CATALYSTS AND MAGNETS
BUILT ENVIRONMENT EFFECTS ON BICYCLE COMMUTING

Jessica Schoner\textsuperscript{1,2} (corresponding author)
(612) 412-4273
schon082@umn.edu

Xinyu Cao\textsuperscript{1}

David Levinson\textsuperscript{2}

\textsuperscript{1} Humphrey School of Public Affairs
University of Minnesota
301 19\textsuperscript{th} Avenue South
Minneapolis, MN 55455

\textsuperscript{2} Department of Civil Engineering
University of Minnesota
500 Pillsbury Drive SE
Minneapolis, MN 55455

6223 + 1 figures + 3 tables

August 1, 2013
ABSTRACT
What effects do bicycle infrastructure and the built environment have on people’s decisions to commute by bicycle? While many studies have considered this question, commonly employed methodologies fail to address the unique statistical challenge of modeling such a low mode share. Additionally, self selection effects that are not adequately accounted for may lead to overestimation of built environment impacts.

This study addresses these two key issues by using a zero-inflated negative binomial model to jointly estimate participation in and frequency of commuting by bicycle, controlling for demographics, residential preferences, and travel attitudes. The findings suggest a strong self selection effect and modest contributions of bicycle accessibility: that bicycle lanes act as “magnets” to attract bicyclists to a neighborhood, rather than being the “catalyst” that encourages non-bikers to shift modes. The results have implications for planners and policymakers attempting to increase bicycling mode share via the strategic infrastructure development.
INTRODUCTION

The relationship between bicycling and the built environment, particularly dedicated bicycle lanes and trails, has captivated the attention of researchers and planners for decades. In a state of the practice and research needs paper, Porter et al. (1) identified critical questions about the role of bicycle infrastructure: how to forecast use of new facilities, how to estimate mode shift due to building new facilities, and how these new facilities may affect mobility, congestion, and air quality (1). Despite many advances in the field, their questions about the impacts of infrastructure are still salient today.

Common strategies for researching and evaluating transportation projects fail to address the nuances of bicycling. The utility of bicycling, more so than any other mode, is strongly affected by weather phenomena and day-to-day variation in travel needs, such as hauling cargo or goods. As a consequence, many bicyclists are in fact multi-modal travelers (2). Distinguishing between participation and frequency is critical for being able to model impacts of bicycling (2).

Because bicycling has such a small mode share, standard survey and data collection strategies, especially those that assume people tend to stick to a single mode throughout the week such as the American Community Survey (ACS), underestimate its prominence. Many also employ research design strategies that skew the sample in favor of people who are already prone to bicycling, producing coefficients that are not accurate for modeling behavior among the general population.

What effect do bicycle infrastructure and the built environment have on people’s decisions to commute by bicycle, and are some people more inclined to be “bikers” than others? This study explores the gap in research spanning both participation in bicycling and frequency of bicycle commuting, with aims of expanding the understanding of bicycling and the built environment. Existing survey data from Minneapolis, MN and a Zero-Inflated Negative Binomial Regression (ZINB) model are employed to jointly estimate participation in and frequency of bicycle commuting as a function of the built environment, controlling for demographics, residential preferences, and travel attitudes.

This research is significant because, while the magnitude and direction of the coefficients are consistent with other studies, the unique structure of the zero-inflated model provide deeper insight to the relationships between individual preferences, self-selection, and the built environment. When interpreted in this framework, it is easy to identify ways of harnessing the residential self-selection effect to increase rates of bicycling.

The extent to which bicycling infrastructure acts as a “catalyst” to induce mode shift among non-bicyclists to biking is unknown, given the difficulty of establishing causality in cross-sectional studies (4). However, the evidence of a self-selection effect suggests that certain infrastructure types function as “magnets” for people who are already prone to bicycling for work, due to their demographic, residential preference, and travel attitude profiles. Combined with evidence from variables used to predict frequency after controlling for residential self-selection, the findings from this study can be used to locate new bicycling infrastructure strategically for providing housing choices for current and would-be bicyclists, and maximizing the number of bicycle trips they choose to make.

This paper is organized as follows: Section 3 reviews literature about how studies manage low numbers of bicyclists among the general population. Section 4 describes the survey administration, data, and modeling procedure. Section 5 presents findings from using a zero-inflated negative binomial model to predict participation in and frequency of bicycle commuting among urban residents. Finally, Section 6 concludes with a discussion of the implications of this research.
LITERATURE REVIEW
Complications of Modeling Bicycling

Bicyclists, more so than other mode users, are distinctly “multi-modal” (2). More so than driving and even transit, bicyclists are vulnerable to day-to-day changes in weather or varying travel needs. Having to make additional stops, carry groceries or other bulky items, or travel when it is dark all decrease the utility of bicycling. Many bicyclists, therefore, can be thought of as “part-time” bicyclists.

This phenomenon results in conventional survey questions underestimating bicycling. Surveys that ask about a single primary commute mode, such as the ACS, miss people who bike only 1-2 days per week, or only for non-work purposes. These questions tell us how many people are bicycling frequently, but do not tell us how many people are biking infrequently and how many trips this translates to.

Surveys that ask what mode was used “yesterday” in theory should average out over the whole population to a representative value of the amount of bicycling being done, but this assumes bicyclists choose their biking days randomly and that sample sizes are large enough to reflect the ground truth of bicycling. With small sample sizes and such a low mode share, these types of questions have low chances of catching a part-time bicyclist on their biking days.

Much of the literature on bicycling employs binary logit models that predict who is a bicyclist in any capacity, and do not tell us how much bicycling is actually being done. Heinen et al. (2) describes this problem in this way: “It is of interest to distinguish between (1) mode choice in general, that is to say, the bicycle is at least one of the modes used; and (2) daily choice, in terms of frequency. The latter is useful because many bicycle commuters choose not to cycle every day.”

