The Journal of Transport and Land Use jtlu.org
Vol. 19 No. 1 [2026] pp. 279-304
https://doi.org/10.5198/jtlu.2026.2890

The impact of urban parks on staying-based activities in transit-oriented development: Causal evidence from large-scale mobile phone data

Xueqing Bo (corresponding author), Kochi University of Technology, bo.xueqing@kochi-tech.ac.jp
Shintaro Terabe, Tokyo University of Science, terabe@rs.tus.ac.jp
Hiroaki Nishiuchi, Kochi University of Technology, nishiuchi.hiroaki@kochi-tech.ac.jp

Abstract: The integration of transit-oriented development (TOD) and green urbanism has been increasingly recognized as a sustainable policy for the future development of cities. Although green amenities such as urban parks play a crucial role in promoting walkability and enhancing urban vitality, the causal impacts within TOD are unclear. In particular, it remains to be seen whether the observed increases in urban vitality are directly attributable to parks or are instead driven by other factors such as population density and land use. To elucidate this point, this study investigates the causal effects of neighborhood parks on visiting populations around railway stations in the Tokyo metropolitan area. Using mobile phone location data to capture staying-based activities, we first apply the propensity score matching method to estimate the impact of neighborhood parks on staying-based activity around stations and then employ a causal forests model to clarify how different locational characteristics influence park effectiveness. Our results show that neighborhood parks significantly increase visiting populations, with stronger effects on weekends than on weekdays. Furthermore, the positive impacts are more pronounced in areas with high densities of commercial facilities and greater transport accessibility. These findings provide empirical evidence on the role of urban parks in enhancing vibrant environments, offering policy-relevant insights for integrated land use and transportation planning, particularly in TOD.

Keywords: Staying-based activities, transit-oriented development, neighborhood park, vitality

1 Introduction

Transit-oriented development (TOD) is an urban-development strategy that concentrates residential, commercial, and office functions around public transport hubs such as railway stations, forming compact, high-density, and highly accessible urban spaces that encourage urban activities. This development model has been widely recognized in urban planning as a key strategy for urban vitality. At the same time, with the growing emphasis on sustainable urbanism, another important concept—green urbanism—has attracted increasing attention. It highlights ecological and environmental sustainability by introducing green infrastructure such as parks and greenways into urban environments, aiming to achieve harmony between people and nature within urban spaces. Although the two approaches differ in their focus, they share the common goal of fostering vibrant and livable urban spaces. Therefore, the integration of TOD and green urbanism has increasingly been regarded as a sustainable policy framework for future urban development (Cervero & Sullivan, 2011).

Received April 5, 2026; Revised July 22, 2026; Accepted August 19, 2026; Available online August 28, 2026

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

However, whether the integration of TOD and green urbanism actually produces urban vitality remains unclear. In practice, land-use planning in many cities has prioritized the concentrate commercial functions to attract visitors, while giving insufficient attention to the green infrastructure. Although this strong emphasis on commercial development may temporarily increase visitors, the lack of green infrastructure can reduce the comfort of public urban space. Visitors may leave soon after completing activities such as shopping or dining, leading to monotonous patterns of urban activity. At the same time, insufficient green infrastructure may limit the ability of urban areas to address environmental challenges, such as the urban heat island effect, air pollution, and stormwater management.

As an important element of green infrastructure, urban parks have demonstrated the multifaceted benefits. Wu et al. (2023) found that urban parks can help mitigate heat island effects, improve the ecological environment, and enhance overall environmental quality. Arena et al. (2017) reported that parks contribute to public health and well-being by providing spaces for physical exercise, leisure, and psychological relaxation. Moreover, several researchers have emphasized that parks serve as important venues for social interaction and community life, where people gather, communicate and engage in recreational activities (Peters et al., 2010). Despite the well-established environmental and social benefits of urban parks, their role in enhancing urban vitality remains insufficiently understood, particularly in station areas developed under TOD principles. Because TOD areas are typically characterized by high-density development and intense land-use competition, it remains unclear whether the provision of urban parks can still promote human activity and improve the attractiveness of station surroundings.

Against this backdrop, this study aims to empirically examine the potential of integrating TOD and green urbanism to enhance urban vitality. Specifically, the objective of this study is to evaluate the effects of urban parks on urban vitality around railway stations. Accordingly, the following hypotheses are proposed.

Hypothesis 1 (H1): Urban parks have a positive effect on the urban vitality around stations.

Hypothesis 2 (H2): This effect depends on the built environment of the stations, including land-use composition, commercial agglomeration, and transport accessibility.

2 Literature review

2.1 Urban vitality

Urban vitality is generally regarded as a multidimensional concept and has been evaluated from various perspectives. Previous studies have used indicators related to economic activities, such as retail activity, as proxies for urban vitality (Porta et al., 2012; Yamaguchi et al., 2025), while others have focused on social activities and the use of public spaces (Chen et al., 2025; Eisazadeh & Vahdani, 2017). Among these indicators, human flow in public urban space is often regarded as an appropriate proxy (Li et al., 2022; Montgomery, 1998), as it provides a deeper understanding for vitality.

However, measurements of human flow have traditionally relied on field surveys and manual counting. Although these approaches have provided valuable insights, they are often limited in terms of spatial coverage, temporal continuity, and ability to distinguish individual attributes. Owing largely to these technical limitations, Gerike et al. (2021) noted that most previous studies have focused primarily on movement-related indicators, while comparatively less attention has been paid to people’s staying activities. Recent advances in information technology and big data analytics, particularly the use of mobile phone location data, have made it possible to observe and analyze stationary activity. Wu and Niu (2019) also showed mobile phone data are a reliable proxy for measuring urban activity at fine spatial and temporal scales.

Numerous studies also have investigated the factors that influence urban activities. Ewing et al. (2015) demonstrated that the proportion of street-facing windows, active frontages, and the number of street amenities are positively associated with pedestrian volumes. Hajrasouliha and Yin (2015) used pedestrian counts of 302 street segments and structural equation modeling to highlight the relationships between street network connectivity, built environment, and pedestrian volumes. Their findings suggested that both the conventional metric-based measure of physical connectivity and the geometric-based measure of visual connectivity have significant positive impacts on pedestrian volumes as well as job density and land-use mix. Ozbil et al. (2011) modeled the distribution of pedestrian movement by street segment in three areas of Atlanta and analyzed its relationship with measures of street connectivity and land use. In the study area of this manuscript, Japan, both government reports (MLIT, 2023) and other research (Hamana et al., 2010) have emphasized the importance of measuring and analyzing movement patterns in the field of urban development.

2.2 Urban parks and urban vitality

As essential components of green infrastructure, urban parks have been examined on urban vitality. Some studies have reported a positive association between park proximity and pedestrian activity, suggesting that areas located farther from parks may exhibit greater pedestrian activity. Similarly, greening ratios have also been found to be negatively correlated with pedestrian activity, possibly because the analyzed parks were located mainly in urban fringe areas, where their contribution to pedestrian activity is limited (Li et al., 2022). In contrast, another study highlighted the positive synergies between parks and surrounding land uses. Che et al. (2024) demonstrated that urban parks frequently coexist with commercial facilities, implying that their spatial integration may reinforce pedestrian activity and enhance vitality. These findings suggest that the impacts of parks on urban activities are not uniform but rather dependent on environmental and locational conditions. This reasoning also forms the basis for **H2** in this study, which seeks to further clarify how different locational characteristics around stations influence the effects of parks on urban activities.

Some studies have also investigated the internal factors of the parks themselves. Zhu et al. (2020) found that features such as water bodies and facility density enhanced park vitality, while the surrounding population density exerted a positive influence. Mu et al. (2021) reported that park facilities such as sandpits, plazas, waterfronts, and activity pavilions attracted larger users, thereby generating vitality. These findings suggest that both external conditions and internal design characteristics underlie parks' contribution to urban activity.

2.3 Position of this study

Many previous studies have focused primarily on pedestrian volumes to identify factors that contribute to urban activities. Although these studies have provided insights into how the built environment influences pedestrian movement, analyses of staying behavior remain insufficiently developed. Staying behavior reflects how long and how actively people engage with urban spaces, offering a more direct measure of urban vitality compared with pedestrian flow. With recent advances in digital technologies, it has become possible to capture both the staying behavior and individual attributes over wide spatial and temporal scales. In this study, mobile phone location data are utilized to provide new insights to evaluate urban vitality from staying behavior.

Furthermore, most previous studies have been limited by a reliance on regression-based correlations or observational analyses, while highlighting associations between parks and urban activity. These methods make it difficult to disentangle the effects of other factors such as population density, land use, or transport accessibility on urban vitality, which can lead to biased estimation results. For example, areas with higher population density tend to have urban parks as well as more urban activities. Without controlling for the influence of population density on urban vitality, one might incorrectly attribute increases in activity solely to the presence of parks. To clearly identify the impacts of urban parks on staying-based activities, this study employs the propensity score matching (PSM) method to control for the influence of these factors on staying-based activities and evaluate the net effects of parks, thereby clarifying whether urban parks influence staying-based activities within TOD areas.

Previous findings have also suggested that the influence of parks on urban activities can vary depending on locational conditions (Che et al., 2024; Li et al., 2022). Although PSM offers a robust method for estimating the overall effects of parks on urban activities, it does not account for spatial heterogeneity across different urban environments. To further explore these spatial differences, this study applies the causal forest method, which identifies how the influence of parks differs across TOD areas with distinct spatial characteristics. This approach enables understanding of the conditions under which urban parks exert stronger or weaker impacts on staying-based activities within TOD areas.

