Travel Behavior in TODs vs. non-TODs: Using Cluster Analysis and Propensity Score Matching

Keunhyun Park
Department of City & Metropolitan Planning
University of Utah
375 S 1530 E, Salt Lake City, UT 84112
Tel: 801-803-9547; Fax: 801-581-8217; Email: keunhyun.park@utah.edu

Reid Ewing, PhD
Department of City & Metropolitan Planning
University of Utah
375 S 1530 E, Room 235, Salt Lake City, UT 84112
Tel: 801-581-8255; Fax: 801-581-8217; Email: ewing@arch.utah.edu

Brenda Case Scheer
Department of City & Metropolitan Planning
University of Utah
375 S 1530 E, Room 235, Salt Lake City, UT 84112
Tel: 801-824-0359; Fax: 801-581-8217; Email: scheer@arch.utah.edu

Shabnam Sifat Ara Khan
City of Spearfish, South Dakota
625 Fifth Street, Spearfish, SD 57783
Tel: 605-642-1335; Fax: 605-642-1337; Email: shabnam.khan@auw.edu.bd

Word count: 6,199 words text + 5 tables/figures x 250 words (each) = 7,449 words
Tables: 5

Submission Date: July 31, 2016.

ABSTRACT

Transit-oriented development (TOD) has gained popularity worldwide as a sustainable form of urbanism by concentrating developments near a transit station so as to minimize auto-dependency and maximize ridership. Existing travel behavior studies in the context of TOD, however, are limited in terms of small sample size, lack of consistency in TOD classification method, and failure to control for residential self-selection. This study examines various travel outcomes – VMT, auto trips, transit trips, and walk trips – in different types of station areas in eight American urban areas using cluster analysis and propensity score matching methods. From cluster analysis with three built environment factors commonly referred as D variables – activity density, land use diversity, and street network design – in ½ mile (about 800 meter) buffer, this study classifies existing station areas as TOD, TAD (transit-adjacent development), and Hybrid types. Literally hundreds of studies have related D variables to household travel outcomes (Ewing & Cervero, 2010). After controlling residential self-section, the result shows that a TOD motivates its residents to walk more and take transit more while driving less. The VMT, however, is not significantly different between TOD and TAD households, implying that a TOD may convert only internal auto trips into walk or transit trips. Travel behavior in the Hybrid type is also examined for the potential outcome of gradual and practical change.

Keywords: Transit-oriented development, Transit adjacent development, Station area types, Residential self-selection

INTRODUCTION

Expenditures on transportation have increased from the sixth largest share (less than 2%) of household budgets in 1917 to the second largest share since the 1970s (17% in 2014; U.S. Bureau of Labor Statistics, 2014). Under this circumstance, transit-oriented development (TOD) has gained popularity worldwide as a sustainable form of urbanism by concentrating developments near a transit station so as to minimize auto-dependency and maximize ridership. A TOD project should give people more transportation options and in turn, decrease their transportation cost.

Much of the literature verifies that TODs enhance the use of public transport and reduce car usage (Cervero, 1993, 2004; Langlois et al., 2015; Nasri & Zhang, 2014; Olaru & Curtis, 2015; Venigalla & Faghri, 2015). Existing TOD studies, however, have limits in terms of 1) small numbers of study sites, 2) lack of systematic methodology to distinguish TOD from other types of station areas, and 3) lack of control for the impact of residential self-selection on travel behavior. There are many exceptions to the above limitations, but no study overcomes all three limitations. As a result, it is hard to generalize the findings from the literature to other regions. Also, the current distinctions between TODs and TADs limit the practical implication for transit officials and planners. Finally, when it fails to control self-selection effect, the result might overestimate the impact of TOD urban form on travel behavior.

Thus, this study asks two research questions. First, how can we distinguish between Transit oriented development (TOD) and Transit-adjacent development (TAD)? Second, how do travel behaviors vary between TODs and TADs? To answer these questions, we utilize cluster analysis to classify station area types and propensity score matching to control for residential self-selection. For better generalizability, the data is collected from household travel surveys in eight urban areas across the U.S. with exact XY coordinates for households and trip ends. For greater policy relevance, the data is used to analyze various travel outcomes such as automobile, transit, and walk trips at household level for different station area types. By doing so, this study seeks to examine the pure impact of living in TOD – versus a TAD – on travel behavior.

There is broad interest in the planning and policy communities for accurate tools to predict the consequences of TOD on the generation of transit ridership and reduction of automobile usage. Our analysis will help guide transportation planners and decision makers to evaluate TOD projects relative to their economic, social, and environmental impacts.

LITERATURE REVIEW

TOD/TAD Classification

Bernick and Cervero (1997, p.5) define TOD as "a compact, mixed-use community, centered around a transit station that, by design, invites residents, workers, and shoppers to drive their cars less and ride mass transit more." Kamruzzaman et al. (2015) state that TOD is a neighborhood that is served by public transit services and offers amenities such as density, walkable, well-connected street patterns, and diversified land uses. TAD is often defined as a failure of a TOD. A TAD is a non-compact, segregated neighborhood development that calls for auto uses instead of inviting walk trips (Belzer & Autler, 2000; Cervero and Duncan, 2002; Dittmar & Ohland, 2004). This study defines Transit-oriented development (TOD) as any area of dense, mixed-use, and walkable development around a transit station, and Transit-adjacent development (TAD) as any area of low density, single use, and car-dependent development around a station area.

