Journal of Transport Geography 55 (2016) 11–21

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

# Estimating pedestrian and cyclist activity at the neighborhood scale

Deborah Salon

*School of Geographical Sciences and Urban Planning, Arizona State University, 975 S. Myrtle Ave, Tempe, AZ 85287, United States*

ARTICLE INFO

*Article history:*
Received 16 June 2015
Received in revised form 17 June 2016
Accepted 28 June 2016
Available online 16 July 2016

*Keywords:*
Walk
Bicycle
Neighborhood type
Road safety
Small area estimation
Cluster analysis

ABSTRACT

In most parts of the U.S., data on bicycle and pedestrian activity at the neighborhood scale are sparse or non-existent, despite the importance of such data for local planning. Here, a simple small-area estimation method is used to pair travel survey with land use and census data to estimate cyclist and pedestrian activity for census tracts in the state of California. This method is an improvement on fixed per-capita estimates of activity based only on regional or statewide averages. These activity estimates are then used to calculate the intensity of road use by cyclists and pedestrians, and crash rates for these road users. For California, the intensity of pedestrian and cyclist road use in urban census tracts is double that found in suburban tracts, while use in suburban tracts is an order of magnitude greater than that found in rural tracts. Per-capita estimates would suggest substantially smaller differences between neighborhood types. On the safety side, although non-severe crashes involving cyclists and pedestrians are much more likely in more urban areas, severe crash rates for the non-motorized modes exhibit no clear spatial pattern. The method used is simple and easily replicable, potentially filling a critical need for bicycle and pedestrian planners.



## 1. Introduction

Good estimates of the total amount of bicycle and pedestrian activity on our roads are needed for two main purposes. First, knowing how much cyclists and pedestrians are using roadways can inform where investments in bicycle and pedestrian infrastructure are needed. Second, estimates of total cyclist and pedestrian activity can serve as the denominator for calculation of cyclist and pedestrian crash rates, which, in turn, help to identify locations for road safety investment. While estimates of vehicle activity are readily available from routinely collected traffic counts as well as travel demand forecasting models, spatially detailed estimates of bicycle and pedestrian activity rarely are, as few communities conduct regular counts of pedestrians or bicyclists and few models generate estimates of the use of these modes.

This paper describes and implements a simple small-area estimation method for estimating cyclist and pedestrian activity in census tracts based on a combination of travel survey, census, and land use data. Cluster analysis is used to categorize census tracts into neighborhood types, and these neighborhood types are used to aggregate spatially sparse travel survey observations in a meaningful way to obtain estimates of travel activity for each tract. Two sets of activity estimates are calculated based on two different household-based travel surveys recently conducted in California, providing a robustness check on the results. The results are a substantial improvement over fixed per-capita estimates of activity based only on regional or statewide averages.

These tract-level activity estimates then are used to calculate two important policy indicators: intensity of road use by cyclists and pedestrians, and crash rates for these road users. The results show that roads are used most intensively for cycling and walking in the most densely populated neighborhoods of the state. The intensity of pedestrian and cyclist road use in urban census tracts is double that found in suburban tracts, which is again double that found in rural tracts. On the safety side, although non-severe crashes involving cyclists and pedestrians are much more likely in more urban areas, severe crash rates for the non-motorized modes exhibit no clear spatial pattern. The method presented is purposefully simple, and could be implemented by pedestrian and bicycle planners themselves.

## 2. Background

Estimation of total bicycle and pedestrian activity is hampered by a lack of basic data. The main sources of bicycle and pedestrian data are household-based travel surveys. One problem with these surveys is that they lack full spatial coverage. For example, at the geographic resolution of the census tract, there are more than 2500 tracts in California that were not sampled at all by the 2009 National Household Travel Survey (NHTS), and only 15 of the sampled tracts include more than 30 household observations. The 2010–12 California Household Travel Survey (CHTS) has impressive coverage of the state's census tracts, with

*E-mail address:* deborah.salon@asu.edu.

http://dx.doi.org/10.1016/j.jtrangeo.2016.06.023
0966-6923/© 2016 Published by Elsevier B.V.

zero observations in only 550 out of 8057 total tracts in the state (USDOT, 2011; CDT, 2013). However, even this large sample only includes 52 tracts in which the number of household observations is 30 or greater. This sparse spatial coverage is especially problematic for understanding bicycle and pedestrian activity, which itself is relatively sparse.

