Using Big and Open Data to Analyze Transit Oriented Development: New Outcomes and Improved Attributes

By

Dr. Jiangping Zhou [1#]

zhoujp@hku.hk

&

Dr. & Professor Chris Webster[1]

cwebster@hku.hk

&

Ms. Yuling Yang[1]

yuling93@connect.hku.hk

[1] Department of Urban Planning and Design, Faculty of Architecture (FoA), The University of Hong Kong (HKU), Hong Kong SAR, China

[#] Corresponding author

Bio:

Dr. Jiangping Zhou is Associate Professor, FoA, HKU.

Dr. Chris Webster is Professor and Dean, FoA, HKU.

Ms. Yuling Yang is a research assistant, FoA, HKU.

ABSTRACT

Problem, research strategy, and findings: This study investigates how to exploit big and open data (BOD) to quantitatively examine the relationships between “TOD attributes” and “TOD outcomes”. Here, TOD attributes are measurable or perceivable attributes that TOD proponents cherish. TOD outcomes are the targeted outcomes, e.g., increased ridership, associated at least partially with TOD attributes. Based on BOD from Shenzhen, China, this study creates indicators to measure both TOD attributes and outcomes. It explores the associations of TOD attributes, including centrality of a TOD site, travel time to the central business district, density, destination, diversity, and design, with TOD outcomes. It identifies the TOD attribute that best predicts TOD outcomes such as metro ridership, frequent riders, people co-located in a station area, and ratios derived from these outcomes. It finds that special neighborhoods, specific metro lines, and age of the district significantly influence TOD outcomes. The study’s limitations are: one, the BOD used do not directly measure TOD attributes and proxies must be employed; two, BOD used contain little information about “why”, “who”, and “how”, e.g., why people rode transit, who they were, and how they perceived/appreciated various TOD attributes.

Takeway for practice: BOD-derived variables allow planners to revalidate existing planning guidelines and principles concerning TOD and adapt them to local contexts. BOD can also be used to formulate new metrics to evaluate different TOD plans or projects in ways not achievable with traditional data alone. In short, BOD can and should be used to refine TOD analytics, design, and to implement corresponding theories in pursuit of TOD.

Key words: Transit-oriented development; Open data; Big data; TOD attributes; TOD outcomes

Transit-oriented development (TOD) is a strategy to integrate transit services and land use in areas surrounding transit stations, i.e., station areas. Existing TOD studies tend to focus on two types of variables: TOD attributes and outcomes. TOD attributes refer to measurable or perceivable qualities such as **d**ensity, **d**iversity, and **d**esign (3Ds) (Cervero and Kockelman, 1997), which reflects "the orientation of land use towards the use of transit" (Singh et al., 2017, p.96). TOD outcomes refer to phenomena such as stability and growth in transit ridership and increased share of transit use, which are commonly pursued as policy or planning goals through TOD (e.g., see Noland et al., 2014; Renne, 2007). If a station area is a thriving community, it should have continuous and large incoming and outgoing flows of people, a wide variety of socioeconomic activities around the clock, and a decent concentration of people therein regardless of hour of the day (c.f., Jacobs, 1961). Therefore, when improving station areas, planners attempt to achieve some threshold of use in one place simultaneously, whether it is about activities or flows of people. Such simultaneity was generally termed "co-location" (Goffman, 1966).

Even without any formal public interventions, co-location can be observed as having evolved naturally over time. Enhancing or reproducing such urban milieu around transit infrastructure via public interventions, e.g., government-led TOD projects, typically require expensive investment and substantial coordination among stakeholders (e.g., see Cervero, 1998; ITDP, 2017; Suzuki et al., 2013; TCRP, 2007). Such investment is made in the belief that it will engineer or foster TOD attributes and optimize TOD outcomes, co-location in particular. This paper asks: how can we effectively quantify TOD attributes and outcomes, and how can we best measure and investigate the relationships between the two based on big and open data (BOD) in a way that helps us design better station areas? Do TOD attributes always contribute to TOD outcomes across different kinds of station areas, i.e., both TOD sites and non-TOD sites? If so, which of the TOD attributes contributes the most across station areas?

Prior to the emergence of BOD, traditional data dominated TOD studies. Consequently, few existing studies were able to include a large number of station areas, unless significant efforts were undertaken to establish a database that combined data from multiple sources (e.g., see Chatman et al., 2014; Renee, 2009; Renne et al., 2016). Traditional data sources are known to have several limitations. Census or survey data, for instance, are collected infrequently or as a one-off and typically have small sample sizes, and limited temporal and/or spatial coverage. They may capture coarse changes between two waves of data collection at best and obtain incomplete and inaccurate results at worst. They can easily overlook changes between two waves of data collection and samples might purposefully or accidently be extreme or exemplars instead of randomly selected, representative, or commonplace or if not representative, heterogeneous enough to capture significant patterns between TOD attributes and outcomes.

In our study, we illustrate how to use BOD to examine whether and by how much TOD attributes contribute to TOD outcomes across metro station areas in Shenzhen, China. We find that BOD can feasibly be used to replace traditional data for measuring both TOD attributes and TOD outcomes. Furthermore, BOD provide opportunities for planners to develop new metrics for TOD outcomes that characterize thriving communities; for example, co-location of people. We validate previously reported empirical findings, now established as TOD doctrine. Specifically, the 3Ds can significantly predict TOD outcomes across station areas. Regional metro accessibility quantified by measures of a metro station’s systemic centrality, is predictive of TOD outcomes. Such information can arguably help prioritize the limited investment in the transport system to realize more TOD outcomes. We discover the counterproductive and multiplier effects of TOD attributes on TOD outcomes. In addition, context-specific TOD attributes in Shenzhen, such as presence of urban village(s) (traditional informal densely occupied housing areas), a ring line, and a younger district are all significantly and positively correlated to TOD outcomes.

The remainder of the paper is organized as follows. Section 2 reviews the existing literature on TOD attributes, TOD outcomes, and their relationships. Section 3 presents our empirical models, which produce results that are generally consistent with the existing studies based on traditional data, but also yield new insights from BOD analysis. Section 4 concludes and stresses that TOD planners and analysts can revisit the data and indicators they currently use to analyze and assess TODs and keep updating their respective inventories of the data and indicators, which should now include BOD and various BOD-based indicators. BOD can enable them to measure certain TOD attributes and outcomes (e.g., destinations in station areas and frequent riders) more efficiently, continuously, and systematically.

TOD Attributes, TOD Outcomes, and Their Relationships

At some risk of oversimplifying the complexity embedded in the existing literature, we divide the literature into three streams: those that define and/or quantify TOD attributes and outcomes; those that categorize TODs; and those that explore relationships between TOD attributes and outcomes.

TOD attributes

Seminal studies include Calthorpe’s (1993) description of TOD-ness such as mixed land use and short distance to transit services and Cervero and Kockelmans’ (1997) simplification of

TOD-ness into the 3Ds (density, diversity, and design). The 3Ds can be quantified using variables such as accessibility to jobs, land use mix, and number of road intersections. Based on those studies, more scholars have used the 3Ds or introduced other ***D*** attributes, such as ***D****istance to transit* and ***D****esination accessiblity* (e.g., see Bertolini et al., 2012; Carlton; Cervero et al., 2002; Cervero et al., 2004; Olaru and Curtis, 2015; Renne, 2009; Wang and Woo, 2017). Some other scholars have identified/quantified a wider range of TOD attributes (See Table A2 in Technical Appendix). Many efforts, however, have remained "at a thematic level" or "rested on subjective qualitative appraisal, or critique of design or built-form outcomes" (Hale, 2014, p. 492).

TOD outcomes

Renne and Wells (2005) identified 56 indicators of TOD outcomes. Renne (2007) measured TOD outcomes in Perth, Australia based on 30 indicators in six categories. Zamir et al. (2014) measured TOD outcomes in the U.S., focusing on trip generation, trip length, and mode share. The relationships between TOD attributes (perceived and objective) and actualized TOD outcomes are complex, as with all urban dynamics. People perceive different levels of "effective quality and accessibility of the infrastructure (inputs)" across station areas and make different choices of travel and housing accordingly (Laaly et al., 2017, p.31). Such choices in turn influence such measurable aggregate outcomes as car usage and transit ridership. Singh et al. (2017) demonstrated how a single index can be used to measure TOD attributes such as population density, and outcomes such as passenger load simultaneously in the Netherlands. For stations with low scores, the score distribution can be used to identify unique problems concerning TOD outcomes.

Categorizing TODs

Indicators have been used to categorize station areas in order to help decision-makers to formulate more context-sensitive plans or policies for TODs. Focusing on redevelopments near rail stations in Europe, Bertolini (1996, 1999) proposed the node-place model (NPM) to differentiate among station areas and to guide related project evaluation and policymaking. The NPM epitomizes the dual nature of station areas: transportation interchanges ("nodes") and activity poles ("places"). It has been used in many customized applications (e.g., Reusser et al., 2008; Chorus and Bertolini, 2011; Kamruzzaman et al., 2014; Lyu et al., 2016). Few of these used BOD, and most only analyzed designated TOD sites, which are a small subset of station areas. More recently, Zhou et al. (2018) used BOD (population heatmaps, points of interest [destinations], and social media check-in's) to classify metro station areas in Shenzhen and to evaluate the performance of these areas using population per hour per destination. Drawing on the TOD indices by Gu et al. (2018), Zhou et al. (2019) examined how these indices, regional-level TOD attributes, and station area characteristics influenced metro ridership and population by station area and the ratio of the two in Shenzhen. Their study confirms that TOD attributes of metro station areas could well predict metro ridership. The magnitude of the impacts of those attributes, however, was smaller than that of regional-level TOD attributes.

Relationships between TOD attributes and outcomes

Most published studies claim that TOD attributes bring about at least some TOD outcomes. Table A2 in Technical Appendix synthesizes 36 studies selected from the Web of Science and Google Scholar databases using the forward reference searching method. The bibliographic search in both databases started with two classics on the topic: Calthorpe (1993) and Cervero and Kockelman (1997). Studies citing them and their derivatives were identified, collected and read, yielding 50+ studies. Table A2 contains a subset of the studies (n=36) that specifically address relationships between TOD attributes and TOD outcomes. Based on this subset, only seven used BOD to quantify TOD attributes and outcomes.

Strong positive relationships have been widely found between density attributes and transit ridership. Other factors such as mixed land use, proximity to transit, walkability, block per square mile, and streetscape design are also positively correlated with ridership. This confirms more general findings of the most cited literature on built environment and travel (e.g., Cervero and Kockelman, 1997; Ewing and Cervero, 2010). Cervero and Kockelman (1997) found that density, land-use diversity, and pedestrian-oriented designs generally reduce auto trip rates and encourage non-auto travel in statistically significant ways. Through meta-analyses, Ewing and Cervero (2010) found that bus and train use are positively related to services and street design variables. But population and job densities were only weakly associated with mode choice.

The seven studies using BOD developed some new insightful metrics to quantify TOD attributes, TOD outcomes, and their relationships. Smartcard data, for instance, can be used to get timely and reliable transit ridership across more temporal resolutions and more sites (Sung and Oh, 2011; Zhou et al., 2019). OpenStreetMap data can be utilized to operationalize the regional accessibility of rail stations with lower costs (Papa and Bertolini 2015). Open data sources such as Google Map and Walk Score can be exploited to quantify more TOD attributes across all station areas in a city (Lyu et al., 2016). In this study, we show how we synthesize various BOD to formulate indicators to measure more TOD attributes and outcomes across nearly all the metro station areas in a mega-city and to quantitatively assess the relationships between the two.

Our study of Shenzhen can be viewed as extension and/or improvements of Singh et al. (2017), Gu et al. (2018), and Zhou et al. (2018, 2019). Unlike Singh et al. (2017), it uses indicators instead of indices to measure TOD attributes and outcomes simultaneously. Thus, impacts of individual TOD attributes on different TOD outcomes can be singled out and compared. Similar to Singh et al. (2017), Gu et al. (2018) and Zhou et al. (2019) used indices to measure multiple TOD attributes (e.g., street network density, expressway density, and ground-floor retail density) simultaneously. They were thus unable to differentiate those TOD attributes' respective impacts on TOD outcomes, including the sign and magnitude of the impacts. However, policy-makers care about the sign and magnitude when making decisions to improve TOD attributes.

Our study has introduced more BOD-derived indicators to measure individual TOD attributes (e.g., ratios of various destinations, land use mix, walkability, and bikability) by station area than Gu et al. (2018) and Zhou et al. (2018, 2019). By exploiting extra BOD and developing new analytics, we also formulate indicators for "new" TOD outcomes, e.g., frequent riders produced in a station area and the ratio of these riders to the ridership produced in the station area. These new TOD outcomes allow us to better differentiate riders to and from metro station areas and design and improve metro station areas by accounting for more nuanced differences between different rider groups.

Last but not least, we use two sets of BOD of the same week (May 15 to 19, 2017) when operationalizing TOD outcome indicators. In Zhou et al. (2019), the two sets of BOD are not of the same week (May 20 to 26, 2017 vs. May 17 to 19, 2019), which could influence the accuracy and validity of corresponding TOD outcome indicators and analytical results and findings based on those indicators. Table A1 in Technical Appendix compares our study and the four papers cited above more systematically.

TOD attributes, TOD outcomes, and Their Relationships in Shenzhen

Study site

We chose Shenzhen, one of the four first-tier cities in China, as our study site. It was a small fishing village until China's opening-up and reform in the late 1970s. In about 40 years, Shenzhen has accumulated an official population of 11.9 million across a jurisdiction of 1,997 square kilometers (SSB and NBSSOS, 2017). Unofficial population estimates, which include internal migrant workers not registered in the city, may reach 25 million (China Mobile, 2017). As of August 2017, Shenzhen had 166 metro stations (one station has missing values in this study and is omitted) on eight metro lines with a total length of 285 kilometers. Many areas around these stations (metro station areas) were purposely designed as TODs. In our study, we define a metro station area as an area within 800 meters of a metro station, which is taken to identify a community better served in terms of all modes of access (including walking and biking) to metro services compared to other communities.