The most frequently studied factors for distinguishing a TOD from other types of station areas has been residential and employment density (Renne & Ewing, 2013; Kamruzzaman et al., 2015; Laaly, 2014; Pollack et al., 2014; Jeihani & Zhang, 2013; Canepa, 2007; Cervero & Kockelman, 1997; Cervero & Gorham, 1995), land use diversity (Renne & Ewing, 2013; Kamruzzaman et al., 2015; Vale, 2015; Jeihani & Zhang, 2013; Cervero & Kockelman, 1997; Cervero & Gorham, 1995), street network design or street connectivity (Renne & Ewing, 2013; Vale, 2015; Pollack et al., 2014; Laaly, 2014; Ngo 2012; Kamruzzaman et al., 2014; Brown & Werner, 2011; Werner et al., 2010). Recent studies trying to classify TOD and TAD deal with all three factors in the analysis (Renne & Ewing, 2013; Kamruzzaman et al., 2015; Jeihani & Zhang, 2013). There are several ways to distinguish TOD from TAD, such as cluster analysis (Kamruzzaman et al., 2015; Vale, 2015) or scoring system (Jeihani & Zhang, 2013; Laaly, 2014; Pollack et al., 2014; Renne & Ewing, 2013).

Existing studies differentiating TOD from TAD in terms of their performance are limited. First, most studies cover only single or few regions. Although Renne and Ewing (2013) study 54 regions across the US, their point-based system is arbitrarily constructed, and the outcome variable is not comprehensive travel behavior, but only the percentage of people who commute via public transportation. In contrast, the present study includes eight metropolitan areas with varying geographic and socioeconomic conditions in the U.S. to examine various travel outcomes. Second, unlike existing studies relying on straight-line catchment areas (Vale, 2015) or simple scoring systems (Renne & Ewing, 2013), this study utilizes network distance from each station, and cluster analysis. Finally, while Kamruzzaman et al.(2015) uses a robust method of classification, their study analyzes all neighborhoods in a single city, Brisbane, Australia. Instead, we use the station-based approach as a focus of TOD and TAD because we deal with built environments of station areas and their impact on travel behavior, which has more direct implications for planning practice.

TOD and Travel Outcomes

Potential benefits of TOD are multiple from promoting active modes of transportation to improving access to opportunities such as jobs or entertainment, to offering alternative mobility options and affordable housing, to reducing greenhouse gas emissions (Center for Transit-Oriented Development, 2011; Noland et al., 2014). Thus, TOD serves interrelated goals of making communities socially, economically and environmentally more robust and sustainable. In order to achieve these multiple goals, a TOD should first create settings which prompt people to drive less and ride public transit more (Cervero, 2004). The Center for Transit Oriented Development (2010) identifies vehicle miles travelled (VMT) as the key performance measure for TOD. Lower VMT means that people walk, bike, and use transit more and have more transportation options.

An extensive literature indicates that TODs enhance the use of public transport and reduce car usage (Cervero, 1993, 2004; Langlois et al., 2015; Nasri & Zhang, 2014; Olaru & Curtis, 2015; Venigalla & Faghri, 2015). Based on data from 17 TOD projects, Cervero and Arrington (2008) show that residents living in TOD areas are two to five times more likely to commute by transit than their non-TOD counterparts. Nasri and Zhang (2014) find that people living in TOD areas tend to drive less, reducing their VMT by around 21-38%, compared to the residents of the non-TOD areas. Olaru and Curtis (2015) confirm that better biking and pedestrian infrastructure results in higher bike and walk mode shares along with higher transit ridership in TOD precincts.

Cervero (2004) finds evidence that many TOD ridership gains are a result of self-selection – individuals who wish to drive less may select transit-oriented environments. Many studies have found associations between attitudes and travel choices as evidence of residential self-selection (Cao, Mokhtarian, & Handy, 2009; Mokhtarian & Cao, 2008; Handy, 2005). Thus, individuals' attitudes may confound the relationship between the TOD-type urban form and travel choices, and in turn the effect of the built environment on travel may be overestimated (Ewing & Hamidi, 2015).

From the review of 38 empirical studies, Cao et al.(2009) examine nine methodological solutions to self-selection bias: direct questioning, statistical control, instrumental variables, sample selection, propensity score, joint discrete choice models, structural equations models, mutually dependent discrete choice models, and longitudinal designs. Among the methodologies, propensity score matching (PSM) method is highly recommended in a non-randomized observational study (Cao et al., 2009). The propensity score approach has recently been applied in travel behavior research (Boer et al., 2007; Cao, 2010; Cao et al., 2010; Cao & Fan, 2012; Cao & Schoner, 2014), but not in the context of station areas yet. Detailed explanation of the PSM will be presented in the Research Design section.