To overcome this limitation, most studies in the travel safety literature aggregate pedestrian and cyclist activity by metropolitan area (McAndrews, 2011), state (McAndrews et al., 2013; Teschke et al., 2013), or even the national level (Beck et al., 2007; Mindell et al., 2012; Dhondt et al., 2013). The focus of these studies is to estimate the relative safety of different modes of travel by gender, age, and ethnicity. They compare the safety results obtained using different measures of total travel activity (e.g., population, number of trips, distance traveled, and time spent traveling). Zhu et al. (2008) offer an exception, using the 2001 NHTS data to estimate pedestrian activity in four types of built environments in New York State. However, the built environment types in Zhu et al. (2008) are identified at the geographic scale of the Metropolitan Statistical Area (MSA).

A sizable number of studies modeled pedestrian and cyclist volumes at the level of the intersection or roadway link, based on original pedestrian count data collection at a sample of locations in an area (e.g., Pulugurtha and Repaka, 2008; Griswold et al., 2011; Miranda-Moreno et al., 2011; Hankey et al., 2012; Schneider et al., 2012). They used regression analysis of pedestrian and cyclist counts along with characteristics of the count locations to estimate a model that can predict volumes at all locations in a city. New work in this line of research augments the intersection count data for cycling with GPS cycle route data volunteered by users of the Strava cycle fitness application (Jestico et al., 2016).

The above-referenced studies predict intersection-specific pedestrian and cyclist volumes, but do not take the next step to use these data to estimate total exposure measures such as distance traveled or time spent traveling. Intersection and link volumes can help identify where cycle infrastructure and pedestrian signals could be most useful, and they can be used to estimate intersection-level crash rates. However, area-wide exposure measures are needed to estimate area-level crash rates. Molino et al. (2012) extended this method to generate distance-based exposure measures for crash rate analysis in Washington, D.C. The model is data-intensive and, to my knowledge, this method is not yet implemented in practice.

Similar to the work presented here, Turner et al. (1998) estimated the census tract-level spatial patterns of total pedestrian and cyclist activity, using spatially sparse data to estimate activity rates and census data to extrapolate these rates to tracts. However, this study employed only socio-demographic information to estimate walking and bicycling rates, rather than using socio-demographic information together with neighborhood typologies, as proposed here.

Where there is good spatial coverage of count data, both volumes and exposure from analyses of these data can be estimated at high levels of spatial resolution. Unfortunately, there is not good spatial coverage of count data in most areas, and the methods for translating sparse count data into full volume and exposure estimates are complex and data-intensive, requiring both count data and detailed measures of the built environment. The advantage of the approach presented in this paper is that the data are readily available for most jurisdictions and the method is computationally simple; the disadvantage is that the result lacks the spatial resolution possible with direct counts.

## 3. Method

In the absence of comprehensive counts of bicyclists and pedestrians, the method presented in this paper relies on data for bicycle and pedestrian activity from two household-based travel diary surveys: the 2009 National Household Travel Survey (NHTS) and the 2010–2012 California Household Travel Survey (USDOT, 2011; CDT, 2013). Reliance on household-based surveys means that this method produces estimates of walking and biking by the residents of each census tract, regardless of where these trips are made, rather than estimates of miles walked and biked within the geographic area of each tract. In other words, the specific research question the method is designed to answer is: How many total miles are walked by pedestrians and biked by cyclists living in each census tract in California? However, because most walk and bike trips are short and begin or end at home (e.g., 76% of NHTS walk trips and 87% of NHTS bike trips), the estimates derived from the method should be highly correlated with actual miles walked and biked within the geographic area of each tract. Notable exceptions to this include downtowns, university campuses, major employment centers, and other areas with high volumes of walking by commuters or visitors.