Shenzhen has three advantages as a study site: First, a high percentage of its metro station areas were master-planned and built according to the concepts of TOD. Often, international TOD experts were invited to work with local counterparts on Shenzhen's TOD projects. Thus, we can expect an unusually high amount of internationally endorsed TOD attributes in Shenzhen's metro station areas. We claim, therefore, that Shenzhen is something of an international laboratory for a BOD-enabled study of TOD attributes and outcomes.

Second, given the population size and significance of Shenzhen in China, many BOD datasets for Shenzhen have been produced, verified, and used by different stakeholders. More importantly, some of these stakeholders have been willing to share their data with researchers and scholars for the greater public goods. The archived anonymous bike-sharing data used in our study, for instance, were kindly provided by OfO, a company that occupies a large share of the bike-sharing market in Shenzhen.

Third, compared to most well-established metropolises, Shenzhen is much younger and has more underutilized land in metro station areas that can be (re)developed. This means that our study results and findings stand a good chance of being directly used to guide the enhanced design of TODs in the city. Those results will also provide a useful reference for those developing cities with metro lines. In China alone, there are already more than 40 cities that have at least one metro line. Thousands of metro station areas have emerged as a result.

Data used

BOD have rarely been able to measure patterns related to TOD outcomes such as co-location. In TCRP (1995), for instance, census and survey data were used to quantify how population density influenced transit ridership. These data only captured impacts of estimated population density for a given year on average transit ridership during survey days. One could not tell from this how daily, weekly, or monthly average population densities influence average transit ridership for a particular day, week, or month. BOD, by contrast, are collected continuously and automatically. They also cover larger samples or even full population of study subjects. BOD can therefore be disaggregated to unravel dynamics and changes across more sub-groups and spatiotemporal resolutions (Batty 2013). Furthermore, because of continuousness, higher numbers, and heterogeneity, BOD allow more segmentation of the samples and/or sensitivity analysis of corresponding results.

BOD, however, do have several drawbacks too. First, it could be costly and challenging for analysts to acquire and process BOD. In the US, for instance, archived cellular network data are so expensive that few in the planning domain have used them for their analyses. Second, BOD may only provide a partial picture. For example, smartcard data collected by most transit agencies only contain information of a rider's origin, destination, start/end time of her/his trip, and route choice. Third, BOD might not be representative of the population. Smartphone data, for instance, overlook those users who do not own a smartphone.

In our study, we are aware of the above pros and cons of BOD and we try to exploit those pros to construct some BOD-based indicators for the empirical study. We currently collected data for 2017 and conducted only cross-sectional but not longitudinal study. We simply designed/decided indicators based on the existing literature and our own local knowledge, instead of consulting local riders and decision-makers. For TOD attributes, we downloaded the latest (2017) Shenzhen road transport network from OpenStreetMap (https://www.openstreetmap.org); retrieved a 2017 point of interest (destination) database (in .txt format) from Weibo, which covered destinations like residential estates, theaters, schools, restaurants, supermarkets, and hospitals; estimated traveling time and distance between metro stations using Baidu Map (the Chinese version of Google Map); and obtained bike-sharing data in 2017 from OfO. More information about these BOD can be found in Technical Appendix.

Based on these BOD sources, we calculated indicators for TOD attributes. Constructing indicators such as an up-to-date and near 100% census of destinations by metro station area would be infeasible had we relied on traditional data from sources such as local land use or tourist maps. Table 1 lists all the indicators that we formulated and used in the study. Most indicators relate to theory and evidence arising from or tested in the existing studies synthesized in Table A2 in Techincal Appendix.

Indicators of regional accessibility/centrality in Table 1 are calculated by the Spatial Design Network Analysis (sDNA) tool and ArcGIS plugin (Cooper et al., 2019), with OpenStreetMap data as input. "Regional metro accessibility" is an indicator measuring how centrally located a metro station is within the local metro network. "Intermediate stations" quantifies the average number of intermediate stations from a station to others. Analogues of intermediate stations have been used in other studies such as Papa and Bertolini (2015), Lyu et al. (2016), and Gu et al. (2018). Network directed-ness is the ratio of the actual network distance to the straight-line distance between a pair of stations on the metro network (Cooper, 2019). Inspired by TCRP (1995), we formulated indicators to measure the weighted distance or time between a metro station area and the Central Business District (CBD) ("distance or time to CBD"). The data and tool used to produce these indicators are Baidu Map and Baidu Map API, respectively. Given that Shenzhen has two officially designated CBDs (Luohu and Futian), we employed the percentage of destinations of each metro station area as weights when calculating the weighted distance or time. Simply put, other things being equal, a metro station area with more destinations would matter more.

We used BOD to indirectly measure (1) design by bikeability and walkability, (2) density and diversity by destination intensity, and (3) distance to transit by bus stop and metro station provision. These are clearly proxy measures because the data were not specifically designed for the purpose that we are using them. Using the number of shared-bike users from OfO, we indirectly measured distance to a metro station from different destinations and bikeability of a metro station area. We assume that the more shared-bike users, the shorter the distance between a metro station and different meaningful destinations and the better bikeability. Using the sDNA tool and ArcGIS plugin, we calculated "betweenness" to measure walkability. In our study, "betweenness" measures the number of times that a midpoint of a link (street segment) lies on the shortest paths between other pairs of midpoints within 5000 meters of the midpoint (Cooper, 2019, also see Figure A2 in Technical Appendix). We chose 5000 meters as a suitable radius for car accessibility in sDNA, following other published studies (e.g., Xiao et al., 2016). Walkability measures the percentage of car-priority streets with either top 30% or 50% values of betweenness in a metro station area. Here, car-priority streets are those streets that have dedicated right-of-ways for vehicles and that are classified as trunk roads for vehicular traffic. Walkability has two similar but slightly different indicators. Yet, in each regression model only one walkability indicator improves the goodness of model-fit, which is kept in the final analysis.

We followed Loo and du Verle (2017) in using a Simpson index to quantify diversity albeit that they used traditional land use maps whereas we used destinations from an open source as input. Similarly, we used destinations to quantify land use intensity. Two indicators measuring distance to transit are "bus stops" and "metro stations", which are the numbers of bus stops and metro stations in a metro station area, respectively. They are comparable to those used by Dill (2008) and Bernick and Cervero (1997) but are derived based on open data too.

Based on our knowledge about the study site, we measured additional context-specific characteristics of each metro station area with other indicators such as the presence of urban villages (homes to Shenzhen large migrant worker population and containing over 50% of the city's population); when a metro station started operation; whether a metro station area is on a special line (i.e., ring or commuter lines); whether a metro station area is in the youngest administrative district (Longhua) established by the Shenzhen Municipal Government; whether a metro station area is in an administrative district that contains a CBD; whether a metro station area is located along the metro line that was constructed in or prior to 2004, the first phase of the local metro system development; and whether a metro station area is located along the metro line that was constructed in or after 2012, the most recent phase of the local metro system development (for more details, see Table 1). These allow us to tease out local context-specific effects.

Data for constructing TOD outcome variables came from two sources: Baidu heatmaps, which have been used in Zhou et al. (2018, 2019) and smartcard data. The data in Zhou et al. (2019) were from two different weeks but in our study, they were from the same week to minimize any biases. Figure 1 is a sample of Baidu heatmaps.

(Figure 1 is here.)

Smartcard data have been used in several existing studies (e.g., Sung and Oh, 2011; Zhou et al., 2019). Compared to methods that use population counts from traditional data such as censuses, the advantages of our estimation method based on the heatmaps include: (a) estimates can be continuously and automatically updated; (b) estimates cover both established built-up areas and new development areas in real time, which is important for fast-changing cities like Shenzhen; (c) it can be easily aggregated to different spatiotemporal units of analysis, e.g., to compute average number of smartphone users by census tract by hour on weekdays, and average number of smartphone users by metro station area per day on weekdays or weekends.

Based on the above data, we formulated six indicators of TOD outcomes (Table 2).

(Table 2 is here.)

Metro ridership produced in a metro station area has commonly been used to measure TOD outcomes (See Table A2 in Technical Appendix). We created five other indicators partially based on BOD, giving six indicators altogether (for more information, please see Table 2):

(1) metro ridership into a metro station area per hour on weekdays ("metro ridership" for shorthand hereafter);
(2) number of people co-located in a metro station area per hour on weekdays ("people per hour" for shorthand hereafter);
(3) number of frequent riders into a metro station area per hour on weekdays ("frequent riders" for shorthand hereafter);
(4) Indicator (3) divided by Indicator (1), i.e., ratio of frequent riders to metro ridership ("the net ratio of frequent riders" for shorthand hereafter);
(5) Indicator (1) divided by Indicator (2), i.e., ratio of metro ridership to people per hour ("ratio of metro riders" for shorthand hereafter); and
(6) Indicator (3) divided by Indicator (2), i.e., ratio of frequent riders to people per hour ("gross ratio of frequent riders" for shorthand hereafter).

The first three indicators measure the co-location of three groups of people in metro station areas while the last three indicators measure the odds of such co-locations. Compared to those indicators listed in Table A2 in Technical Appendix, these six indicators together inform us about additional TOD outcomes, e.g., how different groups of people are attracted to, or co-locate in metro station areas, their sizes, and whether some of them frequently travel by metro. This takes the measurement of TOD outcomes from merely binary success/failure to cover outcomes relating to depth, dynamics, quality, and sustainability of TODs.

Modelling TOD attributes against TOD Outcomes

To quantify the relationships between TOD attributes and TOD outcomes, we fitted six regression models (details explained in Technical Appendix). The models explain 41% to 62% of the variation in the TOD outcomes (details see Table A3 in Technical Appendix). The results verify which TOD attributes are correlated to the metro ridership, a commonly seen TOD outcome in the existing studies. They also show how TOD attributes can predict “new” TOD outcomes such as frequent riders and net ratio of frequent riders that few existing studies have looked at.

Regional accessibility, Ds, and metro ridership

Regional (metro) accessibility has the biggest positive impact on metro ridership. This is in general consistent with findings of existing studies such as TCRP (1995) and Renne et al. (2017). TCRP (1995) found that distance to CBD from a transit station and employment size of the CBD had significant positive impacts on outgoing ridership generated at a station. Renne et al. (2017) showed that network accessibility to jobs and population can increase transit usage. Inspired by Zhou et al. (2019), we also used sDNA to measure regional metro accessibility of a metro station. Such accessibility reflects the average distance between a given metro station to all other metro stations rather than distance from this station to a priori predefined center point (Cooper, 2019). To examine if there is spatial correlation between regional metro accessibility and metro ridership, we also mapped the distribution of the two. It seems that only in the south of the city (Futian District) is there significant spatial correlation between regional metro accessibility and metro ridership (See Figure A1 in Technical Appendix).

Destination intensity, distance to transit, design significantly increase metro ridership. The prediction power of these variables, e.g., bus stops and walkability, however, is smaller than that of regional metro accessibility. Chatman et al. (2014) and Renee et al. (2016) have reported similar findings using indicators derived from traditional data.

Some localized factors significantly increase/decrease the metro ridership. Metro stations built in the first phase of the local metro system development and in the youngest administrative district (Longhua) would enjoy higher metro ridership whereas metro station areas along a commuter line would reduce metro ridership.

Regional accessibility, Ds, and other TOD outcomes

At least one of the regional accessibility indicators significantly predicts the other five TOD outcomes apart from metro ridership: people per hour, frequent riders, net/gross ratio of frequent riders, and ratio of metro riders. For example, fewer intermediate stations and shorter distance to CBD are associated with more people per hour and a lower net ratio of frequent riders. Higher regional metro accessibility is linked to more frequent riders, higher ratio of metro riders and higher gross ratio of frequent riders. These three measures of the regional accessibility: intermediate stations, regional metro accessibility, and distance to CBD, are, however independent of each other. Distance to CBD captures, for example, accessibility to high-level urban functions and services and certain types of employment. It also captures relative land value of each metro station area. Regional metro accessibility measures systemic connectivity and captures accessibility of a metro station area to everywhere else in the city. Intermediate stations measures spatial separation of two metro station areas and the more intermediate stations correspond to more actual travel time of metro riders, who do not face traffic congestion along the track like car drivers. The entry of these distinct measures into the models enables us to how locational attributes of a metro station affect TOD outcomes (More information can be found in Technical Appendix).

Higher proportions of retail and entertainment destinations are associated with more people per hour. However, the more people per hour see a lower percentage of residential destinations. Somewhat to our surprise, the Simpson index, which measures diversity of destinations, did not significantly predict any TOD outcomes. Bikeability, bus stops, and destination intensity all positively influence the number of frequent riders. This is in general consistent with the findings of Cervero and Kockleman (1997) and Ewing and Cervero (2010), which did not differentiate frequent riders and non-frequent riders. Distance to CBD significantly decreases the net ratio of frequent riders the most, followed by the percentage of restaurants. Compared to findings in the existing studies, the negative correlation between restaurant percentage and the net ratio of frequent riders is a new finding. This could mean that the percentage and even the number of restaurants in station areas are not a necessary condition for getting frequent riders using those areas. Most notably, whether a metro station area has one or more metro stations ("one or more metro station") influences the gross ratio of frequent riders the most.

