Title
Pilot Models for Estimating Bicycle Intersection Volumes

Permalink
https://escholarship.org/uc/item/380855q6

Authors
Griswold, Julia B.
Medury, Aditya
Schneider, Robert J.

Publication Date
2011

Peer reviewed
Pilot Models for Estimating Bicycle Intersection Volumes

Julia B. Griswold, SafeTREC,
Aditya Medury, Institute of Transportation Studies, Berkeley, and
Robert J. Schneider, SafeTREC
Pilot Models for Estimating Bicycle Intersection Volumes

Julia B. Griswold (Corresponding Author)
Safe Transportation Research and Education Center
University of California, Berkeley
2614 Dwight Way #7374
Berkeley, CA 94720, U.S.A
Telephone: (510) 643-0566
Fax: (510) 643-9922
Email: juliagris@berkeley.edu

Aditya Medury
Department of Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Berkeley
Email: amedury@berkeley.edu

Robert J. Schneider
Safe Transportation Research and Education Center
University of California, Berkeley
2614 Dwight Way #7374
Berkeley, CA 94720
Telephone: (510) 643-0566
Fax: (510) 643-9922
Email: rschneider@berkeley.edu

Word Count: 5,030 + (4 tables + 1 figure) × 250 = 6,280 words

Submitted to the Transportation Research Board 90th Annual Meeting
Pilot Models for Estimating Bicycle Intersection Volumes

ABSTRACT

Bicycle volume data are useful to practitioners and researchers to understand safety, travel behavior, and development impacts. This paper describes the methodology used to develop several simple models of bicycle intersection volumes in Alameda County, California. The models are based on two-hour bicycle counts performed at a sample of 81 intersections in the Spring of 2008 and 2009. Study sites represented areas with a wide range of population density, employment density, proximity to commercial property, neighborhood income, and street network characteristics. The explanatory variables considered for the models included intersection site, land use, transportation system, and socioeconomic characteristics of the areas surrounding each intersection. Four alternative models are presented with adjusted R-square values ranging from 0.39 to 0.60. The models showed that bicycle volumes tended to be higher at intersections surrounded by more commercial retail properties within 1/10 mile, closer to a major university, with a marked bicycle facility on at least one leg of the intersection, surrounded by less hilly terrain within 1/2 mile, and surrounded by a more connected roadway network. The models also showed several important differences between weekday and weekend intersection volumes. The positive association between bicycle volume and proximity to retail or a large university was greater on weekdays than weekends, while bicycle facilities had a stronger positive association and hilly terrain had a weaker negative association with bicycle volume on weekends than weekdays. Further testing and refinement is necessary before accurate count predictions can be made in Alameda County or other communities.
Bicycling is increasing in many United States communities. Nationwide, the proportion of commuters traveling to work regularly by bicycle grew from 0.45% to 0.55% between 2006 and 2008 (1,2). Cities such as Portland, New York, Seattle, and San Francisco documented increases in bicycle counts over this same time period (3,4,5,6). Bicycling is a convenient and economical mode for commuting, small shopping trips, transit access, and recreation. Estimates of bicycle activity are valuable to planners, engineers, designers, public health professions, and others. However, there are very few tools available to estimate the number of bicyclists that pass by specific locations or use particular intersections within an urban area. A predictive model of bicycle volumes can be used to:

- Quantify bicycle exposure in safety analyses, providing the denominator in the calculation of crash risk;
- Identify priority locations for bicycle facility improvements or safety measures; and
- Estimate changes in bicycle volumes that will occur with new developments, roadway changes, or new transit projects.

STUDY PURPOSE
The purpose of this study is to present several preliminary models that can be used to estimate bicycle intersection volumes. A simple model structure, loglinear regression, was chosen so that the models would be easy to apply using geographic information systems (GIS) and simple spreadsheet software. Since the analysis was conducted in one urban area (Alameda County, California), more research is needed to refine the model equations and determine the applicability of the results for other communities. These models are designed for planning purposes to help identify built environment characteristics associated with higher and lower levels of bicycling and show general differences in bicycle volumes at intersections throughout a city or region. Applications that require precise bicycle volume data for individual intersections should use actual bicycle counts from each location.