Bicycling as a Small Mode Share

Bicycling represents a relatively small mode share, particularly for commuting. In the United States, the ACS estimates that only 0.51% of commuters use a bicycle as their primary commuting mode. While the average is higher when focusing on central cities (0.95% in all Principal Cities, and 3.86% in the City of Minneapolis), the overall rates are still extremely low relative to driving, and even other so-called “alternate” modes such as transit.

A review of the literature on bicycling behavior and the built environment found three distinct strategies for modeling bicycling, given the low mode share. A few studies employed multiple strategies in the same paper. Many studies did not directly address the low mode share concern. The resulting five categories, explained more fully in subsequent sections, are:

1. Inclusion Criteria
2. Strategic Over-sampling
3. Hybrid Inclusion Criteria and Strategic Over-sampling
4. Statistical Distributions
5. No technique
Table 1 summarizes the studies reviewed in each of these five categories. Some studies appear multiple times in the table because the paper includes several components employing one or more of the three techniques.

**Inclusion Criteria**

Inclusion criteria studies apply a filter to general population data to extract a subset sample that applies to their research question. Many, but not all, of the studies select people based on past bicycling behavior (e.g., having bicycled within the past year) or expressed willingness to bicycle. The effect is that these studies model bicycling behavior among a subset of people already expected to have some propensity to bicycle. The results may not be easily extrapolated to the general population.

The clearest example of this technique comes from Wardman et al. (6). Within their large sample of census travel diary records and originally collected survey data, they found that about 60% of people indicate that they would “never contemplate switching to cycling”, so these participants were removed from the dataset. The authors explain, “A model is hardly needed to predict the behavior of such individuals and their actual choices or SP responses would provide little information for modelling purposes.” The authors also filter out trip records from individuals with commutes greater than 7.5 miles (about 12 km) because they were only interested in commute trips where bicycling is a viable option. The authors found that dedicated, separate infrastructure was positively and strongly associated with bicycle commuting. However, the results may be exaggerated by the removal of respondents who would not bicycle under any circumstance because their value of time by mode would be heavily skewed to favor time in the car.

Winters et al. (7), Handy and Xing (8), and Xing et al. (9) all used a similar technique for screening out dedicated non-cyclists (7, 8, 9). Winters et al. (7) considered anyone with access to a bicycle and who has cycled within the past year a “current cyclist”, and anyone who indicates willingness to cycle in the future as a “potential cyclist” (7). Handy and Xing (8) and Xing et al. (9) used a stricter definition, only including people who had bicycled within the past year (8, 9).

Xing et al. (9)’s and Handy and Xing (8)’s studies use the same data about six small cities in California and Oregon and “past year” inclusion criterion (8, 9). Xing et al. (9)’s binary logit model of utilitarian bicycling found that physical environment measures, including shorter distances and more safe destinations, were associated with both a greater share of biking for transportation. While one would expect longer trips to correlate with more mileage biking, the authors explained that long trips are a barrier to making the trip via bicycle in the first place, so the increased distance is offset by reduced probability of making the trip and (presumably) frequency. This finding underscores the importance of understanding both whether people bicycle and how much (duration and/or frequency) they bicycle.

**Strategic Sampling**

Another common strategy for bicycle research and evaluation is to sample deliberately to capture a greater than average proportion of cyclists. Oversampling strategies range from subtle, such as pre-selecting geographies expected to have higher than average rates of bicycling (10, 11), to deliberate, such as snowball sampling bicycle clubs and local bike shops (12) or bicyclist intercept surveys (13).

Like the inclusion criteria studies, over-sampling runs the risk of measuring effects on a concentrated population of people already prone to bicycling. Additionally, the outreach method
for contacting bicyclists heavily biases the types of bicyclists who respond. Specifically, bike club members are more likely to fit the “fearless” category in Geller (14)’s framework (14). These cyclists may be less affected by built environment characteristics than occasional cyclists because they are comfortable bicycling in mixed traffic. For modeling the effects of dedicated bike infrastructure on mode shift or bicycling behavior among infrequent cyclists, this strategy may not be appropriate.

Moudon et al. (11) and Sener et al. (12) model frequency of bike commuting, though their oversampling strategies and model structures differ substantially (11, 12). Moudon et al. (11) pre-selects geographies using Geographic Information System (GIS) that are expected to be conducive to bicycling. They then administered a general population telephone survey via random digit dialing within these geographies. The geography selection component facilitates capturing a higher number of bicyclists than average, while the administration ensures that within those geographies, the sample is relatively representative. Where Moudon et al.’s sampling strategy was subtle, Sener et al. (12) took the opposite approach. The researchers administered the survey using a snowballing technique, sending an online link to bicycling clubs, posting the link in local bike shops, and purchasing ads in local papers. The effort captured bicyclists in over 100 cities in Texas (USA), but clearly is not a representative sample of the general population.

One final strategy for oversampling is the intercept survey. Survey administrators reach out specifically to bicyclists while they are biking, either by stopping them along a trail or corridor, or by attaching a paper survey to parked bicycles. Hunt and Abraham (13), Hunt (15), and Thakuriah et al. (16) all make use of this strategy (13, 15, 16).

Combined Inclusion Criteria and Strategic Sampling
Some studies made use of both of the aforementioned strategies. They first administered a survey using strategic sampling, and then screened their participants using inclusion criteria to focus on a subset of bicyclists. Heinen et al. (17) and Heinen et al. (18) sample employees in cities with high rates of bicycling, as described above (17, 18). However, in some of their analyses, they also filter out non-cyclists. Heinen et al. (17) constructs one binary logit model of whether the participant bike commutes Full-time (FT) or Part-time (PT). Heinen et al. (18) followed up with PT cyclists from the original study periodically over the course of a year to survey them about how they commuted on that particular day. This set of models (Binary Logit, Generalized Estimating Equation (GEE), and Random Coefficient Analysis (RCA)) attempt to explain day-to-day factors that affect a PT cyclist’s choice of mode. From this study, the authors conclude that bicyclists are distinctively multi-modal.

