) of a metro entrance to be a rail-connecting trip.

To protect user privacy, the data had to be de-identified before being shared so as to remove the spatial coordinates of Spin trip origins/destinations while preserving valuable information needed for this research. The de-identification process essentially replaced the spatial coordinates of trip origins and destinations with the ID(s) of the census block group(s) in which the trip started and ended. Thus, the spatiotemporal analysis of the 1,916 transit-connecting shared e-scooter trips will be conducted at the census block group level.

To use the Spin post-ride survey data for validation purposes (i.e., the second research objective discussed in the Introduction section), we further worked with Spin to apply the buffer-zone approach to infer FM/LM rail-connecting trips. In other words, we pretended that the survey component was absent from the Spin post-ride survey data and performed the commonly used buffer-zone approach on the data to infer FM/LM trips. Specifically, we requested Spin to help generate the following variables before removing the spatial coordinates of trip origins and destinations from the data shared with us: two binary variables indicating if the trip origin/destination is within 400 feet of a metro entrance as well as the ID of the metro entrances that a Spin trip origin/destination is potentially connected to.[3] The purpose was to infer FM/LM rail-connecting trips based on the buffer approach (using 400 feet as the threshold) and to identify the metro stops at which the inferred FM/LM trips connected to. This buffer-based approach identified 3,728 out of 33,530 Spin trips (11.1%)

[3]We have also tested using 200 feet as the buffer threshold. The results obtained from using a 200-feet buffer are consistent with those from using a 400-feet buffer. We ended up using the 400-feet buffer for several reasons: First, it is more consistent with the existing literature that commonly used a threshold between 50-300 meters; and 400 feet (133 meters) is in the middle of this range. Second, 400 feet is commonly assumed to be the maximum walking distance to a transit stop that U.S. travelers can comfortably accept. Also, note that when multiple metro entrances were within 400 feet of a trip's origin or destination, the closest metro entrance was identified.

to be FM/LM rail-connecting trips, a proportion much higher than that reported by Spin users (294/7148=4.1%).

### *3.3. Spin trip data*

We also incorporated a Spin trip dataset in this research to increase the robustness of the validation results. This dataset contains all Spin trips (232,129 trips) that took place in June 2021 and June 2022 in Washington DC; in other words, it avoids any survey sampling bias. Also, since it is the type of data commonly used in prior research to infer FM/LM trips, incorporating it in the validation process can enhance the generalizability of our study findings. A similar de-identification process was implemented before Spin shared the data with us. Unlike the Spin post-ride survey data, however, we did not get access to the information regarding trip distance or duration for this dataset. The final dataset contains the timestamp of the trip start time, whether its origin and destination are within 400 feet of a metro entrance, and the ID of the metro entrances closest to its origin and destination. We identified a total of 51,432 FM/LM rail-connecting trips from a total of 232,129 trips based on the buffer approach.

Table 1 presents some basic information about the datasets used in this study. We show the number of trips contained in each subset of the data, the mean and median trip distance and duration, as well as the percentage of trips that happened during morning peak hours (6-9 am) and afternoon peak hours (4-7 pm). The mean trip distance and duration are larger than the corresponding median values due to the existence of outliers (some trips are extremely long in distance and duration).

Table 1: Overview of datasets used in this study

| Data Category | *n* | Distance | | Duration | | %morning peak | %afternoon peak |
|---|---|---|---|---|---|---|---|
| | | Mean | Median | Mean | Median | | |
| **Spin post-ride survey data** | | | | | | | |
| All trips | 33,530 | 1.5 | 1.0 | 18.5 | 9.8 | 6.5% | 30.8% |
| Trips responded to survey | 9,781 | 1.6 | 1.1 | 20.4 | 11.2 | 6.5% | 29.7% |
| Inferred FM/LM *rail-connecting* trips | 3,728 | 1.6 | 1.1 | 22.1 | 12.1 | 6.9% | 32.1% |
| Reported transit-connecting trips | 1,916 | 1.6 | 1.1 | 20.3 | 10.9 | 6.9% | 29.6% |
| Reported FM/LM *rail-connecting* trips | 294 | 1.7 | 1.2 | 22.8 | 12.8 | 6.9% | 32.0% |
| **Spin trip data** | | | | | | | |
| All trips | 232,129 | | | | | 10.1% | 31.1% |
| Inferred FM/LM *rail-connecting* trip | 51,432 | | | | | 11.9% | 32.4% |

## 4. Spatiotemporal characteristics of transit-connecting e-scooter trips

In this section, we address the first research objective discussed above by examining the spatiotemporal patterns and attributes of reported FM/LM transit-connecting trips from the Spin e-scooter post-ride survey data. Our analysis will distinguish FM trips and LM trips, with the former defined as trips with the destination located within 400 feet of a transit stop (metro entrance or bus stop) and the latter defined as trips with the origin located within 400 feet of a transit stop.[4] A total of 1695 trips out of 1916 trips will be

[4]If a reported FM/LM trip has both its origin and destination located within 400 feet of a transit stop, which account for 36.7% of all transit-connecting trips, the trip will be counted as both an FM trip and LM trip (but each will have a weight of 0.5 rather than 1 to avoid double counting). Note that we also tested removing these trips from the analyses but found that the results were consistent.

examined here because some trips have both of their trip ends fall outside of the 400-feet radius of a transit stop.

### *4.1. Temporal patterns*

To examine the temporal patterns, we aggregated the reported FM/LM e-scooter trips by hour of day based on the trip start time. Figure 2 shows the temporal patterns of FM and LM trips separately. The FM trips' and LM trips' usage patterns are quite similar, with two pronounced peaks: one in the morning and one in the afternoon. In detail, the figure shows that the majority of FM/LM trips occurred in the evening, the peak being between 4-7 pm; also, there was a rise in FM/LM trips in the morning peak hours (6-9 am), even though the trip frequency is much smaller compared to the afternoon peak. These results imply some use of shared e-scooters to connect with transit for commuting trips. The results also correspond with findings from a temporal analysis of e-scooter trips in Indianapolis (Liu et al., 2019). Finally, compared to the LM trips, the FM trips have relatively higher frequencies during the morning peak and lower frequencies in the afternoon peak. These differences can be a result of Washington DC's urban form: e-scooter riders' employment destinations tend to be closer to transit stops than their home locations, which drive more FM e-scooter use in the morning and more LM e-scooter use in the afternoon.