PREVIOUS STUDIES
Several techniques have been developed to analyze bicycle activity in different locations. Overlay mapping techniques, or sketch-plan methods, are useful for planning and prioritization (7,8,9). However, they are not typically calibrated to actual bicycle counts. Most previous efforts to model bicycle activity have used zonal areas or the individual as the unit of analysis. Some regional travel demand models have been adjusted to estimate bicycle mode share within a certain area (10), but these models are usually estimated at the traffic analysis zone (TAZ) level and are not fine-grained enough to capture intersection-level bicycle activity. Models that examine behavior at the level of the individual are valuable for identifying factors that determine mode and route choice, but they are difficult to use for volume estimates. Most of these models are based on detailed household travel survey data, which can be expensive to collect.

Two recent studies in Southern California have developed predictive models of bicycle volume, using linear regression with bicycle counts as the dependent variable and demographic and land use measures for the independent variables. Four land use and transportation system explanatory variables (afternoon bus frequency, land use mix, density of residents under age 18, and proximity to the bicycle network) were used to predict weekday afternoon peak hour bicycle volumes at intersections in the City of Santa Monica (11). However, the ordinary least squares regression equation used for the model could produce negative predictions. This modeling issue was addressed in San Diego County by using the natural logarithm of the dependent variable, but the San Diego model was limited to two explanatory variables (employment density and length of nearby multi-use trails) for predicting weekday 7 a.m. to 9 a.m. bicycle volumes (12).

An additional bicycle modeling study explored applying Space Syntax measures in a bicycle model. A Space Syntax measure representing direct paths in the street network was combined with land use variables to predict morning peak hour bicycle volumes at intersections and street segments in Cambridge, MA (13). The model performed fairly well, but the sample size was small (n=16) and the...
land use variables (population and employment density) explained most of the variation in the bicycle volumes. In addition, the Space Syntax measure required special software to calculate.

Table 1 summarizes the factors that have been associated with bicycle activity in these modeling efforts and other bicycle studies.

### Table 1. Factors Associated with Bicycling in Previous Research

<table>
<thead>
<tr>
<th>Variable</th>
<th>Relationship with Bicycle Activity</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environmental and Land Use Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearby population density</td>
<td>+</td>
<td>McCahill and Garrick (2008) (13)</td>
</tr>
<tr>
<td>Proximity to downtown</td>
<td>+</td>
<td>Dill and Voros (2007) (14)</td>
</tr>
<tr>
<td><strong>Transportation System Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearby street connectivity</td>
<td>+</td>
<td>Dill and Voros (2007) (14)</td>
</tr>
<tr>
<td>Proximity to a freeway</td>
<td>+</td>
<td>Dill and Voros (2007) (14)</td>
</tr>
<tr>
<td>Amount of bicycle lanes nearby</td>
<td>+</td>
<td>Dill and Carr (2003) (17)</td>
</tr>
<tr>
<td>Amount of multi-use trails nearby</td>
<td>+</td>
<td>Jones et al. (2010) (12)</td>
</tr>
<tr>
<td>Afternoon peak hour bus frequency</td>
<td>+</td>
<td>Haynes and Andrzejewski (2010) (12)</td>
</tr>
<tr>
<td>Major arterial</td>
<td>-</td>
<td>Stinson and Bhat (2003) (18)</td>
</tr>
<tr>
<td>Parallel parking permitted</td>
<td>-</td>
<td>Stinson and Bhat (2003) (18)</td>
</tr>
<tr>
<td>Smooth pavement</td>
<td>+</td>
<td>Stinson and Bhat (2003) (18)</td>
</tr>
<tr>
<td><strong>Socioeconomic Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age 18-55</td>
<td>+</td>
<td>Dill and Voros (2007) (14)</td>
</tr>
<tr>
<td>Age</td>
<td>-</td>
<td>Xing, Handy, and Buehler (2008) (19)</td>
</tr>
<tr>
<td>% of nearby population under age 18</td>
<td>-</td>
<td>Haynes and Andrzejewski (2010) (12)</td>
</tr>
<tr>
<td>Education level</td>
<td>+</td>
<td>Xing, Handy, and Buehler (2008) (19)</td>
</tr>
<tr>
<td>Income</td>
<td>+</td>
<td>Dill and Voros (2007) (14)</td>
</tr>
</tbody>
</table>