RESEARCH DESIGN

Study Regions

This study includes eight metropolitan regions meeting two criteria. First, they must have household travel survey data with XY coordinates for households and trip ends. Second, they must have had a rail-based transit system before the survey was conducted. For the eight regions (Table 1), household travel surveys were conducted between 2006 and 2012. In these regions, there are 549 rail-based transit stations according to the national TOD Database (Center for Transit Oriented Development, http://toddata.cnt.org/). Transit types include heavy rail (109 stations), commuter rail (148 stations), and light rail (272 stations). Boston has the greatest number of stations (n=239), followed by Portland (n=94) and Miami (n=50), and Minneapolis-St. Paul has the smallest number (n=20).

TABLE 1 Study Regions and Transit Stations[1)]

NO	REGION	YEAR (SURVEY)	HEAVY RAIL	COMMUTER RAIL	LIGHT RAIL	TOTAL	HOUSEHOLD (½ MILE)
1	Atlanta, GA	2011	38	0	0	38	138
2	Boston, MA	2011	49	121	72	239[2)]	1604
3	Denver, CO	2010	0	0	36	36	152
4	Miami, FL	2009	22	4	24[3)]	50	26
5	Minneapolis-St. Paul, MN	2010	0	4	16	20	97
6	Portland, OR	2011	0	7	87	94	307
7	Salt Lake City, UT	2012	0	1	36	37	115
8	Seattle, WA	2006	0	11	25	35[2)]	16
	Total		109	148	272	549	2455

1) This study includes only transit stations which had opened before the survey.
2) The total number of stations is not equal to the sum of the columns because there are some stations having two or more types of transit systems.

3) Miami's People Mover, an automated guideway transit, is included under the LRT category.

Data

Following the definition of TOD and the literature review, this study includes 'activity density', 'land use diversity', and 'street network design' to classify station area types. For 'density' variable, population and employment data for traffic analysis zones (TAZ) were acquired from regional MPOs and summed to compute an overall activity density per square mile. Activity density is sum of population and employment within the station area, divided by gross land area (Ewing et al., 2015). For 'diversity' variable, we computed an entropy index. [1] Each region provided parcel maps so that we could calculate the proportion of the area of each land use type – residential, commercial, and public – in a ½ mile (about 800 meter) buffer from each station. For the 'street network design' variable, we computed the number of intersections per square mile from street network shapefiles. Because these three built environment variables – activity density, land use entropy, and intersection density - vary in range, we scaled the data by standardizing each variable to a mean of 0 and a standard deviation of 1.

In addition, we measured 'distance to transit' variable as a network distance from a household to the rail station because that might be an important determinant of transit trips. Also, regional accessibility is another important variable to predict travel behaviors (Ewing et al., 2015). That variable is defined as the percentage of jobs that can be reached within 30-minute by transit, which tends to be highest at central locations and lowest at peripheral ones. We used travel time skims and TAZ-level employment data acquired from regional MPOs.

From the travel survey data in eight regions, we calculated vehicle miles traveled (VMT), automobile trips, transit trips, and walk trips by individual households. The survey data include demographic variables such as household size, the number of employed, household income, and the number of personal vehicles per person. The number of total households living within half-mile from stations was 2,455 in the eight regions.

Research Process and Methods

Step 1. TOD/TAD Classification: Cluster Analysis

Because the built environments around transit stations fall within a TOD-TAD spectrum not a simple dichotomous scale, and there is no certain agreement of ideal built environments for TOD, identifying TODs and distinguishing them from TADs could be a difficult but important research step. Cluster analysis has been a preferred method for generating TOD typologies in previous studies (Atkinson-Palombo & Kuby, 2011; Kamruzzaman et al., 2014; Vale, 2015).

Using cluster analysis, this study classifies station area types based on three built environment factors – activity density, land use diversity, and street network design. This approach enables to group existing station areas based on their actual built environment characteristics, rather than theoretical criteria of TOD or TAD. To be specific, this study uses

[1] The entropy index measures balance between three different land uses. The index ranges from 0, where all land is in a single use, to 1 where land is evenly divided among the three uses. Values are intermediate when buffers have more than one use but one use predominates. The entropy calculation is:
entropy= – [residential share * ln(residential share) + commercial share * ln(commercial share) + public share * ln(public share)]/ln(3)
, where ln is the natural logarithm of the value in parentheses and the shares are measured in terms of total parcel land areas (Ewing et al., 2015).

hierarchical clustering algorithm with Ward D2 distance measure. To determine the optimal number of clusters in a data set, this study utilizes "NbClust" package in R 3.3.1 software, which provides 26 validation indices of clustering such as Calinski and Harabasz index and Silhouette index (Charrad et al., 2014).