Like Akar and Clifton (10), Rodríguez (19) surveys affiliates of a university campus to oversample bicyclists (10, 19). They further filter their results by selecting only respondents within certain municipal boundaries that are considered “close enough" to be bicycling distance.

Statistical Techniques
While relatively uncommon among the literature on bicycling and the built environment, statistical techniques can be used to address the relatively low mode share for bicycling among general population surveys. Buehler (20) models bicycling for any given commute trip in a large regional travel diary survey using Rare Events Logistic Regression (RELogit) (20). For these individual commute trips, the authors found strong associations with bicycle facilities provided at work, including bike parking. Free car parking was negatively associated with making the commute trip by bicycle.
<table>
<thead>
<tr>
<th>Citation</th>
<th>Data Source</th>
<th>Technique</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Specific Technique Used - General Population</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cao et al. (21)</td>
<td>Original Survey</td>
<td>-</td>
<td>SURE(^1) - Bike/Walk Frequency</td>
</tr>
<tr>
<td>Krizek and Johnson (22)</td>
<td>Regional Survey</td>
<td>-</td>
<td>Logit - Bike trip(s) in travel diary</td>
</tr>
<tr>
<td>Parkin et al. (23)</td>
<td>Census</td>
<td>-</td>
<td>Logit - Bike commute share</td>
</tr>
<tr>
<td><strong>Inclusion Criteria to Select Bicycling Subset</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handy and Xing (8)</td>
<td>Original Survey</td>
<td>Biked within past year</td>
<td>Logit - Primary bike commute</td>
</tr>
<tr>
<td>Winters et al. (7)</td>
<td>Census &amp; Survey</td>
<td>Current/Potential Bicyclists</td>
<td>MNL(^6) - Mode Choice</td>
</tr>
<tr>
<td>Parkin et al. (23)</td>
<td>Census</td>
<td>-</td>
<td>Multilevel Logistic - Bike (vs. car) trip</td>
</tr>
<tr>
<td>Xing et al. (9)</td>
<td>Original Survey</td>
<td>Biked within past year</td>
<td>Logit - Utilitarian v. Recreation Biking</td>
</tr>
<tr>
<td>Xing et al. (9)</td>
<td>Original Survey</td>
<td>Biked within past year</td>
<td>OLS(^2) - Log-miles of Utilitarian Bike</td>
</tr>
<tr>
<td><strong>Strategic Survey to Oversample Bicyclists</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Akar and Clifton (10)</td>
<td>Original Survey</td>
<td>University Affiliates</td>
<td>MNL - Mode Choice</td>
</tr>
<tr>
<td>Heinen et al. (17)</td>
<td>Original Survey</td>
<td>High biking cities</td>
<td>Logit - Bike commute</td>
</tr>
<tr>
<td>Hunt and Abraham (13)</td>
<td>Original Survey</td>
<td>Bicyclists</td>
<td>Logit - SP experiment</td>
</tr>
<tr>
<td>Moudon et al. (11)</td>
<td>Original Survey</td>
<td>Suitable geography</td>
<td>Logit - Biking at least weekly</td>
</tr>
<tr>
<td>Sener et al. (12)</td>
<td>Original Survey</td>
<td>Bicyclists</td>
<td>OLogit(^3) - Bike commute frequency</td>
</tr>
<tr>
<td>Thakuriah et al. (16)</td>
<td>Original Survey</td>
<td>Bicyclists</td>
<td>Binary GMM(^4) - Former captive car user</td>
</tr>
<tr>
<td><strong>Inclusion Criteria &amp; Strategic Survey</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heinen et al. (17)</td>
<td>Original Survey</td>
<td>High biking cities &amp; Cyclists</td>
<td>Logit - FT vs. PT Bike Commute</td>
</tr>
<tr>
<td>Heinen et al. (18)</td>
<td>Original Survey</td>
<td>High biking cities &amp; PT Cyclists</td>
<td>GEE/RCA Logit - Mode Choice</td>
</tr>
<tr>
<td>Rodríguez (19)</td>
<td>Original Survey</td>
<td>City &amp; University Campus</td>
<td>MNL, Nested, &amp; HEV(^5) - Mode Choice</td>
</tr>
<tr>
<td><strong>Statistical Techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Buehler (20)</td>
<td>Regional Survey</td>
<td>RELogit</td>
<td>RELogit - Bike commute</td>
</tr>
</tbody>
</table>

\(^1\)Seemingly Unrelated Regression Equations (SURE)  
\(^2\)Ordinary Least Squares Regression (OLS)  
\(^3\)Ordered Logistic Regression (OLogit)  
\(^4\)Generalized Mixed Model (GMM)  
\(^5\)Heteroscedastic Extreme Value Model (HEV)  
\(^6\)Multinomial Logistic Regression (MNL)
Methodology Review
While use of specific statistical techniques for modeling low bicycle mode share is infrequent, examples from related fields suggest that these techniques are an opportunity for bicycling research.

Handy et al. (24) and Schoner and Cao (25) use Negative Binomial regression to estimate frequency of utilitarian and recreational walking. The Negative Binomial model relaxes the Poisson distribution’s strong assumption of variance equal to the mean by adding a separate dispersion parameter. Large numbers of people who do not bike appear in a dataset as excess “zeroes”, which in effect appears as overdispersion (24, 25).

Sample selection models use a logit or probit function to predict participation separately from frequency. Cao et al. (26) employ the Heckman Selection method to predict participation in trip-chaining behavior on the evening commute and, given participation, the number of stops comprising that trip chain (26).

Zero-inflated models function similarly to the Heckman Selection model, but they use a logit model to predict non-participation (excess zeroes) and either a Poisson or Negative Binomial model to predict frequency. While a traditional Negative Binomial model would treat excess zeroes simply as overdispersion, the zero-inflated model assumes a separate process generates the extra zeroes so it can be modeled using different parameters. This type of model is used commonly for modeling infrequent events such as traffic crashes (27).

