Causality vs. Correlation: Rethinking Research Design in the Case of Pedestrian Environments and Walking [1]

Zhan Guo
Robert F. Wagner Graduate School of Public Service
Rudin Center for Transportation Policy and Management
New York University
Room 3038, 295 Lafayette St,
New York, NY 10012, United States
Tel: +1 212 998 7510; Fax: +1 212 995 4162
E-mail address: zg11@nyu.edu
http://wagner.nyu.edu/guo

Summary

This paper investigates the causal effect of pedestrian environments on walking behavior and focuses on the issue of research design. The paper differentiates between two types of research designs: treatment-based and traveler-based. The first approach emphasizes the variation of the treatment (pedestrian environments), and generally compares distinct neighborhoods, such as urban vs. suburban or transit-oriented vs. auto-dependent. The second approach emphasizes the homogeneity of subject (pedestrians), and aims at the same pedestrian under different environments normally due to home relocation, or the improvement of pedestrian environments. The first approach can easily identify a correlation between the pedestrian environment and walking, but proving it causal is a challenge. The second approach may not even find a correlation, but if it does, such a correlation is more likely to be causal. Which approach is better depends on whether the first approach can effectively control for the unobservable personal heterogeneity, and whether the second approach can find sufficient variation in the pedestrian environments experienced, arguably, by the same person.

Most studies used the first approach but produced inconsistent results in terms of whether self-selection exists and if it does, whether it nullifies a causal relationship. This paper supports the second approach but argues that the few existing studies failed to capture sufficient variation of pedestrian environments in their research design. It then follows a traveler-based research design, and proposes a new method based on pedestrians' path choice. By comparing the preference from the same pedestrian towards multiple walking paths with different pedestrian environments, this research is able to control the personal heterogeneity while still retain a sufficient variation in the pedestrian environments, thus represents a quasi-experimental design. It is able to do so because the path-based measure is sensitive enough to capture even minor differences in the environment experienced by pedestrians. More importantly, path choice is less likely to correlate with job and housing location choices, and therefore largely avoids the self-selection problem.

In the empirical analysis, the paper targets subway commuters' egress path choice from a station to their workplaces in downtown Boston. The results confirm a causal relationship and suggest that the pedestrian environment can significantly affect a person's walking experience. The perceived utility change of walking, due to the sidewalk amenities, averages between 21 and 31 percent. In several street segments in downtown Boston, walking is actually perceived to have a positive utility instead of as a derived demand. Methodological issues regarding the method and generalizability of the findings are also discussed.

Keywords: causality, research design, path choice, pedestrian environment, walking, Boston

[1] Published as "Does the Built Environment Affect the Utility of Walking? A Case of Path Choice in Downtown Boston" in *Transportation Research D: Transport and Environment*, Vol. 14, pp. 343-352

1. Introduction

The causal relationship between the built environment (BE) and travel behavior has been a heated topic in the planning literature over the past ten years. The motivations behind such research are clear: researchers have largely confirmed a correlation between BE and travel behavior (Boarnet and Crane 2001), and causality is seemingly the next step of exploration, but establishing causality is nevertheless a more difficult task. More importantly, a proven causal relationship is critical for justifying proactive policies like smart growth and active living that have been adopted in many communities (Bhat and Guo 2007).

The core issue in identifying causality is to prove that no third factor creates a spurious relationship between BE and travel behavior (Handy et al. 2006). The oft-cited spurious effect entails self-selection, or households who select residential locations at least partly based on their travel preferences. For example, people who prefer to walk choose to live in a neighborhood with a pedestrian-friendly environment. If this is true, behavioral differences are largely caused by differences in travel preferences, not BE, and BE-based policies are unlikely to produce the expected outcomes.

Despite various methods used to address self-selection, there is still lack a consensus on whether self-selection actually exists and, if it does, whether it nullifies a causal relationship between BE and travel (Cao et al. 2006). This problem is partly caused by the inappropriate research designs in many studies. Theoretically, in order to prove causality, an experimental design should be adopted, whereby the difference in experimental units between the control and treatment groups is eliminated through random sampling. However, in reality, empirical studies often choose to enlarge the difference of treatments (BE) instead of minimizing the difference of experimental units (travelers). They often *first* select neighborhoods with distinct BEs (e.g., an urban vs. a suburban neighborhood), and *then* compare the travel behavior of their residents. Because residents in distinct neighborhoods often come from distinct social groups, their behavioral differences might be caused by unobservable personal differences like attitudes and preferences towards travel. In other words, such a design induces a self-selection problem by sacrificing causality for correlation.

This research contributes to the literature in two ways. First, it follows an alternative research design that first selects *similar* travelers and then "*puts*" them in different BEs to examine their behavior differences. This approach resembles the experimental design, and is more likely to identify a causal relationship than is the existing method. Second, the research investigates BE impacts on transit commuters' walking path choices in a non-residential neighborhood. Compared with modal choice, path choice is less likely to correlate with job and housing location choices. Compared to the zone-based BE measure, the path-based measure is more sensitive to BE differences even within the *same* neighborhood. Therefore, this research is able to avoid the self-selection problem while still secure sufficient variation in BEs for statistical investigation.

The rest of this paper is structured as follows. Section 2 compares the treatment-first and traveler-first research designs, and proposes a new method based on the latter. Section 3 presents the path choice approach in the context of downtown Boston. Section 4 describes data and variables. Section 5 presents the empirical analysis and interprets the results. Section 6 discusses policy implications and concludes.

2. Rethinking Research Design: Treatment vs. Traveler

In a metropolitan region, residents are often grouped based on their social status, ethnic background, lifestyle preferences, and other factors. Distinct groups tend to live in different neighborhoods, and distinct neighborhoods often comprise different social groups. Such reality poses a methodological challenge to causal BE-travel research: comparisons across distinct neighborhoods raise the question of self-selection, while comparisons among similar residents, who often live in neighborhoods with similar BEs, result in a lack of "treatment" in research design. The first design can easily identify a correlation between BE and travel, but proving that this correlation is causal is a challenge. The second design may not even find a correlation, but if it does, such a correlation is more likely to be causal. Which method is better depends on whether the first design can effectively control for unobservable personal differences, and whether the second design can secure sufficient variation in the treatment (BE).