3 Methodology

3.1 Causal effect estimated by propensity score matching

3.1.1 Propensity score matching

Previous findings have also suggested that the influence of parks on urban activities can vary depending on locational conditions (Che et al., 2024; Li et al., 2022). Although PSM offers a robust method for estimating the overall effects of parks on urban activities, it does not account for spatial heterogeneity across different urban environments. To further explore these spatial differences, this study applies the causal forest method, which identifies how the influence of parks differs across TOD areas with distinct spatial characteristics. This approach enables understanding of the conditions under which urban parks exert stronger or weaker impacts on staying-based activities within TOD areas. Staying-based activities around stations are commonly influenced by factors such as passenger volume, surrounding land use, and population characteristics, many of which are also correlated with the presence of parks. For example, stations with higher passenger volumes are more likely to have urban parks and, at the same time, attract greater concentrations of commercial facilities and staying-based activities. Without accounting for the effects of these factors, increases in staying-based activities may be mistakenly attributed to the presence of parks, leading to biased conclusions. To control for these factors, this study employs the PSM method.

PSM is a statistical approach originally developed in the field of medicine, it has been widely adopted in causal analyses of urban planning (Li et al., 2024; Wang & Li, 2022). The core idea is to construct two comparable groups, a treatment group (stations with parks nearby) and a control group (stations without parks nearby). At the initial stage, the two groups usually exhibit differences in background characteristics, including socioeconomic, land-use, and accessibility characteristics, around stations. In PSM, these background characteristics are regarded as confounding variables. If the outcome variables (staying-based activities) of two groups are compared directly, the observed differences cannot be attributed to the presence of parks. These differences might be caused by variations in the confounding variables between the two groups. To eliminate the influence of confounding factors on staying-based activities between two groups, it is necessary to use a matching method. In PSM, for each station in the treatment group (stations with parks nearby), a corresponding station in the control group (stations without parks nearby) is selected that has nearly identical values across all confounding variables. By using matching method, the influence of confounding factors is minimized because the distributions of confounding variables between the two matched groups become nearly identical. The only difference between the matched groups is the presence or absence of a park nearby. Therefore, any difference observed in the outcome variable (staying-based activities) can be interpreted as the causal effect of the urban park itself.

However, if key confounding variables are omitted, then the observed differences in staying-based activities might be attributable to other unobserved factors rather than the parks themselves. Therefore, it is important to include as many confounding factors as possible to minimize bias (Chen et al., 2021). Through this process, PSM can provide a rigorous analytical process for identifying and quantifying the causal impacts of urban parks on staying-based activities. As shown in Table 1, three types of variables are commonly used in PSM.

Table 1. Variables used in propensity score matching

Variable	Definition
Z	A dummy variable indicating the treatment assignment: treatment group: ($Z = 1$): stations with a park nearby control group: ($Z = 0$): stations without a park nearby
Y	The outcome variable (staying-based activities around station)
X	Confounding variables (variables that influence treatment assignment and the outcome variable, used to balance the comparison between groups)

When matching the treatment and control groups, if the number of confounding variables is small, researchers can easily find similar station in the control group for each station in the treatment group. However, as the number of confounding variables increases, this approach becomes very difficult. To address this issue, multiple confounding variables can be combined into a single index known as the propensity score (PS). PS is defined as the probability of being assigned to the treatment group given a set of confounding variables X. It is commonly modeled with a logistic regression model (Rosenbaum & Rubin, 1983). The estimation formula is expressed as follows:

$$(X) = P(Z = 1|X) = \frac{\exp(\alpha + \beta^Z X)}{1 + \exp(\alpha + \beta^Z X)} \tag{1}$$

where $e(X)$ represents the PS, Z is the treatment-assignment indicator (1 if a station has a park nearby, 0 otherwise), X denotes the set of confounding variables, α is the intercept, and β is the vector of coefficients for X.

After estimating the PS, we proceeded to the matching. The objective is to pair stations in the treatment and control groups that have similar PS values so that the distributions of confounding variables are the same in the matched groups. Among the available methods, 1:1 nearest-neighbor matching is the most commonly used approach, the PS of each station in the treatment group, a station from the control group with the closest PS value was selected. Furthermore, to avoid pairing stations that are too dissimilar, a caliper is introduced—a tolerance threshold that limits how far apart the PS values of two matched stations can be. If the caliper is too large, inappropriate matches may occur, whereas if set too small, an insufficient number of matched pairs may be obtained. Following previous studies, A caliper of 0.2 is commonly recommended (Austin, 2011), meaning only stations with very close PS values (≤ 0.2) are paired, while others are excluded to improve the credibility of the comparison. Therefore, in this study, the caliper was applied as 0.2.

It is also essential to verify whether the two matched groups have actually become balanced in their confounding variables after matching. In well-matched results, the distributions of confounding variables between two matched groups should be similar, ensuring that differences in outcome variables are no longer influenced by confounding factors. To evaluate whether the confounding variables between the two groups are balanced after matching, the absolute standardized mean difference (ASMD) is commonly used (Guo et al., 2020). When all confounding variables have ASMD values below 0.25, the groups are considered well balanced, indicating that confounding bias has been sufficiently reduced (Stuart, 2010).

3.1.2 Average treatment effect on the treated after propensity score matching

After confirming that the confounding variables are balanced between the matched groups, we quantified the impact of urban parks on staying-based activities. the average treatment effect on the treated (ATT) is a widely used evaluation indicator after PSM. It estimates the average treatment effect specifically for the treatment group, measuring the difference between the observed staying-based activities at stations with parks nearby and the hypothetical outcomes those same stations would have exhibited in the absence of parks. Formally, ATT can be defined as follows:

$$ATT = \frac{1}{N_T}\sum_{Z_i=1}(Y_i(1) - Y_i(0)) \tag{2}$$

Where $Z_i = 1$ represent the treatment group (the station area has an urban park), and N_T denotes the number of treated stations in the treatment group.

The counterfactual outcome $Y_i(0)$—that is, the staying-based activities that stations originally surrounded by parks would have exhibited under the counterfactual scenario in which the parks did not exist—cannot be observed. To approximate this unobservable scenario, PSM identifies comparable non-park stations that share similar background characteristics and uses their weighted average outcomes as substitutes for the counterfactuals. In practice, ATT is obtained by comparing each station with staying-based activities observed in a nearby park ($Y_i(1)$) with the weighted average staying-based activities of its matched stations without parks nearby.

However, PSM alone cannot provide information about the statistical significance of the estimated effects. Therefore, a regression analysis was also performed, using the matched samples as a supplementary. In this regression model, staying-based activities were used as the dependent variable, and the treatment assignment Z was included as a binary independent variable. By examining the p-value of the Z, the statistical significance of the ATT estimate was verified. The overall process of estimating the ATT effect using PSM is illustrated in Figure 1.

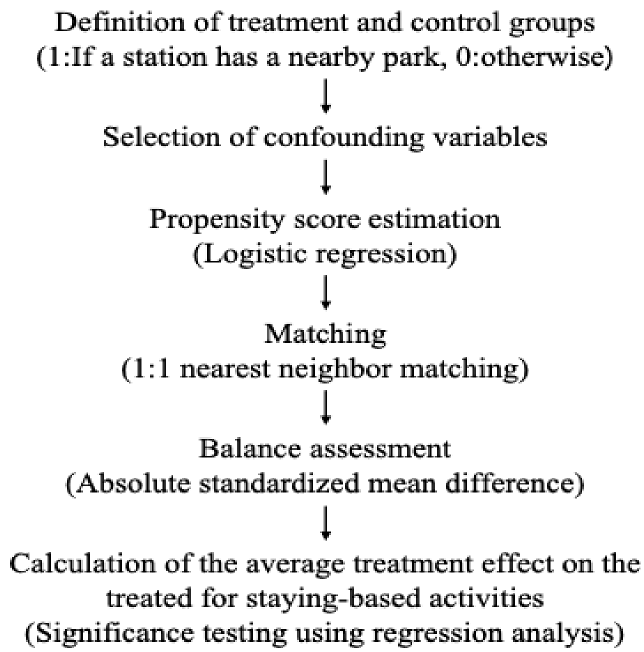

Figure 1. Overall process of estimating the average treatment effect on the treated effect using propensity score matching

3.2 Causal forest

Although ATT provides an overall estimate of the average influence of urban parks on staying-based activities, the impact of urban parks is unlikely to be uniform; rather, it varies depending on the surrounding characteristics of station areas (Che et al., 2024; Li et al., 2022). To further investigate these variations, this study employed the causal forest method to explore the conditions under which urban parks exert stronger or weaker impacts on staying-based activities (Hypothesis 2).

A detailed explanation of the causal forest method has been reported previously (Susukida et al., 2025). In brief, the causal forest method is an extension of the random forest framework designed to estimate causal effects at the individual level. It builds a large number of "causal trees," each trained on a randomly selected subsample. Within each tree, the data are recursively divided through binary splits based on single confounding variables. Unlike conventional regression trees that aim to maximize prediction accuracy, causal trees are designed to maximize differences in estimated treatment effects between subgroups, thereby identifying how treatment effects vary under different conditions. Each sample is assigned to a terminal leaf within a tree, where the treatment effect is estimated as the difference in average outcomes between treatment and control units within that leaf. By averaging these local estimates across many trees, the causal forest produces a robust estimate of the conditional average treatment effect (CATE) of park on urban activities, reflecting how the influence of urban park differs across various spatial contexts.