**METHODOLOGY**

This section describes the study area, the bicycle count collection, and the development of the explanatory variables. The model is based on bicycle counts taken at 81 intersections along arterial and collector streets in Alameda County, California (Figure 1).
Alameda County is located on the eastern side of the San Francisco Bay across the Bay Bridge from San Francisco, and it stretches east and south towards the San Joaquin Valley. Its population in 2008 was estimated to be 1.47 million (2). The western portion of the county was largely developed in the early to mid-twentieth century and contains several small downtown commercial areas and many streetcar suburbs with grid street patterns. The hill areas and the eastern portion of the county were mostly developed in the second half of the twentieth century; these areas are more suburban in nature, with curvilinear street patterns. The largest city in the county is Oakland, with an estimated population of approximately 366,000 (2). In 2008, estimates are that 48 percent of the county population was white, 25 percent Asian, 22 percent Hispanic or Latino, and 13 percent black (2).

Bicycle Counts
The bicycle counts for this study were collected in the spring of 2008 and 2009 in conjunction with a study of pedestrian exposure and crash risk. While the pedestrian study determined that the counts would be at intersections, the count locations were also appropriate for understanding bicycle exposure because approximately half of bicycle crashes occur at intersections (20). Counts were taken at 50 intersections in 2008 and 31 in 2009. All intersections were along arterial or collector roadways. A strategic process was used to select intersections, ensuring that the set of study sites represented areas with a wide range of population density, employment density, proximity to commercial properties, neighborhood incomes, and other socioeconomic characteristics. In addition, the study intersections had a variety of roadway characteristics. Of the 81 intersections, 16 (20%) had a marked bicycle facility (e.g., bicycle lane, bicycle
boulevard, or shared lane marking) on at least one approach. The choice of intersection locations is
described in detail in Schneider, Arnold, and Ragland (21) and Schneider et al. (22).

Each bicycle was logged according to the movement that was made at the intersection—straight,
right turn, or left turn from one of the four intersection legs—which made a total of 12 possible
movements. Bicycles being ridden either on the street or on the sidewalk were included, but bicycles
being walked were logged as pedestrians. The counts for each movement were summed together to find
the total bicycle volume for the intersection. Counts were conducted for two 2-hour periods at each
intersection. One was taken on a Tuesday, Wednesday, or Thursday, and one was taken on a Saturday.
The time of day for the counts varied depending on data collector scheduling. Counts were taken during
12 p.m.-2 p.m., 2 p.m.-4 p.m., 3 p.m.-5 p.m., or 4 p.m.-6 p.m. on weekdays and 9 a.m.-11 a.m., 12 p.m.-2
p.m., 3 p.m.-5 p.m., or 4 p.m.-6 p.m. on weekends. Additional information about the count methodology
is provided in Schneider, Arnold, and Ragland (21) and Schneider et al. (22). Table 2 includes summary
statistics for the bicycle count data.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>All Counts</th>
<th>Weekday</th>
<th>Weekend</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Counts</td>
<td>162</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Maximum</td>
<td>343</td>
<td>343</td>
<td>171</td>
</tr>
<tr>
<td>Median</td>
<td>23.5</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>Mean</td>
<td>35.8</td>
<td>38.6</td>
<td>33.0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>41.4</td>
<td>50.3</td>
<td>30.1</td>
</tr>
</tbody>
</table>

ANALYSIS

The data analysis process for the model estimation included three steps: (a) explanatory variables were
screened to eliminate those weakly correlated to the dependent variable; (b) the remaining variables were
screened for collinearity to avoid including strongly correlated variables in the same model; and (c) four
alternative model structures were chosen for strong goodness-of-fit and significant coefficients.

Regression Modeling

Loglinear ordinary least squares (OLS) regression was used to estimate a model of bicycle intersection
volume. In loglinear regression, the dependent variable is transformed using the natural logarithm. This
is an appropriate method for modeling count data because when the natural log predictions are
transformed back to counts using the exponential, there are no negative values. A negative binomial form
of the regression model was also tested during the analysis process, and the model coefficients were
similar. The loglinear form was selected because it is easy to apply using spreadsheet software and
because it is easier to interpret the model goodness-of-fit and independent variable coefficients. One
intersection had a count of zero bicyclists. This count was changed to 0.1 to allow the log to be
computed. The model formulation is:

\[ Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_j X_{ji} + \epsilon_i \]

where:

- \( Y_i \) = bicycle counts at intersection \( i \)
- \( X_{ji} \) = value of explanatory variable \( j \) at intersection \( i \)
- \( \beta_j \) = model coefficient for variable \( j \)