METHODOLOGY
Survey Administration
The data for this study came from a self-administered ten-page survey mailed in May 2011 to households in five corridors in the Twin Cities as part of a study on the effects of Light Rail Transit (LRT) and associated built environment on travel behavior (28). These corridors were selected with the help of local planners because they had similar demographic trends. Three of the corridors are located in South Minneapolis: Nicollet Avenue, Bloomington Avenue, and Hiawatha Avenue from Lake Street to 50th Street. The two remaining corridors were in suburban communities outside the City of Minneapolis: Coon Rapids, 12 miles north of downtown Minneapolis, and Bloomington, 17 miles south of downtown.

For each corridor, we purchased two databases of residents from AccuData Integrated Marketing (http://www.accudata.com), a commercial data provider: a database of “movers” and a database of “nonmovers.” The “movers” included all current residents who had moved to the corridor after 2004. From this database, we drew a random sample of about 1,000 residents from the Hiawatha corridor and about 500 residents from each of Nicollet, Bloomington, Coon Rapids, and Burnsville corridors. The database of “nonmovers” consisted of a random sample of about 1,000 residents from the Hiawatha corridor and about 500 residents from each of the four corridors, who were not included in the “movers” list for each corridor.

The survey was pretested by students and staff members at the University of Minnesota and neighbors and friends of the investigators. Survey content was revised based on the feedback from pre-testers. The survey and two reminder postcards (1 and 2 weeks later) were mailed in May 2011. Ten $50 gift cards were provided as the incentive for the survey. The original database consisted of 6,017 addresses but only 5,884 were valid. The number of responses totaled 1,303, equivalent to a 22.2% response rate based on the valid addresses only. This is considered reasonable for a survey of this length, since the response rate for a survey administered to the general population is typically 10-40% (29).
Sample Characteristics
This study focuses specifically on residents in the three urban corridors. All three urban corridors exhibit traditional urban development patterns: a well-connected street grid, high levels of transit service (LRT in the Hiawatha corridor and bus in Nicollet and Bloomington), a variety of land uses and housing types, and similar built environment context, as shown in Figure 1. All three corridors contain bicycle infrastructure.

Since this study focuses specifically on bicycling for the journey-to-work commute trip, survey participants who indicated that they are non-employed students or not working (e.g., retired, homemaker, or unemployed) were removed from the sample.

Table 2 compares characteristics of survey respondents and the working sample with the 2011 ACS. Overall, homeowners and households with children are overrepresented among survey respondents due to the oversampling of residents who have lived in their homes since before 2004. Respondents have a higher than average income and less likely to live in a zero-vehicle household. These are typical results for voluntary self-administered surveys.
### TABLE 2 Demographics & Commute Mode Split for Respondents and General Population

<table>
<thead>
<tr>
<th></th>
<th>2011 ACS</th>
<th>2011 Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>City</td>
<td>Tracts</td>
</tr>
<tr>
<td>Population</td>
<td>381,833</td>
<td>113,614</td>
</tr>
<tr>
<td>Pct. Female</td>
<td>49.8 %</td>
<td>50.0 %</td>
</tr>
<tr>
<td>Avg. HH Size</td>
<td>2.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Pct. HH with Kids</td>
<td>24.2 %</td>
<td>28.8 %</td>
</tr>
<tr>
<td>Pct. Owner Occupied</td>
<td>50.4 %</td>
<td>57.5 %</td>
</tr>
<tr>
<td>Median Income</td>
<td>$47,478</td>
<td>$50,231</td>
</tr>
<tr>
<td>Pct. Fulltime</td>
<td>58.1 %</td>
<td>62.6 %</td>
</tr>
<tr>
<td>Pct. Part time</td>
<td>23.9 %</td>
<td>21.1 %</td>
</tr>
<tr>
<td>Pct. Not Working</td>
<td>17.9 %</td>
<td>16.3 %</td>
</tr>
<tr>
<td>Avg. Vehicles/HH</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Pct. Zero-Vehicle HH</td>
<td>8.8 %</td>
<td>8.1 %</td>
</tr>
</tbody>
</table>

#### Percent primarily commuting by:

<table>
<thead>
<tr>
<th></th>
<th>City</th>
<th>Tracts</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOV</td>
<td>61.4 %</td>
<td>60.9 %</td>
<td>63.7 %</td>
</tr>
<tr>
<td>Carpool</td>
<td>8.5 %</td>
<td>11 %</td>
<td>4.2 %</td>
</tr>
<tr>
<td>Transit</td>
<td>14 %</td>
<td>16.1 %</td>
<td>14.4 %</td>
</tr>
<tr>
<td>Bike</td>
<td>3.9 %</td>
<td>4.2 %</td>
<td>8.5 %</td>
</tr>
<tr>
<td>Walk</td>
<td>6.4 %</td>
<td>2.6 %</td>
<td>3.8 %</td>
</tr>
<tr>
<td>Other</td>
<td>0.9 %</td>
<td>0.9 %</td>
<td>1.3 %</td>
</tr>
<tr>
<td>Telecommute</td>
<td>4.9 %</td>
<td>4.4 %</td>
<td>5.6 %</td>
</tr>
</tbody>
</table>

* Results displayed here do not sum to 100 % due to estimation procedure used to make survey results more comparable to ACS.*
Commuting
The dependent variable in this study is a measure of bicycle commuting frequency. Survey respondents were asked, “In a typical week with good weather, how many days do you use each of the following as your primary means of transportation between home and work/school?” Available modes included teleworking, driving alone, carpooling, transit, walking, biking, and other. For each mode, they were presented with six ordinal categories: (1) “Never”, (2) “Less than once per month”, (3) “1-3 days per month”, (4) “Once per week”, (5) “2-3 days per week”, and (6) “4-5 days per week”.

These categories were recoded into a scale variable using the midpoints of each category (0, 0.5, 2, 4.3, 10.75, and 19.35 times per month respectively) to represent commute trips per month made by each mode. Participants were asked to report additional details about their commute, including the distance in both miles and minutes and whether their employer/school provides free parking.

Participation in commuting by bicycle was defined as any respondent who indicated that they commute by bike at least infrequently (categories 2 through 6).