2.1 Treatment (BE)-First Design

Following this design, studies tend to target neighborhoods with distinct BEs, such as urban vs. suburban neighborhoods, transit-oriented vs. auto-dependent neighborhoods, conventional vs. neo-traditional neighborhoods, or diverse neighborhoods in a metropolitan region (Table 1).

Various methods have been used to control for unobservable personal differences and self-selection. Cervero and Duncan (2002) used a joint location and travel choice model and found that self-selection accounts for 40% of the probability of making a decision to commute by rail. Using a different joint-choice model, Bhat and Guo (2007) did not find any evidence of self-selection and confirmed that BE affects automobile ownership. Using instrumental variables, Greenwald and Boarnet (2001) concluded that certain characteristics of BE do promote walking, even while taking into account the possibility of self-selection. Khattak and Rodriguez (2005) drew a similar conclusion using survey data from Chapel Hill, NC. However, using a similar method, Boarnet and Sarmiento (1998) found that predicted BE measures were not significantly related to individual non-work auto trip frequency after taking into account the influence of self-selection. Using structural equations, Bagley and Mokhtarian (2002) found that when attitudinal, lifestyle, and socio-demographic variables are accounted for, neighborhood type has little influence on travel behavior. Based on the same method for the same region, Cao et al. (2007) found that BE still affects auto ownership, even after controlling for self-selection. After incorporating personal attitudes and preferences into their analysis, Handy et al. (2006) showed that self-selection at least partially explained walking behavior, but that BE still had an impact as well. Following a similar method, however, Chatman (2005) found self-selection to be a negligible factor, as those with strong mode preferences seemed to be less sensitive to BE variables, while those with weak mode preferences seemed to be more responsive to differences in BE.

Such inconsistency is especially striking given that these studies are based on similar methods and sometimes the very same dataset, as well as on the same metropolitan region during similar times (Table 1). It is hard to believe that such inconsistency reflects different self-selection styles at the metropolitan level. Rather, the inconsistency is likely caused by inefficiency in the studies' methods to control for unobservable personal differences, especially given the limits of current BE measurement, available data, and statistical models.

2.2 Traveler-First Design

The main requirements of the second method are (1) to find travelers with similar attitudes and preferences towards travel, and (2) to secure sufficient variation in the BEs experienced by travelers. A few studies following this method have produced consistent results showing that causality exists despite of the “threat” of self-selection. Such research designs include longitudinal research, intervention design, and matched attitude (Table 1).

Longitudinal research examines the BE impact on movers’ travel behavior before and after a move. Targeting the same person before and after a treatment is surely an advantage for a causality study, but the longitudinal design also raises two questions: First, if a move is caused by a preference change, the mover self-selects and is therefore no longer essentially the “same person” before and after moving. Second, many movers relocate close to their initial locations (Krizek 2003), meaning that the variation of BE before and after the move might not be large enough to allow a statistical analysis. Sometimes researchers must combine movers and non-movers together in their model estimation, which essentially becomes a cross-sectional research design (Krizek 2003).

Intervention design investigates residents’ behavioral change before and after a major BE investment in a neighborhood. Because most residents remain the same, this method solves the residential self-selection problem, but still raises similar questions as in longitudinal design: First, the investment might be self-selected. For example, a neighborhood that is more favorable to pedestrian activities might be more likely to request an improvement, and thus more likely to get it. Studies on the Safe Routes to School projects in California indicated that investments were more effective in school areas with already moderate or high levels of walking, but are insufficient to affect travel modes in schools with previously low levels of walking or bicycle travel (Boarnet et al. 2005). Second, BE improvements are often marginal (such as improved sidewalks and crosswalks, bike paths, traffic signals, speed bumps, and other improvements), and thus may not be strong enough to induce behavioral changes (Meurs and Haaijer 2001).

The matched attitude method finds individuals with similar preferences towards travel but who live in neighborhoods with distinct BEs. If they travel differently, such a difference should be largely caused by a BE difference. This approach is best represented by Schwanen and Mokhtarian’s study of dissonant residents (2004, 2005). They compared dissonant urban residents with consonant suburban residents, and found the former used cars less frequently than did the latter but more frequently than did the consonant urban residents. However, the method assumes that dissonant urban residents share similar travel and living preferences to consonant suburban residents, which is not always true. Being unsatisfied with urban living does not mean these residents want to live in suburban neighborhoods.

Despite these various methods, traveler-first design tends to produce consistent results that confirm the causal impact of BE on travel behavior. This design’s merit might not be an accident. The same logic is used in other study areas with the similar methodological concerns. For example, when researchers investigate how the environment affects children’s behavior and achievement, researchers often use twins (Horwitz et al. 2003) or siblings (Aaronson 1998) from the same household to control for unobservable personal and household attributes.

The comparison between these two designs suggests that controlling for unobservable personal differences is more difficult than is finding BE differences between similar travelers. Unfortunately, only a very few studies adopted the traveler-first design, and their particular methods are not without question. This study follows this design logic, but proposes a new method based on path choice to untangle the causal impact of BE on travel behavior.

3. Methodology: Path Choice

Path choice investigates travelers' decisions over multiple path options between the same origin and destination. It has been widely used in transportation planning to assign flows within a transportation network, but has rarely been used in a BE-travel study (Lee and Moudon 2006). The path choice approach has a great advantage for solving the self-selection problem, but it is also technically very challenging, which prohibits its application in broad areas.