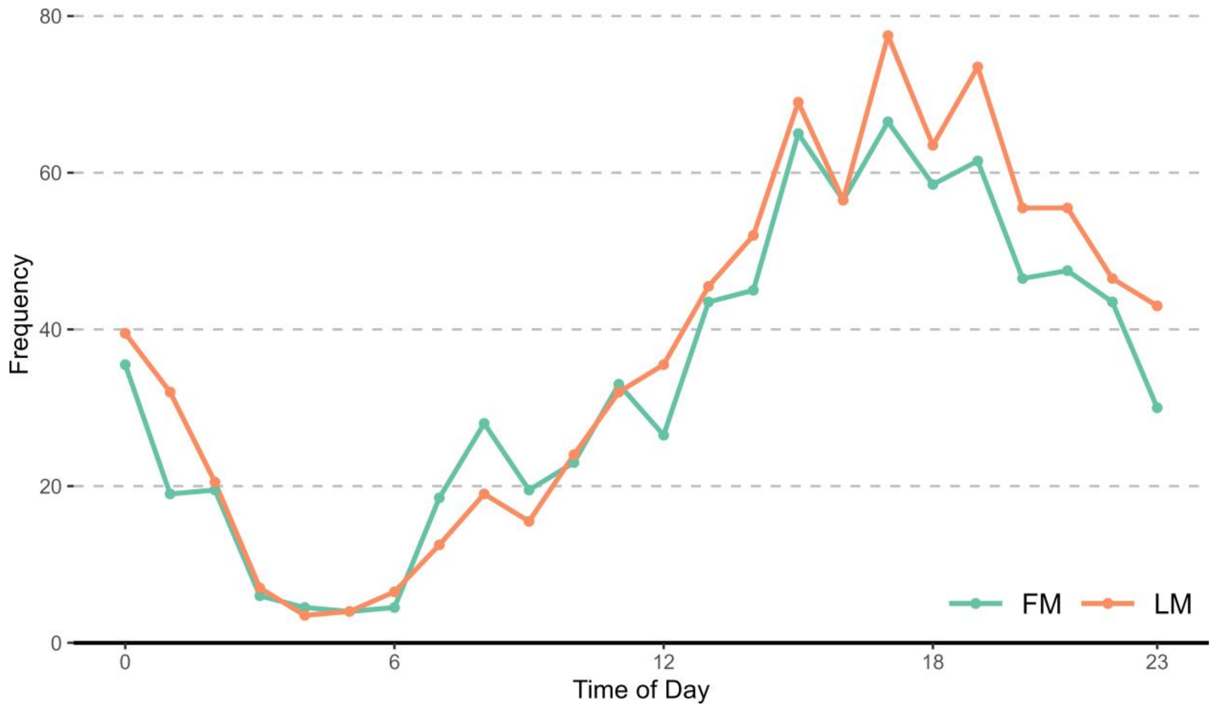


Figure 2: Reported FM/LM trip frequency by hour of day

### 4.2. Spatial patterns

To explore the spatial patterns of FM and LM trips, we aggregated their trip origins and trip destinations at the census block group level. Figure 3 shows the spatial distribution of reported FM/LM trips, which shows that the spatial patterns of FM and LM trips were largely consistent. The highest density of shared e-scooter trips was found near the National Mall, the central business district, and the Potomac River, suggesting that e-scooters were often used for recreational purposes (McKenzie, 2019; Merlin et al., 2021). Also noteworthy was the unevenness in the trip distribution, with a very low share of trips happening at the periphery of the district where some traditionally underserved neighborhoods are located (especially those located in the southeast part of the district). This could be a result of a lack of demand for shared e-scooters among residents of these neighborhoods due to the high user costs, which indicates a potential equity problem. To shed light on this issue, Su et al. (2022) proposed a novel metric, idle time, to evaluate the spatial equity of micromobility systems by accounting for both the service supply and demand.