The lower half of Table 2 focuses specifically on the journey to work mode for survey respondents and the general population. The ACS asks respondents to indicate their single primary mode of transportation for the commute trip (30), assuming that people use only a single mode each and every day. This assumption is problematic for all commuters, and particularly for bicyclists because they experience barriers such as weather events or the need to transport goods that change from day to day (17).

Independent Variables
Table 3 presents the hypothesized relationship and descriptive statistics for all independent variables considered in this study, and the following sections explain how each variable is measured.

Built Environment
Built environment characteristics were measured in two ways: (1) through a set of survey questions asking respondents to indicate how true each of 29 neighborhood characteristics was of their neighborhood, and (2) using a GIS to objectively measure the infrastructure and land use around their homes.

Perceived Built Environment  Respondents rated how true each of 29 characteristics, such as “Large back yards” and “Easy access to transit stop/station”, was of their current neighborhood. The ordinal scale ranged from (1) “Not at all true” to (4) “Entirely true”. The two primary characteristics included in this study are “Good bicycle routes beyond the neighborhood” and “Close to where I work”. Two additional characteristics, “Low crime rate within neighborhood” and “Low level of car traffic on neighborhood streets,” were tested but found to be insignificant.

Objectively Measured Built Environment  To evaluate the built environment and its impacts on travel behavior, we constructed network distance buffers around each participant’s homes at about 400—, 800—, and 1,600—meter (\(\frac{1}{4}\)—, \(\frac{1}{2}\)—, and 1—mile) distances. The network distance buffer includes only areas that the respondent could actually walk to.

Bicycle facilities around each respondent’s home were measured by summing the total distance of bike lanes in meters within each respondent’s network distance buffers.
### TABLE 3 Variables with Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Hypothesis</th>
<th>Mean</th>
<th>(S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Number of days in a typical month with good weather that respondent commutes by bicycle overall</td>
<td>Dependent Variable</td>
<td>1.90</td>
<td>(4.58)</td>
</tr>
<tr>
<td>Yb</td>
<td>Days per month commuting by bicycle for bicyclists</td>
<td></td>
<td>7.26</td>
<td>(6.43)</td>
</tr>
<tr>
<td>A</td>
<td>Jobs accessible by bike within 10 minutes (1000’s)</td>
<td>+</td>
<td>0.61</td>
<td>(0.52)</td>
</tr>
<tr>
<td>C</td>
<td>Number of children under 12 in household</td>
<td>-</td>
<td>0.43</td>
<td>(0.83)</td>
</tr>
<tr>
<td>D</td>
<td>Respondent has college degree or higher</td>
<td>+</td>
<td>0.74</td>
<td>(0.44)</td>
</tr>
<tr>
<td>Ed</td>
<td>Commute distance in km</td>
<td>-</td>
<td>13.37</td>
<td>(13.17)</td>
</tr>
<tr>
<td>Ep</td>
<td>Employer provides free parking</td>
<td>-</td>
<td>0.69</td>
<td>(0.46)</td>
</tr>
<tr>
<td>Et</td>
<td>Respondent works part time</td>
<td>-</td>
<td>0.15</td>
<td>(0.36)</td>
</tr>
<tr>
<td>Fb</td>
<td>Pro-biking Factor</td>
<td>+</td>
<td>0.29</td>
<td>(1.06)</td>
</tr>
<tr>
<td>Fd</td>
<td>Pro-driving Factor</td>
<td>-</td>
<td>−0.12</td>
<td>(1.15)</td>
</tr>
<tr>
<td>Fu</td>
<td>Pro-travel Factor</td>
<td>+</td>
<td>−0.03</td>
<td>(1.26)</td>
</tr>
<tr>
<td>G</td>
<td>Respondent’s age in years</td>
<td>-</td>
<td>45.14</td>
<td>(12.64)</td>
</tr>
<tr>
<td>H</td>
<td>Land use entropy within 1600 m</td>
<td>+</td>
<td>0.35</td>
<td>(0.11)</td>
</tr>
<tr>
<td>Ib</td>
<td>Residential preference for “Good bicycle routes”</td>
<td>+</td>
<td>3.03</td>
<td>(1.07)</td>
</tr>
<tr>
<td>Ic</td>
<td>Residential preference for “Living unit on cul-de-sac”</td>
<td>-</td>
<td>1.36</td>
<td>(0.74)</td>
</tr>
<tr>
<td>Iw</td>
<td>Residential preference for “Close to where I work”</td>
<td>+</td>
<td>3.16</td>
<td>(0.88)</td>
</tr>
<tr>
<td>K</td>
<td>Income ($1000)</td>
<td>-</td>
<td>7.56</td>
<td>(3.34)</td>
</tr>
<tr>
<td>L</td>
<td>Respondent has a limitation that makes biking difficult</td>
<td>-</td>
<td>0.05</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Nl</td>
<td>km of bike lane within 1600 m</td>
<td>+</td>
<td>3.37</td>
<td>(2.45)</td>
</tr>
<tr>
<td>Nt</td>
<td>km of bike trail within 1600 m</td>
<td>+</td>
<td>3.84</td>
<td>(2.79)</td>
</tr>
<tr>
<td>N4w</td>
<td>Intersection density within 400 m</td>
<td>+</td>
<td>0.19</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Pb</td>
<td>Perception: Good bicycle routes beyond the neighborhood</td>
<td>+</td>
<td>3.66</td>
<td>(0.62)</td>
</tr>
<tr>
<td>Pc</td>
<td>Perception: Living unit on cul-de-sac rather than through street</td>
<td>-</td>
<td>1.13</td>
<td>(0.52)</td>
</tr>
<tr>
<td>Pw</td>
<td>Perception: Close to where I work</td>
<td>+</td>
<td>2.87</td>
<td>(1.06)</td>
</tr>
<tr>
<td>V</td>
<td>Fewer than 1 car per driver in household</td>
<td>+</td>
<td>0.18</td>
<td>(0.38)</td>
</tr>
<tr>
<td>W</td>
<td>Respondent is female</td>
<td>-</td>
<td>0.51</td>
<td>(0.50)</td>
</tr>
</tbody>
</table>
An additional measure of bicycle cumulative opportunity accessibility summed the number of jobs accessible by bicycle within 10 minutes using OpenTripPlanner Development Team (31), assuming an average bicycling speed of 16.1 kilometers per hour (10 miles per hour). Data inputs were 2010 Longitudinal Employer-Household Dynamics (LEHD) job estimates by census block and an OpenStreetMap (OSM) street network shapefile.