For a given causal tree b, the treatment effect estimates for sample i evaluate at x_i is computed as follows:

$$\hat{\tau}(x_i) = \frac{1}{|i:Z_i=1,X_i\in L|}\sum_{i:Z_i=1,X_i\in L} Y_i - \frac{1}{|i:Z_i=0,X_i\in L|}\sum_{i:Z_i=0,X_i\in L} Y_i \quad (3)$$

where i indexes samples (stations). Y_i denotes the outcome variable, and Z_i is a binary treatment indicator. X_i represents the vector of confounding variables, and L denotes a leaf node of a causal tree. $\{i: Z_i = 1, X_i \in L\}$ denotes the set of stations in the treatment group whose confounding variable values fall within leaf L, while $\{i: Z_i = 0, X_i \in L\}$ denotes the set of stations in the control group whose confounding variable values fall within leaf L. $\hat{\tau}(x_i)$ denotes the estimated treatment effect evaluated at the confounding variable values x_i. It represents the conditional average treatment effect (CATE) for samples with confounding variable profiles similar to that of sample i.

The final treatment effect estimate is obtained by averaging the estimates across all B trees:

$$\hat{\tau}(x_i) = \frac{1}{B}\sum_{b=1}^{B} \widehat{\tau_b}\,(x_i) \tag{4}$$

In addition, the causal forest provides importance scores that measure the contribution of each confounding variable to the effect heterogeneity. Variables that frequently appear in split nodes and lead to larger differences in estimated effects across subgroups receive higher importance scores. These scores help to identify which station-area characteristics play the most critical roles in shaping the impact of urban parks on staying-based activities.

4 Study area and data

4.1 Study area

The Tokyo Metropolitan Area—comprising Tokyo, Chiba, Saitama, and Kanagawa prefectures—was selected as the study area. Tokyo, Japan’s capital, covers approximately 2,200 km² and had a population of 14.2 million as of January 2025. Chiba, Saitama, and Kanagawa prefectures are geographically adjacent to Tokyo (Figure 2), and many residents share overlapping spheres of daily life and economic activity with the capital. Tokyo has a day–night population ratio of 119.2, with approximately 3.36 million people commuting into the city each day—95.2% of whom come from these adjacent prefectures. In addition to commuters, a large number of residents from these areas also travel to Tokyo for leisure and shopping. Given the strong spatial, social, and economic interconnections between Tokyo and its neighboring prefectures, these four areas are often treated as a single integrated urban system in Japanese metropolitan studies. Accordingly, this study also considers Tokyo, Chiba, Saitama, and Kanagawa collectively as an integrated metropolitan region.

Figure 2. Map of the Tokyo metropolitan area

4.2 Analysis target

Railway infrastructure has been steadily developed in the Tokyo metropolitan area, with rail transport accounting for approximately 30% of the modal share, reflecting its important role in public transportation. As shown in Table 2, compared with other major cities, Tokyo exhibits a relatively high station density (2nd) and a densely developed railway network. Accordingly, urban development in the Tokyo metropolitan area has strongly followed the TOD, where urban functions, population, and public infrastructure are highly concentrated around railway station. By studying such a "super TOD," it is possible to examine the extent to which urban parks can further enhance staying-based activities even in areas where walkability and transit access are already well developed.

In Japan, municipalities are administratively classified into cities (Shi), towns (Cho), villages (Son), and special wards (Ku). These administrative units differ substantially in terms of governance structures, population density, and urban functions. Therefore, to ensure consistency and comparability across the study area, this study focused on cities (Shi) in the Tokyo Metropolitan Area. Accordingly, all railway stations (N=689) located within cities (Shi) in the Tokyo metropolitan area were selected as the analysis targets.

Table 2. Railway network infrastructure in major global cities

City	Population density (1,000/km^2)	Number of railway stations	Railway length (km)	Station density (stations/km^2)
Hong Kong	6.53	87	221	0.08 (7th)
London	5.28	659	894	0.42 (5th)
New York	10.72	569	587	0.73 (3rd)
Paris	21.65	363	208	3.46 (1st)
Singapore	7.54	109	151	0.15 (6th)
Seoul	17.17	373	406	0.62 (4th)
Shanghai	3.75	336	674	0.05 (8th)
Tokyo	14.61	661	760	1.06 (2nd)

4.3 Data

Before collecting data on the areas surrounding railway stations, it was necessary to define the station catchment area. First, we examined the number of treated stations under different catchment-area definitions. The number of stations with neighborhood parks was 21 within 400 m, 54 within 600 m, and 187 within 800 m. The smaller treatment group samples under the 400-m and 600-m definitions would reduce statistical power and make the causal estimates less stable. Furthermore, previous studies have shown that an 800m radius is widely used as a practical standard for rail-transit catchment areas in TODs (Chen et al., 2018; Guerra & Cervero., 2013). Therefore, this study defined each station catchment as a circular zone with a radius of 800 m, corresponding to an approximate 10-min walking distance from the station. All relevant data within the 800-m catchment area of each station were collected and integrated for analysis. Table 3 summarizes the datasets used in this study. All data used, except for the staying-based activity, were obtained from open data sources provided by the Japanese government.

Regarding the selection of data years, it is ideal to align all variables to the same reference year to eliminate potential biases caused by temporal discrepancies. However, to control for confounding factors, this study integrated multiple datasets. Because many of these datasets are not updated annually, it is difficult to standardize all data to a single year. Therefore, based on the availability of the data, 2014 was selected as the reference year. For variables lacking data for 2014, the value closest to 2014 within a range of 4 years before or after was used.

Table 3. Overview of the data

	A set of variables	Variable description	Year
Urban facilities	Medical facilities	Numbers of hospitals, dental centers, and clinics	2014
	Post office	Number of post offices	2013
	Library	Number of libraries	2013
	Childcare facilities	Number of childcare facilities	2011
	Older adult welfare facility	Number of older adult welfare facilities	2011
	School	Numbers of elementary schools, middle schools, high schools, and colleges	2013
	City hall	Number of city halls, City hall main building dummy	2014
	Retail facilities	Numbers of retail faculties of convenience stores, supermarkets, department stores, medical and beauty products stores, entertainment products stores, household goods stores, beauty salons, clothing stores, vehicle stores, and food stores	2018
	Visitor-attracting facilities	Numbers of cinemas, community halls, theaters, exhibition venues, gymnasiums with spectator seating	2014
Tourism resources	Tourism resources	Numbers of the nature resources, the historical and cultural resources, the hot springs and health resources, the sports and recreation resources, and the urban tourism (shopping and dining) resources	2014
Urban Planning policies and land use	Urbanized area policies	Areas of the urbanization promotion areas, urbanization control areas	2011
	Zoning policies	Areas of residential, commercial, and industrial zones	2011
	Land use	Areas of fields, other agricultural land, forests, wasteland, road land, building land, railway land, rivers and lakes, seashore, coastal areas, golf courses and other land	2014
Transportation factors	Passengers	Number of daily passengers visiting the station	2014
	Distance to the bus stop	Straight-line distance from station to the nearest bus stop	2010
	Bus stops	Number of bus stops	2010
	Station plazas	Area of the station plazas	2014
	Roads	Number of roads	2010

	A set of variables	Variable description	Year
	Accessibility	Network analysis indicators (integration and betweenness) within each station's area	-
Residential population	Residential population	Number of residents	2015
Landscape	Park	Areas of each type of park	2011
	Green space	Area of green space	2011
Urban vitality	Staying-based activities	Number of visiting populations	2018

4.3.1 Treatment assignment dummy

In Japan, urban parks are classified into many types. Block parks are the smallest (~0.25 ha), followed by neighborhood parks (2 ha), district parks (4 ha), general parks (10–50 ha), sport parks (15–75 ha), wide-area parks (50 ha), and national parks (≥300 ha). The influence of parks on staying-based activities varies according to their type. If all urban parks are treated as a single category in an analysis rather than distinguished by type, the estimated causal effect may become biased. Therefore, it is essential to identify the specific type of park for the analysis.

In the Tokyo metropolitan area, large parks such as district parks and general parks are relatively uncommon. If these were adopted, the number of treatment samples in the PSM would become too small, thereby reducing the generalizability of the results. Moreover, because these large parks occupy substantial land areas, they often dominate the spatial configuration around stations, making it difficult to discuss other urban facilities within the 800-m area. In contrast, very small parks such as block parks tend to have limited attractiveness and exert only a minimal influence on urban vitality. For this reason, this study focused on the impact of neighborhood parks, which are frequently used in daily life and provide a sufficient number of samples for PSM.

Specifically, stations with a neighborhood park located within the 800-m area were defined as the treatment group (Z = 1, N=187), while those without neighborhood parks were defined as the control group (Z = 0, N=502) in the PSM analysis.

4.3.2 Staying-based activities

Although previous studies have focused primarily on pedestrian volumes to identify factors that contribute to urban activities, analyses of staying behavior around station areas remain insufficiently developed. To address this gap, this study employed GPS-based mobile phone data from the KDDI Location Analyzer (KLA), a tool developed by one of Japan's major telecommunications companies, to capture stay-based activities around railway stations. To identify users' place of residence and work, KLA uses the most frequent location during nighttime (22:00–5:00) and daytime (8:00–19:00) over the past month. From these data, individuals can be categorized as residents, workers, or visitors within the 800-m area around each station.