The method used here is one of the simpler techniques in the family of small-area estimation, a version of the Broad Area Ratio Estimator with Auxiliary Data (see ABS (2006) for an accessible overview of small area estimation). It requires four steps. First, cluster analysis is used to assign census tracts to neighborhood types based on built environment characteristics. Second, daily miles biked and walked are calculated for each travel survey respondent. Third, each survey respondent is assigned to a category based on their age, gender, and home neighborhood type, with daily average miles biked and miles walked calculated for each respondent category. Finally, these category averages are used to generate estimates of bicycle and pedestrian activity for a given area by multiplying the average miles by the population in each category, as reported in census data. This paper presents estimates of total daily miles of walking and bicycling for all census tracts in California circa 2010.

### 3.1. Classifying census tracts into neighborhood types

To classify census tracts into neighborhood types, k-means cluster analysis is used. This method takes multiple pieces of information about each census tract as the input and organizes the tracts into groups that are similar to each other. The analyst chooses how many groups to create and which variables to use as the input data, and these choices are informed by the analyst's judgment and by a process of testing a variety of input variable forms and numbers of groups.

Here, ten variables representing different aspects of the built environment in each census tract in California are used as inputs. These 10 variables were chosen collectively to represent physical characteristics of the tracts: two types of density, two representing local accessibility, one representing regional accessibility, one representing bicycle and pedestrian friendliness, three characterizing the housing stock, and one providing an indicator of housing values. Most of these variables are self-explanatory, but two that deserve further explanation are local and regional job access. The data used to create the two job access variables are census block group counts of total jobs from the 2010 Longitudinal Employer-Household Dynamics (LEHD) dataset. Local job access is captured by the inverse distance-weighted sum of the jobs within five miles, and regional job access is the inverse distance-weighted sum of all jobs between 5 and 50 miles from a tract. All variables are standardized prior to cluster analysis.

From the cluster analysis of these 10 variables for California's census tracts, four neighborhood type clusters emerge. The ten variables and their data sources are listed in Table 1, along with the means of standardized versions of each of these variables for each neighborhood type cluster. Standardized variables have means of zero for the full sample, so looking at means of these variables for each cluster provides information about how that neighborhood type's census tracts are different from the average for the whole state. For instance, the first row of Table 1 indicates that the cluster of tracts labeled “Suburb” is slightly less dense than the state average, that “Urban” tracts are substantially denser than the state average, that “Rural” tracts are substantially less dense than the state average, and that “Central City” tracts are much denser than the state average.

**Table 1**
Mean values of standardized variables within each neighborhood type.

| | Source | Rural (N = 2042) | Suburb (N = 3776) | Urban (N = 1978) | Central city (N = 180) |
|---|---|---|---|---|---|
| Population density | 2010 Decennial Census | −0.69 | −0.18 | 0.76 | 3.41 |
| Road density | ESRI North America Detailed Streets | −1.13 | 0.11 | 0.88 | 1.09 |
| Restaurants within 10 min walk | MapQuest Point Of Interest | −0.29 | −0.18 | 0.25 | 4.41 |
| Local job access | 2010 LEHD | −0.69 | −0.23 | 0.81 | 3.83 |
| Regional job access | 2010 LEHD | −0.88 | −0.06 | 0.99 | 0.41 |
| Pct. walk/bike commuters | 2011 5-year ACS | −0.17 | −0.18 | 0.24 | 2.57 |
| Pct. single family detached | 2011 5-year ACS | 0.53 | 0.16 | −0.67 | −1.98 |
| Pct. old housing | 2011 5-year ACS | −0.45 | −0.40 | 1.09 | 1.43 |
| Pct. new housing | 2011 5-year ACS | 0.99 | −0.34 | −0.37 | −0.06 |
| Median house value | 2011 5-year ACS | −0.39 | 0.03 | 0.21 | 0.73 |

Fig. 1 depicts the spatial distribution of the neighborhood type clusters for three urban regions in California: the San Francisco Bay Area, Los Angeles, and Sacramento. The neighborhood types cluster spatially and largely appear as expected. A small number of tracts were not assigned to a neighborhood type due to missing data.