3.1 Path Choice: Advantages and Challenges

The biggest advantage of path choice is that it is less likely to correlate with residential location decisions. Other travel decisions, like mode choice, VMT, trip frequency, and car ownership, are long-term lifestyle decisions and are therefore associated with housing location choice. However, path choice is a sub-level decision given location choice--fixed origin and destination. It is hard to believe that the path one chooses from a subway station to a workplace is correlated with either housing or job location choices.

Unfortunately, path choice is very difficult to model for two reasons. First, researchers only know the chosen path, not alternative paths considered by a traveler when the decision is made (Bovy and Stern 1990). These alternative paths must be generated based on various decision rules, which is very difficult and hard to validate (Hoogendoorn-Lanser 2005). Second, these multiple paths often overlap with one another because they begin at the same origin and end at the same destination, violating the assumption of Independent and Identical Distribution (i.i.d.) for discrete choice models (Ben-Akiva and Lerman 1985). The overlap can be very complicated and makes the correcting effort extremely difficult.

To fully address the two issues is beyond the scope of this research. Instead, a simplified sub-path choice is adopted, which avoids the modeling hardship while still retaining the merit of the path choice approach.

3.2 Sub-path Choice: Egress from Public Transit

A path often consists of a hierarchical structure, i.e., freeways, arteries, and collectors in a road network, as well as feeder and long-haul services in a public transit network. The sub-path choice models the path decision at the access or egress segment, conditional on the mode and service selected by the traveler for the trunk portion. Figure 1 illustrates such a situation.

A, B, and C are stations on two separate subway lines, and D is the trip destination. Suppose a traveler enters the area on Line 1 from the south. When the traveler reaches station A, he or she has two options to get to destination D, which is closer to station C. The traveler can leave the system at A and walk to D, or can continue traveling on Line 1 to Line 2, transfer at B, exit at C, and then walk a shorter distance to D. Therefore, the traveler has two possible egress sub-paths: ABCD or AD. Which path is better depends on four path attributes: extra in-vehicle time spent on path ABC, transfer convenience at station B, the street walking time saved (AD - CD), and the pedestrian environment (PE) along AD and CD.

Such a sub-path choice situation is simple--we do not need to define the attributes of the full path from the origin to the destination. Alternative paths are easy to identify because they are attached to a few transit lines. The problem of path correlation is greatly reduced because ABCD and AD are unlikely to overlap with one another.

3.3 Commuting Trips to Urban Center

The above situation is set for commuting trips to an urban center rather than home trips to suburban neighborhoods for two reasons. First, it avoids the possible correlation with residential or destination location choice. A walking-conducive neighborhood may have more walking path options within the neighborhood. Shoppers may choose their stores based on the attractiveness of a walking path. Setting the sub-path choice in a non-residential neighborhood for commuters solves both problems. Workers are unlikely to choose their working place based on the attractiveness of the walk path to the location.

Second, such a setting benefits path choice modeling: the high ridership in an urban center guarantees a large sample size; concentrated transit stations and job locations provide ample egress path options; the PE is diverse; and there is no competing egress mode other than walking[2].

3.4 Summary of Methodology

The sub-path choice approach meets the two requirements of the traveler-first design. First, it is reasonable to assume that subway commuters to an urban center share a similar attitude and preference towards travel. Second, when the PE is measured along a path, it is more sensitive to PE differences experienced by pedestrians than is the traditional measure at the zone level (e.g., neighborhood, census tract, TAZ). The zone-based approach often measures the PE incompletely and disproportionally: Only a portion of the path is measured, and short-distance trips are weighted more than long-distance trips if the zone size is fixed. In other words, the zone-based approach measures the “treatment” differently for different “experimental units,” which violates the principles of experimental design.

This proposed method focuses on the pedestrian environment (PE) and walking because walking is an indispensible part of most modes of travel; it is universally available to almost everybody; it exposes people directly to the BE; and it is the most common and preferred form of physical activity for the general population (Badland and Schofield 2005).

A subway survey in Boston reveals that multiple egress options exist, and passengers indeed made the trade-off between the two sub-paths, ABCD and AD. Among the 6,500 subway trips ending in downtown Boston, half of them have the ABCD option, and one third of this group finally chose the sub-path ABCD over AD to get to their destinations.

4. Boston Case Study: Data and Variables

Downtown Boston is a 1.5-square-mile area on the Boston peninsula including Back Bay, Beacon Hill, the North End, the financial district, and other neighborhoods. The cityscape ranges from traditional colonial style to contemporary art, with plenty of parks, bars, churches, stores, schools, restaurants, and street performances (in good weather). Twenty-one subway stations, including four major transfer stations (Park Street, Government Center, Downtown

[2] In downtown Boston, 98% of subway riders walk to their destination, only 1.5% take buses after leave the system.

Crossing, and State Street), are concentrated in this area, and about half of all trips in this area are on foot (Boston Foundation 2004).

The main datasets used in this analysis are the most recent subway on-board surveys conducted by the Massachusetts Bay Transportation Authority (MBTA), the Boston 1996 Assessor Parcel Database, and the Road Inventory Database from the Massachusetts Highway Department (MHD). For each of the 2,748 commuting trips ending in this area, four types of variables are created for the two paths (ABCD and AD): transit attributes, street walking time, trip and personal characteristics, and the PE. Transit attributes include transfer walking and waiting time, extra in-vehicle time, and escalator presence during transfer along the segment ABC. Street walking is assumed to follow the shortest path in the street network (with a speed of three miles per hour), which is a dominant characteristic of pedestrian movement within a city (Zacharias 2001). Weather is another factor that may affect the path decision (Guo and Wilson 2007), but is not included due to data unavailability.

The 2,748 observations contain more females and young professionals with slightly higher incomes than the average subway commuter has. A great number of demographic and trip characteristics are investigated (age, driver's license, auto availability, occupation, gender, household size, income, car ownership, trip time, purpose, frequency, and fare types), and, as expected, most of these are insignificant in model estimations.