Step 2. Household Sample Selection: Propensity Score Matching
Propensity score matching (PSM) has been widely used to overcome nonrandom assignment of treatment in the evaluation of social programs (Oakes & Johnson, 2006). Evaluation studies are often based on observational data, in which the assignment of treatment is not random. Accordingly, individuals in the treatment group are likely to differ systematically from those in the control group. For example, households living in suburban regions could be more affluent than their counterparts in downtown, a result of residential self-selection. Therefore, the observed difference in behavioral outcomes between the groups is confounded by residential self-selection. Statistically, it generates a biased estimate of treatment effect.

The propensity score is defined as the conditional probability of assignment to a particular treatment given a vector of observed covariates (Rosenbaum & Rubin, 1984). In the context of TOD and TAD, the treated group is households living in TOD station areas while the control group is those living in either TAD or Hybrid areas. The propensity score matching was implemented in R 3.3.1 using MatchIt package. First, we develop a binary logit model to estimate propensity score using the subsample of households living in TOD (treatment) and TAD (control). We chose household characteristics as independent variables – household size, the number of workers, the number of vehicles per person, household income, distance to nearest transit station, regional job accessibility, and the regions – as potential sources of residential self-selection and confounding factors in travel outcome. Second, we match each household living in TOD with those in TAD based on the propensity score. Caliper length of 0.03 is used for matching, meaning that for a treatment observation, we search a match in control observations whose propensity scores are within 0.03 of the score of the treatment observation (Austin, 2009). Third, we evaluate whether the matched residents in TOD are systematically different from those in TAD. We use t-test to assess whether demographics and locational factors are balanced between the matched groups.

The final goal of PSM is to compute the ''true" impact of TOD/TAD on travel behavior. Once the matching was complete, we calculated the average treatment effects (ATE) of station area type on VMT, transit trips, and walk trips. For the illustration example below, the ATE is computed as the mean travel factors of the matched TOD households minus those of the matched TAD households. The observed influence of living in TOD on travel behavior is same calculation but using the original samples in TOD and TAD before matching.

TOD/TAD CLASSIFICATION

By using the NbClust package in R 3.3.1 software, which generates 26 validation indices of clustering, this study could determine the optimal number of clusters in the data set. As a result, thirteen of the 26 indices suggest that three is the optional number of clusters.

Table 2 shows the result of hierarchical clustering. The first cluster (n=107) is titled as 'TAD' because it has the lowest level of density, diversity, and intersection density. The second and largest one (n=382) is classified as 'Hybrid' which has low level of activity density and intersection density, but highest entropy index. The final cluster (n=60) is named as 'TOD' in terms that it has highest activity density and intersection density, and high level of land use mix level.

TABLE 2 Cluster Analysis Result and Descriptive Statistics

Cluster type	Number of Stations	Activity Density (/sq.mi.)		Entropy Index		Intersection Density (/sq.mi.)	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
TAD	107	10,319	11,751	0.30	0.19	110	58
Hybrid	382	21,210	19,764	0.75	0.15	194	79
TOD	60	135,327	51,025	0.70	0.24	386	110
TOTAL	**549**	**31,559**	**43,821**	**0.66**	**0.24**	**199**	**108**

Sample households were selected as those living within half-mile network distance from stations. We allotted individual households to their nearest stations based on network distance in order to assign the station types. TAD type has 251 households while TOD and Hybrid type have 311 and 1,893 households, respectively (Table 3).

Table 3 shows that households living in TADs have more household members, more workers, more vehicles, and higher incomes than those living in TODs or Hybrids. ANOVA analysis shows that the differences are significant. Regarding travel behavior, TAD households have much higher VMT and auto trips and lower transit and walk trips than those in TODs and Hybrids. The hybrid type is in the middle, except for their lowest household incomes and highest level of transit trips on average.

Post-hoc comparisons using Tukey's Honest Significant Difference (HSD) method show that all three groups are significantly different with each other in vehicle per capita, single-family housing, auto trips, and walk trips variables while only TAD and Hybrid show no significant difference in household size, number of workers, and VMT variables. TOD and Hybrid are not different with each other in terms of transit trips (Table 3).

TABLE 3 Household Characteristics and Travel Behavior by Station Area Types: Average and ANOVA Analysis

Cluster type	No. of Stations	HH samples	HH size	HH workers	Vehicle per cap	HH Income ($1000)	VMT	Auto Trips	Transit trips	Walk trips
TAD	107	251	2.19	1.28	0.66	88.15	21.56	6.06	0.72	1.91
Hybrid	382	1,891	2.15	1.22	0.56	81.04	18.85	4.91	1.46	3.88
TOD	60	313	1.52	0.96	0.48	86.33	15.21	2.04	1.35	4.77
Total	**549**	**2455**	**2.07**	**1.19**	**0.56**	**82.44**	**18.87**	**4.66**	**1.37**	**3.79**
F-statistic (ANOVA)	-	-	39.5 ***	14.6 ***	7.92 ***	2.43 *	6.4 ***	48.1 ***	12.3 ***	29.9 ***
No difference (Tukey HSD)	-	-	TAD-Hybrid	TAD-Hybrid	none	all	TAD-Hybrid	none	TOD-Hybrid	none

***: p<.01, **: p<.5, *: p<.1

HOUSEHOLD SAMPLE SELECTION: PROPENSITY SCORE MATCHING

As shown above, in the context of station areas, households living in TAD tend to be more affluent, have more cars, live in a larger household, and be more auto-oriented than their counterparts in TOD. Residential self-selection theory says, however, that the households living in TAD might live there because they are auto-oriented. Therefore, the true difference in travel outcomes between TOD and TAD is estimated here by matching samples using PSM.