The modeling process examined the relationship between the bicycle intersection volume and the
built environment surrounding the intersection, such as nearby land use, transportation system, and site
characteristics. Some of the measures were evaluated at different scales, including buffer radii of 0.1 mi
This study took advantage of explanatory variables previously developed for the 81 intersections for pedestrian volume and pedestrian crash models (21, 22). Several additional factors used in previous studies were also included, bringing the total to more than 70 variables considered in the analysis. Network distance to UC Berkeley Campus and Oakland City Center were included to account for two of the major trip attractors in the county. Connected node ratio was calculated using GIS. This measure represented the ratio of three- and four-way intersections to dead-end streets in the area around each study intersection. Areas with lower connected node ratios had higher proportions of dead ends. The new variables and any existing variables that were used in the final bicycle volume models are described in Table 3. Descriptions of the other variables considered for the analysis can be found in Schneider, Arnold, and Ragland (21) and Schneider et al. (22).

It was necessary to narrow down the number of independent variables to avoid overfitting the model. Models with nearly as many independent variables as the number of observations they are based on predict volumes poorly. First, correlation coefficients were estimated between each explanatory variable and the bicycle counts. Those variables with weak correlation ($\rho < 0.2$) were eliminated from the analysis. Next, the remaining independent variables were screened for collinearity; pairs of variables with moderate to strong correlation ($\rho > 0.3$) were not included in the same model.

Table 3. Description of Selected Explanatory Variables Considered in the Modeling Process

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Variable Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SlopeT</td>
<td>Average slope (degrees) of terrain within 1/10 mile</td>
</tr>
<tr>
<td>SlopeQ</td>
<td>Average slope (degrees) of terrain within 1/4 mile</td>
</tr>
<tr>
<td>SlopeH</td>
<td>Average slope (degrees) of terrain within 1/2 mile</td>
</tr>
<tr>
<td>BikeSym</td>
<td>Presence of bicycle markings on any approach (dummy)</td>
</tr>
<tr>
<td>NComPropT</td>
<td>Number of commercial properties within 1/10 mile</td>
</tr>
<tr>
<td>UCBDist</td>
<td>Network distance to UC Berkeley Campus edge (mi.)</td>
</tr>
<tr>
<td>lnUCBDist</td>
<td>Natural log of network distance to UC Berkeley Campus edge (mi.)</td>
</tr>
<tr>
<td>TransDist</td>
<td>Network distance to nearest BART station (mi.)</td>
</tr>
<tr>
<td>CCDist</td>
<td>Network distance to Oakland City Center (mi.)</td>
</tr>
<tr>
<td>IDensT</td>
<td>Intersection density within 1/10 mile</td>
</tr>
<tr>
<td>IDensQ</td>
<td>Intersection density within 1/4 mile</td>
</tr>
<tr>
<td>IDensH</td>
<td>Intersection density within 1/2 mile</td>
</tr>
<tr>
<td>CNRT</td>
<td>Connected node ratio within 1/10 mile</td>
</tr>
<tr>
<td>CNRQ</td>
<td>Connected node ratio within 1/4 mile</td>
</tr>
<tr>
<td>CNRH</td>
<td>Connected node ratio within 1/2 mile</td>
</tr>
<tr>
<td>Count09</td>
<td>Indicates counts conducted in 2009 (dummy)</td>
</tr>
</tbody>
</table>

1. 1 mile = 1.61 km
2. Bicycle markings include on-street bicycle facilities: bicycle lanes, bicycle boulevard markings, or shared lane markings. They do not include bicycle route signs on approaches without other street markings. This variable does not represent the association between multi-use trails and bicycle volume. Multi-use trail proximity was tested in a different variable that was not significant in the predictive models.

Factors Associated with Intersection Bicycle Volumes

Several options were considered during the model specification process. The models discussed are shown in Table 4. While the adjusted R-square values varied among the models, the F-statistics showed that all models were significant. Model A was the best fit model specification based on all counts ($N = 162, F = 33.9$). Number of retail commercial properties within 1/10 mile and presence of on-street bicycle facilities both had positive coefficients in this model, indicating that retail activity and bicycle facilities
were associated with higher levels of bicycle activity at the 81 study intersections. The average slope within 1/2 mile of each study intersection had a negative coefficient, indicating that cyclists tended to avoid hilly terrain. These results are consistent with previous research.