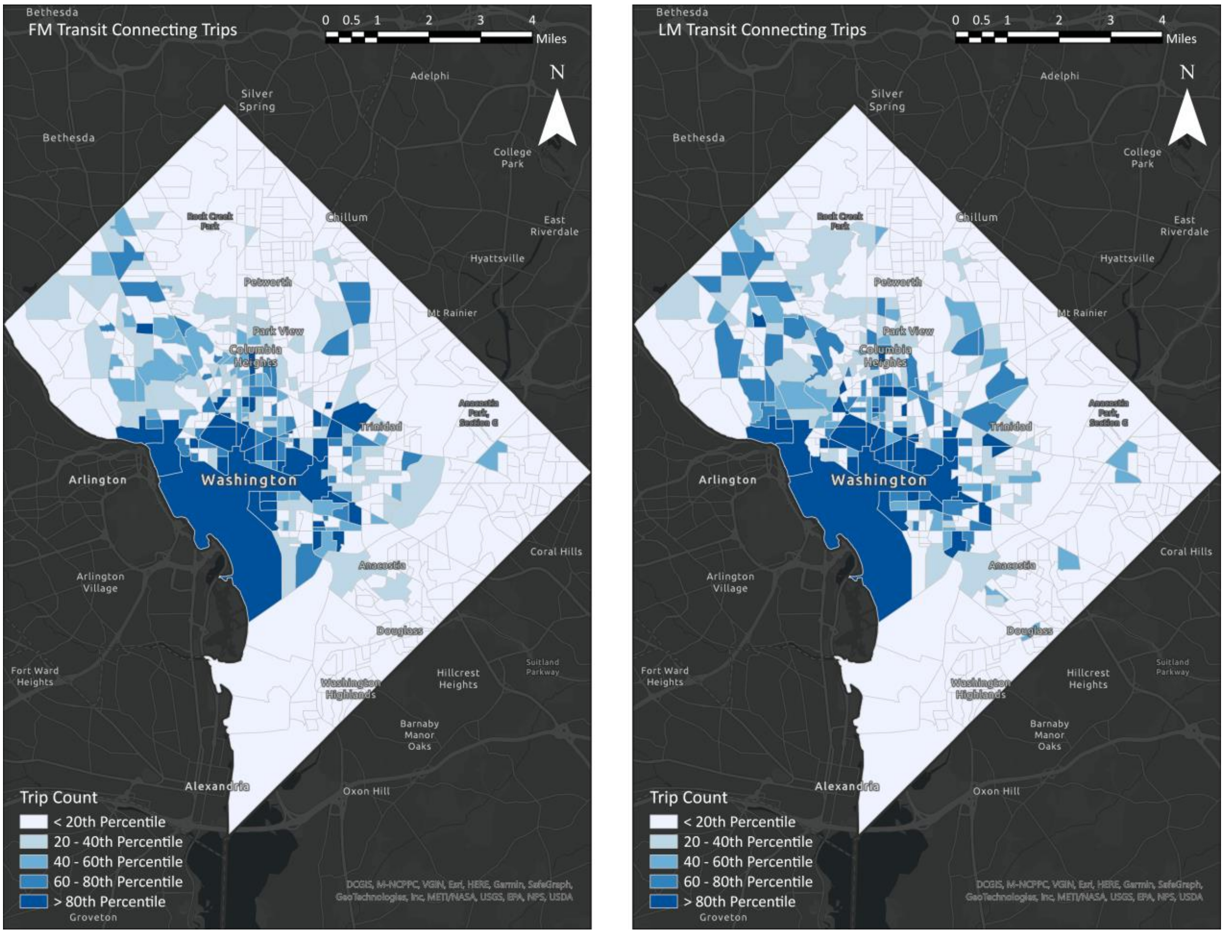


Figure 3: Spatial Pattern of Reported FM/LM trips in Washington DC

### 4.3. Trip attributes

Table 2 presents some key trip attributes of the reported FM and LM transit-connecting trips. The median trip distance of these trips was 1.0 miles, and their median duration was 10.3 min. About 7.3% and 30.3% trips, respectively, happen in morning peak hours (6-9 am) and afternoon peak hours (4-7 pm).

Table 2: Descriptive statistics and test results of FM/LM transit-connecting e-scooter trips

| Data Category | $n$ | Distance (miles) | | Duration (minutes) | | %morning peak | %afternoon peak |
|---|---|---|---|---|---|---|---|
| | | Median | Mean | Median | Mean | | |
| Reported FM/LM Transit-connecting trips | 1,695 | 1.0 | 1.5 | 10.3 | 20 | 7.3% | 30.3% |
| FM trips | 806 | 0.9 | 1.5 | 9.5 | 19.7 | 8.8% | 30.2% |
| LM trips | 889 | 1.1 | 1.6 | 11.1 | 20.3 | 6.0% | 30.5% |
| FM trips vs LM trips | | Mood's median test (Two-sided) | | Two-sample t-test (One-sided) | | Two-proportions Z-test (One-sided) | |
| *p-value* | | 0.002 | 0.019 | 0.273 | 0.512 | 0.001 | 0.440 |

We also computed separate metrics for the FM and LM trips. The median distance and duration of LM trips were slightly longer than those of FM trips. The Mood's median test results indicated that these differences were statistically significant at the 0.05 level. Compared to LM trips, a greater share of FM trips happened during morning peak hours whereas a slightly lower share of them happened in the afternoon peak hours. The one-sided two-proportion Z-tests suggested that while the former difference was significant at the 0.05 level, the latter difference was not statistically significant. In sum, these results suggest that compared to LM trips, FM transit-connecting trips were slightly shorter and disproportionately happened during the morning peak.

## 5. Validation of buffer-based FM/LM micromobility trip inference approach

### 5.1. Validation framework

To validate if and to what extent the commonly applied buffer-zone approach can infer FM/LM micromobility trips accurately (the second research objective of this paper), we developed a validation framework as shown in Figure 4. We apply a set of statistical tests to examine if the inferred FM/LM trips and the reported FM/LM trips have different *spatial* and *temporal* patterns and if their *trip attributes* (i.e., trip distance and duration) are different. If no statistical differences are found, we can conclude that the buffer-based FM/LM micromobility trip inference approach will generate results that can reflect actual FM/LM micromobility trips' spatiotemporal patterns and trip characteristics. If test results demonstrate statistical differences, we can conclude that the commonly applied buffer-zone approach to identify FM/LM micromobility trips will lead to inaccurate results. The approach used here resembles methods applied in McKenzie (2019), which conducted a spatiotemporal comparative analysis of scooter-share and bikeshare usage patterns.