Human flow in public urban space is often regarded as an appropriate proxy. Due to limitations in the KLA dataset, it is not possible to distinguish whether the stays of residents or workers occur inside their homes/workplaces or within public urban spaces. Therefore, the staying population of residents may include individuals who are simply at home rather than engaging in activities in public spaces, and the same issue applies to workers. For this reason, this study focuses specifically on visitors. We defined "visiting population" as visitors who stayed within an 800-m radius of a station for 15 min or longer. This indicator reflects the attractiveness around station areas and was thus used as a proxy for staying-based activities.

The data period selected for this study, October 1–7, 2018, corresponds to the week of Japan's national census. This period was also chosen because the contribution of parks to staying activities may vary by season, with effects likely diminished during the extreme heat of summer or cold of winter. October, corresponding to autumn in Japan, offers comfortable temperatures, which may help reduce the influence of extreme seasonal conditions on outdoor staying activities. In addition, the observation period from October 1 to 7 did not include any national holidays, so the data represent a typical week and are less likely to be strongly affected by holiday-specific events.

Because neighborhood parks may influence visiting populations differently during the day than at night, we separately analyzed the hourly average visiting population for a daytime period (13:00–16:00) and a nighttime period (18:00–21:00). Specifically, the following variables within an 800-m radius of each station were used as indicators of staying-based activities and as outcome variables Y in the PSM analysis. The daily visiting population is calculated as the total number of staying individuals aggregated on a daily basis, whereas the hourly visiting population is calculated by aggregating visiting population at an hourly level.

- Daily visiting population
- Hourly average visiting population during the daytime on weekdays/weekends
- Hourly average visiting population during the nighttime on weekdays/weekends

4.3.3 Confounding variables

To accurately assess the pure effect of neighborhood parks on the visiting population, it was necessary to control for the influence of other factors that might affect urban activity. In this study, we included 65 built-environment variables—covering urban facilities, tourism resources, urban planning policies and land use, transportation factors, residential population, landscape—within an 800-m radius around each station as confounding variables X in the PSM analysis.

a) Urban facilities and tourism resources

Urban facilities are an essential component of urban infrastructure and contribute substantially to daily urban life. The density of urban facilities around railway stations, particularly commercial facilities, often has a substantial influence on the vitality of the station area. For visiting populations, not only urban facilities but also tourism resources might also attract visitors.

b) Urban planning policies and land use

In Japan, urbanized area policies classify land into "urbanization promotion areas" and "urbanization control areas." Promotion areas are those where urban development already exists or should proceed in a planned manner within about 10 years, while control areas are those where development should be restricted. Zoning policies classify city areas into residential, commercial, and industrial zones, with strict regulations on land use and building forms such as floor area ratio.

Because this policy-based planning reflects both current and future urban development intentions and is closely related to spatial structure, the potential influence on urban activities must be controlled for. In addition to planning policies, the actual land-use can also affect urban vitality (Wu et al., 2018).

c) Transportation factors

Previous studies have shown that factors such as road density, bus-stop density, road network, and distance to bus stop/railway station/plazas significantly influence urban vitality (Li et al., 2022). Therefore, their potential influence on urban activities should be controlled in the PSM analysis. Because the analysis targets in this study were railway stations, and in Japan, station plazas are generally more common than general public plazas around railway stations, the variables "area of station plazas" and "Number of daily passengers" were considered more appropriate compared with "distance to plaza/railway station."

In addition, this study calculated two network analysis indicators—the average values of "closeness centrality" and "betweenness centrality" across all nodes—based on the street network within an 800-m radius of each station. Closeness centrality (c_c) is used to characterize the distance relationships between a node and all other nodes in the network, reflecting how easily the node can reach other nodes.

$$C_c(i) = \frac{n-1}{\sum_{j=1}^{n} d_{ij}} \quad (5)$$

where n denotes the total number of nodes in the network, and d_{ij} represents the shortest path distance between node i and node j.

Betweenness centrality (c_b) extent to which a node forms part of the shortest paths between other pairs of nodes, thereby reflecting its role within the network.

$$C_b(i) = \sum_{i \neq j \neq k} \frac{g_{jk}(i)}{g_{jk}} \quad (6)$$

where $g_{jk}$is the total number of shortest paths between nodes j and k, and $g_{jk}(i)$ denotes the number of those shortest paths that pass-through node i.

d) Landscape

Vegetation index is also recognized as a factor in analyses of urban vitality (Li et al., 2022). Although the present study focuses on examining the impact of neighborhood parks on urban vitality within station areas, the other landscape environment (e.g., other urban parks and green spaces), might also attract urban activities. Therefore, the areas of other types of urban parks (excluding neighborhood parks) and green spaces within an 800-m radius of each station were included as confounding variables in the PSM analysis.

e) Residential population

The residential population forms the foundation of urban activities and is widely recognized as an essential factor in analyses of urban vitality (Hajrasouliha &Yin, 2015). Therefore, the potential influence of the residential population on urban activities was included as a confounding variable in the PSM analysis.

5 Results

5.1 Average treatment effect on the treated of neighborhood parks for visiting population

5.1.1 Propensity score estimation by logistic regression

As described above, to eliminate the influence of confounding variables on staying-based activities, this study applied PSM to estimate the net effect of neighborhood parks. Specifically, stations with a neighborhood park located within an 800-m radius were defined as the treatment group (Z = 1, N = 187), while those without such neighborhood parks were defined as the control group (Z = 0, N = 502). Variables other than the presence of neighborhood parks that might affect staying-based activities were treated as confounding variables, while the visiting population within an 800-m radius of each station was used as the outcome variable for comparison. To address the imbalance in confounding variables between the treatment and control groups, this study introduced the PS, which condenses information from multiple confounding variables into a single measure, thereby improving the comparability of the treatment and control groups during the matching process.

In this study, PS was estimated using logistic regression, with 65 confounding variables included as independent variables and the dependent variable defined as the presence or absence of a neighborhood park within an 800-m radius of the station (N=689). In PSM, PS serves only as a summary index that condenses the information of numerous confounding variables into a single value; therefore, its model specification differs from that of the traditional logistic regression used for statistical inference. Specifically, in the calculation of the PS, the statistical significance and multicollinearity of the confounding variables are not critical concerns. On the contrary, excluding certain confounding variables merely because they are statistically insignificant or exhibit multicollinearity might lead to the omission of important confounding variables that influence the outcome variable, thereby reducing the reliability of the estimated effects (Chen et al., 2021).

Figure 3 illustrates the distributions of the estimated PS for the treatment group (1: bottom) and control group (0: top). The horizontal axis represents the estimated PS, while the vertical axis indicates probability density. If the distributions of the two groups are completely separated, it implies that there are no stations with similar characteristics shared between the groups, making a valid comparison impossible. However, the overlapping region of the distributions indicates that samples from both groups have comparable treatment probabilities, which can serve as suitable matches. Figure 3 shows that the distribution of the control group is concentrated around the range of 0–0.3, while the treatment group exhibits a wider spread between 0.2 and 0.7. Although differences remain in the overall PS distributions between the two groups, there is some overlap, indicating basic comparability between the two groups for the subsequent PSM analysis.

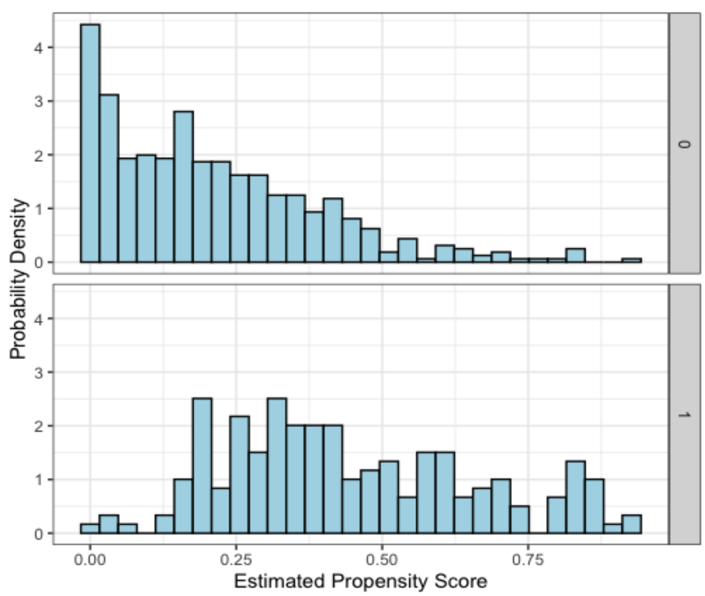

Figure 3. Estimations of propensity score before matching

5.1.2 Propensity score matching

As shown in Figure 3, the PS distribution differed between the treatment and control groups, indicating low comparability between them. To adjust the imbalance in confounding variables between the two groups, nearest-neighbor matching was performed using estimated PS for each station. Specifically, for each station in the treatment group, the station with the closest PS in the control group was selected to form a one-to-one match.