The natural log of the distance to the UC Berkeley Campus had a negative coefficient, indicating that the further the intersection was located from the campus, the less likely it was to have a high bicycle volume. The high numbers of bicycles counted at intersections near campus were not surprising since UC Berkeley is the largest employer in Alameda County, the largest university in the Bay Area, and has limited space for on-campus automobile parking. The natural log form of the variable was used because beyond a certain distance the likelihood of bicycling to campus drops dramatically. This variable may also be a proxy for general urban, bicycle-friendliness. Berkeley has a well-connected bicycle boulevard system, and Oakland has more than 80 miles of bicycle lanes and routes (23, 24). Both cities are located in the relatively dense, urban, northwestern corner of Alameda County. When tested, the distance to Oakland City Center variable was also correlated with the bicycle counts, but it was strongly correlated with the UC Berkeley Campus variable. Since the distance to UC Berkeley variable was a better predictor of bicycle volumes, it was used instead of distance to Oakland City Center.

A dummy variable indicating counts performed in 2009 was also included in the model because it was highly significant. This variable is of less interest in this study because it is probably unrelated to the built environment; it could reflect changes in behavior due to gas price shock of Summer 2008, the recession of 2009, or differences in the sample of intersections where counts were conducted in 2008 and 2009.

Table 4. Alternative Bicycle Model Specifications

<table>
<thead>
<tr>
<th>Model Variable</th>
<th>Model A: All Counts</th>
<th>Model B: Weekday</th>
<th>Model C: Weekend</th>
<th>Model D: Weekday Alt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent Variable = 2-hr Intersection Bicycle Count</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>3.776</td>
<td>0.185***</td>
<td>3.899</td>
<td>0.262***</td>
</tr>
<tr>
<td>NComPropT</td>
<td>0.024</td>
<td>0.007***</td>
<td>0.030</td>
<td>0.010***</td>
</tr>
<tr>
<td>BikeSym</td>
<td>0.477</td>
<td>0.163***</td>
<td>0.437</td>
<td>0.230*</td>
</tr>
<tr>
<td>lnUCBDist</td>
<td>-0.458</td>
<td>0.059***</td>
<td>-0.546</td>
<td>0.083***</td>
</tr>
<tr>
<td>SlopeH</td>
<td>-0.517</td>
<td>0.073***</td>
<td>-0.659</td>
<td>0.103***</td>
</tr>
<tr>
<td>CNRH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count09</td>
<td>0.811</td>
<td>0.127***</td>
<td>1.002</td>
<td>0.180***</td>
</tr>
</tbody>
</table>

Overall Model

| Sample size (N) | 162 | 81 | 81 | 81 |
| Adjusted R^2    | 0.505 | 0.600 | 0.386 | 0.450 |
| F-test          | 33.87*** | 24*** | 11.08*** | 17.38*** |

Note: Coeff. = coefficient and Std. Err. = standard error. *** = significant at 99% ( p < .01); ** = significant at 95% ( p < .05); * = significant at 90% ( p < .10). Model variables are defined in Table 3.

Weekday and Weekend Bicycle Volume Models

Although all of its coefficients have intuitive signs and are significant at the 99 percent confidence level, Model A is problematic because the sample includes two counts from each intersection (one from a weekday and one from a weekend time period). Datasets contain unobserved heterogeneity when multiple observations are considered for the same location, and as a result, the standard errors are evaluated incorrectly, which could lead to some variables being seen as more significant than they are. A first approach to avoid this problem, was to estimate Model A with a dummy variable representing weekday versus weekend counts, but it was not statistically significant. Therefore, a second approach was used. Separate models were estimated for the weekday (Model B) and weekend (Model C) counts.
As Table 4 shows, the weekday model (Model B) has the highest adjusted R-square (0.6) of the models presented. The coefficients of Model B are similar to Model A, but the coefficient for the bicycle markings is only significant at 90 percent. The weekend model (Model C), on the other hand, has a lower adjusted R-square than Models A or B. While presence of bicycle markings is highly significant with a larger coefficient, number of commercial properties is less significant with a smaller coefficient than in the previous models. This result may have occurred because recreational bicycle trips are more common on weekends, so the proximity of retail establishments may be less important for attracting weekend bicycle activity. Bicycle markings could have a greater influence on weekends because bicyclists may have more flexibility to seek routes with bicycle facilities or because less experienced riders who tend to prefer marked bicycle facilities are more likely to ride on weekends. Model C also has smaller coefficients for the UC Berkeley variable and the slope variable. This is consistent with the expectation that fewer people would see the campus as a destination on the weekend and recreational riders might seek out hilly terrain.