Metro station areas with fewer destinations can expect a higher ratio of metro riders and gross/net ratios of frequent riders. A plausible explanation for the negative relationship between destinations and ratio of metro riders is that more destinations generate more people per hour from all sources, not just metro riders. This is a novel finding that we denote the TOD's counterproductive or multiplier effects. A counterproductive-effect example is that many metro station areas have attracted high-end restaurants and shopping malls, which offer customers free or cheap on-site parking. This could reduce the number of metro riders to and from these metro station areas. A multiplier-effect example is that a metro station area with much open space and few destinations can attract frequent riders, pedestrians, and cyclists simultaneously, raising the vibrancy of station areas but reducing ratio of metro riders. The metro riders and frequent riders seem to prefer metro station areas with different sets of destinations. The latter are more likely to reside in a metro station area with more restaurant destinations whereas the former prefer a residence in a metro station area with fewer entertainment and restaurant destinations.

Some context-specific factors also significantly influence the five BOD-measured TOD outcomes. For instance, a community in the youngest administrative district (Longhua) or a community served by metro services since 2004 have more frequent riders. In addition, urban villages significantly increase the net ratio of frequent riders. These findings reiterate the importance of accounting for affordable housing in TODs (c.f., Boarnet and Crane, 1998). It also suggest that the impacts of TOD attributes on the elasticity of demand for metro journeys may be higher for low-income dense neighborhood—an important evidence of policy relevance, especially for Shenzhen/China where there is currently a policy to retain rather than demolish these residential areas at the same time as active metro-line expansion. If a metro served area is served by Lines 4 or 5, a line running between the north and the south and a ring line, it can expect a higher gross ratio of frequent riders.

Academics and Planners: What They Can Learn

We conclude that BOD adds value to TOD technical and academic studies and to practices. Our study illustrates that BOD can supplement and even replace traditional data to measure TOD attributes and outcomes. Both planning academics and practicing planners should not overlook that value. And the most notable implications from our study are as follows.

First, compared to traditional data, BOD have advantages such as currency and fineness of spatiotemporal grain. It is costly and even impossible to build indicators such as people per hour and frequent riders across all metro station areas in a city using traditional data alone. In addition, (novel) indicators based on BOD can enable (a) validation of existing knowledge about TOD attributes, TOD outcomes, and their relationships, and (b) production of new insights into the relationships, as summarized below.

Second, regional metro accessibility significantly adds explanatory power in models explaining TOD outcomes (c.f., TCRP, 1995 and Renne et al., 2017). In general, the more connected a station area is in the urban/regional transit network, the better the TOD outcomes if we improve TOD attributes of the station area. In other words, we should pay as much attention to the systemic context of a station area when formulating TOD policies and implementing TOD projects as we pay to the local variables such as the Ds. Choice of candidate TODs for scarce investment should also be driven by network analytics at the first stage.

Third, the TOD's counterproductive and multiplier effects that we identified indirectly show the value of planning, which should holistically account for causes and effects. The property market, for instance, might not necessarily attract as many as destinations and customers in station areas that are conducive to a higher transit usage rate. Instead, it can incentivize private car usage by offering so-called "free-parking" for customers. There are cases that fewer destinations and more open space can boost walking and cycling in a station area, which in the end can contribute to more and frequent transit usage—transit usage, walking, and cycling can be mutually supportive of one another.

Fourth, the number of shared-bike users (as a proxy of "bikability"), which can serve as a proxy of the time-distance from a metro station to destinations in and around a station area and urban design quality, increases metro ridership, frequent riders and people per hour. However, bikability could not predict ratio of metro riders or the net ratio of frequent riders. This means that if we want to increase these ratios, the promotion of bike-sharing alone may not work. It is highly likely that frequent metro riders would use metro regardless of the availability of a shared-bike program or not. Few frequent metro riders will ride frequently simply because of the availability of a shared-bike program. Bikability is more likely to be the TOD attribute that increases the demand for occasional use of a particular metro station rather than promote regular use.

Fifth, the number and the percentage of different destinations, which indirectly measure land uses and various Ds, are significantly associated with all the six TOD outcomes. However, their respective impacts on different TOD outcomes are mixed. Thus, we must be cautious if our aim is to improve a set of TOD outcomes by changing destinations (and even land uses) and their composition in or around metro station areas. The latter might not always bring about all these intended improvements. More importantly, "disparate stakeholders" of metro station areas can have different perspectives about such improvements (Renne, 2009, p.241). It is therefore necessary to fully engage these stakeholders and understand their respective perspectives and, eventually reach agreement about the improvements and possible tradeoff among them. By exploiting BOD to study various metro station areas, this study reconfirms the existence of such tradeoff, for instance, increasing metro ridership versus stabilizing net ratio of frequent riders.

Sixth, despite the fact that many variables might be specific to Shenzhen, such as the presence of urban village(s), whether a metro station area is in the youngest administrative or not, whether a metro station area is served by a ring line, and when a metro station started to serve a community, the results concerning these variables as a whole suggest that history and cost (e.g., urban villages offer cheap housing) cannot be overlooked in TODs. In addition, we can say that location and line characteristics matter when we aim to improve TOD outcomes. We have also suggested that the urban village effect may in fact be a more generally applicable income effect.

What Can Be Done Next?

TOD analysts and planners can revisit the data they currently use to analyze and assess TODs and keep updating their respective inventories of the data, which should now include various BOD. More potentials of BOD are described below.

First, BOD can be used to monitor the impacts of improved TOD attributes on TOD outcomes of particular interest to decision-makers across virtually all metro station areas over time. BOD such as smartcard data and population heatmaps, for instance, can be used to help decision-makers to see how metro station areas' various people-related indicators, e.g., people per hour and net ratio of frequent riders change after the introduction of shared bikes and more destinations. By combing a natural experiment approach (e.g., see Meyer, 1995) and BOD, we can further explore whether specific improvements of TOD attributes change particular TOD outcomes and if so, to what degree.

Second, practitioners may attempt to link BOD to rider survey and interview data in order to better understand socioeconomic attributes of riders. For example, riders can be asked to voluntarily give their smartcard number in a rider survey so that we can know more about who the rider is, why s/he travels, and how s/he feels about her/his trips. In addition, focus-group meetings—including those who might not even own a smartcard—will be useful when designing or implementing BOD-facilitated research and policymaking. They will help groundtruth discrepancies between BOD-generated rider information and its real-world counterpart, in particular, riders' purpose, feelings, and perception (c.f., Harten, et al., 2018).

Third, as more BOD platforms and vendors emerge recently, it might be easier for the public to access BOD. For example, bike-sharing data for some cities can now be obtained from Bike Share Research (See: https://bikeshare-research.org/). In addition, Uber also started sharing with the public more anonymous trajectory data (See: https://movement.uber.com/cities?lang=en-US). Based on those data, we can update related indicators continuously to evaluate whether and to what degree certain station areas lose or gain transit ridership or people per hour because of the entry of new shared mobility modes.

Finally, by combing all the above, we could establish comprehensive and real-time performance monitoring systems and smartphone apps for all metro station areas in a city (Renee, 2009). Those systems and apps would automatically and continuously measure crucial indicators for TOD attributes and TOD outcomes that are of relevance to various stakeholders, e.g., transit agencies, riders, elected officials, and businesses. In the case of Shenzhen, for instance, given the fact that retail, entertainment and restaurant destinations are significantly increase people per hour in a metro station area, we can collect real-time and comprehensive information about these destinations. Such information could be made publicly available via smartphone apps or the Internet. The selection of appropriate indicators would need to be determined through deliberative communication among various stakeholders. Scholars may play a valuable role in this process by providing stakeholders with knowledge and insights generated in scientific BOD-enhanced data, methods, theory, and practice guidelines.

References:

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274-279. DOI: 10.1177/2043820613513390.

Bernick, M and Cervero, R. (1997) Transit Villages for the 21st Century. New York: McGraw-Hill.

Bertolini, L. (1996) Nodes and places: Complexities of railway station redevelopment. *European Planning Studies* 4(3), 331-345. DOI: 10.1080/09654319608720349.

Bertolini, L. (1999) Spatial development patterns and public transport: The application of an analytical model in the Netherlands. Planning Practice and Research 14(2), DOI:199-210. 10.1080/02697459915724.

Bertolini, L., Curtis, C. and Renne, J. (2012) Station area projects in Europe and beyond: Towards transit oriented development? *Built Environment* 38(1), 31-50. http://www.jstor.org/stable/23289699.

Boarnet, M. and Crane, R. (1998) L.A. Story: A reality check for transit-based housing. Journal of the American Planning Association, 63(2), 189-204. DOI: 10.1080/01944369708975914.

Calthorpe, P. (1993) The Next American metropolis: Ecology, Community, and the American Dream. Princeton, NJ: Princeton Architectural Press.

Carlton, I. (2007) Histories of Transit-Oriented Development: Perspectives on the Development of TOD Concept (WP-2009-02). University of California, Berkeley: Institute of Urban and Regional Development.

Cervero, R. (1998) Transit Metropolis: A Global Inquiry. Washington DC: Inland Press.

Cervero, R. and Kockelman, K. (1997) Travel demand and the 3Ds: Density, diversity, and design. Transportation Research D 2(3) 199-219.

Cervero, R., Ferrell, C., and Murphy, S. (2002) Transit Cooperative Research Program, Research Results Digest Number 52. Washington, D.C.: Transportation Research Board.

Cevero, R., Murphy, S., Ferrell, C., Goguts, N., Tsai, Y., Arrington, G.B., Boroski, J., Smith-Heimer, J., Golem, R., Peninger, P, Nakajima, E., Chui, E., Dunphy, R., Meyers, M., McKay, S. and Witenstein, N. (2004) Transit-oriented Development in the United States: Experiences, Challenges, and Prospects. Washington D.C.: Federal Transit Administration.

Chatman, D. G., Cervero, R., Moylan, E., Carlton, I., Weissman, D., Zissman, J., ... & Tischler, D. (2014) Making Effective Fixed-Guideway Transit Investments: Indicators of Success. Washington, DC: The National Academies Press. DOI:10.17226/22355.

China Mobile. 2017. Report on Shenzhen's estimated population. Guangzhou: Guangdong Innovation Center for Mobile Phone Data, China Mobile.

Chorus, P. and Bertolini, L. (2011) An application of the node place model to explore the spatial development dynamics of station areas in Tokyo. *Journal of Transport and Land Use* 4(1), 45–58. DOI: 10.5198/jtlu.v4i1.145.

Cooper, C. (2019) Spatial Design Network Analysis (sDNA) version 4.0 Manual. Cardiff University. Available, http://www.cardiff.ac.uk/sdna/software/documentation

Cooper, C., Chiaradia, A. and Webster, C. (2019) Spatial Design Network Analysis software, version 4.0, Cardiff University, http://www.cardiff.ac.uk/sdna/

Dill, J. (2008) Transit use at transit-oriented developments in Portland, Oregon, Area. *Transportation Research Record*, 2083, 159-167. DOI: 10.3141/2063-19.

Ewing, R. and Cervero, R. (2010) Travel and the Built Environment: A Meta-Analysis. *Journal of American Planning Association*, 76(3), 265-294. DOI: 10.1080/01944361003766766.

Goffman, E. (1966) [1963] Behavior in Public Places, New York: Free Press.

Gu, P., He, D., Chen, Y., Zegras, P.C., and Jiang, Y. (2018) Transit-oriented development and air quality in Chinese cities: A city-level examination. *Transportation Research D*. DOI: 10.1016/j.trd.2018.03.009.

Hale, C. (2014) TOD versus TAD: The great debate resolved...(?) *Planning, Practice and Research*, 29(5): 492-507. DOI: 10.1080/02697459.2012.749056.

Harten, J. G., Kim, A. M. and Brazier, C. (2018) Hidden informality in urban China: A mixed method approach to Shanghai's market for overcrowded housing. A working paper for the USC Lusk Center for Real Estate. Available at: https://pdfs.semanticscholar.org/040a/20a4708a894761bc936c90b2e40df7e05eb1.pdf. Accessed December 22, 2019.

Institute for Transport and Development (ITDP). (2017) TOD Standards Version 3.0: New York: ITDP.

Jacobs, J. (1961) The Life and Death of great American Cities. New York: Random House.

Kamruzzaman, M., Baker, D., Washington, S. and Turrell, G. (2014) Advance transit oriented development typology: Case study in Brisbane, Australia. *Journal of Transport Geography*, 34, 54-70. DOI: 10.1016/j.jtrangeo.2013.11.002.

Laaly, S., Jeihani, M. and Lee, Y. (2017) A multiscale, transit-oriented development definition based on context-sensitive paradigm. *Transportation Research Record*, 2671, 31-39. DOI:10.3141/2671-04.

Loo, B. P.Y. and du Verle, F. (2017). Transit-oriented development in future cities: Towards a Two-level Sustainable Mobility Strategy. *International Journal of Urban Sciences*, 21(sub 1), 54-67.

Lyu, G., Bertolini, L., and Pfeffer, K. (2016) Developing a TOD typology for Beijing metro station areas. *Journal of Transport Geography*, 55, 40-50. DOI: 10.1016/j.jtrangeo.2016.07.002.

Meyer, B. D. (1995) Natural and quasi-experiments in Economics. *Journal of Business and Economic Statistics*, 13(2), 151-161. DOI: 10.2307/1392369.

Nasri, A. and Zhang, L. (2014) The analysis of transit-oriented development (TOD) in Washington D.C. and Baltimore metropolitan area. *Transport Policy*, 32, 172-179. DOI: 10.1016/j.tranpol.2013.12.009.

Noland, R.B., Ozbay, K., DiPetrillo, S. and Lyer, S. (2014) Measuring Benefits of Transit Oriented Development. San Jose, CA: Mineta National Transit Research Consortium, San Jose State University.

Olaru, D. and Curtis, C. (2015) Designing TOD precincts: accessibility and travel patterns. *European Journal of Transport and Infrastructure Research* 15(1), 6-26.