Residential Preferences
Respondents rated how important each of 29 characteristics, such as “Large back yards” and “Easy access to transit stop/station”, was when they were last looking for a place to live. The ordinal scale ranged from (1) “Not at all important” to (4) “Extremely important”. Respondents could also choose “I never considered it”. For this same list of characteristics, respondents rated how accurately each of the statements represents their current neighborhood on a scale from (1) “Not at all true” to (4) “ Entirely true”. In this study, the two characteristics considered are “Good bicycle routes beyond the neighborhood” and “Close to where I work”.

Travel Attitudes
To measure attitudes regarding travel, the survey asked respondents whether they agreed or disagreed with a series of 21 statements on a 5-point scale from “strongly disagree” (1) to “strongly agree” (5). Factor analysis was then used to extract the fundamental dimensions spanned by these 21 items, since some of the items are highly correlated. Seven underlying dimensions were identified: pro-drive, pro-walk, pro-bike, pro-transit, safety of car, status of car and pro-travel.

Pro-bike ($F_b$) and pro-travel ($F_u$) were selected for use in both models to control for travel attitudes.

Modeling Approach
The zero-inflated Poisson model was developed to account for count data that are overdispersed due to an excess number of zeroes in the dataset (32). A binary logistic function is used to predict the probability of the dependent variable assuming a value of zero. Given this probability, the Poisson model is then fit to the non-zero data. These two equations are jointly modeled using Maximum Likelihood Estimation (MLE).

A zero-inflated negative binomial distribution is used to model count data with excess zeroes that otherwise would violate the assumption of equal mean and variance in Poisson (33). The distribution of zeroes follows a binary logistic distribution, similar to zero-inflated Poisson, but the values of non-zero observations are generated by a negative binomial process, as shown in equation 1:

$$ Y_i \sim \begin{cases} 0 & \text{with probability } p_i \\ \text{NB}(\lambda_i) & \text{with probability } 1 - p_i \end{cases} (1) $$

In this study, $Y_i$ is the number of days in a typical month with good weather on which person $i$ commutes by bicycle. Zero outcomes ($Y_i = 0$), indicating non-participation in bicycle commuting, occur with probability $p_i$. The remaining count outcomes ($Y_i > 0$) occur with probability $1 - p_i$, and they follow a Negative Binomial distribution with expected frequency $\lambda$.

Thus for any value of $y$, the probability $P(Y_i = y) =$:
\[ P(Y = 0) = p + (1 - p)t^k, \quad 0 < p < 1 \]
\[ P(Y = y) = (1 - p) \binom{y + k - 1}{y} t^k (1 - t)^y, \quad y = 1, 2, \ldots \] (2)

RESULTS

Model Results

Table 4 shows the results from modeling participation and frequency of commuting by bicycle using Zero-Inflated Negative Binomial Regression (ZINB). First, we explain how to read the table. Step 1 includes only built environment predictors, and the subsequent models add controls for demographics and individual commute characteristics (Step 2), residential preferences (Step 3), and travel attitudes (Final Model). The two components of the ZINB model are shown separately, with participation on top and frequency below.

In unformatted ZINB results from statistical software, the coefficients represent “zero inflation”, or the relationship between that variable and the probability of the dependent variable assuming a value of 0. Coefficients have been reversed here for clarity. A positive coefficient indicates a positive association with the probability of participating in bicycle commuting.

Since these results come from the binary logit process in the ZINB model, the a one-unit increase in the independent variable can be interpreted as an increase in the log-odds of being a bicycle commuter (at any frequency) by the value of the coefficient. Using the final model (Model 4) as an example, a 1-unit increase in \( N_l \) (kilometers of bicycle lanes within 1600 meters of the respondent’s home) is associated with a 0.15 increase in the log-odds of being a bicycle commuter. Exponentiating this gives us an odds ratio of 1.15 (not shown).

Model Fit

The third section of the table shows the \( \ln(\alpha) \) and \( \alpha \) parameters from the negative binomial regression. Significant values indicate that \( \alpha \) differs statistically from 1. Insignificant values, such as the model in Step 2, indicate a lack of overdispersion and suggest that a Poisson process would be sufficient. The \( \alpha \) parameter is significant at the 0.05 level for steps 1 and 3, as well as the final model. Finally, tests of model fit are shown. A significant value for the Likelihood Ratio Test (LR Test) is another indicator that the negative binomial model is a better fit than a Poisson model. A significant value in the Vuong Test indicates that the zero-inflated model is a better fit than the traditional Negative Binomial Regression (NBREG). These tests are significant in all steps of the model.

The Pseudo-\( R^2 \) value is a McFadden’s Adjusted pseudo-\( R^2 \). There is no directly comparable measure in negative binomial or binary logistic regression to the classic \( R^2 \) used in OLS, which represents the percent of variation in the dependent variable that can be explained by the independent variables. The McFadden’s Adjusted pseudo-\( R^2 \) takes on values from approximately 0 to 1, but it represents the relative improvement of this model’s log likelihood over that of a null or constant-only model, with a penalty for additional parameters in the model.