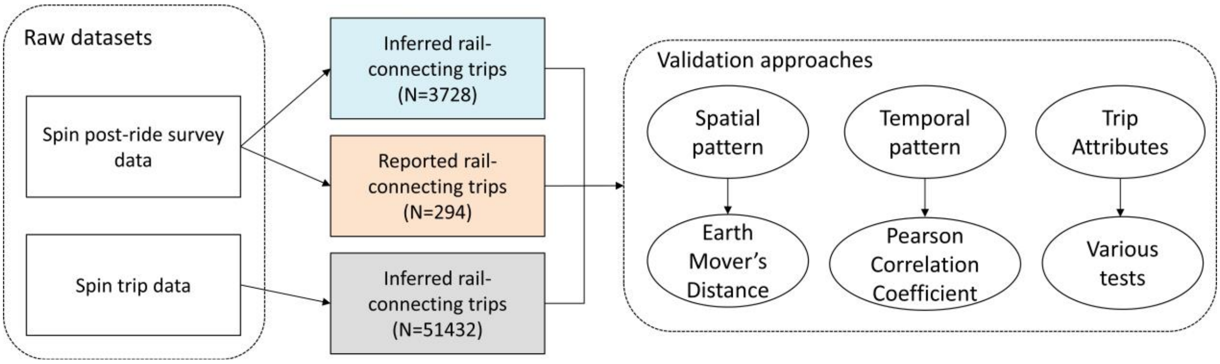


Figure 4: Validation framework

The Spin post-ride survey data are the main data source used for the validation. As discussed in the Data section, the data contains actual transit-connecting e-scooter trips reported by Spin users, which allows us to identify a total of 294 rail-connecting trips that happened between May 2021 and September 2021. We also inferred 3,728 rail-connecting trips from the 33,530 trips that happened in the same period. Comparing the reported rail-connecting trips and the inferred rail-connecting trips from the Spin post-ride survey data in terms of their spatial patterns, temporal patterns, and trip characteristics is the key focus of the validation.

The Spin post-ride survey data may contain potential sampling bias and response bias (We do not know which Spin users or trips were sampled for the survey and the corresponding response rate). Therefore, to address the potential survey bias, we incorporated into the validation procedure a second dataset for comparison, i.e., the Spin trip data. The dataset includes all Spin trip records (232,129 trips) that occurred during the two months of the study period, based on which we applied the buffer-zone approach to infer a total of 51,432 rail-connecting trips. These trips were compared against the inferred and the reported rail-connecting trips from the Spin post-ride survey in terms of their spatial patterns, temporal patterns, and trip attributes. These comparisons can enrich the study findings and enhance the robustness of the validation results.

### *5.1.1. Comparison of temporal patterns*

To determine whether there is a significant difference in the temporal patterns, we performed a Watson's two-sample test of homogeneity (Watson, 1961) using R package "circular" (Agostinelli and Lund, 2022). This test provides a means to reveal the likelihood of two temporal distributions such as time of day variations coming from the same population (McKenzie, 2019). In addition, we aggregated the FM/LM trips by the trip start time using a 15-minute interval and computed Person correlation coefficients between the three samples to further quantify the degree to which their temporal distribution are similar to each other. Figure 5) shows the temporal distribution of the data samples: the inferred rail-connecting trips from the Spin post-ride survey data, the reported rail-connecting trips from the Spin post-ride survey data, and the inferred rail-connecting trips from the Spin trip data.


Figure 5: Temporal distributions of FM/LM rail-connecting trips

*5.1.2. Comparison of spatial patterns*

We analyzed the spatial patterns of the three groups of rail-connecting e-scooter trips using the Earth Mover's Distance (EMD) implemented with the R-package "emdist" (Urbanek and Rubner, 2022). The algorithm was first developed by Rubner et al. (2000) to compare images, and it has been used recently in spatial ecology (Kranstauber et al., 2017). The EMD calculates the cost of converting one multidimensional matrix to the other of equal size and determines how similar the two matrices are. A higher distance indicates that more costs or efforts are needed to convert one trip's spatial distribution to the other by moving the weights which are the trip count percentage at each metro station in this analysis (McKenzie, 2019). In short, greater distances mean less similarity between the two trips' spatial distributions.

Two-dimensional EMD analysis of FM/LM e-scooter trips aggregated at each metro station (see Figure 6) was applied to evaluate the spatial differences between the inferred and reported rail-connecting trips. Specifically, we used the easting (x) and northing (y) as the coordinates of metro stations. For the weighting, we used the percentage of e-scooter trips–as opposed to the number of e-scooter trips–at each metro station to account for the differences in the number of observations between the three groups of FM/LM rail-connecting trips. A higher distance generated from the EMD analysis would imply less similarity between the two samples' spatial distributions.