Finally, in consideration that Models A through C include a variable that is particular to this Alameda County study area (lnUCBDist), a fourth model specification is presented. Model D is an alternative weekday model that excludes the UC Berkeley variable. Model D also adds connected node ratio within 1/2 mile and excludes the commercial property variable. As expected, connected node ratio has a positive coefficient, meaning that areas with fewer dead end streets and cul-de-sacs have more bicycle activity. Population density and employment density were included in a variety of model specifications, but they did not perform well. Although these variables have been used in previous studies, they were not included in the final set of bicycle volume models.

PRACTICAL APPLICATION
Practitioners wishing to apply an intersection bicycle volume model should select among Models B, C, and D based on the type of counts being predicted (weekday or weekend) and their understanding of the built environment in the study area. To extract counts from the models above the following formulation should be used:

\[
Y_i = \sum_{j=1}^{n} X_{ji} \beta_j
\]

where:
- \( Y_i \) = predicted two-hour bicycle count at intersection \( i \)
- \( X_{ji} \) = value of explanatory variable \( j \) at intersection \( i \)
- \( \beta_j \) = model coefficient for variable \( j \)

When working with the loglinear form of a regression model, it is important to understand how to interpret the coefficients correctly. They can be interpreted as the fractional increase in the bicycle volume when the explanatory variable increases by 1 unit. Unlike OLS regression, the relationship between coefficients, explanatory variables, and the predictions is nonlinear. For example, consider the application of Model D to the median variable values for the set of 81 sample intersections. This model produces a count prediction of 15 bicycles within a two-hour period on a typical weekday afternoon. The median value of the connected node ratio variable (CNRH) in the sample is 0.85. Increasing the coefficient of this variable from 4.634 to 5.634 brings the prediction up to 35.1 bicycles, an increase of 134 percent. Another unit increase in the coefficient raises the prediction to 82.4, more than five times the original estimate. Along the same lines, examining the median values in Model B, a 20 percent increase or decrease in the median value for SlopeH changes the predicted counts by -3 percent and 4 percent, respectively. Identical modifications to the median value for NComPropT change the predictions by 36 percent and -26 percent, respectively. This sensitivity of the coefficients should be considered when interpreting models.
CONSIDERATIONS FOR FUTURE RESEARCH

The models presented in this paper are preliminary and are not intended to be a final product. They represent a first attempt at modeling bicycle volumes in one county in Northern California. Bicycle activity in other communities is not expected to have an identical relationship to the built environment. Additional models in other states and countries are needed to provide a more complete picture of how surrounding land use, transportation system, and socioeconomic characteristics are related to bicycle volumes at specific locations. One of the next steps for refining the Alameda County models is to collect counts at additional locations and conduct model validation.

Ideally, the models would have been estimated using count data that was collected during the same time-of-day at each location. In this case, however, counts were taken during different times of the day on weekdays and Saturdays. In theory, the number of bicyclists passing through an intersection between 12 p.m. and 2 p.m. on a given day is likely to be different than the number of bicyclists passing through between 4 p.m. and 6 p.m. To account for these time-of-day differences, dummy variables indicating the count time period were included in several versions of the models. However, the coefficients of these variables were highly insignificant, indicating that controlling for time of day provided little additional value for predicting bicycle volumes at the sample of 81 intersections. Therefore, the time-of-day variables were not included in the models. Prior research on pedestrian volumes in Alameda County dealt with this issue by using several months of automated pedestrian counter data from several locations to extrapolate two-hour counts to weekly counts (21). Unfortunately, since automated bicycle counters must be installed in a roadway, it is more difficult to collect counts from a variety of locations or to capture both directions of a roadway. The available automated bicycle count data from two locations in Alameda County could not be used for extrapolation without making gross assumptions.

Weather effects were not incorporated into the final bicycle volume models. Dummy variables indicating extreme temperatures and cloud cover proved highly insignificant in all model specifications. This is likely due to the lack of variation in weather during the bicycle count periods. The count data were collected in the spring months, and counts were rescheduled if it was raining. Alameda County has a temperate climate, and there were only two counts on days where the temperature was greater than 90°F (32°C) or less than 50°F (10°C). Weather is expected to show a greater effect on bicycle activity in areas where the climate varies more.