With the explanatory variables - household size, the number of workers, the number of vehicles per person, household income, distance to nearest transit station, regional job accessibility, and the regions, household pairs in three area type pairs (TOD-TAD, TOD-Hybrid, and TAD-Hybrid) are matched. The PSM generates 82 household pairs (164 in total) in TOD-TAD pair, 161 pairs in TOD-Hybrid pair, and 189 pairs in TAD-Hybrid pair.

After matching, we first evaluate whether the chosen residents in one type are systematically different from those in another type. If they are different in terms of demographics, self-selection is still a concern. Table 4 shows differences of household characteristics before and after matching. Unlike unmatched samples, t-test results for matched samples show that residents in TOD and TAD do not differ by all covariates. Those variables are not statistically different in both TOD-Hybrid and TAD-Hybrid pairs as well (results are not shown).

TABLE 4 Mean Differences of Observed Covariates between TOD and TAD in Unmatched and Matched Samples

Variables		Before Matching			After Matching		
		TAD (n=251)	TOD (n=313)	Mean Difference[1)]	TAD (n=82)	TOD (n=82)	Mean difference
Household size		2.19	1.52	0.66***	1.98	1.74	0.23
Number of workers		1.28	0.97	0.32***	1.20	1.09	0.11
Vehicle per capita		0.66	0.48	0.17***	0.60	0.67	-0.07
Household income ($1000)		88.15	86.33	1.81	90.80	80.32	10.48
Distance to station (mile)		0.33	0.28	0.05***	0.32	0.32	0.00
Regional job accessibility		39.95	44.88	-4.93***	45.58	46.36	-0.77
Number of Station Areas by Region	Atlanta	32	12	-	11	12	-
	Boston	75	210	-	27	19	-
	Denver	40	29	-	15	19	-
	Miami[2)]	-	-	-	-	-	-
	Minneapolis	67	23	-	22	23	-
	Portland	31	39	-	7	9	-
	Salt Lake City	6	-	-	-	-	-
	Seattle[2)]	-	-	-	-	-	-

1) ***: p<.01, **: p<.5, *: p<.1 (T-test results)

2) All stations in Miami and Seattle were classified as 'Hybrid' type.

Once the matching was complete, we calculated the average treatment effects (ATE), the observed differences, and the ratio between them on VMT, auto trips, transit trips, and walk trips for each area pair. As an example for TOD-TAD pair, the observed difference is the mean travel factors of all TOD households minus that of all TAD households in the original sample. The ATE is the difference in mean travel factors between the matched samples in TOD and TAD.

From the 3rd to 7th columns, Table 5 shows observed difference in mean in the original sample, ATE in matched sample, ratio of ATE over observed difference, mean value of control group after matching (TAD in the first and third pair and Hybrid in the second pair), and ratio of ATE over control mean. Thus, after controlling for self-selection, on average, TAD households tend to drive 2.36 miles more than TOD residents. The difference in VMT between two groups, however, is not statistically significant, implying that residential location itself may dominate the observed influence of living in TOD on VMT. On the other hand, the mean differences in automobile trips between TAD and TOD households is 3.13, which is highly significant. That is, if a randomly-selected household moves from a TAD to TOD, we expect a decrease in the number of driving by 3.13 trips per day. On overage, the matched sample households in TAD drove 5.48 times per day. Thus, the effect of living in TOD itself represents a 57% decrease in daily auto trips, which is considerable.

In addition, the probability to walking or taking transit significantly decreases from TAD to TOD. For all automobile, transit, and walk trips, the effect of living in TOD accounts for approximately 80% of the observed influence, that is, 20% of the observed difference may result from residential self-selection. The ATE/control ratio of walk trips in TAD-TOD pair (-0.98) means that after accounting for self-selection, walk trips are almost twice in TOD area than in TAD area.

When we compare Hybrid to TOD areas, only the number of automobile trips is high in Hybrid areas, but the difference is less than TAD-TOD pair. The ATE accounts for 31% of the observed influence of living in TOD on daily auto trips, comparing to living in Hybrid. In the case of TAD-Hybrid pair, only the number of walk trips is slightly, but statistically significantly low in Hybrid area. The ATE accounts for 31% of the observed influence of living in Hybrid on daily walk trips, comparing to living in TAD. For both cases, residential self-selection may explain approximately 70% of the observed differences.