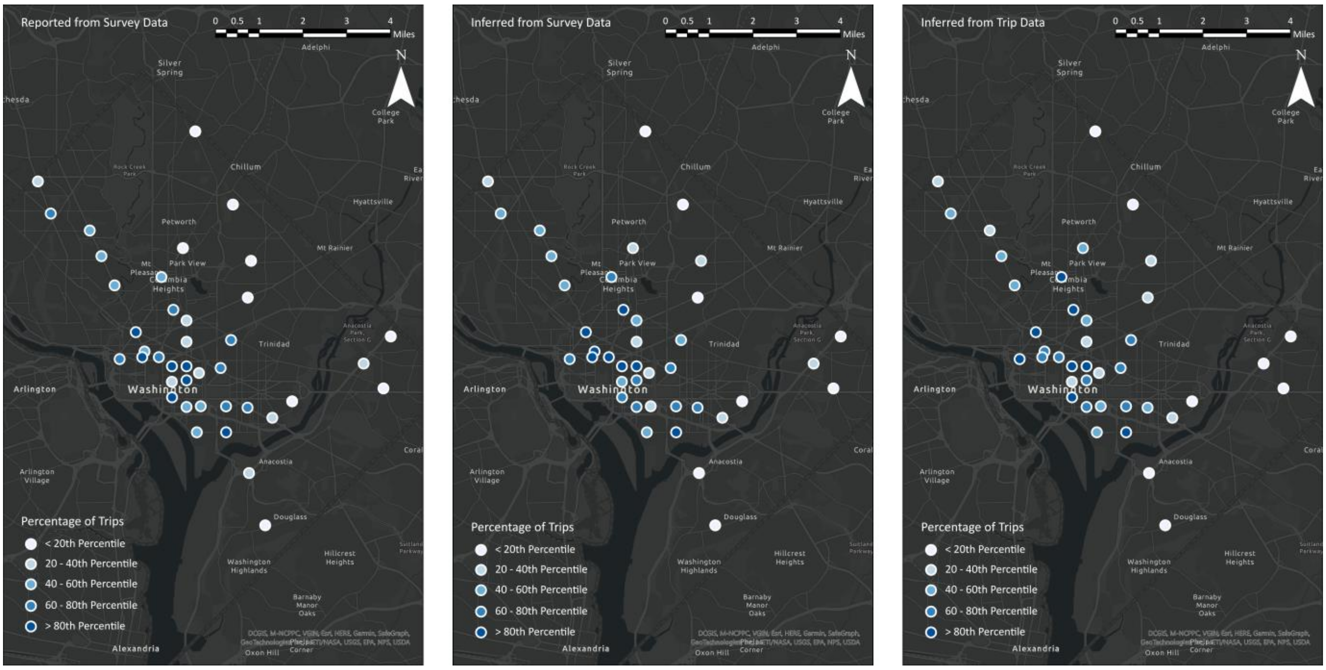


Figure 6: Spatial distributions of FM/LM rail-connecting trips

#### *5.1.3. Comparison of trip attributes*

We applied several statistical tests to further examine whether there were differences in the attributes of inferred and reported FM/LM trips. Four non-parametric tests were used and consisted of: Mood's median test (Mood, 1954), Wilcoxon rank sum test (Wilcoxon, 1945), Fligner-Killeen test (Fligner and Killeen, 1976), and Kolmogorov-Smirnov test (Massey, 1951). Mood's median test is a non-parametric test that compares the medians of the two groups. It counts how many observations in each group are above or below the grand median and tests whether they are independent of the group membership. Wilcoxon rank sum test is a non-parametric alternative to the unpaired two-sample t-test. It compares the ranks of the observations in each group and tests whether they are equally likely to be larger or smaller than each other. Fligner-Killeen test assesses the homogeneity of variances and is robust to non-normal data. It compares the absolute deviations of each observation from the median of its group. Kolmogorov-Smirnov test measures the maximum difference between the empirical cumulative distribution functions of the two samples. It tests whether this difference is larger than expected by chance. It evaluates whether the two samples are drawn from the same population distribution.

These were collectively used to check whether there were any significant differences between the attribute distributions' centers, spreads, or shapes: the two trip attributes examined in this study were trip distance and duration. Since the two attributes are not available in Spin trips that happened in June 2021 and June 2022 data, we only conduct these tests on inferred and reported FM/LM trips from the Spin e-scooter post-ride survey data. We decided to use the non-parametric tests because these two attributes don't follow a normal distribution. As a last step, we applied a Pearson correlation coefficient analysis on the

Table 3: Test results for spatiotemporal patterns and characteristics

| | | | | | | |
|---|---|---|---|---|---|---|
| Temporal Pattern | Pearson Correlation Coefficient | | % FM/LM Trips at 15-minute interval | | | |
| | | Sample 1[1] vs Sample 2[2] | 0.833 | | | |
| | | Sample 1 vs Sample 3[3] | 0.781 | | | |
| | | Sample 2 vs Sample 3 | 0.931 | | | |
| Spatial Pattern | Earth Mover's Distance | | Weighting Variable % FM/LM Trips | | | |
| | | Sample 1 vs Sample 2 | 1618.597 | | | |
| | | Sample 1 vs Sample 3 | 1029.708 | | | |
| | | Sample 2 vs Sample 3 | 835.210 | | | |
| Trip Attributes Sample 1 vs Sample 2 | | Trip Distance (meters) | | Trip Duration (seconds) | | |
| | | Test Statistics | *p-value* | Test Statistics | *p-value* | |
| | Mood's Median Test | 0.485 | 0.628 | 0.363 | 0.716 | |
| | Wilcoxon Rank Sum Test | 529883 | 0.344 | 539742 | 0.666 | |
| | Fligner-Killeen Test | 1.171 | 0.279 | 1.396 | 0.237 | |
| | Kolmogorov-Smirnov Tests | 0.056 | 0.357 | 0.036 | 0.881 | |

Notes: 1. Reported FM/LM trips from the Spin post-ride survey data 2. Inferred FM/LM trips from the Spin post-ride survey data 3. Inferred FM/LM trips from the Spin trip data

three datasets used in this research to determine to what extent the trips were correlated.

## *5.2. Validation results*

This section presents the test results for the validation, which are shown in Table 3.

Regarding the temporal patterns, we observed noticeable differences between inferred and reported transit-integrated FM/LM trips in Figure 5. Compared to the reported trips in the post-ride survey (Sample 1, the orange curve), the inferred FM/LM trips from both the post-ride survey (Sample 2, the green curve) and the Spin trip data (Sample 3, the grey curve) have a relatively lower density of morning-peak trips and an elevated density of mid-afternoon trips. Despite the observed differences, the Watson's Two-Sample tests produced insignificant results at the 0.05 level (results not shown here), suggesting that we cannot reject the null hypothesis that the temporal distributions of the reported FM/LM trips and the inferred FM/LM trips are from the same population. On the other hand, the Pearson Correlation coefficients suggest that the temporal distribution of reported FM/LM trips (Sample 1) is less similar to that of the inferred FM/LM trips (Sample 2 and Sample 3), even though Sample 1 and Sample 2 were both drawn from the dataset (Spin post-ride survey data) whereas Sample 2 and Sample 3 were not. Overall, these results suggest that the temporal pattern of the reported FM/LM trips and that of the inferred FM/LM trips have some differences but these differences are not qualitative.

The EMD analyses compare the spatial patterns between samples, with a large distance indicating more different spatial patterns. We found that the distance between the reported FM/LM trips from the Spin post-ride survey data (Sample 1) and the inferred FM/LM trips from the same dataset (Sample 2) was approximately 1,600 units, which was nearly twice that between the two inferred FM/LM trip samples (Sample 2 and Sample 3). Also, the EMD distance between the reported trips from the Spin post-ride survey (Sample 1) and the inferred trips from the Spin trip data (Sample 3) was larger than that between Sample

2 and Sample 3. Overall, these results tell us that there are relatively large differences in the spatial patterns between the inferred FM/LM trips and the reported FM/LM trips.