For PE attributes, only the design variables (streetscape, sidewalk, planting, open space, and so on) are considered. Distance-based variables [3] like density, mixed use, and accessibility are not considered because they may be heavily correlated with location choices (housing, job, firm, and so on). Five design variables were selected because others either were not recorded in existing datasets or had little variation in the study region. The five variables are: the density of pedestrian-friendly parcels (retail, bookstores, and so on) along the path, average sidewall width along the path, the density of intersection along the path, and two dummy variables on open space (Boston Common) and topography (Beacon Hill). For detailed descriptions of these variables, see Guo and Ferreira (2008).

The descriptive statistics for all variables are shown in Table 2.

5. Model Development: Capturing the Utility of the Pedestrian Environment

A subway commuter always chooses the path with a higher utility between ABCD and AD. Their path utilities are determined by the transit and walking experiences (for path ABCD), or solely the walking experience (for path AD). The walking experience is further determined by two factors: path length (walking time) and the PE along the path. If the PE affects the path choice, it does so by affecting the utility of walking along either AD or CD, and then the causal relationship between PE and walking is justified. The follow-up question would be how much the utility of walking is affected by the PE because, if the magnitude is small, such a causal relationship does not make sense for policy intervention, even if it is statistically significant. Therefore, the investigation results in two questions:

[3] Density, the most widely used measure of the built environment, is largely a proxy for average distances. Mixed use measures a set of collective distances to different activities. Accessibility combines travel impedance with spatially distributed activities. Because travel impedance is often measured as a function of distance, accessibility is actually a more general term for mixed use.

Question 1: Does the PE affect the utility of walking? Alternatively, are PE variables statistically significant in the path choice model after controlling for other variables?

Question 2: Is the causal effect significant enough to justify policy intervention? Or, what is the percentage change in walking utility caused by the PE?

The two questions are elaborated in mathematical form below. Suppose paths ABCD and AD are recoded as paths 1 and 2, respectively, and path 2 (AD) is treated as the base, then the modeling structure takes a binary logit form (Ben-Akiva and Lerman 1985):

$$P_n(1) = \frac{e^{-\mu V_{1n}}}{e^{-\mu V_{1n}} + e^{-\mu V_{2n}}} \quad (1)$$

$$V_{1n} = C + \alpha S_{1n} + \beta W_{1n} + \gamma PE_{1n} + \delta K_{1n} \quad (2)$$

$$V_{2n} = \beta W_{2n} + \gamma PE_{2n} \quad (3)$$

where $P_n(1)$ is the probability of person n selecting path ABCD, V_{1n} and V_{2n} are the systematic components of the utility for paths ABCD and AD for person n, μ is the positive scale parameter, C is the alternative specific constant for path ABCD, S_{1n} is a vector of transit characteristics for path ABCD for person n, W_{1n} and W_{2n} are egress walking times for path ABCD and AD for person n, PE_{1n} and PE_{2n} are vectors of PE variables along the path of CD and AD for person n, K_{1n} is a vector of all other attributes for path ABCD for person n, and α, β, γ and δ are the coefficients to be estimated. All these coefficients are vectors, except β.

When the PE effect is not considered, the utility of walking is solely determined by walking time. The initial utility of walking may either increase or decrease after the PE is included, depending on PE quality. The utility change can be measured as the equivalence of walking time by calculating the coefficient ratio between walking time and the PE variables. Denoting the initial utility of AD or CD for street segment i as U_i^0, and the combined utility as Ui^1, then the mathematical forms are

$$U_i^0 = W_i \quad (4)$$

$$U_i^1 = W_i + \frac{\gamma}{\beta} PE_i \quad (5)$$

$$R_i = \frac{U_i^1 - U_i^0}{U_i^0} \quad (6)$$

where W_i is the walking time on pedestrian path i, and PE_i are the PE variables along path i. R_i is the percentage of walking utility change due to the PE effect, indicating the magnitude of the impact for each pedestrian path i.

Two types of models are developed with and without the PE variables, and the estimation results are summarized in Table 3.

5.1 Modeling Results

After other variables are controlled, all five PE variables are still significant at the 5% level with expected signs. The improvement test from the base to the expanded model is highly significant ($P < 0.0001$), which suggests that the PE indeed affects the utility of walking. The higher the density of PFPs, and the wider the sidewalk along the path CD relative to the path AD, the more likely a commuter is to choose CD. If there is a higher density of intersections along the path CD, riders are more likely to take that path, which implies an “encouraging” instead of a “deterring” effect of intersection density on walking. Similarly, Boston Common has an encouraging effect, while Beacon Hill has a deterring effect on walking.

Next, the magnitude of the effect is quantified using equations 4-6. For example, on an AD path, one more PFP parcel per 100 meters increases the utility of walking by an equivalence of 0.5 minutes of walk [4], while the utility increase from one more intersection per 100 meters is 0.3 minutes. If the average sidewalk width increases by 6 feet, the walking utility also increases by 0.5 minutes. If the path passes through Boson Common, while the alternative path does not, the utility increases by 2.9 minutes. However, if the path passes through the hilly Beacon Hill, the utility decreases by 3.5 minutes. In other words, a pedestrian is willing to walk 2.9 minutes longer if the path is through Boston Common, but 3.5 minutes shorter if it is through Beacon Hill, all else being equal. The result suggests that the PE affects walk utility primarily through attractions (parks) and barriers (hilly topology and sidewalks). Parcel types and intersection density play an important but smaller role in path choices.

The overall PE effect is calculated from all five PE variables, and the percentage of utility change is computed. For example, for passenger ID=14301, her AD walk is 6.9 minutes, while the PE along this walk path has a total utility equivalent to 2.5 minutes. Therefore, the combined utility for path AD is 6.9-2.5 = 4.4 minutes, a 36 % increase. For the whole sample, the average changes are 2.4 (AD) and 2.8 (CD) minutes, representing 33 % (AD) and 21% (CD) of the utility increase. Figure 2 shows the distribution of such utility change for all 2,748 passengers. Note a few passengers actually have a negative value, which suggests that their utility of walking decreases due to the “bad” PE along their path. A screen of these trips shows that they all pass through Beacon Hill. A few other passengers at the far right side of the distribution experience an increase in utility equal to or larger than the actual walking time, which suggests that the PE along their walk paths is good enough to cancel out the time spent on walking and generate a positive utility. In other words, walking is not a derived demand but itself has a utility for these passengers. These are normally short-distance trips along Boylston Street, a commercial street with plenty of amenities for pedestrians.