Papa, E. and Bertolini, L. (2015) Accessibility and transit-oriented development in European metropolitan areas. Journal of Transport Geography, 47, 70–83. DOI: 10.1016/j.trangeo.2015.97.003.

Renne, J.L. (2009) Measuring the success of transit oriented development. In: Curtis, C., Renne, J.L. and Bertolini, L. (ed.) Transit Oriented Development: Making it Happen. Burlington, VT: Ashgate Publish Company.

Renne, J. L. (2007) Measuring the Performance of Transit-oriented Developments in Western Australia. Planning and Transport Research Centre. Perth, Australia: The Government of Western Australia.

Renne, J. L. (2009) From transit-adjacent to transit-oriented development. *Local Environment*, 14(1), 1-15. DOI: 10.1080/13549830802522376.

Renne, J. L., Hamidi, S. and Ewing, R. (2017) Transit commuting, the network accessibility effect, and the built environment in station areas across the United States. *Research in Transportation Economics* 60, 35-43.

Renne, J. and Wells, J. (2005) Transit-oriented Development: Developing A Strategy to Measure Success—Research Results Digest 294. Washington, D.C.: Transportation Research Board.

Reusser, D.E., Loukopoulos, P., Stauffacher, M. and Scholz, R.W. (2008) Classifying railway stations for sustainable transitions–balancing node and place functions. *Journal of Transport Geography*, 16 (3), 191–202. DOI: 10.1016/j.jtrangeo.2007.05.004.

Shenzhen Statistics Bureau (SSB) and National Bureau of Statistics' Survey Office in Shenzhen (NBSSOS). (2017). Shenzhen Statistical Year Book. Beijing: China Statistics Press.

Singh, Y.J., Lukman, A., Flacke,J. Zuidgeest, M. and van Maarseveen, M. (2017) Measuring TOD around transit nodes--towards TOD policy. *Journal of Transport*

Geography, 56: 96-117. DOI: 10.1016/j.tranpol.2017.03.013.

Sung, H. and Oh, J.T. (2011) Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*, 28 (1), 70-82. DOI: 10.1016/j.cities.2010.09.004.

Suzuki, H., Cervero, R. and Iuchi, K. (2013) Transforming Cities with Transit: Transit and Land-Use Integration for Sustainable Urban Development. Washington, D.C.: The World Bank.

Transit Cooperative Research Program (TCRP). (2007) Transit Oriented Development Traveler Response to Transportation System Changes, Report 95. Washington D.C.: Federal Transit Administration.

TCRP, 1995. Regional Transit Corridors: The Land Use Connection. Washington D.C.: TCRP.

Xiao, Y., Orford, S. and Webster, C.J. (2016) Urban configuration, accessibility, and property prices: A case study of Cardiff, Wales. *Environment and Planning B*, 43(1), 108-129. DOI: 10.1177/0265813515600120.

Wang, K. and Woo, M. (2017) The relationship between transit rich neighborhoods and transit ridership: Evidence from the decentralization of poverty. *Applied Geography*, 86, 183-196. DOI: 10.1016/j.apgeog.2017.07.004.

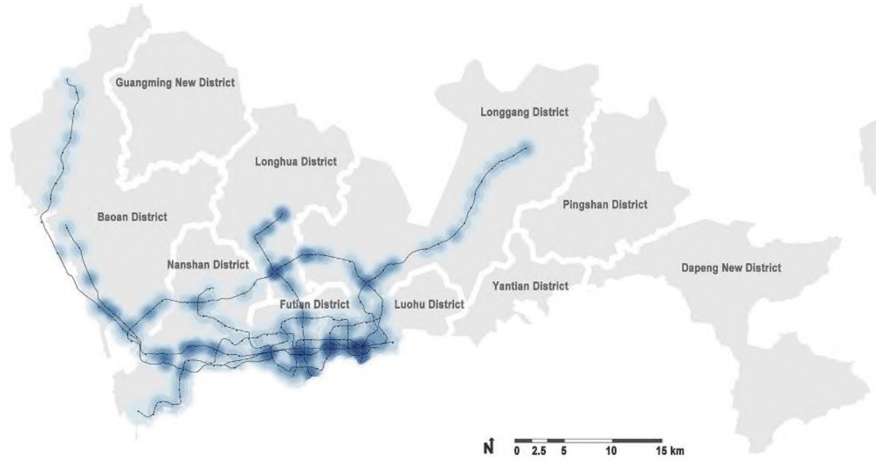
Zamir, K.R., Nasri, A. Baghaei, B., Mahapatra, S. and Zhang, L. (2014) Effects of transit-oriented development on trip generation, distribution, and mode share in Washington, D.C., and Baltimore, Maryland. *Transportation Research Record*, 2413, 45-53. DOI: 10.3141/2413-05.

Zhou, J., Wang, Q. and Liu, H. (2018) Evaluating transit-served areas with non-traditional data. *Journal of Transport and Land Use*, 11(1): 1323–1349. DOI: 10.5198/jtlu.2018.1296.

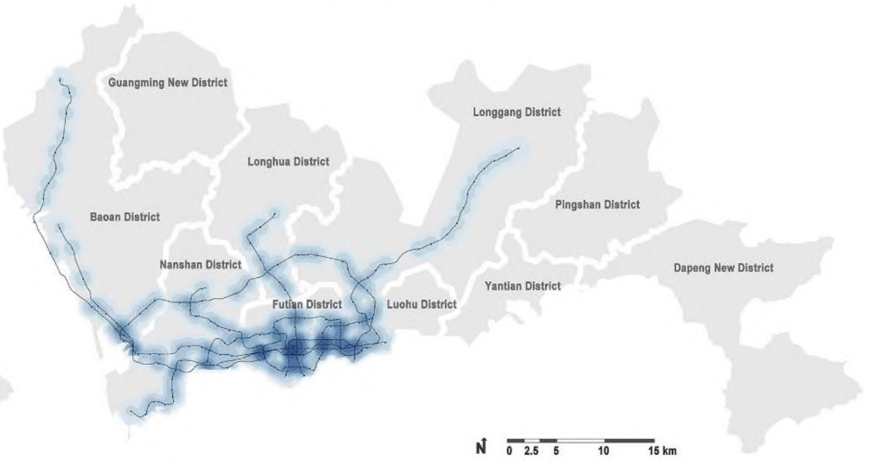
Zhou, J., Yang, Y., Gu, P., Yin, L., Zhang, F., Zhang, F. and Li, D. (2019) Can TODness improve (expected) performances of TODs? An exploration facilitated by non-traditional data. *Transportation Research D*, 74, 28-47. DOI:10.1016/j.trd.2019.07.008.