*3.2. Calculating daily miles walked and biked from travel surveys*

Both to provide a point of comparison and to check the robustness of the results, two travel survey datasets were used to obtain two independent estimates of pedestrian and cyclist activity in California: the California portion of the 2009 National Household Travel Survey (NHTS) and the 2010–2012 California Household Travel Survey (USDOT, 2011; CDT, 2013). The two surveys are similar in most respects, and both include a full 24-hour travel diary in which every person in surveyed households provided the full details of their travel and activities for an assigned day.

The results here focus on those individuals who were surveyed on a weekday, provided sufficient information for analysis, and who were not outliers in their walking or bicycling distances. Outliers were identified as any person who reported walking more than nine miles in one day, and any person who reported bicycling more than 30 miles in one day. These two distance thresholds are roughly equivalent to spending 3 hours traveling by these modes. Dropping these outliers removed less than 1% of the observations from each survey.

The final samples included nearly 32,000 individuals from the NHTS and nearly 68,000 individuals from the CHTS surveys whose data were used to calculate total miles walked and miles biked for each survey respondent on the travel-diary day. The CHTS included calculated road network trip distances for most trips, and these distances were calculated for NHTS trips using the MapQuest API. For approximately 10% of NHTS trips, self-reported distances are used. For approximately 2.5% of CHTS trips, distances are estimated using self-reported travel times and assuming an average walking speed of three miles per hour and an average biking speed of 10 miles per hour. Survey respondents were dropped from the analysis if they walked or made a bike trip and did not provide exact origin and destination information, a self-reported trip distance, or a self-reported travel time.

*3.3. Averages by category*

In this step, each respondent was assigned to a socio-demographic and neighborhood type category, and the average daily walking and biking distance per person was calculated for each category. Here, I describe a key adjustment that was implemented for the CHTS data, define the categories, and present summary results.