In summary, the answers to questions 1 and 2 are; Yes, the PE does indeed affect the utility of walking, and the effect is significant enough to call for policy intervention, at least in downtown Boston.

5.2 Discussion

---

[4] Since time is a disutility, the increased utility should be deducted from the original walking time to get the perceived utility after the PE effect is considered.

Several concerns about the method remain. First, self-selection may still exist at the path level. For example, if a group of people lives in a particular neighborhood, or work in a particular downtown district served by a particular subway line, where ABCD is more convenient than AD due to the network configuration, then these people will always choose path ABCD. In this case, the path choice decision is correlated with housing and job location choices, which would either over- or underestimate the PE effect on walking. To check for this possibility, the demographic attributes by line and destination zone in downtown Boston were examined, but no systematic pattern was identified. Most destination zones are well served by all four lines, except for Back Bay, which is better served by the Green and Orange Lines. Therefore, the correlation between demographic groups, employment zones, and the subway network is not evident, and the self-selection concern at the path level is cleared.

Second, the shortest path assumption from a subway station to a destination may not be true. To relax this assumption, a group of destinations that do not involve street directional changes from a station (such as D in Figure 1) were examined, and the results did not contradict the original finding. Another assumption is the uniform walking speed of 3 mile per hour for all pedestrians. Unless walking speed is correlated with the PE, which we know little about empirically and theoretically, this assumption is unlikely to result in a systematic bias in the estimation result.

Finally, there are two issues regarding the generalizability of findings. On the one hand, the adopted research design tends to underestimate the PE impact on walking: the PE effect would have been stronger if the comparison were across different neighborhoods, used a more comprehensive measure of the PE, or covered non-commuting trips. On the other hand, the PE impact is likely to be weaker in a modal choice than in a path choice. For example, the elasticity of sidewalk width is approximately 0.20 in this research, but only 0.09 in a prior study based on modal choice (Cervero and Kockelman 1997). Case studies in other metropolitan regions for other travel types are necessary to provide a complete picture of the causal relationship.

6. Conclusion

This paper examines the causal effect of the pedestrian environment (PE) on the utility of walking. It describes the merits and weaknesses of two types of research designs used in prior studies: treatment-first and traveler-first research designs. The first focuses on enlarging the difference of treatment (the built environment) by comparing travel behavior across *distinct* neighborhoods, while the second focuses on finding *similar* travelers and then exposing them to different built environments. The treatment-first design can easily find a correlation, but is unable to prove that this correlation is causal. The traveler-first design may not find a correlation, but if it does, that correlation is more likely to be causal. Which method is better depends on whether the first design can effectively control for travelers' attitudinal difference toward travel, and whether the second design can find enough differences in the built environment for statistical analysis. Inconsistent results from studies based on the treatment-first design suggest that the traveler-first design might be more effective to untangle the causal effect of the built environment on travel.

This research utilizes the travel-first research design by targeting a group of people with similar attitudes toward travel--subway commuters--and exposing them to different PEs along multiple path options. The path choice approach avoids the residential self-selection problem while still being able to capture minor differences in the PE experienced by pedestrians. The

Boston case study confirms that PE indeed *affects* the utility of walking: "good" PE can increase the utility of walking while "bad" PE reduces this utility. The average effect in downtown Boston is equivalent to 2.4 – 2.8 minutes of walking, which represents a 21% to 33 % increase in walking utility. This effect is significant enough to justify policy intervention.

The findings and methodology have important implications for transit and pedestrian planning. First, improved PE will attract more people to public transit and non-motorized modes of transportation like walking. When the utility of walking increases, the willingness to walk a longer distance also increases, which increases the length, frequency, and mode share of walking and enlarges the catchment area of a transit system. Second, the path-based approach could improve the efficiency of pedestrian planning. By measuring the utility of the PE in monetary terms at the street segment level, this approach helps planners identify which streets should be improved, what improvements are needed, and how investments could be justified. This approach could also make pedestrian planning better positioned in the current funding system by improving its project evaluation practice, which heavily relies on accident reduction and additional users while mainstream transportation evaluation methods are based on time savings for all users.

Reference

Aaronson, D. (1998) Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes. *Journal of Human Resources*, 33 (4), 915-946

Badland, H., Schofield, G. (2005) Transport, urban design, and physical activity: an evidence-based update. *Transportation Research Part D* 10, 177–196

Bagley, M., Mokhtarian, P. (2002) The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *The Annals of Regional Science* 36, 279-297

Ben-Akiva, M., Lerman, S, R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA

Besser, L. M., Dannenberg, A. L. (2005) Walking to public transit: steps to help meet physical activity recommendations. *American Journal of Preventive Medicine* 29 (4) 273-280.

Bhat, C., Guo, J. (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B* 41, 506–526

Boarnet, M. G., Sarmiento, S. (1998) Can land use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics. *Urban Studies* 35 (7), 1155-1169.

Boarnet, M., Crane, R. (2001) *Travel by Design: The Influence of Urban Form on Travel*. Oxford University Press, New York, NY

Boarnet, M., Day, K., Anderson, C., McMillan, T., Alfonzo, M. (2005) Can street and sidewalk improvements enhance walking and bicycling to school? evaluating California's Safe Routes to School Program. *Journal of the American Planning Association* 71 (3), 301-317.