Technical Appendix

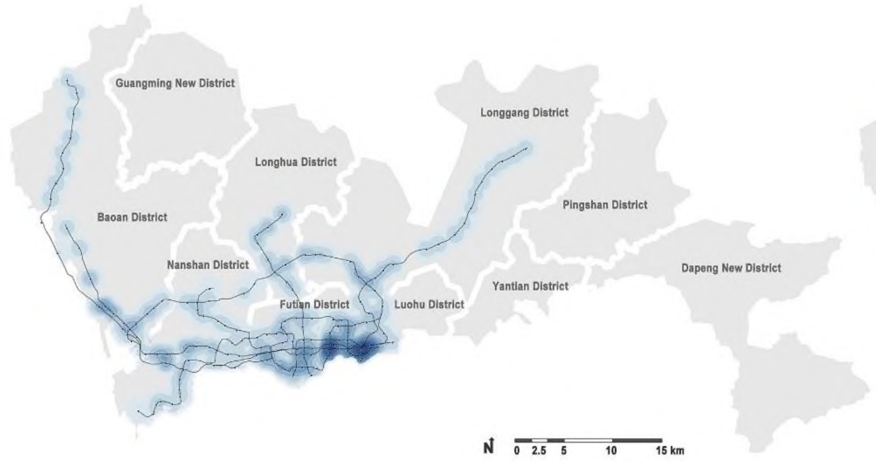
Metro Ridership



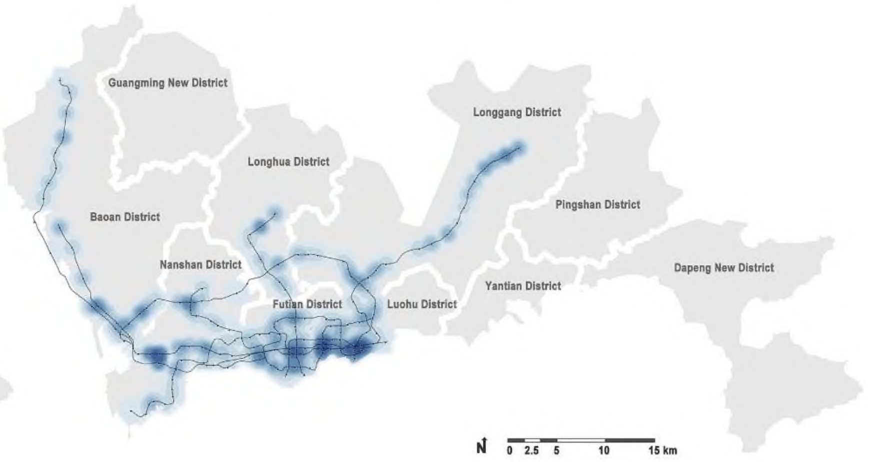
Regional Metro Accessibility



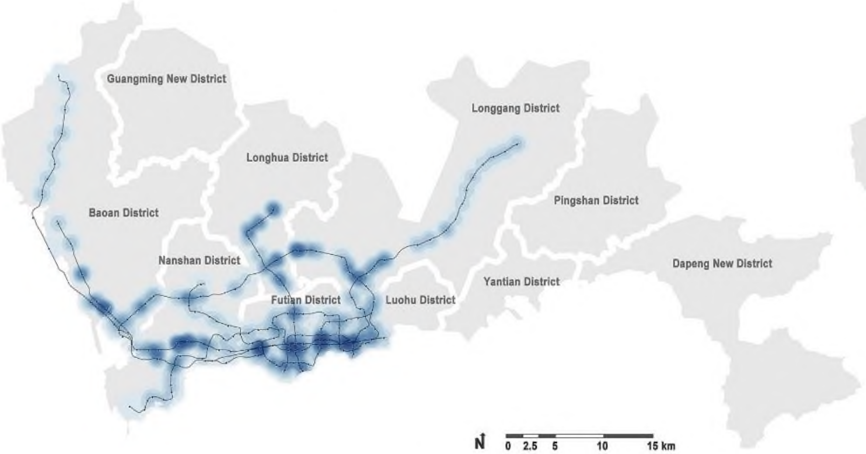
People per Hour



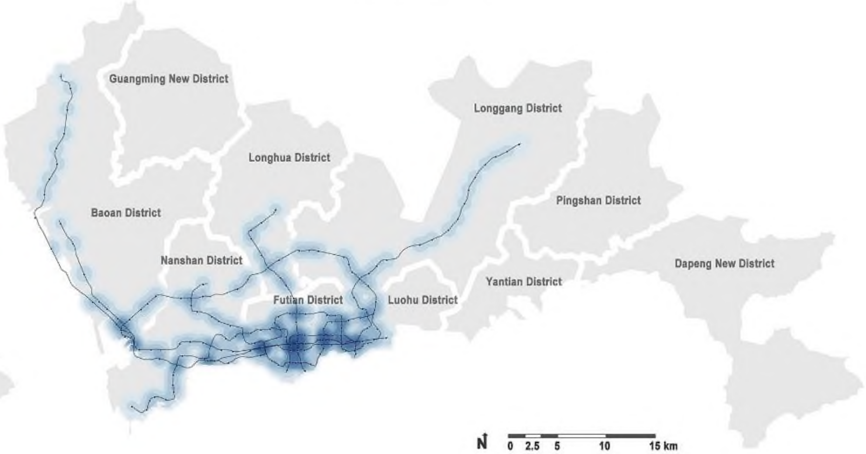
Time to CBD



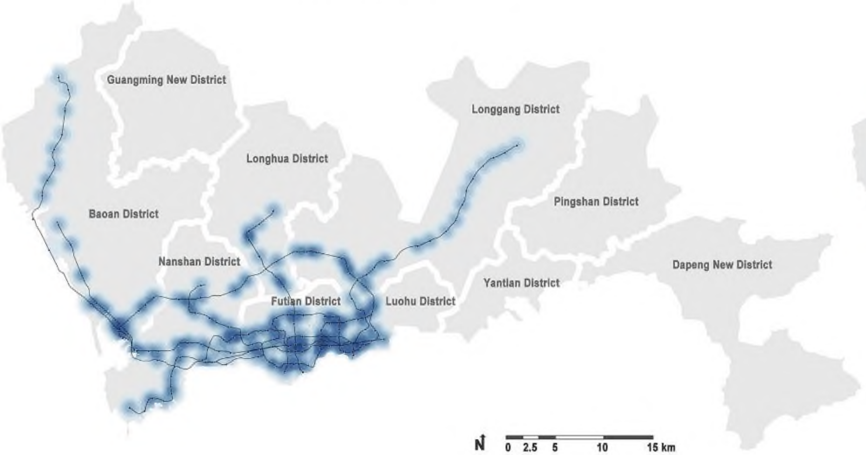
Frequent Riders



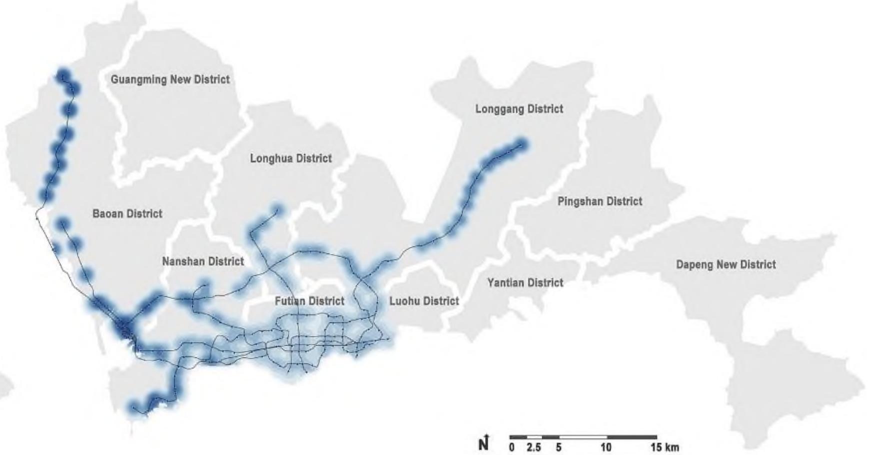
Regional Metro Accessibility



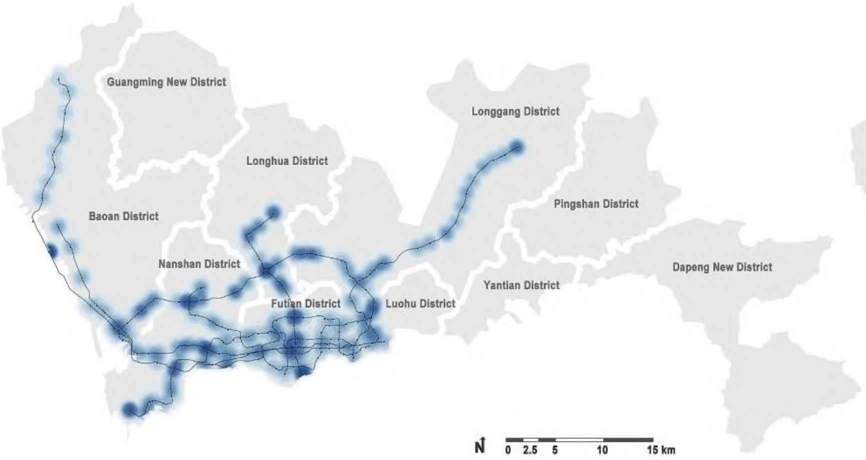
Net Ratio of Frequent Riders



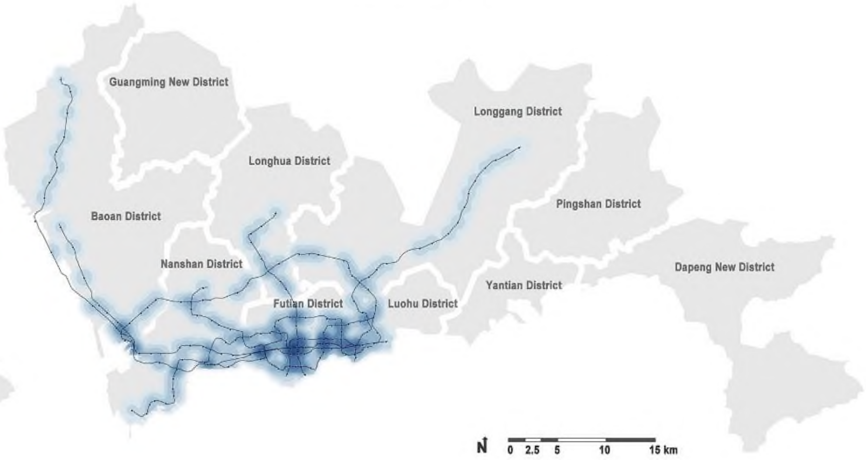
Distance to CBD



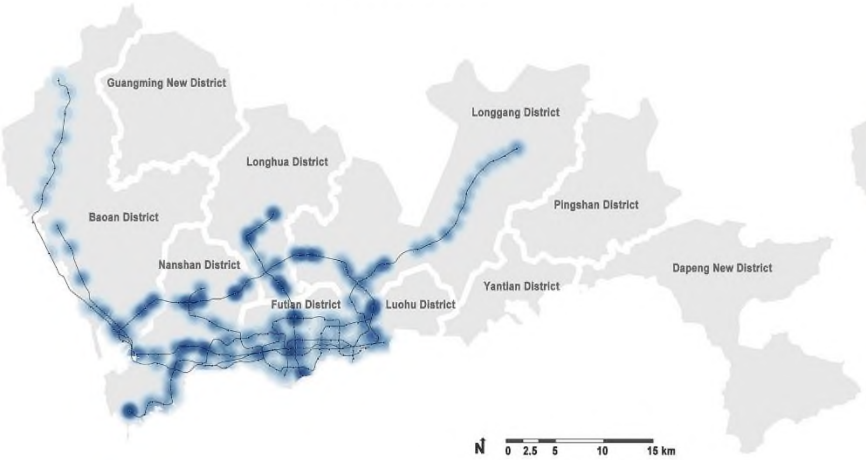
Ratio of Metro Riders



Regional Metro Accessibility



Gross Ratio of Frequent Riders



All Destination Intensity

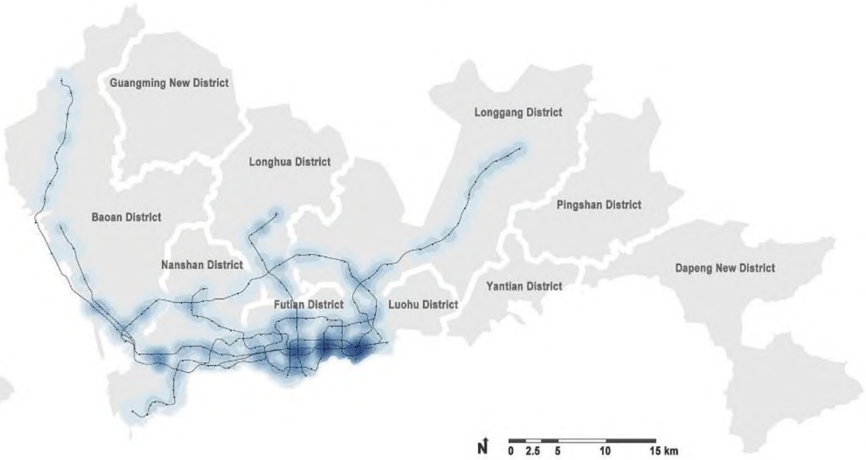


Figure A1: Distribution of Dependent Variables & Their Most Influential Independent Variables*

Notes: The darker the color, the larger the dependent variables or the more favorable the independent variables (except Distance to CBD), e.g., the higher the accessibility and the less time to CBD

Table A1: Comparisons of This Study and Other Key Papers

Paper	**Hypotheses**	**Key/New Variables**	**Methods**	**Findings**
This study	Individual TOD attributes affect TOD outcomes	TOD Attributes*: The regional centrality of a station within the local metro network; *Time to CBD;* *Distance to CBD;* *Destination density (all and category-specific);* *Destination diversity;* *Walkability;* *Bikability;* Number of bus stops; Special metro line (Dummy variable); Phase of metro line (Dummy variable); *Urban village (Dummy variable)* TOD outcomes*: (1) the people per hour**; *(2) the frequent riders,* *(3)/(4) the net/gross ratios of frequent riders, and* (5) the ratio of metro riders	Geovisualization; Regression with heteroskedasticity-robust standard error; Elasticity estimation.	The regional centrality has the biggest impact on the metro ridership; At least one of the regional accessibility indicators significantly predicts TOD outcomes (2) to (5). Some context-specific factors (e.g., urban village) also significantly influence TOD outcomes (2) to (5).
Singh et al. 2017	Levels of TOD of station areas can be quantified using an index.	TOD attributes*: Urban Densities; Land use diversity; *Transit systems' carrying capacity;* *User friendliness of transit systems;* Centrality of a station in the transit system;	Multiple Criteria Analysis	Bigger stations' levels of TOD are higher; Subjectively decided weights won't change the ranking of stations' level of TOD; the above findings can be used to guide policymaking regarding how to improve different station areas.

		Frequency of transit services; Parking for cars and bicycles. TOD outcomes: Ridership; Level of economic development.		
Gu et al., 2018	Overall TOD level and various Ds’ of station areas can be operationalized as an index, which can predict air quality at the city level.	_Density_*: Population density; Employment density; _Density gradient;_ _Diversity_*: Land use mix; _Job-housing imbalance;_ _Design_*: Street network density; Expressway density; Ground-floor retail density; Number of parking facilities; _Accessibility to Destinations_*: Distance to passenger transport terminal; Number of bus lines; Number of bus stops; _Distance to municipal public service facilities._	Spatial regression	Rail-based TOD is associated with better air quality
Zhou et al., 2018	Points of interest (destinations) can predict population per hour in a station area; History and existing land uses will also predict population per	TOD attributes: Points of interest (destinations). TOD outcomes: Population per hour**.	Descriptive statistics; Case study	Destinations cannot always predict population per hour in a station area; History and existing land uses predict population per hour in those well-established station areas.

	hour in a station area.			
Zhou et al., 2019	The local TOD and D indices by Gu et al. (2018), regional-level TOD characteristics, and station area attributes jointly influence TOD outcomes.	Local TOD attributes*: The indices/variables by Gu et al. (2018); The regional centrality of a station within the local metro network; New district or not (Dummy variable); Years in operation; Interchange station (Dummy variable); Number of stations in the station area. Regional TOD Attributes: *TOD attributes within 30 minutes' metro travel from the metro station in question.* TOD outcomes: (1) Metro ridership; (2) Population per hour**.	Regression (in the log form); Spatial regression.	Local TOD levels are good predictors of metro ridership on both weekends and weekdays; Regional-level TOD attributes have the biggest impacts on metro ridership on both weekends and weekdays; Density has the biggest impacts on population per hour and Design's impacts are larger on weekdays.

Notes:
* New variables or indices based on BOD as compared to the other existing studies are shown in italic.
** All were derived from Baidu population heatmaps.
Sources: The authors.

Table A2: Relationships between TOD Attributes and Outcomes in Existing Studies

Source	**Claimed (C) or Quantified (Q) Relationships**	**Measurement**		**Input Data**	
		TOD attributes	***TOD outcomes***	***TOD attributes***	***TOD outcomes***
Calthorpe, 1993	C: TOD attributes would engender many benefits such as decreased car ownership, auto trips, etc.	"Comfortable walking distance" for most (p.56); Mixed-use	Decreased car ownership, auto trips, etc.	n/a	n/a
TCRP, 1995	Q: On average, 1% increase in population density, 0.6% of transit ridership at 216 rail stations across the US. (Other two factors	Population density	Transit ridership	Censuses; Surveys	

	influencing transit ridership: distance to CBD and employment at CBD.)				
Cervero and Kockelman, 1997	Q: Density, land-use diversity, and pedestrian-oriented designs generally reduce trip rates and encourage non-auto travel in statistically significant ways, though their influences appear to be fairly marginal.	Accessibility to jobs; Dissimilarity index; Entropy; Vertical mixture of land uses; Intensity of land uses; Activity center mixture; Commercial intensities; Proximity to commercial-retail uses; Street patterns; Ped./Bike facility provision; Site design (esp. parking); Population/ Employment density	Mode share; Vehicle miles traveled (VMT); Trip rates;	Land use records; Field surveys	Travel diary
Bernick and Cervero, 1997	Q: Proximity to rail is associated with employment, population and residential building growth; Residential/Retail attributes (e.g., housing density) and rail station area' street patterns influence mode split.	Housing density; Percentage of single-family homes; Design factors (e.g., block per sq. mile; average block length)	Non-work trips by mode choice; Share of non-auto trips	Censuses; Field surveys	Censuses
Boarnet and Crane, 1998	Q: Proximity to rail station may entice other land uses rather than residential housing	Proximity to rail station	Supply of transit-based housing	Zoning data near stations	Residential land uses around stations
Cervero et al., 2002	C/Q: (based on evidences from secondary sources): TOD can increase transit ridership and rent premium.	Compactness; Mixed use; Transit connectivity; Site design	Transit patronage; Rent premium	n/a	n/a
Caltrans, 2002	C: TOD has a wide range of TOD outcomes.	Development density; Walk time to transit	Transit use; Sprawl trend; Safety; VMT; Disposable household income; Air pollution; Energy consumption; Infrastructure costs; Affordable housing	n/a	n/a

Wells and Renne, 2003	C: TODs meeting certain criteria would perform well in TOD outcomes, which can be measured using indicators.	Comprehensive transit-village criteria and related actions by the government (e.g., see: https://goo.gl/Do9tEP)	14 indicators about housing, business, public spaces, parking, transit ridership, pedestrians, public investment and public perception	Building permit; Governmental data; Surveys	
Cervero et al., 2004	C/Q: (based on evidences from literature review and secondary sources): TOD has a wide range of TOD outcomes.	n/a	Transit ridership and associated revenue gains; Congestion relief; Land conservation; Reduced outlays for roads and improved road safety	n/a	n/a
Dittmar and Ohland, 2004	C: Sufficient density, walkability, diversity, location-efficient development and convenient transit services at TODs can improve quality of life and metro economies.	Land use at station areas (based on case studies)	Quality of life and metro economies	Data from secondary sources	n/a
NCHRP, 2005	C: TOD can mitigate sprawling developments, preserving land, reducing traffic congestion, etc.	Density-population/housing; Quality of streetscape design; Mode connections; Parking configuration	Vary depending on survey responses, some of the most recommended: Transit ridership; Ped. activity/safety; Public perception.	(Governments need to invest into such data and update it “yearly or less frequently” [p.2])	
Renne, 2007	C: TOD can yield many benefits and evaluation of them is subjective. But the evaluation should be cross-sectional and longitudinal.	Population/Housing density; Street quality; Public space; Land cover/use Pedestrian accessibility; Parking inventory	5 categories, 21 indicators, ranging from travel behaviors to the policy context	Official data plus ad-hoc surveys	