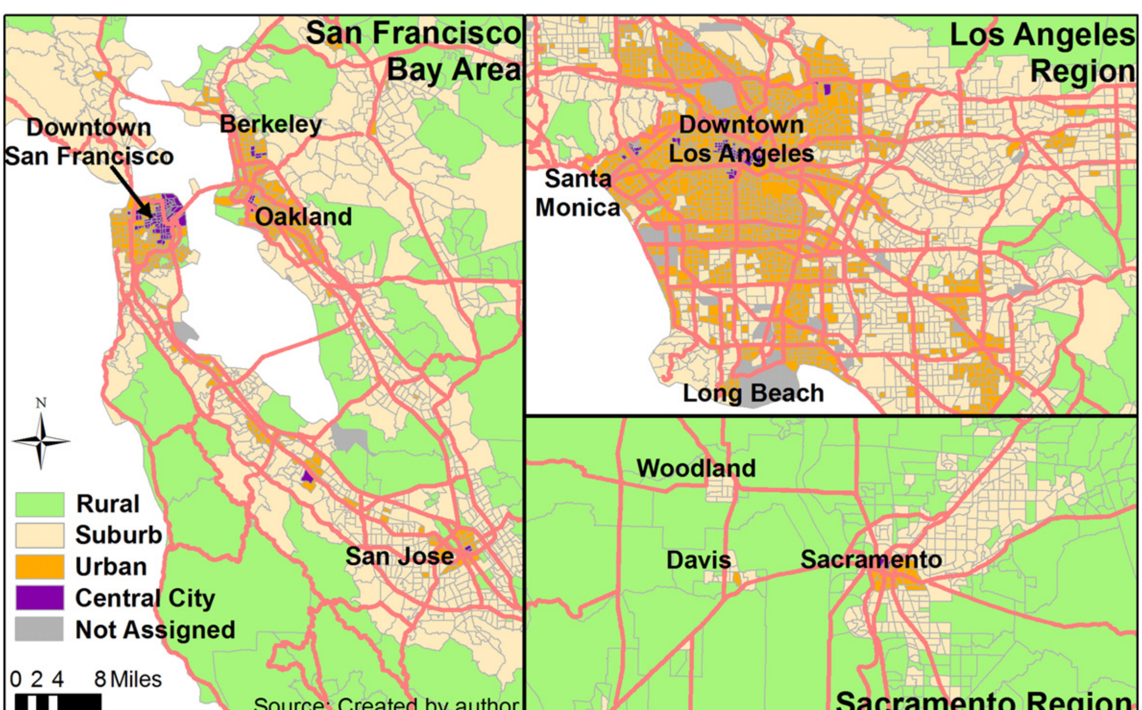


**Fig. 1.** Neighborhood type map of three urban regions in California.

*3.3.1. Key CHTS data adjustment*

An important difference between the surveys is in the percentage of respondents who reported making zero trips on the travel diary day. Overall, this fraction was 12% among weekday respondents to the NHTS, and 21% among weekday respondents to the CHTS. Evidence from the existing literature suggests that the lower percentage of immobiles in the NHTS is likely to be closer to the actual immobile rate in the population (Madre et al., 2007). In addition, the subset of weekday CHTS respondents who used a wearable GPS device (nearly 12,000 individuals) reported an immobility rate of 14%, which is substantially lower than the 21% immobile rate in the full sample. To address this discrepancy and make the results comparable across the surveys, the CHTS results presented here are adjusted such that the percent of immobiles in each gender-age category is equal to the percent of immobiles for that category in the subset of the CHTS respondent sample that used a wearable GPS device.

*3.3.2. Category definition*

Age categories are based on the travel survey data; ages with similar levels of biking and walking activity are grouped together. Two genders, five age categories, and four neighborhood types were identified, producing a total of 40 gender-age-neighborhood type categories for calculating average miles walked and biked. With so many categories, the question then arises of whether the number of survey respondents in each category is sufficient to yield a robust estimate for the population in that category. All categories in the neighborhood types Urban, Suburb, and Rural contained a large number of individual observations.

However, in the Central City neighborhood type, the number of individual survey respondents in each age-gender category was much smaller. For this reason, the number of categories in the Central City neighborhood type was reduced to only three for walking (ages 5–17, 18–74, and over 74), categorized by age only. Similarly for cycling, the number of categories was reduced to two but, in this case, the split was by gender. These decisions were made by examining the actual distributions in the data, and pooling the original set of categories together where their average values were similar. The final set of averages by category is available online (Salon and Handy, 2014).

*3.3.3. Average walking and biking by age, gender, and neighborhood type: Results for California*

The first finding of note is that neither walking nor bicycling are undertaken at all by most respondents to either survey. As is evident from Table 2, less than 20% of people reported any walking, and less than 3% of people reported any biking. These respondents thus traveled zero miles by foot or bike, and these zeroes are included in the calculation of average distances by category.

For those NHTS respondents who reported at least one walk or one bike trip, Fig. 2 illustrates the distributions of distances walked and biked. The majority of pedestrians and cyclists don't walk or cycle very far. Approximately half of all pedestrian respondents to the survey walked less than one mile on the travel diary day, and a third of cyclists biked less than two miles on that day. Distributions based on the CHTS survey data are similar.