Examining the Pseudo-\( R^2 \) values for these four models provides a simple comparison of the relative importance of each set of variables. Model 1, with built environment variables only, has a very low pseudo-\( R^2 \) (0.005). Adding demographics increases this by an order of magnitude, and residential preferences cause another modest increase. Adding the travel attitude factors, however,
TABLE 4 Results of Zero-Inflated Negative Binomial Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Participation</th>
<th>Frequency</th>
<th>( \alpha )</th>
<th>LR Test</th>
<th>Vuong Test</th>
<th>Pseudo—R²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>SE</td>
<td>Coeff.</td>
<td>SE</td>
<td>Coeff.</td>
<td>SE</td>
</tr>
<tr>
<td>( N_t )</td>
<td>0.13 *** (0.047)</td>
<td>0.13 ** (0.058)</td>
<td>0.15 ** (0.065)</td>
<td>492.75 ***</td>
<td>1.91 **</td>
<td>0.005</td>
</tr>
<tr>
<td>( E_d )</td>
<td>-0.086*** (0.016)</td>
<td>-0.095*** (0.019)</td>
<td>-0.11 *** (0.022)</td>
<td>431.66 ***</td>
<td>4.65 ***</td>
<td>0.044</td>
</tr>
<tr>
<td>( G )</td>
<td>-0.041*** (0.0096)</td>
<td>-0.047*** (0.012)</td>
<td>-0.031*** (0.013)</td>
<td>422.29 ***</td>
<td>4.18 ***</td>
<td>0.060</td>
</tr>
<tr>
<td>( F_b )</td>
<td>1.86 *** (0.24)</td>
<td>0.20 * (0.12)</td>
<td>0.21 (0.51)</td>
<td>200.38 ***</td>
<td>5.28 ***</td>
<td>0.220</td>
</tr>
<tr>
<td>( F_u )</td>
<td>0.83 *** (0.12)</td>
<td>0.15 ** (0.062)</td>
<td>0.62 (0.13)</td>
<td>1.91 **</td>
<td>4.65 ***</td>
<td>0.044</td>
</tr>
</tbody>
</table>

*** Significant at the p < 0.01 level  ** Significant at the p < 0.05 level  * Significant at the p < 0.1 level
causes the pseudo-$R^2$ to more than triple, suggesting that these are the most influential components in the model. The final model has a pseudo-$R^2$ of 0.220.

**Built Environment**

Of all the built environment variables hypothesized to have a relationship with bicycle commuting, only bike lanes ($N_l$), job accessibility ($A$), and living on a cul-de-sac ($P_c$) are significant. Interestingly, bike lanes appear in the participation portion of the equation. Bicycle lanes are more strongly associated with whether a person bike commutes at all than how much they do so. This result is likely connected to residential self-selection of people who like bicycling into suitable neighborhoods. People who like bicycling choose to live into neighborhoods with bike lanes, which enables them to commute by bicycle.

The coefficient for job accessibility ($A$) suggests that close proximity to jobs is an important predictor in how frequently one can make that commute trip by bicycle. It also probably serves as a proxy for other built environment variables that were pushed out of the model due to multicollinearity, such as density.

The cul-de-sac variable ($P_c$) is intuitive in that cul-de-sacs represent interruptions in the street grid, but the number of cul-de-sacs in the portion of South Minneapolis where this study occurred are limited. Further exploration of what street characteristics are associated with people’s perceptions of living on a cul-de-sac is warranted. Additionally, the perceived built environment measure of living on a cul-de-sac was significant, where other street grid measures, such as number of cul-de-sacs within 400 meters of the respondent’s home, were not. Given the low number of cul-de-sacs in South Minneapolis, it is reasonable to believe that only close proximity to this kind of street network interruption has a depressive effect on bicycling.

**Demographics and Commute Characteristics**

The coefficients on demographic and commuting variables are all intuitive, given existing literature on bicyclists, but the interpretation is not necessarily as straightforward on some of them. Longer commutes ($E_d$) decrease the probability that someone will commute by bicycle. The interpretation of this variable is unambiguous: each additional kilometer of commute distance represents a decrease in the log-odds of commuting by bicycle by 0.11.

Free parking ($E_p$) is associated with a decrease in the frequency with which a respondent commutes by bicycle. Given the spatial distribution of free parking in Minneapolis, however, it may be that this variable is simply serving as a proxy for working in Downtown Minneapolis or at the University of Minnesota since the respondents’ actual work locations are not available in a geocoded format. Bicycle infrastructure connectivity to Downtown and the University of Minnesota is very strong, with several major north-south bike lanes and the Hiawatha LRT Bike Trail connecting the study area to Downtown. The rest of the demographic variables are consistent with literature.

**Residential Preferences**

The positive coefficients on the importance of strong bike routes beyond the neighborhood ($I_b$) and living close to work ($I_w$) further suggest the self-selection effect. Curiously, however, these variables fit best in the frequency portion of the model. This reinforces the interpretation of free parking as a proxy for Downtown or University employment: a person who works at one of these pay-for-parking destinations is already prone to bicycling, but if they value (and then presumably
self-select into) a neighborhood that is both close to work and has good bike routes beyond the neighborhood, these South Minneapolis corridors provide adequate infrastructure for bicycling often. If work locations could be secured, a network analysis may confirm this finding or provide a stronger interpretation.

Travel Attitudes
Both the Pro-bicycling factor \( (F_b) \) and the Pro-travel factor \( (F_u) \) are positively and significantly associated with participation and frequency. This is unsurprising, but still noteworthy because they represent such a large contribution in this model’s explanatory power. The pro-travel factor contained sentiments about enjoying the journey as much as reaching the destination and valuing time spent in travel (versus believing it to be wasted). Utility theory for predicting mode share assumes that people will choose the mode that minimizes their cost and time investments, but these findings suggest that bicyclists derive value specifically from their commute. This is consistent with literature on positive utility of commuting in which people value mental separation from work and self-report non-zero ideal commute distances (34). Further supporting this finding is a recent study by Paige Willis et al. (35), in which bicyclists in particular derive satisfaction from many aspects of their commute.

DISCUSSION AND CONCLUSIONS
Catalysts & Magnets
The model results show that bicycle commuting participation and frequency are associated with different built environment measures. However, interpreting these results requires some measure of caution. As discussed extensively in other literature, establishing causality between travel behavior and the built environment is challenging (3, 4). The Zero-Inflated Negative Binomial Regression (ZINB) model helps control for residential self-selection effects that confound an observed relationship between built environment characteristics and commute mode choice by predicting participation (self-selection) separately from frequency.