Boston Foundation (2004) *Boston Indicator Project (BIP): Boston Transportation Trends.* http://www.livablestreets.info/files/mos/stats.pdf

Bovy, P. and Stern, E. (1990) *Route Choice: Wayfinding in Transport Networks*. Kluwer Academic Publishers, Norwell, MA

Cao, X., Mokhtarian, P., Handy, S. (2006) *Examining the impacts of residential self-selection on travel behavior: methodologies and empirical findings*. Working Paper, Institute of Transportation Studies, University of California, Davis

Cao, X., Mokhtarian, P. L., Handy, S. L. (2007). Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership. *Environment and Planning A* 39 (4), 830-847

Central Transportation Planning Staff (CTPS) (1994) *MBTA Rapid-Transit system-wide passenger survey*, Massachusetts Bay Transportation Authority (MBTA), Boston.

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

Cervero, R. and Duncan, M. (2002). *Residential self-selection and rail commuting: A nested logit analysis*.Working paper #604, University of California Transportation Center. Available at http://www.uctc.net/papers/604.pdf, accessed on March 18, 2007

Chatman, D. (2005) *How the built environment influences non-work travel: theoretical and empirical essays*. Dissertation, University of California, Los Angeles.

Cubukcu, E., Nasar, J. L. (2005) Influence of physical characteristics of routes on distance cognition in virtual environments. *Environment and Planning B: Planning and Design*. 32, 777-785.

Dongjoo, P. and Rilett, L. (1997) Identifying Multiple and Reasonable Paths in Transportation Networks: A Heuristic Approach. *Transportation Research Record* 1607, 31-37.

Ewing, R., Cervero, R. (2001) Travel and the built environment: a synthesis. *Transportation Research Record* 1780, 87-114.

Handy, S. (1996) Urban form and pedestrian choices: A study of Austin neighborhoods. *Transportation Research Record* 1552, 135-144.

Handy, S. (2005) Critical assessment of the literature on the relationships among transportation, land use, and physical activity. *TRB Special Report 282: Does The Built Environment Influence Physical Activtiy? Examining the Evidence*.

Handy, S., Cao, X., Mokhtarian, P. (2006) Self-selection in the relationship between the built environment and walking: empirical evidence from northern california. *Journal of the American Planning Association* 72 (1),55-76.

Hess, P., Vernez Moudon, A., Snyder, M., Stanilov, K. (1999) Site design and pedestrian travel. *Transportation Research Record* 1674, 9–19.

Hoogendoorn-Lanser, S. (2005) *Modeling travel behavior for multi-modal transport networks*, TRAIL Thesis Series T2005, TRAIL,UT Delft, The Netherlands

Howard, C., Burns, E.K. (2001) Cycling to work in Phoenix route choice, travel behavior, and commuter characteristics. *Transportation Research Record* 1773, 39-46

Giles-Corti, B., Donovan, R. (2002) Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive Medicine* 35, 601–611.

Giles-Corti, B., Melissa, H., Broomhall, M., Collins, C., Douglas, K., Ng, K., Lange, A., Donovan, R. (2005) Increasing walking how important is distance to, attractiveness, and size of public open space? *American Journal of Preventive Medicine* 28, 2S2 169-176.

Greenwald, M., and Boarnet, M. (2001) The built environment as a determinant of walking behavior: Analyzing non-work pedestrian travel in Portland, Oregon. *Transportation Research Record* 1780, 33-42.

Guo, Z., Ferreira, J. (2008) Pedestrian environments, transit path choice, and transfer penalties: understanding land use impacts on transit travel. *Environment and Planning B*

Guo, Z., Wilson, N. (2008) The Impact of Weather on Transit Ridership in Chicago. *Transportation Research Record* 2034, 3-10

Guo, Z., Wilson, N. (2004) Assessment of the Transfer Penalty: A GIS-based Disaggregate Modeling Approach. *Transportation Research Record* 1872, 10-19

Khattak, A. J., Rodriguez, D. (2005) Travel behavior in neo-traditional neighborhood developments: A case study in USA. *Transportation Research Part A* 39, 481-500.

Kristen, D. (2006) Active living and social justice. *Journal of the American Planning Association* 72 (1), 88-99

Krizek, K. (2003) Residential relocation and changes in urban travel: does neighborhood-scale urban form matter? *Journal of the American Planning Association* 69 (3), 265–281.

Lee, C., Moudon, A. (2006) Environmental correlates of walking for transportation versus recreation purposes. *Journal of Physical Activity and Health* 3 (1), S99-S117.

Levine, J. (1998) Rethinking accessibility and jobs-housing balance. *Journal of the American Planning Association* 64 (2), 133-149

Meurs, H. and Haaijer, R. (2001) Spatial structure and mobility. *Transportation Research Part D*, 6 (6), 429-446.

Ewing, R., Cervero R. (2001) Travel and the Built Environment: A Synthesis. *Transportation Research Record* 1780, 87-114

Rodriguez, D., Joo, J. (2004) The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D* 9, 151–173.

Rodríguez, D., Khattak, A., Evenson, K. (2006) Can new urbanism encourage physical activity? *Journal of the American Planning Association* 72 (1), 43-54.

Schwane, T., Mokhtarian, P. (2004) The extent and determinants of dissonance between actual and preferred residential neighborhood type. *Environment and Planning B: Planning and Design* 31, 759-784

Schwane, T., Mokhtarian, P. (2005) What if you live in the wrong neighborhood? The impact of residential neighborhood type dissonance on distance traveled. *Transportation Research Part D* 10, 127–151
Horwitz, A., Videon, T., Schmitz, M., Davis, D. (2003) Rethinking twins and environments: Possible social sources for assumed genetic influences in twin research. *Journal of Health and Social Behavior* 44 (2), 111-129
Zacharias, J. (2001) Pedestrian behavior and perception in urban walking environment. *Journal of Planning Literature* 16, 2 -18