TCRP, 2007	Q: TOD attributes such as density, diversity, land use mix, site layout and pedestrian-friendly design can have effects on travel demand based on information from secondary sources.	“TOD-index” or “TOD attributes” reflecting distance to transit, walkability, transit services, land use mix, density and parking	Primary transit mode; Car ownership; VMT; Level of congestion	Building data	Travel surveys; Rider count
Kahn, 2007	Q: Proximity to rail transit can increase gentrification at station areas.	Proximity to rail transit	Average housing prices; Ratio of college graduates in a census track	Census data	
Dill, 2008	Q: Transit station proximity influence access mode to the station and transit usage for noncommute travel. (Distance to rail transit and parking pricing influence commute mode choice)	Proximity to rail transit	Access mode to the station and transit usage by trip purpose	Surveys	
Cervero and Murakami, 2009	Q: Rail+Property Development increases ridership and housing prices.	Building area; Scale; Density; Mix-use attributes	Housing prices; Transit ridership	Local official statistics and transit agencies’ reports	
Renne, 2009	Q: TOD attributes can increase transit ridership.	Housing density; Number of street links; Number of nodes; Typical block dimensions; Station design rating; Pedestrian/ Bike accessibility rating	Transit ridership; Mode share	Censuses and a special TOD database	
Loo et al, 2010	Q: Station characteristics has the most significant impacts on rail ridership.	Total commercial/ residential floor area; Parking area; Location of the station (dummy); Mixed land use (dummy); Generalized travel cost to Midtown; Major interchange station (dummy); Number of bus stops	Rail ridership	Transit companies’ data	
Sung and Oh, 2011	Q: TOD attributes such as density, diversity and design of each	Number of bus routes; Average headway;	Ridership	Local transit plan; Local	Smartcard data*

	station have significant impacts of transit ridership. (In Seoul, some factors can be more influential, e.g., transit service network, land-use mix and urban design.)	Distance between stations; Residential/ Commercial density; Commercial/ Business mix index; Total road length; Average road width; Four-way intersection density; Dead end road; Average building area		transit service data; Official land use and building design data	
Chatman et al., 2014	Q: The combination of more concentration of jobs and population and more expensive CBD parking costs can increase ridership at station level. A higher value of population, jobs, and congestion score increases passenger miles at metropolitan level, especially high-wage jobs and leisure jobs.	Job density; Population; CBD parking rate; Transit service; High-wealth jobs; Leisure jobs; Congestion score	Ridership within station areas and passenger miles traveled within metropolitans	Databases from agencies or government; Censuses	Databases from agencies
Nasri and Zhang, 2014	Q: TOD can reduce VMT.	Households in TOD; Household in a rail-accessible zone; Bus stop density; Residential/ Employment Density; Distance from CBD; Average block size	Transit ridership; VMT	GIS shapefiles of traffic analysis zones, census blocks and stations surveys	Household travel
Singh et al., 2014	C: TOD can stimulate sustainable development by encouraging better land use and transport integration, which can be measured by a TOD index.	TOD index (reflecting density, diversity, design and current level of economic development); Transit connectivity	n/a	Secondary spatial and statistical data	n/a
Noland et al., 2014	Q: TOD can have many beneficial effects, e.g., increased share of non-auto trips and cost of travel.	Presence of a train station and redevelopments	Frequency of walking, driving and using transit; Potential health benefits due to transit	Interviews; Focus groups and surveys	

			proximity; Cost of travel; Property value; Train usage and highway congestion		
Papa and Bertolini 2015	C: TOD could deliver multiple benefits, e.g., shaping polycentricism, mitigating sprawl, boosting transit ridership, etc.	Proximity to transit station; Regional rail accessibility of a station	Density around station areas	Official statistics; OpenStreetMap*	
Lyu et al., 2016	Q: Different TODs can be categorized according to morphological characteristics, carrying capacity used, benefits produced, etc. (In other words, different TOD attributes can lead to different levels of TOD outcomes.)	Six indicators for "transit", "development" and "oriented", respectively (not differentiate TOD attributes and TOD outcomes)		On-line open data (e.g., OpenStreetMap, Google Map, Baidu Map, Walk Score*); Census data	
Renne et al., 2017	Q: Network accessibility to jobs and population within station areas yields an increase in transit commuting at all fix-guideway transit stations in USA. Besides, factors such as socioeconomics, walkability, land-use diversity, and transit service, contribute to the increase of transit commuting.	Regional network accessibility; Household income; Intensity of jobs and population; Share of nonwhites; Land-use mix; Walkability; Transit frequency; Transit technology	Transit commute mode share	Censuses; Surveys; Databases from agencies/companies	
Cervero, et al., 2017	C/Q: TOD, together with other strategies can create better communities, environments, and economies.	Density of residents and employees; Average block size; Availability and quality of urban living infrastructure; Access to and connectivity of bikeways and sidewalks; Transit service frequency; Place identity (Narratives only)	VMT; Travel time/cost saving; In-migration	Do not specify	
Griffiths and Curtis, 2017	Q: TOD can reduce car trips.	Density of dwelling units	Car trips	Official maps	Survey data

Loo and du Verle, 2017	Q: The success of TOD can be more than just transit ridership. It should include other benefits, e.g., share of walking, greenery and vibrancy.	Metro/Bus access; Gross floor areas of different land uses; Simpson index of land uses; Road length/total area; share of built open space; Total number of station exists; Presence of covered walkway; Number of retail shops; Population/ Employment density	Mode share	Railway maps	Travel surveys
Rogriguez and Vergel-Tovar, 2017	Q: TOD attributes at bus rapid transit station areas can boost ridership.	38 variables, ranging from facility density to land use mix	Ridership; Pedestrian activities**; Land values**; Affordability**	Self-collected primary data and secondary data from local transit agencies	
Singh et al., 2017	Q: Different TODs can have different levels of TOD attributes and TOD outcomes. (A TOD index can be used to quantify them.)	Population/ Commercial density; Entropy; Mixedness; Total length of walkable/cyclable paths; Density of business establishment***; Employment level***; Impedance pedestrian catchment area; User-friendliness of transit system; Access and accessibility; Parking at station	Capacity utilization of transit;	Spatial/Statistical data from the government or GIS vendors	
Renne, 2018	Q: TODs in proximity to ports have high levels of population, housing, job density, transit usage and walking and lower level of car ownership.	Job density; Percentage of professional jobs; Walkability	Mode share; Car ownership	An ad-hoc national TOD database; The National Transportation Atlas Database	

Gu et al., 2018	Q: TOD attributes can improve air quality of cities	TOD index (reflecting the following aspects: Job-housing mix; Population density; Density gradient; Distance to municipal public services facilities; Distance to passenger transport terminal; Employment density; Expressway density; Ground-floor retail density; Land use mix; Number of parking facilities; Street network density; Number of bus lines; Number of bus stops.)	Air quality index	Open data, e.g., The Worldpop database, Baidu or Gaode Maps	Governmental statistics
Zhao and Li, 2018	Q: TOD can encourage people to live and consume in or around transit. (In addition to land use near stations, available transport services and individual preference influence TOD outcomes, e.g., mode choice.)	Distance to metro station (other TOD attributes such as destination was indirectly measured)	Patrons of facilities/shops; Mode choice; Car ownership	Local land use data	Survey data
Zhou et al., 2018	Q/C: Points of interest and the population per hour in a station area is correlated.	Density of points of interest	Population per hour by station area	Sina Weibo (Chinese version of Twitter)	Baidu population heatmaps
Zhou et al., 2019	Q: TOD indices by Gu et al. (2018) would predict TOD outcomes.	TOD indices by Gu et al (2018); Regional level of TOD; Station area characteristics	Average hourly metro ridership and population by station area and the ratio of the two on weekends and weekdays, respectively	See Gu et al. (2018) above	Smartcard data; Baidu population heatmaps; OpenStreet Map data.

Sources: Information of some literature was adapted from Zhou et al. (2018, 2019).
Notes:

*Big or open data were used.
** Recommended for future studies.
***Authors treated them as both TOD attributes and TOD outcomes.

Table A3: Regression Results

Reg. 1 (Metro ridership & TOD-ness)				***Reg. 2 (People per hour & TOD-ness)***				***Reg. 3 (Frequent riders and TOD-ness)***			
	Coefficient[a]	Std. Coefficient	VIF		Coefficient[a]	Std. Coefficient	VIF		Coefficient[a]	Std. Coefficient	VIF
Constant	2354.17*** [531.71]	n/a	n/a	Constant	27663.39*** [8045.23]	n/a	n/a	Constant	440.01* [231.24]	n/a	n/a
Regional accessibility/centrality:											
Regional metro accessibility	29.64*** [8.15]	0.35	1.87	Time to CBDs	73017.55*** [12376.65]	0.43	1.66	Regional metro accessibility	12.30*** [2.68]	0.40	1.83
			n/a	Intermediate stations	-987.99*** [345.80]	-0.21	2.96				n/a
Design & Distance:											
Walkability	-3868.91*** [1038.56]	-0.23	1.69	Walkability	-33564.30*** [10427.60]	-0.21	2.29	Walkability	-810.90* [477.04]	-0.13	1.77
Bikeability	0.25*** [0.07]	0.27	1.45	Bikeability	1.10** [0.51]	0.12	1.46	Bikeability	0.13*** [0.03]	0.39	1.35
Bus	32.79* [18.18]	0.12	1.16				n/a	Bus	21.72*** [7.13]	0.22	1.21
			n/a	One or more metro stations	8204.60*** [2098.09]	0.22	1.48	One or more metro stations	-402.42*** [95.32]	-0.29	1.56
Density:											
All destination intensity	0.96*** [0.28]	0.25	1.81				n/a	All destination intensity	0.23*** [0.09]	0.17	1.91
Diversity:											
			n/a	Destination intensity (retail)	101349.10*** [26793.51]	0.20	1.26				n/a
			n/a	Destination intensity (residence)	-28140.08*** [9179.89]	-0.15	1.19				n/a

				Destination intensity (entertainment)	122321.80** [51588.70]	0.14	1.25				
Specific factors in Shenzhen:											
Longhua	1874.42* [987.28]	0.22	1.06	CBD	7326.10*** [2151.81]	0.20	2.11	Longhua	467.74*** [176.78]	0.15	1.07
Before 2004	1177.70** [554.31]	0.19	1.56				n/a	Before 2004	302.99** [145.71]	0.14	1.40
Lines 7 & 9	-697.89*** [244.46]	-0.16	1.33				n/a				n/a
R-Square: 0.52				R-Square: 0.62				R-Square: 0.56			

Reg. 4 (Net ratio of frequent riders & TOD-ness)				***Reg.5 (Ratio of metro riders & TOD-ness***				***Reg. 6 (Gross ratio of frequent & TOD-ness)***			
	Coefficient[a]	Std. Coefficient	VIF		Coefficient[a]	Std. Coefficient	VIF		Coefficient[a]	Std. Coefficient	VIF
Constant	0.19* [0.11]	n/a	n/a	Constant	0.19*** [0.05]	n/a	n/a	Constant	0.06*** [0.01]	n/a	n/a
Regional accessibility/centrality:											
Distance to CBD	-2.69x10^{-6}*** [0.00]	-0.33	2.64	Regional metro accessibility	0.001*** [0.00]	0.34	1.74	Regional metro accessibility	0.003*** [0.00]	0.18	1.52
Network directed-ness	0.16** [0.07]	0.16	1.26				n/a				n/a
Distance & Design:											
Walkability	0.21*** [0.07]	0.25	1.57	Walkability	-0.14* [0.08]	-0.17	1.68				n/a
			n/a					Bikeability	2.45x10^{-6}** [0.00]	0.14	1.30
Bus	0.004*** [0.00]	0.28	1.29	One metro station	-0.05*** [0.01]	-0.21	1.35	One or more metro stations	-0.02*** [0.00]	-0.34	1.41
			n/a	Two metro stations	-0.09*** [0.02]	-0.30	1.47				n/a
			n/a	Three or more metro stations	-0.07*** [0.02]	-0.18	1.50				n/a
Density:											
All destination intensity	-2.85x10^{-5}* [0.00]	-0.16	2.21	All destination intensity	-4.08x10^{-5}** [0.00]	-0.22	2.39	All destination intensity	-1.34x10^{-5}*** [0.00]	-0.20	1.90
Diversity:											

Destination intensity (restaurant)	-0.15*** [0.05]	-0.29	1.53	Destination intensity (restaurant)	0.12** [0.05]	0.22	1.36	Destination intensity (entertainment)	-0.18** [0.08]	-0.11	1.16
Specific factors in Shenzhen:											
After 2012	-0.04*** [0.01]	-0.22	1.22	After 2012	-0.03** [0.02]	-0.15	1.27	After 2012	-0.02*** [0.00]	-0.27	1.26
CBD	-0.05*** [0.02]	-0.29	2.86	Lines 4 & 5	0.04** [0.02]	0.18	1.24	Lines 4 & 5	0.02*** [0.01]	0.27	1.24
Urban village	0.03* [0.01]	0.14	1.33				n/a				n/a
R-Square: 0.41				R-Square: 0.44				R-Square: 0.52			

Notes: No. of observations (for all six models: 165)
* Indicators significance at the 90% level.
** Indicators significance at the 95% level.
*** Indicators significance at the 99% level.
[a] Robust standard deviation in brackets.