We further compared the attributes from the inferred FM/LM trips and the reported FM/LM trips from the post-ride survey using several tests, with the goal of assessing similarities and differences. Table 3 shows that the Mood's Median test produced statistically insignificant results ($p > 0.05$) in regard to trip distance and duration. In other words, the median values of the two trip attributes between the inferred FM/LM trips and the reported FM/LM trips are not statistically different from each other. The remaining tests (i.e., Wilcoxon Rank test, Fligner-Killeen test, and Kolmogorov-Smirnov test) all produced statistically insignificant results. Overall, these results suggest that the center, spread, and shape of the trip distance and duration are not statistically different between the reported FM/LM trips (Sample 1) and the inferred FM/LM trips (Sample 2). Furthermore, we discovered that the trip attributes (trip distance and duration) of FM/LM shared e-scooter trips and those of non-FM/LM trips are not very different. However, this may be attributed to the uniqueness of Washington DC: While FM/LM trips are generally expected to be shorter than non-FM/LM trips, the district's densely urban form and the presence of clustered attractions (such as the National Mall) have resulted in a higher proportion of short trips.

In sum, we found that the spatial and temporal patterns of reported FM/LM trips from the Spin post-ride survey data are statistically different from those of inferred FM/LM trips from both the Spin post-ride survey data and the Spin trip data. On the other hand, the trip attributes (i.e., distance and duration) of the reported FM/LM trips and those of the inferred FM/LM trips are not statistically different.

### 5.3. Biases in inferred FM/LM micromobility trips from the buffer-based approach

In this subsection, we further examine to what extent the buffer-based inference approach may lead to biased results regarding FM and LM micromobility trips. Specifically, we conducted two sets of correlation analyses: 1) we computed the correlation coefficients of the number of FM trips and that of LM trips for the same time period (based on trip start time at the 15-minute interval) and at the same metro stop; 2) and we calculated the correlation coefficient of the number of FM/LM trips at each metro stop and the metro ridership.

Table 4 shows the results for the Pearson correlation coefficients. For the inferred FM/LM trips from both the Spin post-ride survey data and the Spin trip data, the correlation coefficients of the number of FM trips and LM trips—regardless if they are aggregated temporally (15-min intervals) or spatially (metro stop)—are very close to one; however, for the reported FM/LM trips from the Spin post-ride survey data, the correlation coefficient was equal to or small than 0.7. This means that although the FM trips and the LM trips are highly correlated at both the spatial and temporal dimensions, inferring FM/LM micromobility trips based on the buffer-zone approach can overestimate the correlation.

Recent studies have established the association between the number of transit-connecting micromobility trips and transit ridership (Ma et al., 2015; Zhao and Li, 2017). For the inferred FM/LM trips, we found that the correlation coefficient of FM/LM trip frequency and ridership is close to 0.65 on average (ranging from 0.59 to 0.71); for the reported FM/LM trips, however, the coefficient was around 0.5. In other words, the inferred rail-connecting trips based on the buffer-based approach may overestimate the magnitude of the association between metro ridership and the number of transit-integrated e-scooter trips.

In sum, the correlation analyses conducted in this subsection suggest that using inferred FM/LM micromobility trips based on the buffer-zone approach can introduce bias to study results. For example, we found that the differences in the spatiotemporal patterns between FM trips and LM trips may be overlooked, and the correlation between FM/LM micromobility trips and transit ridership may be overestimated.

Table 4: Pearson correlation coefficients of some variables

| Data source | FM&LM trip count (temporal)[1] | FM&LM trip count (spatial)[2] | Ridership&FM trip count[2] | Ridership&LM trip count[2] |
|---|---|---|---|---|
| Reported FM/LM trips from Spin post-ride survey data | 0.58 | 0.70 | 0.54 | 0.47 |
| Inferred FM/LM trips from Spin post-ride survey data | 0.90 | 0.98 | 0.65 | 0.60 |
| Inferred FM/LM trips from Spin trip data | 0.92 | 0.99 | 0.71 | 0.67 |

Notes: 1. Grouped into 15-minute intervals in 24 hours
2. Grouped at metro stops level with 400 feet buffers

## 6. Discussion and Conclusion

Serving as an FM/LM feeder mode to enhance transit connectivity has been recognized as one of the most important use scenarios of shared micromobility. Understanding how shared micromobility is being used for accessing public transit services can help improve the overall accessibility of public transit and generate mobility and sustainability benefits. However, empirical data on the integrated use of shared micromobility and public transit are rarely available, and existing data sources (e.g., survey data and shared micromobility trip data) both have some major limitations. To advance knowledge on public transit and shared micromobility integration, this paper leveraged a novel dataset (i.e., the Spin post-ride survey data) that records thousands of transit-connecting shared e-scooter trips reported by Spin users in Washington DC. With this dataset, we examined the spatiotemporal patterns and characteristics of transit-integrated micromobility trips and the utility of applying a commonly used buffer-based approach to identify first-/last-mile transit-connecting shared e-scooter trips.

When analyzing the spatiotemporal characteristics of transit-connecting shared e-scooter trips, we distinguished first-mile (FM) and last-mile (LM) trips. We found that their temporal and spatial usage patterns are largely similar but with some key differences. For both FM and LM trips, a majority of them occurred in the evening, the peak being between 4-7 pm; also, there was a rise in FM/LM trips in the morning peak hours (6-9 am), even though the frequency is much smaller compared to the afternoon peak. These results imply that many users ride shared e-scooters to connect with transit for commuting trips. Meanwhile, compared to LM trips, the proportion of FM trips happening in the morning peak hours is statistically higher. Spatially, the number of transit-connecting shared e-scooter trips tends to decrease as the distance to the urban core increases. Most of the FM/LM trips were found near the National Mall, the central business district, and the Potomac River, whereas a very low share of trips happened at the periphery of the district where some traditionally under-served neighborhoods are located. The result implies that transit and shared micromobility integration has delivered limited equity benefits at present. Targeted policy strategies such as low-income fare programs may be required to promote great use of shared micromobility to connect with transit among low-income populations and minority neighborhoods.