Figure 4 shows histograms comparing the distribution of PS between the treatment and control groups before and after matching. Prior to matching (Z = 1, N=187; Z = 0, N=502), the treatment group distribution was concentrated at higher PS values (above approximately 0.5), while the control group was concentrated at lower scores (below around 0.2), indicating substantial imbalance between groups. It shows that without matching, comparison between the two groups would make it difficult to determine whether the observed differences in visiting population are attributable to an imbalance in confounding variables or to the effect of the neighborhood parks themselves. After applying matching with a caliper, the distributions of PS (Figure 4) in the treatment and new control groups (Z = 1, N=187; Z = 0, N=187) showed greater overlap in the range of 0.2 to 0.8. This improvement in overlap implies that the matching procedure successfully improved the balance in confounding variables, increasing the comparability of the treatment and newly control groups for subsequent analyses.

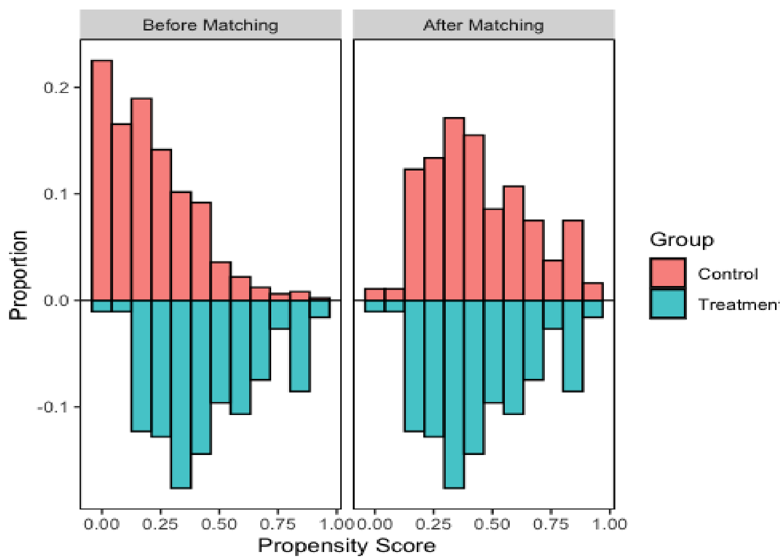

Figure 4. Distributions of PS before and after matching

5.1.3 Balance assessment

After matching, it was also essential to verify whether this adjustment effectively improved the balance of confounding variables between the new two groups. To confirm this, the balance of confounding variables was examined, as shown in Figure 5. Specifically, the distributions of the ASMD before matching (blue) and after matching (red) were compared. ASMD can be used to assess whether an imbalance exists in each confounding variable, and an ASMD of 0.25 or lower is generally considered to indicate a good balance in confounding variables between the treatment and control groups (Wu et al., 2018).

As shown in Figure 5, many confounding variables, including "residential populations" and "area of building land," had ASMD values exceeding 0.25 before matching (Z = 1, N=187; Z = 0, N=502). This indicates that, prior to matching, there was a clear imbalance in these variables between the treatment group (stations with neighborhood parks) and the control group (stations without neighborhood parks). Without adjusting for these imbalances, a direct comparison of the two groups would make it difficult to determine whether the observed differences in visiting population might be attributable to confounding variable imbalances such as residential population or building land area, rather than to the effects of neighborhood parks.

After performing matching, all 65 confounding variables used in this study showed ASMD values below 0.25 between the treatment and the newly constructed control group (Z = 1, N=187; Z = 0, N=187), indicating that a good balance had been achieved. Consequently, apart from the treatment variable itself (i.e., the presence or absence of neighborhood parks around the station), all confounding variables that might influence the visiting population were balanced between the two groups. Therefore, any remaining differences in visiting population can reasonably be attributed to the neighborhood park.

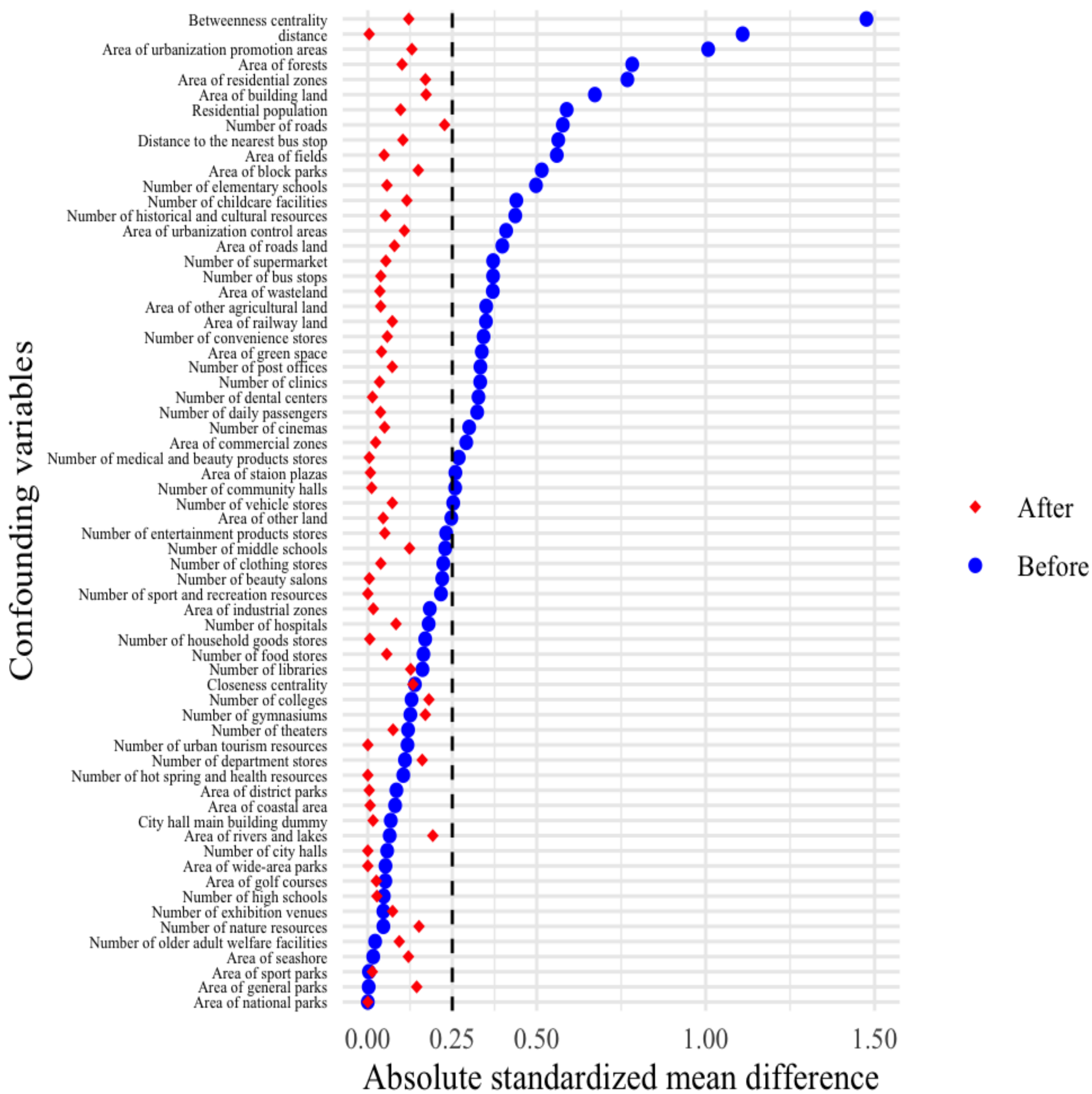

Figure 5. Balance assessment of confounding variables

5.1.4 Average treatment effect on the treated of neighborhood parks for visiting populations

Through the PSM analysis, we found that after matching, the treatment group and the newly constructed control group (Z = 1, N = 187; Z = 0, N = 187) achieved a balance in confounding variables. To quantify the impact of neighborhood parks on staying-based activities, the ATT was estimated using the treatment and newly control groups. In addition, this study employed simple regression analysis to validate the robustness of the results, using each urban vitality indicator as the dependent variable and the neighborhood park development dummy within the station area as the independent variable.

The results are presented in Table 4. Based on the results averaged over 1 week, the daily visiting populations around stations with neighborhood parks were 4,743 higher than those around stations without neighborhood parks, which is statistically significant at the 10% level and very close to the 5% level, indicating a certain degree of reliability. This is because neighborhood parks are inherently attractive spaces for relaxation, social interaction, and strolling, and they encourage visitors to extend their visiting duration around station areas. In addition, stations are often surrounded by a variety of commercial facilities alongside parks, which enables people to combine activities such as shopping and dining with leisure time in the neighborhood park, thereby promoting multi-purpose visits and stays.

The impact of neighborhood parks on visiting populations is considered to differ between daytime and nighttime on weekdays and weekends. The results show that the ATT of neighborhood parks on hourly daytime visiting populations on weekdays increased by 1,362 persons, a statistically significant difference. In contrast, the effect during nighttime on weekdays increased by 1,048 persons, a positive value but not reach the 5% statistically significance. On weekends, the effects on visiting population were more pronounced: an increase of 2,501 persons during daytime and 1,690 persons during nighttime, both of which are statistically significant at the 5% level. This shows that the impact of neighborhood parks on visiting populations was generally higher on weekends than on weekdays overall. The results also indicate that the impact of neighborhood parks on visiting populations exhibits temporal variation. On weekdays, a statistically significant effect was observed during daytime, but the effect was limited at night. This can be interpreted as a result of weekday nighttime being dominated by commuting and return-home trips, during which parks are less likely to serve as destinations, thereby reducing their impact. In contrast, people have more free time on weekend, therefore neighborhood parks, leisure activities, and interactions with commercial facilities likely contributed to the increase in visiting populations around stations during both daytime and nighttime. Therefore, the **Hypothesis 1** was verified in this analysis.