BOD INFORMATION

In terms of sources of BOD used in this study, we either obtained them from the Internet or solicited them from a primary data owner. Specifically, for the OpenStreetMap and Baidu Map data and relevant secondary information, we simply downloaded or queried it from their websites via Python codes. For the Weibo data, we used Application Programming Interface (API) to obtain them from Weibo's websites by following Weibo's user guidelines (http://open.weibo.com/wiki/Points of interest/add_poi). Regarding the bike-sharing data, we solicited them from the OfO company, which was willing to share with selected scholars the numbers of OfO users to and from different metro station areas—such information was not deemed sensitive.

Compared to traditional data, BOD are collected or updated continuously and cover large samples of various groups. Weibo, for example, routinely updates and calibrates its point of interest database since accuracy and currency of related information are vital for its business model. We do not know exactly the percentage of actual points of interest in Shenzhen that Weibo covers, but it is reasonable to assume that because of fierce location-based service (LBS) market competition and Weibo's leading position in the market, the percentage is high enough, the heterogeneity is wide enough and bias is low enough for reliable studies. Regarding OpenStreetMap, Baidu Map data, and secondary information derived from these sources, such as travel distances from one metro station to another, they are widely considered to be reliable and have been extensively used in other refereed publications (e.g., see Papa and Bertolini, 2015; Lyu et al., 2016; Gu et al., 2018; Zhou et al., 2018, 2019). Thus, they are considered reliable and accurate too.

The Shenzhen smartcard data were obtained from Shenzhen Urban Planning and Land Resource Research Center (SUPLRR). To pilot the open-government policy, SUPLRR allowed authorized users to access and analyze various non-sensitive data stored within a workstation. This workstation was disconnected to the Internet and any peripheral hardware. Via authorized access to this workstation, we analyzed five-weekdays' (May 15 to 19, 2017) smartcard data of Shenzhen. The analysis was straightforward as the smartcard data are structured data. The processing of the Baidu heatmaps, however, was more time consuming because the heatmaps are not structured data but rather a series of digital maps. Figure 1 is a sample of Baidu heatmaps.

(Figure 1 is here.)

Baidu heatmaps are generated by Baidu, an IT giant in China. Anyone who installs a Baidu Map app on a smartphone or desktop can view these maps in vector format, easily scalable to accommodate the resolution of a wide variety of output devices. The maps cover Mainland China and are updated every 20 minutes. They show density and locations of smartphone users who used various apps supported by Baidu's LBS. At the city level, Baidu allows these maps to be crawled (downloaded) at the scale of 1/1000. This gives 7500*6000 pixels per tile and identifies spatial units of data as small as 0.1 meter in size (Survey and Mapping Services, 2018). We used these heatmaps to estimate average number of people per hour on three weekdays (May 17 to 19, 2017) across the Shenzhen metro station area, an area of 2,009,600 square meters.

Baidu heatmaps use colors to represent different densities of smartphone users. Given that the market share of Baidu's LBS is 62 percent in Mainland China (Analysys, 2018) and that Shenzhen has a smartphone market penetration rate of 90 percent (Shenzhen Association of Online Media, 2018), we assume that Baidu heatmaps capture a significant percentage of smartphone users in Shenzhen throughout the day. Unlike the commercial heatmaps that Baidu sells, public domain Baidu heatmaps do not allow us to directly count the exact number of smartphone users by metro station area. We thus re-vectorized and re-georeferenced the heatmaps to derive number of smartphone users by metro station area, taking advantage of a series of geoprocessing tools and procedures such as "create signature" and "maximum likelihood classification", embedded in ArcGIS 10.5.1. In a nutshell, these tools and procedures enable us to (a) divide the public domain heatmaps into 100*100-meter geo-tagged grids and, (b) assign a distinct color, which represents the averaged Baidu density of smartphone users, to each grid.

Across countries, such density map of smartphone users has become increasing available given the emergence of LBS and corresponding data. One can even estimate and map out the dynamic population across different locales at the global level based on LBS data (e.g., see Deville et al., 2014). In principle, any LBS providers such as Google, Facebook, and Twitter can produce and share various density maps of smartphone users who use their respective LBS. Using this spatial re-aggregation method, we summed up the numbers of smartphone users of all grids within a metro station area to estimate the number of smartphone users in the complete metro served area. We assume that the estimate is proportional to the total number of people within a metro station area.

BETWEENNESS EXAMPLE

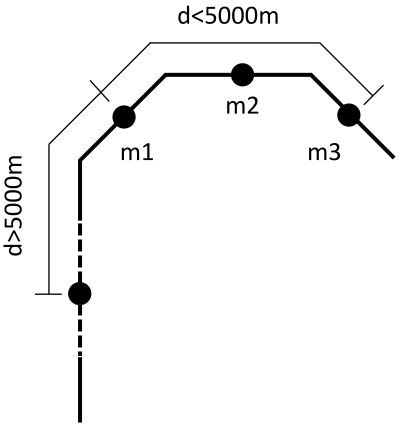


Figure A2: Betweenness Example

Endpoints	Points passing through
m1,m1	1/3 m1
m1,m2	1/2 m1, 1/2 m2
m1,m3	1/2 m1, 1/2 m3, m2
m2,m1	1/2 m2, 1/2 m1,
m2,m2	1/3 m2
m2,m3	1/2 m2, 1/2 m3
m3,m1	1/2 m3, 1/2 m1, m2
m3,m2	1/2 m3, 1/2 m2
m3,m3	1/3 m3

Betweenness values by endpoint:
m1 Betweenness= 1/3 + 1/2 + 1/2 + 1/2 + 1/2 = 2.33
m2 Betweenness = 1/2 + 1 + 1/2 + 1/3 + 1/2 + 1 + 1/2 = 2.33
m3 Betweenness = 1/2 + 1/2 + 1/2 + 1/2 + 1/3 = 2.33

Notes:
m_i is the midpoint of a link.
Analysis radius is 5000m.
The contributions of self-betweenness is 1/3.

REGRESSION MODELS

Model Overview

All our regression models were validated by the collinearity and multicollinearity tests, where the variance inflation factor (VIF) for each independent variable is less than 4. We were able to address most part of spatial autocorrelation in all the other regressions by simply introducing independent variables measuring the spatial location of a metro station area. Moreover, we employed heteroskedasticity-robust standard error to correct the standard deviation of different coefficients. Heteroskedasticity-robust standard error is commonly used to address the variability of a dependent variable that is unequal across the range of values of an independent variable (e.g., see Wooldridge, 2006).

Estimating Impacts of TOD attributes on TOD outcomes

To revalidate the existing relationships and to explore new relationships between TOD attributes and TOD outcomes, we fitted six regression models, most variables of which were based on BOD. For the revalidation model about the metro ridership and TOD attributes, we were fortunate to have some existing studies as references (e.g., TCRP, 1995, 2007) for our initial selection of independent variables. For the exploratory models, we did not have good existing references to guide us, except those references sporadically touching on various Ds and TOD outcomes. To expedite the processes of selecting the most statistically relevant independent variables, we applied the automatic linear modelling function in IBM SPSS 24.0 and used the trail-and-error method. Tables 1 and 2 in the article present descriptive statistics of variables we formulated to quantify TOD attributes and TOD outcomes in the regressions, respectively.

Table A3 (see above) presents the estimates of coefficients, including their OLS standard deviations, robust standard errors, standard coefficients, and corresponding VIF values in the six regressions. After performing Breusch-Pagan and Koenker-Bassett tests, we found that the residuals of all model display heteroskedasticity, and thus we introduced the robust standard errors for adjustment (Wooldridge, 2006). After the adjustment, all independent variables are still statistically significant after introducing the robust standard errors. As for the Jarque-Bera test, except for the net ratio of frequent riders (Regression 4), other residuals are not normally distributed. However, the assumption for normal distribution can be ignored where there is a sample size that is more than 30 (Wooldridge, 2006). All our six regressions have 165 samples and therefore the normal distribution assumption can be relaxed. Overall, we consider the six regression models are robust.

Overall, all regressions with BOD in general produce results that are consistent with the existing studies, which were primarily based on traditional data (e.g., surveys and censuses) and were synthesized in Table A2. Regional accessibility/centrality (measured by indicators such as the time/distance to CBD, intermediate stations, and regional metro accessibility), density/all destination intensity (measured by the number of all/different points of interest), diversity/destination intensity (measured by the percentage of various points of interest) and/or design (measured by an sDNA indicator, e.g., walkability, and bikeability, measured by the number of shared-bike users to and from a metro station area), for instance, are statistically significant predictors of TOD outcomes, which are measured by the metro ridership, the frequent riders, the people per hour, the net/gross ratios of frequent riders, and the ratio of metro riders.

In addition, distance to transit (measured by the numbers of bus stops and metro stations in a metro station area), and some localized factors (e.g., whether a metro station area is served by a ring line) also play important roles in influencing the TOD outcomes. However, the signs and magnitudes of corresponding coefficients of different indicators vary across different models. This indicates that the effect of TOD attribute could vary on different TOD outcomes. More details about the regression results are as follows.

TOD attributes and the metro ridership: Regression 1 is basically a validation model as the existing studies ran similar regression models before. We found that regional metro accessibility has the biggest positive impact on the metro ridership.

We denote this the regional network accessibility (RNA) TOD effect:
$RNA = \delta MR/\delta C$ TOD effect [1]
Where:
MR=outgoing metro ridership produced in a metro station area;
C= centrality of the station in a metro network, which can be sDNA-based indicators such as regional metro accessibility and intermediate stations and conventional indicators measuring CBD proximity such as time/distance to CBD, and the average travel time or distance to the two CBDs (Futian and Luohu). Regional metro accessibility measures a metro station's systemic centrality. Intermediate stations quantifies the average number of intermediate stations from a station to others.
RNA is the elasticity of outgoing weekday ridership with respect to regional network accessibility.

RNA TOD effect is in general consistent with findings of the existing studies, e.g., Chatman et al. (2014), Renne et al. (2016), TCRP (1995), and Papa and Bertolini (2015). TCRP (1995) found that the distance to CBD from a transit station and the employment size of the CBD had significant impacts on ridership generated at a station. Papa and Bertolini (2015) claimed that regional transit accessibility can boost the overall transit ridership in cities. Regional network accessibility, however, is an improvement on 'distance to CBD' as a measure of centrality, since 'connectedness' is measured endogenously with reference to network topology rather than distance to an exogenously defined a priori center point.

The Regression 1 results indicate that density/destination, distance to transit, and design are significantly correlated to the metro riders. In our study, design, to some degree, walkability (an sDNA-based indicator and measured by how much percentage of car-priority streets) are negatively associated with the metro riders. We denote the above the Ds' TOD effect:
$R^* = \delta MR/\delta D^*$ TOD effect [2]

Where:
D*= quality of TOD attributes of a metro station area;

R* is the elasticity of outgoing weekday ridership with respect to different Ds such as density/destination, distance, and design. * can be 1,2,3,…n.

We also found that factors that are probably contextually specific to Shenzhen such as whether a metro station area is located in the youngest administrative district (Longhua) and when metro services were first introduced in a metro station area significantly influence the metro riders. Longhua used to rural areas and is the youngest administrative district established by the Shenzhen Municipal Government after most agricultural land was urbanized and most farmers became urban residents.

TOD attributes and other TOD outcomes: Regressions 2 to 6 can be viewed as models exploring the relationship of TOD attributes and new TOD outcomes that are measured by BOD: the people per hour (Regression 2), frequent riders (Regression 3), the net/gross ratios of frequent riders (Regression 4/6), and the ratio of metro riders (Regression 5). As a whole, Regressions 2 to 6 results indicate that TOD effects [1] [2] are still at work. However, in Regressions 2 to 6, Regional network accessibility and various Ds were sometimes measured with different indicators as compared to Regression 1. Across the regressions, the entry of regional metro accessibility in the model gives a measure of the elasticity of the TOD outcomes with changing unit of connectivity to decentralized (non-CBD) urban centers. Similarly, the entry of intermediate stations adds additional explanation by capturing the new travel time landscape and distribution among different metro station pairs or locales near a metro station brought about by metro services. Such explanations have not been given in any of the existing studies listed in Table A2 in this Technical Appendix.

Specifically, Regression 2 results inform us that the time to CBD, diversity/destination (measured by the percentage of retail and entertainment points of interest) and design/bikability have positive impacts on the people per hour. As expected, a higher ratio of retail and entertainment points of interest would attract more people. Interestingly, more residential points of interest could decrease the people per hour. Similarly, intermediate stations and walkability are negatively correlated to the people per hour. If a metro station area is in one of the two CBDs (Luohu or Futian), two administrative districts that house the two local CBDs, respectively, it can expect significantly more people per hour.

Regression 3 results are in line with those of Regression 1 in terms of independent variables and signs of coefficients, indicating that the metro ridership and the frequent riders are influenced by a similar set of independent variables. Interestingly, if there are more than one metro stations in a metro station area the metro station area could expect fewer frequent riders into a metro station. In other words, the more metro stations the less likely that (frequent) metro riders would stick to one of them.

Regressions 4 and 6 results are about the net/gross ratios of frequent riders. The net ratio of frequent riders is positively associated with the number of bus stops, walkability, and network directed-ness. Interestingly, we found that the lower network directed-ness, the higher the proportion of frequent riders. It is likely that this captures a pattern of many frequent riders residing in communities in comparatively remote locations (attracted by lower housing prices).

They are frequent because of the lack of alternative routes and modes. We denote this the network normalized travel distance (NNTD) TOD effect:

$NNTD= \delta FR/\delta(\text{Network directed-ness})$ TOD effect [3]

Where:

Network directed-ness= the length of journey into a station normalized by the straight line distance from the station to all other stations on the network;

FR=the frequent riders into a metro station;

NNTD is the elasticity of inward-ridership with respect to the length of actual journeys coming into the station standardized by total network length.