Moreover, we found that the buffer-zone approach commonly used to infer transit-connecting e-scooter trips may not be as reliable as widely assumed. The buffer-based approach is likely to significantly overestimate the number of rail-connecting trips: in Washington DC, we found that the buffer-based approach inferred 11.1% of the Spin trips to be rail-connecting whereas the proportion reported by Spin users was only 4.1%. Results from various statistical tests suggest that the inferred FM/LM trips and the user-reported FM/LM trips have different spatial and temporal patterns. Notably, inferring transit-connecting e-scooter trips based on the buffer-zone approach will likely lead to imprecise hot-spots of such trips. It should also be noted that using the buffer-zone approach to infer FM/LM transit-connecting trips can obscure differences in the spatiotemporal patterns of FM and LM trips, and may result in an overestimation of the correlation between FM/LM shared micromobility trips and transit ridership. In other words, the predictive power of transit ridership in determining the use of micromobility services in certain locations may be weaker than some studies have suggested.

The study results are based on an empirical analysis of the Spin e-scooter data collected from Washington DC. While we believe that the validation results regarding the buffer-based approach are generalizable across urban and transportation contexts, some of the empirical findings are only applicable to cities similar to DC—a dense U.S. city with an extensive transit network and robust micromobility systems. Notably, the findings on the spatiotemporal patterns of FM/LM transit-connecting trips may not be transferable to other U.S. cities with less transit and shared-micromobility coverage or non-US cities (e.g., Chinese cities) where pricing for shared micromobility is much lower. Future studies conducted across a diverse range of urban environments are needed to enrich the empirical findings on transit and shared micromobility integration. Moreover, while the Spin post-ride survey data used in this study have many advantages over existing data sources in shedding light on FM/LM micromobility trip patterns, the sample size of rail-connecting trips (n=294) is relatively small. To enhance the validity of statistical tests and strengthen the robustness of study findings, future research should aim to obtain a larger sample size.

In light of the above discussions, this study calls for greater emphasis and improved approaches to collect data on integrated micromobility and transit trips. In the current era when new mobility options such as on-demand ridesharing services and shared micromobility are rising in popularity, understanding where and when integrated use of various shared-used modes such as public transit and shared micromobility are happening is particularly important to inform policy efforts and operational strategies for enhancing multimodal travel. However, collecting such data is a challenging task for transportation agencies and operators because existing efforts focus on getting data for a single mode only. The Spin post-ride survey demonstrates a possible approach to improve multimodal travel data collection efforts. Moreover, with the increased adoption of trip-planning platforms and integrated payment apps/systems, integrated multimodal trip data should become easier to be collected. In cases where such data are not available, and the only alternative is to infer transit-connecting trips based on the buffer-based approach, we recommend researchers and practitioners refine the inference results by validating against other data sources such as the transit scheduling information or the micromobility parking zones at transit stations.

## References


Agostinelli, C., Lund, U., 2022. R package circular: Circular Statistics (version 0.4-95). CA: Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA. URL https://r-forge.r-project.org/projects/circular/

Beale, K., Kapatsila, B., Gris´e, E., 2022. Integrating public transit and shared micromobility payments to improve transportation equity in seattle, wa. Transportation Research Record, 03611981221103233.

Boarnet, M. G., Giuliano, G., Hou, Y., Shin, E. J., 2017. First/last mile transit access as an equity planning issue. Transportation Research Part A: Policy and Practice 103, 296–310.

Cao, Z., Zhang, X., Chua, K., Yu, H., Zhao, J., 2021. E-scooter sharing to serve short-distance transit trips: A singapore case. Transportation research part A: policy and practice 147, 177–196.

Caspi, O., Smart, M. J., Noland, R. B., 2020. Spatial associations of dockless shared e-scooter usage. Transportation Research Part D: Transport and Environment 86, 102396.

Chen, S., Yan, X., Pan, H., Deal, B., 2021. Using big data for last mile performance evaluation: An accessibility-based approach. Travel behaviour and society 25, 153–163.

Cheng, L., Wang, K., De Vos, J., Huang, J., Witlox, F., 2022. Exploring non-linear built environment effects on the integration of free-floating bike-share and urban rail transport: A quantile regression approach. Transportation Research Part A: Policy and Practice 162, 175–187.

Feng, C., Jiao, J., Wang, H., 2022. Estimating e-scooter traffic flow using big data to support planning for micromobility. Journal of Urban Technology 29 (2), 139–157.

Fishman, E., 2016. Bikeshare: A review of recent literature. Transport Reviews 36 (1), 92–113.

Fishman, E., Washington, S., Haworth, N., 2015. Bikeshare's impact on active travel: Evidence from the united states, great britain, and australia. Journal of Transport & Health 2 (2), 135–142.

Fligner, M. A., Killeen, T. J., 1976. Distribution-free two-sample tests for scale. Journal of the American Statistical Association 71 (353), 210–213. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1976.10481517

Foissaud, N., Gioldasis, C., Tamura, S., Christoforou, Z., Farhi, N., 2022. Free-floating e-scooter usage in urban areas: A spatiotemporal analysis. Journal of transport geography 100, 103335.

Guo, Y., He, S. Y., 2020. Built environment effects on the integration of dockless bike-sharing and the metro. Transportation Research Part D: Transport and Environment 83, 102335.

Guo, Y., Zhang, Y., 2021. Understanding factors influencing shared e-scooter usage and its impact on auto mode substitution. Transportation research part D: transport and environment 99, 102991.

Ham, S. W., Cho, J.-H., Park, S., Kim, D.-K., 2021. Spatiotemporal demand prediction model for e-scooter sharing services with latent feature and deep learning. Transportation research record 2675 (11), 34–43.