Table 4. ATT of neighborhood parks on visiting populations

Day type	Indicator	ATT (persons)	P value
Overall	Daily visiting population	4,743	0.062*
Weekday	Hourly daytime visiting population	1,362	0.047**
Weekday	Hourly nighttime visiting population	1,048	0.076*
Weekend	Hourly daytime visiting population	2,501	0.018**
Weekend	Hourly nighttime visiting population	1,690	0.032**

*: p<0.10, **: p<0.05, ***: p<0.01

5.2 Conditional average treatment effect of neighborhood parks based on station-area characteristics

As shown in Table 4, PSM provides insights into the overall average impact of neighborhood parks on visiting populations. However, because ATT reflects only the average treatment effect, it cannot reveal differences in effects across locational conditions. In reality, the impacts of neighborhood parks are likely to be strongly influenced by the locational characteristics of the surrounding station area. For example, when a neighborhood park is located in an area dominated by agricultural land with limited urban infrastructure, its effect on visiting populations may be minimal. In contrast, if a park is situated in a commercially dense district, the effect on visiting populations may be stronger. To address this point, this study employed the causal forests method to examine which station-area characteristics influence the effects of neighborhood parks on visiting populations.

Although Table 4 shows the effects of neighborhood parks on five types of visiting populations, this section focuses specifically on analyzing the "daily visiting population." In this study, the confounding variables incorporated station-area characteristics, including urban infrastructure, land use, and transportation accessibility. Because the impact of neighborhood parks on visiting populations can vary depending on these conditions, we applied the causal forest method to calculate importance scores for all 65 confounding variables, thereby identifying the factors that play critical roles in influencing the effects. The results are shown in Figure 6. Importance scores are the results of comparisons among variables to identify which factors are relatively more influential, rather than provide an absolute threshold for interpreting variable importance scores.

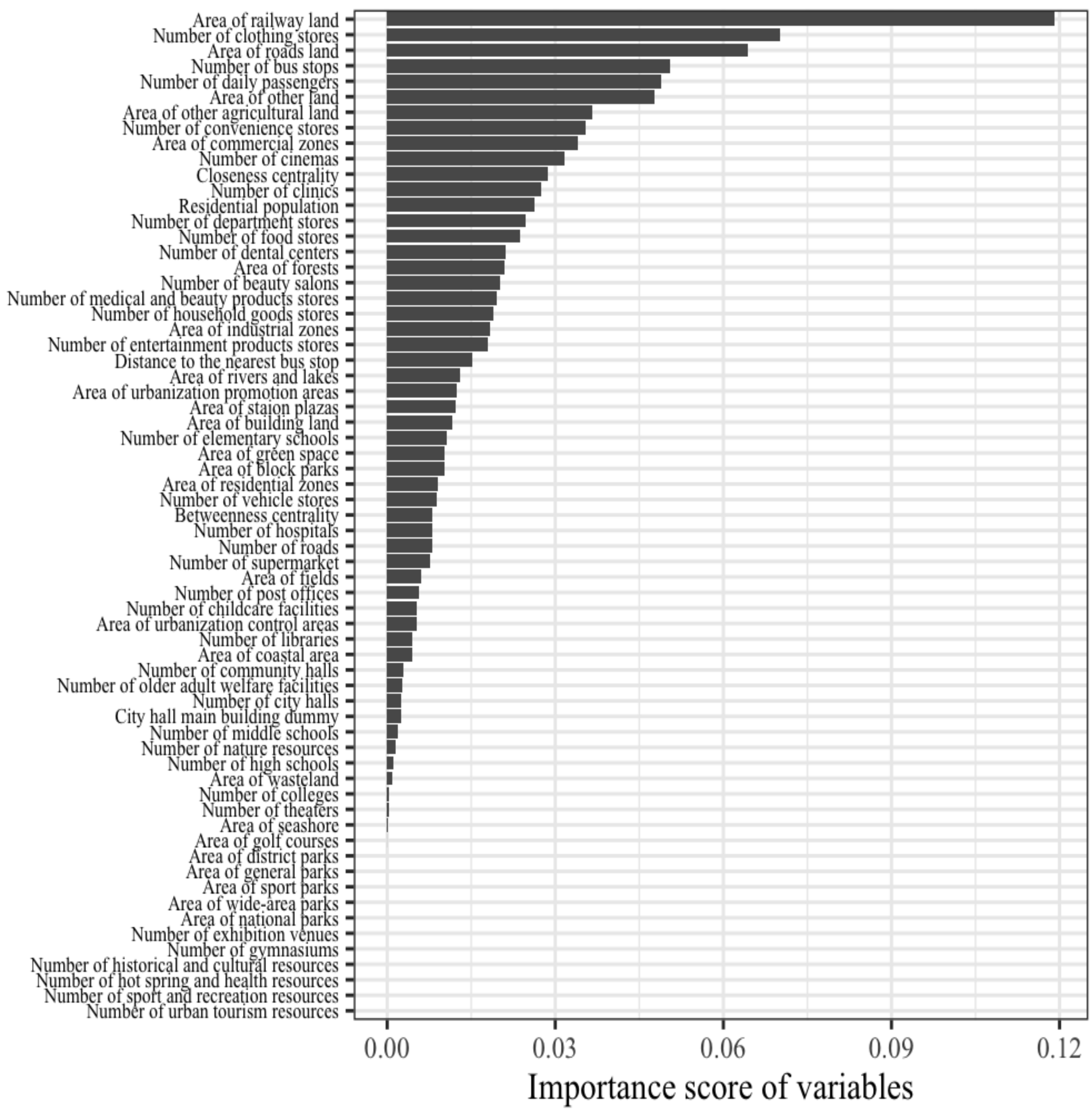

Figure 6. Importance ranking of the impact of neighborhood parks on daily visiting populations

To further illustrate the heterogeneity in neighborhood park effects across different station-area conditions, Figure 7 showed the variation in the effect of neighborhood parks on daily visiting populations for each station. The vertical axis represents the neighborhood park effects estimated using the causal forests model, while the horizontal axis corresponds to the values of each factor. Each dot denotes a station. The blue line depicts the CATE estimated using local regression smoothing, and the shaded gray area indicates the confidence intervals.

As Figure 6 shows, "area of railway land" was the most important variable influencing the effect of neighborhood parks on daily visiting population. The CATE of neighborhood parks on visiting populations also showed a positive trend with an increasing area of railway land (Figure 7). Because the analysis in this study targeted railway stations, a larger area of railway land generally indicates larger station scale and higher passenger volumes, which are likely to influence park effects. In addition, “area of road land” was identified as another important variable, reflecting the transportation accessibility around the station. In such transport areas, it is easy to access the station and neighborhood parks, enhancing neighborhood park impact on visiting population.

Land-use related factors around stations such as “area of other agricultural land” and “area of forest land” also demonstrated high importance. As shown in Figure 7, with the increase in agricultural land area, the impact of neighborhood parks on daily visiting population tended to decrease. These variables reflect the characteristics of suburban and rural contexts; where agricultural and forest land occupies a larger area, urban and transport infrastructure tends to be less developed, and the weaker integration of neighborhood parks with the surrounding area limits its ability to attract and retain visitors. In contrast, the importance of “area of commercial zones” exhibited the opposite pattern. These areas are typically located in urban centers, where commercial and entertainment facilities are highly concentrated and transport infrastructure is well developed, thereby offering greater potential to increase the visiting population around stations. Under such conditions, neighborhood parks can interact more strongly with surrounding commercial facilities. This also explains the results shown in Figure 7 that the effect of neighborhood parks on daily visiting populations is higher in station areas with larger commercial zones than in those with smaller commercial zones.

Retail and attraction facilities related factors around stations, including “number of clothing stores,” “number of convenience stores,” “number of food stores,” and “number of cinemas” showed high importance. The reason is that these facilities, which are frequently used in daily life, exhibit the strong visitor-drawing functions of station area. Visitors may choose to rest in the park after shopping or entertainment activities. Therefore, the impact of neighborhood parks on daily visiting populations showed a positive trend with an increasing number of these facilities in Figure 7.

Transport-related related factors around stations, such as “number of bus stops” and “number of daily passengers” were also found to be highly important (Figure 6). The impact of neighborhood parks on visiting populations showed a positive trend with an increasing number of these facilities (Figure 7). The reason is that these variables generally reflect the transport accessibility around station areas. Because this study focuses on visiting populations rather than residents or commuters, the accessibility plays an important role in determining whether individuals make use of parks.

In addition, “residential population” also demonstrated high importance (Figure 6), and the impact of neighborhood parks on daily visiting populations showed a positive trend with an increasing residential population (Figure 7). The reason is that the location of urban facilities tends to be closely related to the surrounding residential population. Areas with larger populations tend to attract more facility development, thereby influencing the synergies with neighborhood parks.

From these results, we can see that the effects of neighborhood park on the visiting populations around stations largely depend on the conditions of the surrounding station area. This not only provides an explanation for **H2** but also clarifies why previous studies have reported inconsistent findings regarding the impact of urban parks on urban vitality (Che et al., 2024; Li et al., 2022).