TABLE 5 Differences in Travel Behavior between Station Area Types after Matching

Area Type Pair	Travel outcomes	Observed difference	PSM ATE	ATE/observed difference ratio	Mean of Control Group	ATE/control ratio
TAD-TOD n=564(unmatched), n=164(matched)	VMT	6.34***	2.36	.37	18.03	.13
	Auto trips	4.02***	3.13***	.78	5.48	.57
	Transit trips	-.64***	-.56**	.88	0.77	-.73
	Walk trips	-2.86***	-2.28***	.80	2.33	-.98
Hybrid-TOD n=2,204(unmatched), n=322(matched)	VMT	3.64***	-.08	-.02	15.73	-.01
	Auto trips	2.87***	.88**	.31	3.11	.28
	Transit trips	.11	.06	.55	1.58	.04
	Walk trips	-.89***	-.60	.67	4.67	-.13
TAD-Hybrid n=2,142(unmatched), n=378(matched)	VMT	2.70**	3.03	1.12	23.16	.13
	Auto trips	1.15***	.40	.35	7.02	.06
	Transit trips	-.75***	-.28	.37	0.74	-.38
	Walk trips	-1.97***	-.62*	.31	2.05	-.30

***: p<.01, **: p<.5, *: p<.1 (T-test results)

DISCUSSION

The clustering approach in this study classified existing station areas into TOD, TAD, and Hybrid types in terms of built environment factors – density, diversity, and street network design. As a result, 11% of the 549 stations in eight regions were labeled as TOD as being dense, diverse, and walkable. One-fifth were named as TAD as having opposite urban form of TOD. The other

70% of the stations could be classified as Hybrid. Land use mix was a key factor to distinguish TAD from Hybrid while density and intersection density played important roles to differentiate TOD and Hybrid. Statin area types vary among the literature according to classifying methods, factors, and regions. This study has an advantage in terms that we draw the area types from all stations in eight urban areas in the U.S. and utilizes more objective and systematic one – the hierarchical cluster analysis.

Household characteristics and travel behaviors from household travel survey data were matched to each station area type, and this study found that residents living in different types are different with each other. Households in TAD tend to be more affluent, have more cars, live in a larger household, and be more auto-oriented than their counterparts in TOD. Regarding travel behavior, TAD households have much higher VMT and lower walk and transit trips than those in TODs and Hybrids. The average number of daily automobile trips shows the most dramatic differences that TAD households generate the auto trips three times more than TOD households (6.06 vs. 2.04). The big difference in mode share between TOD and TAD (e.g. auto mode shares in TAD and TOD are 68% and 25%, respectively) is observed in other studies (Renne, 2009; Renne & Ewing, 2013), some of which is much less dramatic – approximately 70% (TOD) vs. 85% (non-TOD) in other studies (Kamruzzaman et al., 2013; Jeihani & Zhang, 2013).

In this study, propensity score matching enables the researcher to match samples so as to control for residential self-selection. Although the differences in travel outcomes become less dramatic after controlling self-selection, the matched sample still shows that TOD motivates its residents to walk more and take transit more while using personal vehicles less. On the other hand, non-significant difference between TOD and TAD in VMT means that TOD does not make the personal vehicle trips shorter, but fewer. This implies that in TOD, there might be still needs of long trips such as commuting, but more destinations within walking distance might encourage residents to choose walking or transit instead of driving.

By considering the in-between hybrid type, this study could give practical implications. The result shows that only walk trips are significantly different between TAD and Hybrid, and only auto trips are significantly different between TOD and Hybrid. For example, when a local government and transit authority develop a TAD-type station area which is sprawled, single-use, and not walkable into a Hybrid type mainly by adding different land uses, they could expect an increase in internal walk trips. Then a Hybrid type of station area could be changed into a TOD type by adding density and decreasing block sizes, which would result in less driving by their residents. Then the cumulative change from TAD to TOD could encourage its residents to drive less, walk more, and take transit more, which will have positive impacts on the city's environment, society, and economy.

CONCLUSION

Transit-oriented development is expected to minimize auto-dependency and maximize ridership of its residents. Also, higher mode share by walking and biking is another goal of TOD. This study demonstrates that TOD and TAD are different with each other in terms of not only its urban form but also its impacts on travel outcomes. After controlling for residential self-selection effect, TOD motivates its residents to walk more and take transit more while using personal vehicles less. In addition, TOD makes the personal vehicle trips fewer, but not shorter, implying that more destinations within walking distance in TOD could encourage its residents to choose walking or transit instead of driving the short distance.

This study has mainly three limitations. First, station area classification might generate different results if you change the input – e.g. if you include different regions or different built environment factors. The result depends on the clustering method as well. However, the clustering approach in this study reflects reality better than using hypothetical benchmarks defining TOD and TAD.

Second, propensity score matching works only when all confounding factors are included in the analysis. This study, however, only includes the factors reflecting self-selection indirectly, which are household demographic characteristics and location factors while not having residents' attitude information. The risk of not controlling all confounding factors is that we might under- or over- estimate the effect of residential self-selection on travel behavior. To our knowledge, there is no such attitude data covering multiple regions in the U.S., but the result of this study needs to be checked its external validity by additional TOD studies including residential preference data in specific regions.