Previous studies have found a significant relationship between the socioeconomic characteristics of individuals and neighborhoods and likelihood of bicycling (14,17,25). In Alameda County, there was a moderate correlation (0.3 < |ρ| < 0.5) between the bicycle counts and socioeconomic variables such as surrounding neighborhood income (positive), percentage of rental housing (positive), and percentage of residents under age 18 (negative). However, these variables were less significant than the built environment variables in the final models, so they were not included.

In Alameda County, most retail and employment attractors are in the flatter areas, although the hills are popular with recreational riders. A different set of study intersections could have included more counts along popular recreational routes in the hills. This would likely have produced different model results, particularly for the slope and proximity to retail variables. However, the 81 intersections were selected to ensure representation from areas with different surrounding land uses, transportation infrastructure, and neighborhood socioeconomic characteristics. Intersections were not chosen with the specific intention to capture high bicycle volumes or popular bicycle routes.

The bicycle count dataset included only one zero-count occurrence. As mentioned above, this value was changed to 0.1 so that the natural log could be computed. Since the motivation behind these models is to provide reasonable estimates for planning purposes, rather than absolute accuracy, it is acceptable to assume that a prediction of 0.1 represents no bicycle activity. This approach was expected to have little effect on the models because it only applied to one count in the sample. Study areas with less bicycle activity and multiple zero-count intersections would require a different model structure.

The independent variables included in the final models show statistically significant associations between specific land use and transportation infrastructure characteristics and bicycle volumes. However,
they do not necessarily imply a causal relationship between any of the independent variables and levels of bicycling. For example, adding a bicycle facility to an intersection approach may make the roadway a more attractive place to ride, which could increase the bicycle volume at the intersection. However, communities often add bicycle facilities to roadways that already have high bicycle volumes in order to make conditions more comfortable for existing bicyclists. In this case, a high intersection bicycle volume would precede the bicycle facility. Accordingly, these models are not intended to predict the change in bicycle volume after installation of a bicycle facility. Instrumental variable methods could improve the models by accounting for the potential endogeneity of the bicycle facility variable.

CONCLUSION
The preliminary models presented in this paper are simple tools that can be used to estimate bicycle intersection counts during specific time periods. The analyses performed here contribute to the body of research on relationships between the built environment and bicycle activity. In particular, the models showed that bicycle volumes tended to be higher at intersections:

- Surrounded by more commercial retail properties within 1/10 mile of the intersection
- Closer to a major university
- With a marked bicycle facility on at least one leg of the intersection
- Surrounded by flatter terrain within 1/2 mile of the intersection
- Surrounded by a more connected roadway network

In addition, the models showed several important differences between weekday and weekend intersection bicycle volumes. The models showed that:

- The positive influence of commercial retail establishments on bicycle volumes tended to be greater on weekdays than on weekends
- The positive influence of proximity to UC Berkeley tended to be greater on weekdays than on weekends
- The positive influence of bicycle facilities on bicycle volumes tended to be greater on weekends than on weekdays
- The negative influence of hilly terrain tended to be smaller on weekends than on weekdays

Further refinement and testing in other study areas is necessary to improve count predictions in Alameda County or elsewhere. Bicycle volume counts are an important tool to help understand bicycle crash risk and the impact of the built environment on travel. It is crucial to understand the level of bicycle activity for bicycles to have a greater impact on funding and policy decisions.

ACKNOWLEDGEMENTS
The authors would like to thank Lindsay Arnold and David Ragland of UC Berkeley Safe Transportation Research and Education Center (SafeTREC) for their assistance with the project. Special thanks to Population Research Systems for collecting bicycle counts in Spring 2008 and to all of the volunteers who collected bicycle counts in Spring 2009. This work was partially funded by the California Department of Transportation and the Alameda County Transportation Improvement Authority.

REFERENCES


LIST OF TABLES AND FIGURES

TABLES
Table 1. Factors Associated with Bicycling in Previous Research
Table 2. Summary Statistics for all Two-Hour Bicycle Counts
Table 3. Description of Selected Explanatory Variables Considered in the Modeling Process
Table 4. Alternative Bicycle Model Specifications

FIGURES
Figure 1. Map of Manual Bicycle Count Locations