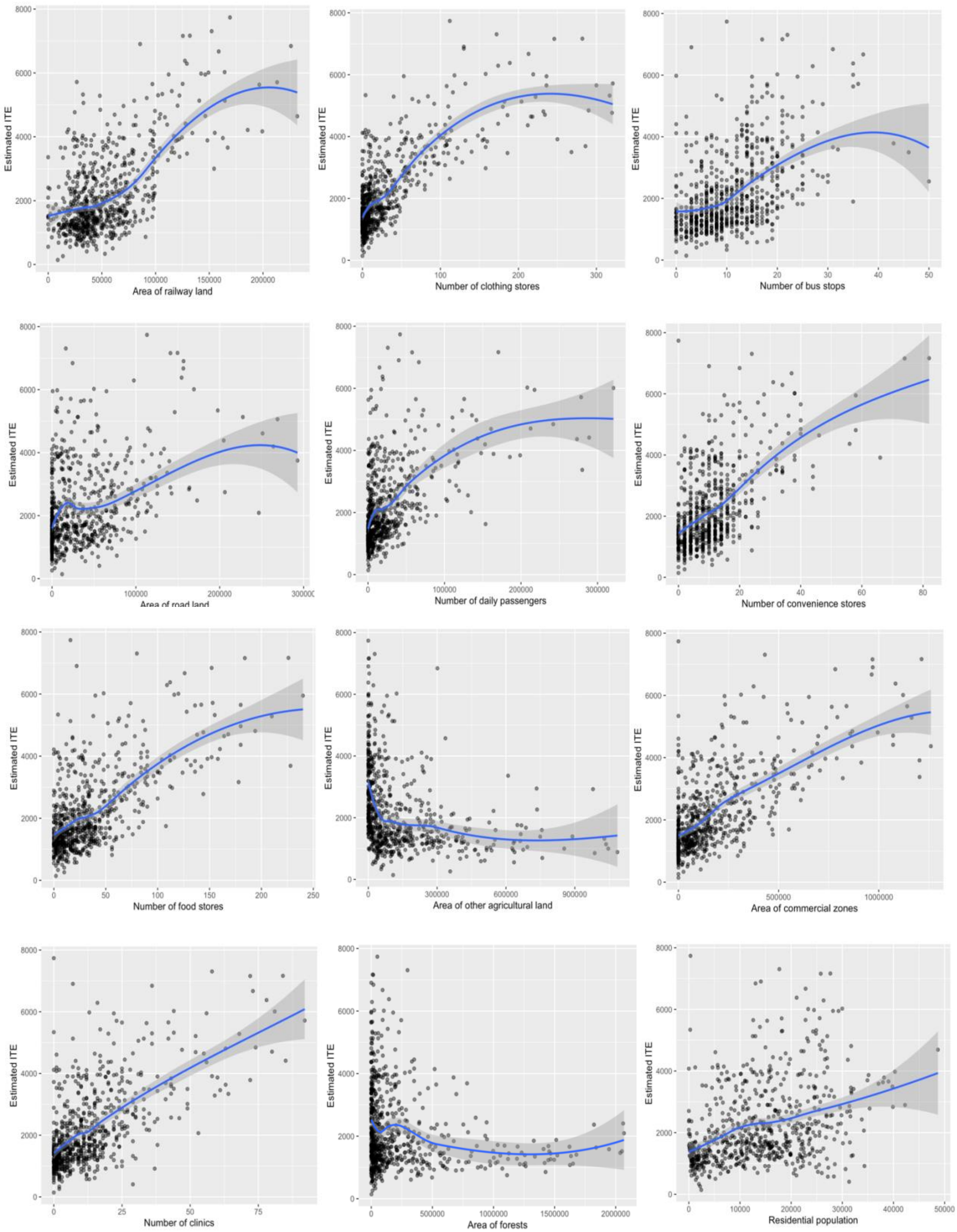

Figure 7. Conditional average treatment effect on neighborhood parks and important factors

6 Conclusions

6.1 Conclusions

The integration of TOD and green urbanism has been increasingly recognized as a sustainable policy for future cities development. Although green infrastructure plays a crucial role in promoting environment and social activities, the impact on stationary vitality is unclear. In particular, it remains to be seen whether the observed increases in urban vitality are directly attributable to parks or are instead driven by other factors such as population density and land-use. To answer this, we focused on TOD areas in the Tokyo metropolitan region and used mobile phone location data to capture staying-based activities in urban spaces, thereby analyzing the effects of neighborhood parks on urban vitality.

Firstly, we estimated the net effects of neighborhood parks on staying-based activities by applying the PSM method. The result indicates that stations with neighborhood parks have an average treatment effect of 4,743 daily visiting population, suggesting that neighborhood parks contribute to enhancing urban vitality. As time-of-day and day-of-week differences, the weekday daytime visiting population was found to increase by 1,362 visiting population, whereas the nighttime number increased by 1,048 (not statistically significant). On weekends, the effects were stronger: a daytime increase of 2,501 and a nighttime increase of 1,690, both of which were significant at the 5% level. These results highlight that neighborhood parks play an important role in urban vitality during weekends, when recreational demand is higher.

At the same time, the impact of neighborhood parks is likely to be influenced by the built environment of the station area. To examine this, we also employed a causal forest method to identify which station-area characteristics shape the effect of neighborhood parks on visiting population. The results show that land-use policy, transport conditions, commercial facilities, and demographics significantly shape the impact of neighborhood parks on visiting populations. In particular, areas with intense commercial activity such as numerous clothing stores, convenience stores, and department stores as well as a large commercial zone, show strong positive synergies with neighborhood parks in enhancing urban vitality. Transportation factors such as the number of bus stops and daily passengers also strengthen the park effect. In contrast, areas dominated by forest or agricultural land tend to show negative associations with park impact.

This study found that neighborhood parks have a positive effect on enhancing urban vitality around station areas, particularly in commercially dense areas and during weekends, where they substantially increase visiting populations. In urban planning, priority can be given to developing neighborhood parks in station areas with strong commercial potential while also improving accessibility through coordination with public transport facilities such as bus stops. Integrating urban parks with commercial functions can help create green, livable, and vibrant urban environments. The results also indicate that the effectiveness of urban parks on urban vitality is influenced by land-use composition. This suggests that policy design should be sensitive to the specific characteristics of stations.

6.2 Discussion

This study quantitatively assessed how urban parks affect stationary activities around railway stations and provided empirical evidence on whether the integration of TOD and green urbanism can contribute to urban vitality. The results show that the presence of urban parks near stations is associated with increased stationary activities among visitors. This suggests that parks do more than improve environmental conditions; they may also encourage people to remain in station areas for longer periods. It offers a useful perspective for understanding how green infrastructure support urban vitality.

The findings also add empirical support to the discussion on combining TOD with green urbanism. Conventional TOD planning has tended to emphasize compact development, high-density land use, public transport accessibility, and the concentration of urban functions around stations. However, when density and commercial development are given excessive priority, station areas may attract visitors without providing comfortable spaces where people are willing to stay. The results of this study indicate that urban parks can encourage longer and more active use of station areas, thereby complementing the functions traditionally emphasized in TOD planning. Providing parks in TOD areas may therefore offer a planning approach that supports both environmental sustainability and urban vitality. This has important implications for the future development of greener and more livable transit-oriented areas.

The study also examined whether the effects of urban parks differ according to the surrounding built environment. The results indicate that the contribution of parks to stationary activities is not uniform across station areas. Instead, it may be strengthened or weakened by local conditions, including land use, commercial facilities, accessibility, transport characteristics, and other area-specific factors. Such contextual variation may help explain why previous studies have reported inconsistent relationships between urban parks and urban vitality. It also provides more specific evidence for the strategic provision of urban parks around railway stations.

6.3 Limitations

In this study, PSM and the causal forest method were employed to estimate the effects of urban parks on urban vitality while controlling for observable confounding factors. Although the results suggest a positive association between neighborhood parks and urban vitality, the identification strategy still relies on the assumption of "no unobserved confounding variables." To mitigate this assumption, a comprehensive set of covariates, including built-environment characteristics, demographic variables, and accessibility indicators, was incorporated into the analysis. However, it remains possible that unobserved factors—such as local policy interventions, historical development processes, or unmeasured social characteristics—may influence urban vitality. Due to data limitations, these factors could not be explicitly controlled for in this study. Therefore, the estimated results should be interpreted as causal effects identified under the given assumptions.

In addition, to ensure temporal consistency as much as possible, 2014 was selected as the reference year. For variables without data available for 2014, values from the closest year within a range of four years before or after were used, under the assumption that these variables did not change substantially within this period. However, such temporal inconsistencies may still influence the interpretation of the results. Future research could further address temporal effects if suitable data become available. However, as noted above, the use of mobile phone location data to capture staying-based activities has only become widespread in recent years, and historical data are generally unavailable. This limitation makes it difficult to examine long-term temporal. Therefore, future studies may consider incorporating alternative indicators or data sources to better account for temporal dynamics.

In this study, staying-based activities around railway stations were defined as visits lasting 15 minutes or longer. Given the high frequency of train departures in the Tokyo metropolitan area, this threshold likely excludes most passengers whose brief presence results from transfers between trains. We also acknowledge, however, that some passengers may still be included within this threshold for transit-related reasons, such as waiting up to 15 minutes for a train, transferring between lines, or meeting other travelers in the station. The observed visiting population should therefore be interpreted as representing sustained presence within station areas rather than exclusively leisure-oriented urban activities. A 30-minute threshold might more effectively distinguish staying-oriented activities from transit-related waiting, but this would also exclude a substantial number of visitors who genuinely stay near the station for less than 30 minutes. As our data do not include individual movement trajectories, future research would integrate other GPS-based data with trajectory information to distinguish among different types of staying behavior.