Third, in theory, the observed covariates in the propensity score equation are measured before the treatment while the outcome is measured after the treatment (Rosenbaum & Rubin, 1983). In the context of this study, the data point for household characteristics and location factors needs to be before the station area was developed, while the travel outcome data should be collected after the development. This requires longitudinal data. However, because the regional household travel surveys are conducted in different years in each region, it is not plausible to put all longitudinal data into one analysis. Although this study uses cross-sectional data to control the temporal differences across regions and stations, further research needs more advanced methods.

Nevertheless, as a first-of-its-kind research using both cluster analysis and propensity score matching in TOD/TAD classification, this study provides an evidence that a TOD and even a Hybrid type of station area could encourage its residents to use more active modes of transportation. An effort to create transit-oriented neighborhood does not have to be a 'mega-project.' Gradual changes of a station area into denser, more diverse, and more walkable environment would compensate us in the form of sustainable travel behavior, which gives more environmental, social, and economic benefits ultimately.

ACKNOWLEDGEMENT

This research was funded by the National Institute for Transportation and Communities (NITC), a program of the Transportation Research and Education Center at Portland State University and a U.S. Department of Transportation university transportation center.

REFERENCE

Atkinson-Palombo, C. and M. J. Kuby. The geography of advance transit-oriented development in metropolitan Phoenix, Arizona, 2000–2007. *Journal of Transport Geography,* Vol. 19, No. 2, 2011, pp. 189-199.

Austin, P. C. Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo simulations. *Biometrical Journal,* Vol. 51, No. 1, 2009, pp. 171-184.

Belzer, D. and G. Autler. *Transit Oriented Development: Moving from Rhetoric to Reality*. Washington, DC: Brookings Institution Center on Urban and Metropolitan Policy. 2002.

Bernick, M. and R. Cervero. *Transit villages in the 21st century*. McGraw-Hill, New York, 1997

Boer, R. et al. Neighborhood design and walking trips in ten US metropolitan areas. *American journal of preventive medicine,* Vol. 32, No. 4, 2007, pp. 298-304.

Brown, B.B., and C.M. Werner. The residents' benefits and concerns before and after a new rail stop: Do residents get what they expect? *Environment and Behavior,* Vol. 43, 2011, pp. 789–806.

Canepa, B. Bursting the bubble: Determining the transit-oriented development's walkable limits. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1992,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 28-34.

Cao, X. J., P. L. Mokhtarian, and S. L. Handy. Examining the impacts of residential self-selection on travel behaviour: a focus on empirical findings. *Transport Reviews,* Vol. 29, No. 3, 2009, pp. 359-395.

Cao, X. J. Exploring causal effects of neighborhood type on walking behavior using stratification on the propensity score. *Environment and Planning A,* Vol. 42, No. 2, 2010, pp. 487-504.

Cao, X. J. Z. Xu, and Y. Fan. Exploring the connections among residential location, self-selection, and driving: Propensity score matching with multiple treatments. *Transportation research part A: policy and practice*, Vol. 44, No. 10, 2010, pp. 797-805.

Cao, X. J. and Y. Fan. Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B: Planning and Design*, Vol. 39, No.3, 2012, pp. 459-470.

Cao, X. J. and J. Schoner. The influence of light rail transit on transit use: An exploration of station area residents along the Hiawatha line in Minneapolis. *Transportation Research Part A: Policy and Practice,* Vol. 59, 2014, pp. 134-143.

Center for Transit-Oriented Development. *Performance-Based Transit-Oriented Development Typology Guidebook.* Report TOD 204, 2010.

Center for Transit-Oriented Development. *TOD 204 Planning for TOD at the Regional Scale: The Big Picture*. Report TOD 204, 2011.

Cervero, R., *Ridership Impacts of Transit-Focused Development in California. Institute of Urban and Regional Development*, University of California, Berkeley, CA (Monograph 45), 1993.

Cervero, R., and R. Gorham, R. Commuting in transit versus automobile neighborhoods. *Journal of the American Planning Association*, Vol. 61, No. 2, 1995, pp. 210-225.

Cervero, R., and K. Kockelman. Travel demand and the 3Ds: density, diversity, and design. *Transportation Research Part D: Transport and Environment*. Vol. 2, No. 3, 1997, pp. 199–219.

Cervero, R. and M. *Duncan Residential Self Selection and Rail Commuting: A Nested Logit Analysis.* Berkeley: UCTC Working Paper, 2002.

Cervero, R. *Transit-Oriented Development in the United States: Experiences, Challenges, and Prospects,* TCRP Report 102. Transportation Research Board, 2004.

Cervero, R. and G. B. Arrington. Vehicle trip reduction impacts of transit-oriented housing. *Journal of Public Transportation,* Vol. 11, No. 3, 2008, pp.1-17.

Charrad, M., N. Ghazzali, V. Boiteau, A. Niknafs, and M. M. Charrad. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software,* Vol. 61, No. 6, 2014, pp.1-36.

Dittmar, H. and G. Ohland. *The new transit town : best practices in transit oriented development,* London: Island Press, Washington, D.C., 2004

Ewing, R., G. Tian, J. P. Goates, M. Zhang, M. J. Greenwald, A. Joyce, ... and W. Greene. Varying influences of the built environment on household travel in 15 diverse regions of the United States. *Urban Studies*, Vol. 52, No. 13. 2015, pp. 2330-2348.