A variable in a model of bicycling participation that has a causal impact on bicycling behavior could be considered a “catalyst”; these variables encourage people who are not already doing so to start bicycling. Conversely, a variable that indicates residential self-selection of people already prone to bicycling into bike-friendly neighborhoods can be called a “magnet”. The presence of one of these magnets in a neighborhood does not cause a non-bicyclist to become a regular bike commuter, but it does provide a neighborhood that meets the travel needs and preferences of bicyclists, giving them better opportunities to bicycle. Given the structure of the model employed here, coefficients in the participation half of the model are likely to have a “magnet” effect that overshadows any possible “catalyst” effect.

Implications for Practice
Bicycle infrastructure, measured as kilometers of bike lane within 1 kilometer of the respondent’s home, is significant in the participation portion of the model. However, subsequent bivariate correlation tests (not shown) between survey respondents’ pro-bike travel attitude factor \( (F_b) \) and their length of tenure in a neighborhood (stratified by age) failed to provide evidence of temporal precedence. Travel attitudes are stronger among residents who moved into their current more recently, suggesting that travel attitudes precede location choice. This suggests that bicycle infrastructure functions more like a magnet than a catalyst. While evidence to infer causality is lacking, it is
evident that people who are more likely to use a bicycle for commuting do in fact live near these facilities. This is an important finding because it implies that placing new bicycle infrastructure around other built environment characteristics that do appear to influence bicycling will magnify their effects by attracting residents with a propensity to commute by bicycle.

In the frequency half of the equation, cul-de-sacs, job accessibility, and free parking all have significant coefficients. The relationships to job accessibility and cul-de-sacs are intuitive: cul-de-sacs interrupt the street network, and respondents with greater accessibility to jobs by bicycling have a higher probability of working within a reasonable bike distance from home.

While it is possible that free parking directly influences how often a respondent commutes by bicycle, this is probably not its only mechanism of action. Pay-for-parking is located primarily Downtown and around the University of Minnesota, whereas free parking is the norm in most other parts of the city. There are some exceptions, but the general trend suggests a probable association between a lack of free parking at the respondent’s work and that workplace being located either Downtown or at the University. Connectivity to Downtown and the University via bicycle is excellent due to several closely-spaced major north-south bicycle routes, whereas connectivity throughout the rest of the city is limited.

The accessibility and parking variables have significant implications for practice, given the finding about bicycle lanes. After controlling for self-selection, which includes self-selection into close proximity of bike lanes, these two variables still have a significant relationship with frequency of commuting by bicycle. If a city aspires to increase bicycling, these results suggest that new bicycle infrastructure should be deployed in neighborhoods with high accessibility to employment, and the routes should be designed to provide connections to major job centers. As new residents self-select into the neighborhood because of the bike lane, these other factors will enable them to bike more frequently.

Limitations and Areas for Further Study

The issue of free parking at a respondent’s work location raises several questions about how the route choices available to a person might influence their mode choice. This study measured built environment characteristics around participants’ homes, but not their work locations nor along possible routes connecting the two. Previous studies have found that characteristics along the route are stronger predictors of nonmotorized travel (7, 36), and the free parking variable in this study demonstrates the need of exploring this issue in greater detail.

Future research should identify respondents’ work locations and construct separate measures for bicycle infrastructure along the routes in between home and work. A measure of job accessibility that uses a modified network with stronger weights on jobs accessible via a route comprised mostly of dedicated infrastructure might serve as a sufficient proxy, if work locations are not available. While this still would not resolve whether infrastructure serves as a catalyst or magnet, it would clarify the contexts that make infrastructure relevant to travel decisions.

We modeled the participation component of bicycle commuting as a choice between being a bicycle commuter, or not being a bicycle commuter who uses any and all other modes. Some studies consider bicycling versus a specific choice, such as driving (8). Focusing specifically on bicycling versus driving removes possible dampening effects on the results from consolidating walking, transit, and driving into a single “non-bicycle” mode category. Built environment characteristics that support walking and transit have more in common with bicycle-friendly spaces than car-supportive environments. However, walking and transit are also small mode shares, so any
dampening effect may not have much of an impact. Additionally, despite research and policy that
collapses all non-auto modes into a broader category of “alternative transportation”, there are dis-
tinct differences in how each mode functions and what needs it serves. Comparing bicycling to all
other modes in aggregate, as was done here, has a lower potential to overestimate results.

Finally, this study was not able to establish causality the relationship between bicycling
and the built environment. Cross-sectional data, as used in this paper, is notably weak in this
regard. As previously mentioned, a bivariate correlation test between the strength of travel attitudes
length of tenure in a neighborhood confirms that temporal precedence is a likely issue in this study.
Nonetheless, the rest of the evidence presented in this study considerably advances the conversation
about self-selection and the built environment.

ACKNOWLEDGEMENTS
The study was funded by the Transitway Impact Research Program (TIRP) in the Twin Cities. We
also thank Hennepin County and the City of Minneapolis planning staff for their help and support.

REFERENCES


[3] Mokhtarian, P. L. and X. Cao, Examining the Impacts of Residential Self-Selection on Travel
Behavior: A Focus on Methodologies. Transportation Research Part B: Methodological,

Selection on Travel Behaviour: A Focus on Empirical Findings. Transport Reviews, Vol. 29,

[5] Metropolitan Council, Travel Behavior Inventory (TBI) Surveys, ????


Transportation Choices: Bicycling Versus Driving. Journal of Urban Health: Bulletin of the

[8] Handy, S. L. and Y. Xing, Factors Correlated with Bicycle Commuting: A Study in Six Small
110.

[9] Xing, Y., S. L. Handy, and P. L. Mokhtarian, Factors Associated with Proportions and Miles
of Bicycling for Transportation and Recreation in Six Small US Cities. Transportation Re-