Hawa, L., Cui, B., Sun, L., El-Geneidy, A., 2021. Scoot over: Determinants of shared electric scooter presence in washington dc. Case studies on transport policy 9 (2), 418–430.

Hosseinzadeh, A., Algomaiah, M., Kluger, R., Li, Z., 2021. Spatial analysis of shared e-scooter trips. Journal of transport geography 92, 103016.

Kong, H., Jin, S. T., Sui, D. Z., 2020. Deciphering the relationship between bikesharing and public transit: Modal substitution, integration, and complementation. Transportation Research Part D: Transport and Environment 85, 102392.

Kranstauber, B., Smolla, M., Safi, K., 2017. Similarity in spatial utilization distributions measured by the earth mover's distance. Methods in Ecology and Evolution 8 (2), 155–160.

Li, H., Yuan, Z., Novack, T., Huang, W., Zipf, A., 2022. Understanding spatiotemporal trip purposes of urban micro-mobility from the lens of dockless e-scooter sharing. Computers, Environment and Urban Systems 96, 101848.

Lin, L., He, Z., Peeta, S., 2018. Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies 97, 258–276.

Liu, M., Seeder, S., Li, H., et al., 2019. Analysis of e-scooter trips and their temporal usage patterns. Institute of Transportation Engineers. ITE Journal 89 (6), 44–49.

Luo, H., Zhang, Z., Gkritza, K., Cai, H., 2021. Are shared electric scooters competing with buses? a case study in indianapolis. Transportation Research Part D: Transport and Environment 97, 102877.

Ma, Q., Xin, Y., Yang, H., Xie, K., 2022. Connecting metros with shared electric scooters: Comparisons with shared bikes and taxis. Transportation Research Part D: Transport and Environment 109, 103376.

Ma, T., Liu, C., Erdoğan, S., 2015. Bicycle sharing and public transit: Does capital bikeshare affect metrorail ridership in washington, dc? Transportation research record 2534 (1), 1–9.

Ma, X., Ji, Y., Yang, M., Jin, Y., Tan, X., 2018. Understanding bikeshare mode as a feeder to metro by isolating metro-bikeshare transfers from smart card data. Transport policy 71, 57–69.

Massey, Jr., F. J., 1951. The kolmogorov-smirnov test for goodness of fit. Journal of the American Statistical Association 46 (253), 68–78.
URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1951.10500769

McKenzie, G., 2019. Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in washington, d.c. Journal of Transport Geography 78, 19–28.
URL https://www.sciencedirect.com/science/article/pii/S0966692319302741

Merlin, L. A., Yan, X., Xu, Y., Zhao, X., 2021. A segment-level model of shared, electric scooter origins and destinations. Transportation Research Part D: Transport and Environment 92, 102709.

Mood, A. M., 1954. On the asymptotic efficiency of certain nonparametric two-sample tests. The Annals of Mathematical Statistics 25 (3), 514–522.
URL http://www.jstor.org/stable/2236833

NABSA, 2022. 2021 shared micromobility state of the industry report. Tech. rep.

Oeschger, G., Carroll, P., Caulfield, B., 2020. Micromobility and public transport integration: The current state of knowledge. Transportation Research Part D: Transport and Environment 89, 102628.

Qin, H., Gao, J., Kluger, R., Wu, Y.-J., 2018. Effects of perception on public bike-and-ride: A survey under complex, multifactor mode-choice scenarios. Transportation research part F: traffic psychology and behaviour 54, 264–275.

Rubner, Y., Tomasi, C., Guibas, L. J., 2000. Earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 40 (2), 99–121, cited By :2987.
URL www.scopus.com

Schwinger, F., Tanriverdi, B., Jarke, M., 2022. Comparing micromobility with public transportation trips in a data-driven spatio-temporal analysis. Sustainability 14 (14), 8247.

Su, L., Yan, X., Zhao, X., 2022. Spatial equity of micromobility systems: A comparison of shared e-scooters and station-based bikeshare in washington dc. arXiv preprint arXiv:2208.09107.

Sun, B., Garikapati, V., Wilson, A., Duvall, A., 2021. Estimating energy bounds for adoption of shared micromobility. Transportation Research Part D: Transport and Environment 100, 103012.

URL https://www.sciencedirect.com/science/article/pii/S1361920921003102

Tuli, F. M., Mitra, S., Crews, M. B., 2021. Factors influencing the usage of shared e-scooters in chicago. Transportation research part A: policy and practice 154, 164–185.

Urbanek, S., Rubner, Y., 2022. emdist: Earth Mover's Distance. R package version 0.3-2. URL https://CRAN.R-project.org/package=emdist

Watson, G. S., 1961. Goodness-of-fit tests on a circle. Biometrika 48 (1/2), 109–114. URL http://www.jstor.org/stable/2333135

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1 (6), 80–83. URL http://www.jstor.org/stable/3001968

Yan, X., Yang, W., Zhang, X., Xu, Y., Bejleri, I., Zhao, X., 2021. A spatiotemporal analysis of e-scooters' relationships with transit and station-based bikeshare. Transportation research part D: transport and environment 101, 103088.

Yang, Y., Heppenstall, A., Turner, A., Comber, A., 2019. A spatiotemporal and graph-based analysis of dockless bike sharing patterns to understand urban flows over the last mile. Computers, Environment and Urban Systems 77, 101361.

Zhao, P., Li, S., 2017. Bicycle-metro integration in a growing city: The determinants of cycling as a transfer mode in metro station areas in beijing. Transportation research part A: policy and practice 99, 46–60.

Zhu, R., Zhang, X., Kondor, D., Santi, P., Ratti, C., 2020. Understanding spatio-temporal heterogeneity of bike-sharing and scooter-sharing mobility. Computers, Environment and Urban Systems 81, 101483.

Ziedan, A., Shah, N. R., Wen, Y., Brakewood, C., Cherry, C. R., Cole, J., 2021. Complement or compete? the effects of shared electric scooters on bus ridership. Transportation Research Part D: Transport and Environment 101, 103098.