The net ratio of frequent riders is negatively influenced by distance CBD, the percentage of restaurant points of interest, and the number of all points of interest. The causality between the percentage of restaurant points of interest and the number of all points of interest and the net ratio of frequent riders, if any, could be bilateral. Restaurants may follow ridership and ridership may follow restaurants. It is more likely that restaurants follow ridership, since ridership tends to be commuter driven. On the other hand, once attractions like restaurants become established, they may generate ridership and people flows into the station, both from other metro stations as well as from surrounding areas by bus, bike and other modes (hence a rise in restaurants may cause a rise in people per hour). The negative association of restaurants with the net ratio is intriguing. In other words, once the ratio of metro riders is accounted for, a rise in the ratio of frequent riders is associated with a fall in the importance of restaurants among points of interest. This suggests, on the one hand, that the frequent riders use restaurants less than other metro riders and on the other that the non-frequent riders often travel by metro to use restaurants. This comparison also lends credence to the interpretation that metro riders use the metro to dine at restaurants on weekdays.

Considering factors that might be contextually specific to Shenzhen, the presence of urban village(s), which means availability of cheaper housing, could increase the net ratio of frequent riders. This may reflect the lower attraction of these neighborhoods to inbound travelers from elsewhere. On the other hand, it may also be due to income and time budget effects. Since the people per hour is controlled for in the model, this result suggests that the elasticity of demand for regular transit services in lower income neighborhoods is higher than for non-regular services. This is entirely in line with the intuitive idea of non-essential travel being a normal economic good, demand for which increases as income rises.

If a metro station area is in CBD or was not served by metro services until 2012, it could expect a lower net ratio of the frequent riders. The gross ratio of frequent riders tends to be affected by a slightly different set of TOD attributes. Regional metro accessibility and the number of shared-bike users positively (“bikability”) influence the gross ratio of frequent riders. The numbers of metro stations and points of interest, whether a metro station area was not served by metro services until 2012 and the percentage of entertainment points of interest are negatively associated with the gross ratio of frequent riders. In addition, if a metro served area is served by Lines 4 or 5, the north-south line, and the ring line in Shenzhen, it can expect a higher gross ratio of frequent riders.

Regression 5 results indicate that the ratio of metro riders is positively related to regional network accessibility, the percentage of restaurant points of interest, whether a metro area is served by

Lines 4 and 5. If a metro station area was not connected to the local metro network until 2012, it could expect a lower ratio of metro riders. The number of metro stations influences the ratio of metro riders negatively.

Somewhat to our surprise, the number of all points of interest is negatively associated with both the net ratio of frequent riders (Regression 4) and the ratio of metro riders (Regression 5). However, a plausible explanation is that more points of interest generate more people per hour from all sources, not just metro riders. The pattern suggests that points of interest cluster around metro station areas, perhaps initially triggered by metro demand, but that they generate their own demand, arriving on other transport modes. The negative coefficient suggests that elasticity of local demand (rise in local demand for a given rise in point of interest supply in a metro station area) is greater than the elasticity of metro demand. This is a novel finding that we denote the TOD localization effect (TLE) and which we denote as:

$$\delta D^{L}/\delta POI > \delta D^{R}/\delta POI \qquad \text{TLE [4]}$$

Where
The left hand term is the elasticity of local demand for points of interest in a metro station area;
The right hand term, elasticity of demand for non-local demand of points of interest.
POI is the number of points of interest in a metro station area;
D^{L}=Local demand for POI;
D^{R}=regional demand for POI;
TEL [4] can have two possibilities. One is counterproductive effect and the other is multiplier effects. A counterproductive-effect example is that metro station areas have attracted high-end restaurants and shopping malls, which offer customers free or cheap on-site parking. This could reduce the number of metro riders to and from these metro station areas. A multiplier-effect example is that a metro station area with much open space and few points of interest can attract more frequent riders, pedestrians and cyclists simultaneously.

In addition, we found that the number of metro stations in a metro station area, which more or less also reflects the average distance to metro services, is correlated to the people per hour (+) (Regression 2), the gross ratio of frequent riders (-) (Regression 6), and the ratio of metro riders (-) (Regression 5).

The lower net ratio of frequent riders may not necessarily mean smaller numbers of frequent riders. Yet, it is likely to be attributed to higher local population in these dense areas. We also note that a high density of metro stations in an area means lower distances and more choice of station. This is also likely to account for the lower proportion of both regular and total weekday riders at any one station in these metro clusters. People change their station throughout the week. Depending on of the extent to which each station's capacity is used, this may indicate an inefficiency in metro locations. A test of this would be if the aggregate total of regular riders across all stations in a metro station cluster was positively correlated with aggregate points of interest.

Other TOD attributes such as design/walkability, which we use the percentage of car-priority streets as a proxy to measure, have mixed impacts on three of the five BOD measured TOD outcomes (Regressions 2 to 6). The higher this percentage measure, the higher net ratio of frequent riders, the fewer people per hour, the fewer frequent riders, and the lower ratio of metro riders. We conclude from this that more car-priority streets (i.e., lower walkability levels) (a) deter the co-location of people in a metro station area; and (b) make the metro station area less attractive to regular and other metro riders, perhaps partly because access by foot, bicycle, and bus are lower. We also note that the negative elasticity of demand by all riders as the supply of car-oriented roads increases is higher than the negative elasticity of demand from frequent riders. This seems entirely reasonable, given total riders include riders for non-essential purposes and non-essential visits are likely to be more elastic in their sensitivity to pedestrian unfriendly design than regular commuting use.

References:

Anlysys. (2018) Annual Report for Location-based Service Market in China. Beijing: Anlysys. Available at: https://www.analysys.cn/article/analysis/detail/1001123, accessed March 13, 2019.

Bernick, M and Cervero, R. (1997) Transit Villages for the 21st Century. New York: McGraw-Hill.

Boarnet, M. and Crane, R. (1998) L.A. Story: A reality check for transit-based housing. Journal of the American Planning Association, 63(2), 189-204. DOI: 10.1080/01944369708975914.

California Department of Transportation (Caltrans). (2002). Statewide Transit-Oriented Development Study: Factors for Success in California. Sacramento, CA: Caltrans.

Calthorpe, P. (1993) The Next American metropolis: Ecology, Community, and the American Dream. Princeton, NJ: Princeton Architectural Press.

Cervero, R. and Kockelman, K. (1997) Travel demand and the 3Ds: Density, diversity, and design. Transportation Research D 2(3) 199-219.

Cervero, R. Guerra, E. and Al, S. (2017) Beyond Mobility: Planning Cities for People and Places. Washington D.C.: Island Press.

Cervero, R., Ferrell, C., and Murphy, S. (2002) Transit Cooperative Research Program, Research Results Digest Number 52. Washington, D.C.: Transportation Research Board.

Cevero, R. and Murakami, J. (2009). Rail and property development in Hong Kong: Experiences and extensions. *Urban Studies* 46(10), 2019-2043.

Cevero, R., Murphy, S., Ferrell, C., Goguts, N., Tsai, Y., Arrington, G.B., Boroski, J., Smith-Heimer, J., Golem, R., Peninger, P, Nakajima, E., Chui, E., Dunphy, R., Meyers, M., McKay, S. and Witenstein, N. (2004) Transit-oriented Development in the United States: Experiences, Challenges, and Prospects. Washington D.C.: Federal Transit Administration.

Chatman, D. G., Cervero, R., Moylan, E., Carlton, I., Weissman, D., Zissman, J., ... & Tischler, D. (2014) Making Effective Fixed-Guideway Transit Investments: Indicators of Success. Washington, DC: The National Academies Press. DOI:10.17226/22355.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D. and

Tatem, A. 2014. Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences of the United States of America. 111(45), 15888-15893. DOI: 10.1073/pnas.1408439111.

Dill, J. (2008) Transit use at transit-oriented developments in Portland, Oregon, Area. *Transportation Research Record*, 2083, 159-167. DOI: 10.3141/2063-19.

Dittmar, H. and Ohland, G. (2004) The New Transit Town: Best Practices in Transit-Oriented Development. Washington, D.C.: Island Press.

Griffiths, B. and Curtis, C. (2017) Effectiveness of transit oriented development in reducing car use: Case Study of Subiaco, Western Australia, *Urban Policy and Research* 35(4), 391-408. DOI: 10.1080/08111146.2017.1311855.

Gu, P., He, D., Chen, Y., Zegras, P.C., and Jiang, Y. (2018) Transit-oriented development and air quality in Chinese cities: A city-level examination. *Transportation Research D*. DOI: 10.1016/j.trd.2018.03.009.

Kahn, M. (2007) Gentrification trends in new transit-oriented communities: Evidence from 14 cities that expanded and built rail transit systems. *Real Estate Economics*, 35(2), 155-182. DOI: 10.1111/j.1540-6229.2007.00186.x.

Loo, B. P.Y. and du Verle, F. (2017). Transit-oriented development in future cities: Towards a Two-level Sustainable Mobility Strategy. *International Journal of Urban Sciences*, 21(sub 1), 54-67.

Loo, B.P.Y., Chen, C. and Chan, E.T.H. (2010) Rail-based transit-oriented development: Lessons from New York City and Hong Kong. *Landscape and Urban Planning*, 97, 202-212.DOI: 10.1080/12265934.2016.1235488.

Lyu, G., Bertolini, L., and Pfeffer, K. (2016) Developing a TOD typology for Beijing metro station areas. *Journal of Transport Geography*, 55, 40-50. DOI: 10.1016/j.jtrangeo.2016.07.002.

Nasri, A. and Zhang, L. (2014) The analysis of transit-oriented development (TOD) in Washington D.C. and Baltimore metropolitan area. *Transport Policy*, 32, 172-179. DOI: 10.1016/j.tranpol.2013.12.009.

National Cooperative Highway Research Program (NCHRP). (2005). Transit-Oriented Development: Developing a Strategy to Measure Success. Washington D.C.: The National Academies Press.

Noland, R.B., Ozbay, K., DiPetrillo, S. and Lyer, S. (2014) Measuring Benefits of Transit Oriented Development. San Jose, CA: Mineta National Transit Research Consortium, San Jose State University.

Papa, E. and Bertolini, L. (2015) Accessibility and transit-oriented development in European metropolitan areas. *Journal of Transport Geography* 47, 70–83. DOI: 10.1016/j.trangeo.2015.97.003.

Renne, J. L. (2007) Measuring the Performance of Transit-oriented Developments in Western Australia. Planning and Transport Research Centre. Perth, Australia: The Government of Western Australia.

Renne, J. L. (2009) From transit-adjacent to transit-oriented development. *Local Environment*, 14(1), 1-15. DOI: 10.1080/13549830802522376.

Renne, J. L. (2018) Transit-oriented development and ports: A national analysis in the United States. *Journal of Transport and Land Use*, 11(1), 297-304. DOI: 10.5198/jtlu.2018.1121.

Renne, J. L., Hamidi, S., & Ewing, R. (2017) Transit commuting, the network accessibility effect, and the built environment in station areas across the United States. *Research in Transportation Economics* 60, 35-43. DOI: 10.1016/j.retrec.2017.02.003.

Rodriguez, D.A. and Vergel-Tovar, E. (2017) Urban development around bus rapid transit stops in seven cities in Latin-America, *Journal of Urbanism: International Research on Placemaking and Urban Sustainability,* 11(2), 175-201. DOI: 10.1080/17549175.2017.1372507.

Singh, Y.J., Fard, P., Zuidgeest, M., Brussel, M. and van Maarseveen, M. (2014) Measuring transit oriented development: a spatial multi criteria assessment approach for the City Region Arnhem and Nijmegen. *Journal of Transport Geography*, 35, 130-143. DOI: 10.1016/j.jtrangeo.2014.01.014.

Singh, Y.J., Lukman, A., Flacke,J. Zuidgeest, M. and van Maarseveen, M. (2017) Measuring TOD around transit nodes--towards TOD policy. *Journal of Transport Geography*, 56: 96-117. DOI: 10.1016/j.tranpol.2017.03.013.

Sung, H. and Oh, J.T. (2011) Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*, 28 (1), 70-82. DOI:

10.1016/j.cities.2010.09.004.

Survey and Mapping Services. (2018) Maps and Services. Available at: https://www.landsd.gov.hk/mapping/en/digital_map/service/product/dop.htm, accessed March 14, 2019.

Transit Cooperative Research Program (TCRP). (2007) Transit Oriented Development Traveler Response to Transportation System Changes, Report 95. Washington D.C.: Federal Transit Administration.

TCRP, 1995. Regional Transit Corridors: The Land Use Connection. Washington D.C.: TCRP.

Wells, J. S. and Renne, J. (2003) Transit Villages in New Jersey: Recommendations for Assessment and Accountability, Alan M. Voorhees Transportation Center, New Brunswick, New Jersey: Rutgers University.

Wooldridge, J.M., 2006. Introductory Econometrics: A Modern Approach. Mason, OH: South-Western, Cengage Learning.

Zhao, P. and Li, X. (2018) Suburbanization, land use of TOD and lifestyle mobility in the suburbs: An examination of passengers' choice to live, shop and entertain in the metro station areas of Beijing. *Journal of Transport and Land Use*, 11(1), 195-215. DOI: 10.5198/jtlu.2018.1099.

Zhou, J., Wang, Q. and Liu, H. (2018) Evaluating transit-served areas with non-traditional data. *Journal of Transport and Land Use*, 11(1): 1323–1349. DOI: 10.5198/jtlu.2018.1296.

Zhou, J., Yang, Y., Gu, P., Yin, L., Zhang, F., Zhang, F. and Li, D. (2019) Can TODness improve (expected) performances of TODs? An exploration facilitated by non-traditional data. *Transportation Research D*, 74, 28-47. DOI:10.1016/j.trd.2019.07.008.