We evaluated the effects of neighborhood parks on staying activities around railway stations and further examined how external factors, such as the surrounding station environment, may influence these effects. However, we also acknowledge that, even within the same park category, the effects of parks on urban vitality may vary depending on their size, internal facilities, maintenance conditions, landscaping, water features, and event activities. Because detailed park-level information on these characteristics was unavailable in the open datasets used in this study, such heterogeneity could not be incorporated into the analysis. Future research should collect more detailed information on individual parks and investigate how differences in their internal characteristics contribute to effects on urban vitality.

Visitors' staying-based activities are strongly influenced by seasonal and weather conditions. In this study, we used data collected in October to minimize the influence of extreme seasonal conditions on outdoor staying activities. In future research, to better understand how the effects of parks on urban vitality vary across different seasons and weather conditions, we will collect new datasets and conduct a more comprehensive analysis.

Acknowledgments

The authors thank ThinkSCIENCE for English language editing.

The staying-based activity data used in this study were obtained from a telecommunications company under a data-use agreement and are subject to contractual restrictions. Therefore, the dataset and associated analysis scripts cannot be made publicly available. For inquiries regarding the data or analysis, please contact the corresponding author.

Author contribution

The authors confirm contribution to the paper as follows: study conception and design: X. Bo; data collection: X. Bo, S. Terabe; analysis and interpretation of results: X. Bo, S. Terabe, H. Nishiuchi; draft manuscript preparation: X. Bo. All authors reviewed the results and approved the final version of the manuscript.

References

Arena, R., Bond, S., O'Neill, R., Laddu, D. R., Hills, A. P., Lavie, C. J., & McNeil, A. (2017). Public park spaces as a platform to promote healthy living: Introducing a HealthPark concept. *Progress in Cardiovascular Diseases*, *60*(1), 152–158. https://doi.org/10.1016/j.pcad.2017.05.010

Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, *10*(2), 150–161. https://doi.org/10.1002/pst.433

Cervero, R., & Sullivan, C. (2011). Green TODs: Marrying transit-oriented development and green urbanism. *International Journal of Sustainable Development & World Ecology*, *18*(3), 210–218. https://doi.org/10.1080/13504509.2011.570801

Che, L., Deng, Y., & Guo, S. (2024). Association between park vitality and commercial vitality: A case study in Chengdu. *Journal of Asian Architecture and Building Engineering*, *24*(4), 3127–3143. https://doi.org/10.1080/13467581.2024.2366826

Chen, F., Wu, J., Chen, X., & Nielsen, C. P. (2021). Disentangling the impacts of the built environment and residential self-selection on travel behavior: An empirical study in the context of diversified housing types. *Cities*, *116*, 103285. https://doi.org/10.1016/j.cities.2021.103285

Chen, Q., Zhao, S., & Higuchi, G. (2018). Analysis on the influence factors of passenger by using small sample size of subway stations. *Japan's Journal of Architecture and Planning, 83*(747), 907–916. https://doi.org/10.3130/aija.83.907

Chen, W., Natapov, A., & Ali, Y. (2025). Can urban vitality be seen? Video analytics of social interaction and land use. *Land Use Policy*, *159*, 107818. https://doi.org/10.1016/j.landusepol.2025.107818

Eisazadeh, N., & Vahdani, H. (2017). The role of urban spaces and structures in increasing the social vitality of citizens with an emphasis on urban design approaches. *Journal of Ecophysiology and Occupational Health*, *17*(3–4), 77–86. https://doi.org/10.18311/jeoh/2017/18296

Ewing, R., Hajrasouliha, A., Neckerman, K. M., Purciel-Hill, M., & Greene, W. (2015). Streetscape features related to pedestrian activity. *Journal of Planning Education and Research*, *36*(1), 5–15. https://doi.org/10.1177/0739456X15591585

Gerike, R., Koszowski, C., Schröter, B., Buehler, R., Schepers, P., Weber, J., Wittwer, R., & Jones, P. (2021). Built environment determinants of pedestrian activities and their consideration in urban street design. *Sustainability*. *13*(16), 9362. https://doi.org/10.3390/su13169362

Guerra, E., & Cervero, R. (2013). Is a half-mile circle the right standard for TODs? *ACCESS Magazine*. https://www.accessmagazine.org/spring-2013/half-mile-circle-right-standard-tods/

Guo, S., Fraser, M., & Qi, C. (2020). Propensity score analysis: Recent debate and discussion. *Journal of the Society for Social Work and Research*, *11*, 463–482. https://doi.org/10.1086/711393

Hajrasouliha, A., & Yin, L. (2015). The impact of street network connectivity on pedestrian volume. *Urban Studies*, *52*(13), 2483–2497. https://doi.org/10.1177/0042098014544763

Hamana, S., Nakagawa, D., Matsunaka, R. & Oba, T. (2010). A study on the relationship of pedestrian space and the life of shopping malls. *Infrastructure Planning Review*, *27*(2), 313–321. https://doi.org/10.2208/journalip.27.313

Li, W., Gu, H., Han, S., & Sun, B. (2024). On the changing of urban spatial structure: A glimpse from Shanghai's relocated residents. *Cities*, *147*, 104795. https://doi.org/10.1016/j.cities.2024.104795

Li, X., Li, Y., Jia, T., Zhou, L., & Hijazi, I. H. (2022). The six dimensions of built environment on urban vitality: Fusion evidence from multi-source data. *Cities*, *121*, 103482. https://doi.org/10.1016/j.cities.2021.103482

Ministry of Land, Infrastructure, Transport and Tourism (MLIT). (2023). *Utilization of digital technologies in measuring urban vitality.* MLIT.

Montgomery, J. (1998). Making a city: Urbanity, vitality and urban design. *Journal of Urban Design*, *3*(1), 93–116. https://doi.org/10.1080/13574809808724418

Mu, B., Liu, C., Mu, T., Xu, X., Tian, G., Zhang, Y., & Kim, G. (2021). Spatiotemporal fluctuations in urban park spatial vitality determined by on-site observation and behavior mapping: A case study of three parks in Zhengzhou City, China. *Urban Forestry & Urban Greening*, *64*. 127246. https://doi.org/10.1016/j.ufug.2021.127246

Ozbil, A., Peponis, J., & Stone, B. (2011). Understanding the link between street connectivity, land use and pedestrian flows. *Urban Design International*, *16*(2), 125–141. https://doi.org/10.1057/udi.2011.2

Peters, K., Elands, B., & Buijs, A. (2010). Social interactions in urban parks: Stimulating social cohesion? *Urban Forestry & Urban Greening*, *9*(2), 93–100. https://doi.org/10.1016/j.ufug.2009.11.003

Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, S., Cardillo, A., Belli, E., Cardenas, F., & Cormenzana, B. (2012). Street centrality and the location of economic activities in Barcelona. *Urban Studies*, *49*(7), 1471–1488. http://www.jstor.org/stable/26150937

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–45. https://doi.org/10.1093/biomet/70.1.41

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. https://doi.org/10.1214/09-STS313

Susukida, R., Amin-Esmaeili, M., Badillo-Goicoechea, E., Nguyen, T. Q., Stuart, E. A., Rosenblum, M., Dunn, K. E., & Mojtabai, R. (2025). Application of causal forest model to examine treatment effect heterogeneity in substance use disorder psychosocial treatments. *International Journal of Methods in Psychiatric Research*, *34*(1), e70011. https://doi.org/10.1002/mpr.70011

Wang, J., & Li, G. (2022). Pursuing educational equality and divergence in the housing market: How do educational equality policies affect housing prices in Shanghai? *Cities*, *131*, 104001. https://doi.org/10.1016/j.cities.2022.104001

Wu, J., Ta, N., Song, Y., Lin, J., & Chai, Y. (2018). Urban form breeds neighborhood vibrancy: A case study using a GPS-based activity survey in suburban Beijing. *Cities*, *74*, 100–108. https://doi.org/10.1016/j.cities.2017.11.008

Wu, W., & Niu, X. (2019). Influence of built environment on urban vitality: Case study of Shanghai using mobile phone location data. *Journal of Urban Planning and Development*, *145*(3). https://doi.org/10.1061/(ASCE)UP.1943-5444.0000513

Wu, W., Yao, Y., & Wang, R. (2023). Green space exposure at subway stations, transportation mode choice and travel satisfaction. *Transportation Research Part D: Transport and Environment*, 122. https://doi.org/10.1016/j.trd.2023.103862

Yamaguchi, K., Bo, X., Terabe, S. & Ajito, M.(2025). Causal effect of neighborhood park development on urban liveliness. *Journal of the Eastern Asia Society for Transportation Studies*, *16*. https://doi.org/10.11175/easts.16.PP4068

Zhu, J.; Lu, H.; Zheng, T.; Rong, Y.; Wang, C.; Zhang, W.; Yan, Y.; Tang, L. (2020). Vitality of urban parks and its influencing factors from the perspective of recreational service supply, demand, and spatial links. *International Journal of Environmental Research and Public Health*, *17*, 1615. https://doi.org/10.3390/ijerph17051615