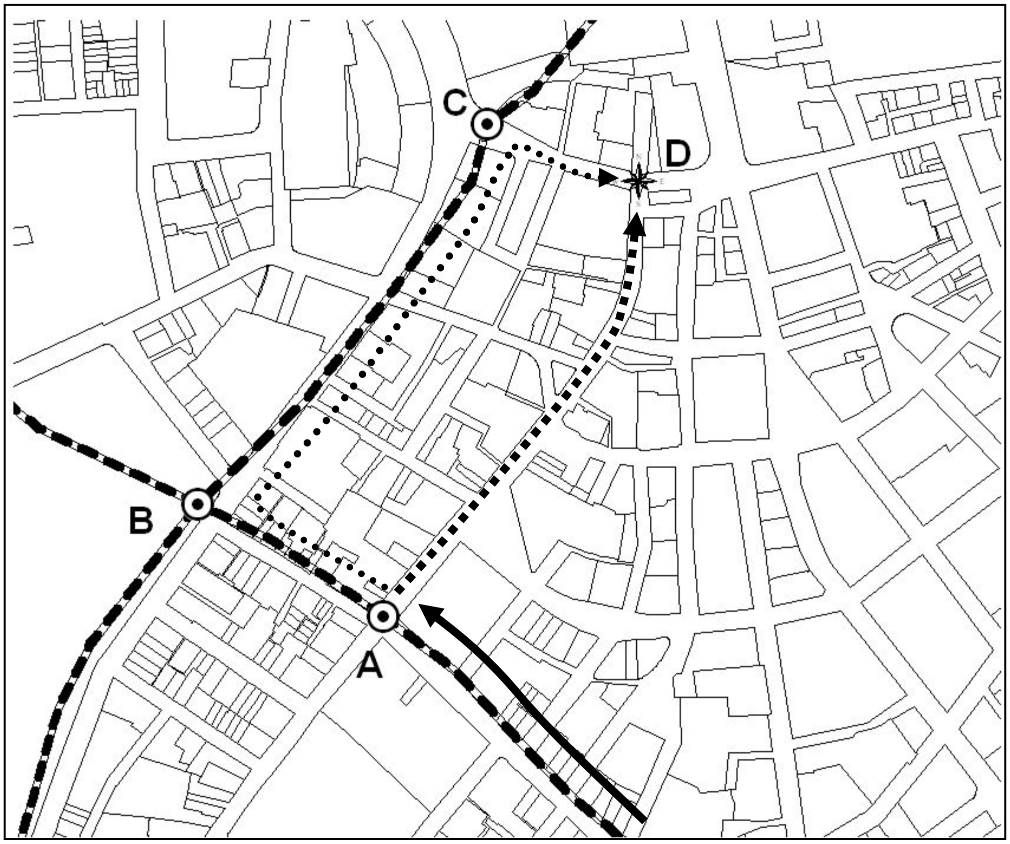

Figure 1 Illustration of Two Egress Paths from Subway to a Destination

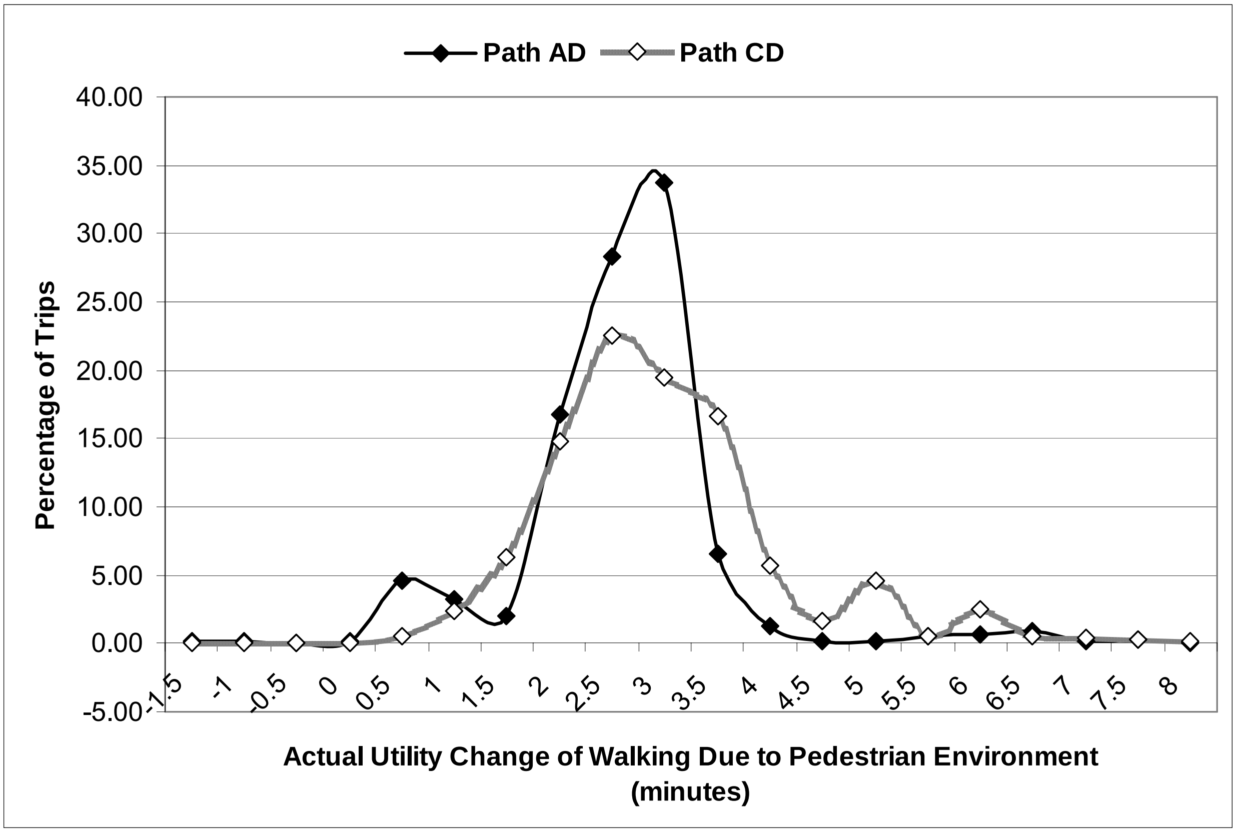

Figure 2 Distribution of the Impact of Built Environments on the utility of Walking