Ewing, R., S. Hamidi, and J. B. Grace. Urban sprawl as a risk factor in motor vehicle crashes. *Urban Studies*, Vol. 53, No. 2, 2016, pp. 247–266.

Handy, S., *Does the Built Environment Influence Physical Activity? Examining the Evidence. Critical Assessment of the Literature of the Relationships Among Transportation, Land Use and Physical Activity,* Transportation Research Board Special Report 282, 2005.

Jeihani, M., L. Zhang, A. Ardeshiri, A. Amiri, A. Nasri, K. R. Zamir, and B. Baghaei. *Development of a Framework for Transit-Oriented Development (TOD),* No. MD-13-SP209B4N, 2013

Kamruzzaman, M., D. Baker, S, Washington, G. Turrell. Residential dissonance and mode choice. *Journal of transport geography*, No. 33, 2013, pp. 12–28.

Kamruzzaman, M., D. Baker, S. Washington, and G. Turrell. Advance transit oriented development typology: case study in Brisbane, Australia. *Journal of Transport Geography*, No. 34, 2014, pp. 54-70.

Kamruzzaman, M., F. M. Shatu, J. Hine, and G. Turrell. Commuting mode choice in transit oriented development: Disentangling the effects of competitive neighbourhoods, travel attitudes, and self-selection. *Transport Policy*, No. 42, 2015, pp. 187-196.

Laaly, S. New definition of transit oriented development (TOD) based on context sensitive paradigm. Doctoral dissertation, Morgan State University, 2014.

Langlois, M., D. van Lierop, R. A. Wasfi, and A. M. El-Geneidy. Chasing Sustainability: Do New Transit-Oriented Development Residents Adopt More Sustainable Modes of

Transportation?. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2531,* Transportation Research Board of the National Academies, Washington, D.C., 2015, pp. 83-92.

Mokhtarian, P. L. and C. Xinyu. Examining the impacts of residential self-selection on travel behavior: A focus on methodologies. *Transportation Research Part B: Methodological,* Vol. 42, No.3, 2008, pp. 204-228.

Nasri, A. and L. Zhang. The analysis of transit-oriented development (TOD) in Washington, D.C. and Baltimore metropolitan areas. *Transport Policy,* Vol. 32, No. 1, 2014, pp. 172–179.

Ngo, V. Identifying Areas for Transit-Oriented Development in Vancouver. *Trail Six: An Undergraduate Journal of Geography*, No. 6, 2012.

Noland, R.B., K. Ozbay, S. Dipetrillo, and S. Iyer. *Measuring Benefits of Transit Oriented Development*. Mineta National Transit Research Consortium Report 12-18, 2014.

Oakes, J. M., and P. J. Johnson. Propensity score matching for social epidemiology. *Methods in social epidemiology,* No. 1, 2006, pp. 370-393.

Olaru, D. and C. Curtis. Designing TOD precincts: accessibility and travel patterns. *Planning Practice and Research*, Vol. 15, No. 1, 2015, pp. 1567-7141.

Pollack, S., A. Gartsman, M. Boston, A. Benedict, and J. Wood. Rating the Performance of Station Areas for Effective and Equitable Transit-Oriented Development, in Transportation Research Board 93rd Annual Meeting (No. 14-0392), 2014.

Renne, J. *Transit Oriented Development: Measuring Benefits, Analyzing Trends, and Evaluating Policy*, Ph.D. Dissertation, Rutgers University, 2005.

Renne, J. L. and R. Ewing. *Transit-Oriented Development: An Examination of America's Transit Precincts in 2000 & 2010.* UNOTI Publications Paper 17, 2013.

Rosenbaum, P. R., and D. B. Rubin. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association,* Vol. 79, No. 387, 1984, pp. 516-524.

Vale, D. S. Transit-oriented development, integration of land use and transport, and pedestrian accessibility: Combining node-place model with pedestrian shed ratio to evaluate and classify station areas in Lisbon. *Journal of Transport Geography*, No. 45, 2015, pp. 70-80.

Venigalla, M. and A. Faghri. A Quick-Response Discrete Transit-Share Model forbTransit-Oriented Developments. *Journal of Public Transportation,* Vol. 18, No. 3, 2015, pp. 107-123.

Werner, C.M., B. B. Brown, and J. Gallimore. Light rail use is more likely on ''walkable'' blocks: Further support for using micro-level environmental audit measures. *Journal of Environmental Psychology,* No. 30, 2010, pp. 206–214.

U.S. Bureau of Labor Statistics. *Consumer Expenditure Survey*, 2014

Zamir, K.R., A. Nasri, B. Baghaei, S. Mahapatra, and L. Zhang. Comparative Analysis of the Effect of Transit-Oriented Development on Trip Generation, Distribution, and Mode Share in the

Washington, D.C. and Baltimore Metropolitan Areas. *Transportation Research Board 2014 Annual Meeting*, 2014, pp. 1-19.