Methods	Studies	Year	Metro Region	Data	Travel Behavior	Causality Exist?
Treatment-First Design						
Instrumental Variables	Greenwald and Boarnet	2001	Portland, OR	ODOT Travel Survey (1994)	Walking frequency	**Yes** (only at neighborhood not zip code level)
	Khattak and Rodriguez	2005	Chapel Hill and Carrboro, NC	Self-conducted household travel survey (2004)	Modal choice, length, frequency	**Yes** (self-selection exists)
	Boarnet and Sarmiento	1998	Los Angeles, CA	Panel Study of S. CA Commuters (1990-1994)	# auto trips and VMT	**No** (except for non-work auto trip frequency)
Joint-choice Models	Cervero and Duncan	2002	San Francisco, CA	Bay Area Travel Survey (2000)	Car ownership, modal choice	**Yes** (self-selection exists)
	Bhat and Guo	2007	San Francisco, CA	Bay Area Travel Survey (2000)	Car ownership	**Yes** (self-selection not exist)
Attitude Control	Handy et al.	2006	San Francisco, CA	Travel Survey by Handy et al. (2003)	Walking and biking frequency	**Yes** (self-selection exists)
	Chatman	2005	San Francisco, CA San Diego, CA	Travel Survey by Chatman (2004)	Modal choice and VMT	**No** (self-selection not exist)
Structure Equation	Bagley and Mokhtarian	2002	San Francisco, CA	Travel Survey sponsored by CA Air Resources Board (1992)	Mileage by car, transit, & walk/bike	**No** (self-selection exists)
	Cao et al.	2007	San Francisco, CA	Travel Survey by Handy et al. (2003)	Car ownership	**Yes** (self-selection exits)
Traveler-First Design						
Longitudinal	Krizek	2003	Seattle, WA	PSTP Panel Survey (1989-1998)	VMT and frequency	**Yes**
	Meurs and Haaijer	2001	Netherlands	Dutch Time Use Study (1990-1999)	# of trips by modes	**Yes**
Intervention Design	Boarnet et al.	2005	CA Statewide	Field survey by authors	Walking/biking count	**Yes**
Matched Attitude	Schwanen and Mokhtarian	2004, 2005	San Francisco, CA	Travel Survey by authors (1998)	Distance traveled overall & by mode	**Yes**

Table 1 Summary of Prior Studies on Self-selection and Causality Investigation

Variables	Mean	Std Dev	Min	Max
Egress Walking Time (Both Paths)				
Surface Walking Time (Transfer path) (seconds)	193	139	0.1	935
Surface Walking Time (Non-transfer path) (seconds)	597	385	58	2380
Transfer Characteristics (Transfer Path)				
Transfer Walking Time (seconds)	59	52	1	217
Transfer Waiting Time (seconds)	112	62	44	270
Extra In-vehicle Time (seconds)	216	157	67	964
Escalator presence	0.18	0.39	0	1
Pedestrian Environment Attributes (Both Paths)				
# of PFPs per 100 meters (non-transfer path)	0.49	0.44	0	2.6
# of PFPs per 100 meters (transfer path)	0.58	0.94	0	13.8
Average Sidewalk Width (non-transfer path) (feet)	21.6	8.9	2.5	56.8
Average Sidewalk Width (transfer path) (feet)	18.9	8	0	59
# of Intersections per 100 meters (non-transfer path)	1.6	0.6	0.5	3.5
# of Intersections per 100 meters (transfer path)	2.2	2.2	0	25.8
Pass through Boston Common (non-transfer path)	0.26	0.44	0	1
Pass through Boston Common (transfer path)	0.03	0.17	0	1
Pass through Beacon Hill (non-transfer path)	0.31	0.46	0	1
Pass through Beacon Hill (transfer path)	0.01	0.08	0	1

Table 2 Descriptive Statistics of Independent Variables

Variables	Base Model		Expanded Model	
	Coefficient	Robust t	Coefficient	Robust t
Walking Time (Both Paths)				
Walking Time (minutes) (both options)	-0.31	-17.2	-0.28	-15.4
Transfer Attributes (Transfer Path)				
Transfer Constant	-0.92	-4.7	-1.19	-5.0
Transfer Walking Time (minutes)	-1.26	-12.5	-1.08	-9.4
Transfer Waiting Time (minutes)	-0.14	-2.6	-0.08	-1.2
Extra In-vehicle Time (minutes)	-0.22	-7.4	-0.23	-7.8
Government Center Station (dummy)	-1.39	-7.2	-1.43	-7.1
Escalator presence (dummy)	0.37	2.3	0.48	3.3
Trip and Demographic (Transfer path)				
Round Trips (dummy)	-0.14	-1.1	-0.14	-1.1
Own =< 1 car (dummy)	0.43	2.8	0.43	2.7
Household >= 3 Members (dummy)	0.20	1.8	0.23	2.0
Pedestrian Environments (Both Paths)				
PFP density (# of PFPs per 100 meters)			0.13	2.0
Average sidewalk width (feet)			0.02	2.7
Beacon Hill (dummy)			-0.90	-3.7
Boston Common (dummy)			0.65	3.2
Intersection Density (# of intersections per 100 meters)			0.11	2.4
Summary Statistics				
N	2748		2748	
Final log likelihood	-1102.26		-1073.69	
Adjusted ρ^2 (pseudo adjusted R square)	0.421		0.436	
Model Improvement Test: (From Base to Expanded Model)	χ^2 = -2 (-1102.26–(-1073.69)) = 57.14, df =5, prob. < 0.0001			

Table 3 Base and Expanded Path Choice Models