Age and gender differences in bicycling rates and, to a lesser extent, walking rates are well documented (e.g., Krizek et al., 2005; Lee and Moudon, 2006a). For the CHTS and NHTS respondents, the walking percentages are similar for men and women, while biking is substantially more likely among men than among women survey respondents. Among those who do walk or bike, however, the average distances traveled are only slightly lower for women than for men. Age differences are also notable for the respondents. The patterns across age groups are similar for the percent of respondents who walked and biked, with children being most likely to walk or bike, and the likelihood of walking or biking declines with age. The distances for both activities are highest in the middle adult years.

Differences in walking and bicycling across geographic settings are also well documented (e.g., Handy et al., 2002; Lee and Moudon, 2006b; Saelens and Handy, 2008). For NHTS and CHTS respondents living in different types of neighborhoods, the patterns are roughly as expected. People living in dense urban neighborhood types are more likely to walk than those in less dense neighborhood types. Among those who walked at all, walking distances do not vary substantially across neighborhood types, though the average distance walked is longest in Central City neighborhoods. The survey sample results diverge in the pattern for biking. NHTS central city dwellers are less likely to bike than Urban neighborhood residents, but both groups are more likely to bike than residents of Suburb and Rural neighborhoods. Based on CHTS

**Table 2**
Summary of walking and biking trips on survey day.

| | N | Adjusted N | Percent walked at all | | Mean miles walked (for walkers) | | Percent biked at all | | Mean miles biked (for bikers) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | NHTS | CHTS | NHTS | CHTS | NHTS | CHTS | NHTS | CHTS | NHTS | CHTS |
| Total | 31,715 | 67,910 | 18.1% | 16.5% | 1.50 | 1.36 | 1.8% | 2.5% | 4.99 | 5.15 |
| *Gender* | | | | | | | | | | |
| Male | 14,903 | 32,984 | 17.8% | 16.1% | 1.52 | 1.36 | 2.7% | 3.5% | 5.28 | 5.58 |
| Female | 16,812 | 34,926 | 18.4% | 16.8% | 1.48 | 1.35 | 0.9% | 1.5% | 4.23 | 4.18 |
| *Age groups (walk)* | | | | | | | | | | |
| 5–10 | 1686 | 4281 | 22.2% | 24.3% | 1.00 | 0.88 | | | | |
| 11–17 | 3276 | 7917 | 24.4% | 26.9% | 1.45 | 1.22 | | | | |
| 18–59 | 16,137 | 36,829 | 18.4% | 15.5% | 1.62 | 1.48 | | | | |
| 60–74 | 6832 | 14,810 | 16.6% | 13.5% | 1.73 | 1.41 | | | | |
| 75+ | 3754 | 4073 | 12.5% | 7.5% | 1.69 | 1.35 | | | | |
| *Age groups (bike)* | | | | | | | | | | |
| 5–10 | 1686 | 4284 | | | | | 3.1% | 2.5% | 2.03 | 2.15 |
| 11–34 | 7025 | 17,557 | | | | | 2.9% | 3.4% | 3.76 | 3.94 |
| 35–59 | 12,388 | 27,271 | | | | | 1.8% | 2.6% | 6.60 | 6.24 |
| 60–69 | 4987 | 11,902 | | | | | 1.1% | 1.8% | 6.13 | 6.46 |
| 70+ | 5629 | 7016 | | | | | 0.5% | 1.0% | 4.70 | 5.28 |
| *Neighborhood types* | | | | | | | | | | |
| Central City | 244 | 1108 | 41.8% | 57.8% | 1.67 | 1.66 | 2.1% | 5.5% | 3.31 | 4.51 |
| Urban | 4043 | 11,405 | 26.0% | 28.9% | 1.40 | 1.42 | 2.3% | 3.6% | 5.90 | 5.31 |
| Suburb | 17,130 | 29,904 | 18.1% | 15.1% | 1.54 | 1.34 | 1.7% | 2.6% | 5.00 | 5.22 |
| Rural | 10,502 | 25,493 | 14.6% | 10.7% | 1.48 | 1.24 | 1.5% | 1.7% | 4.52 | 4.94 |

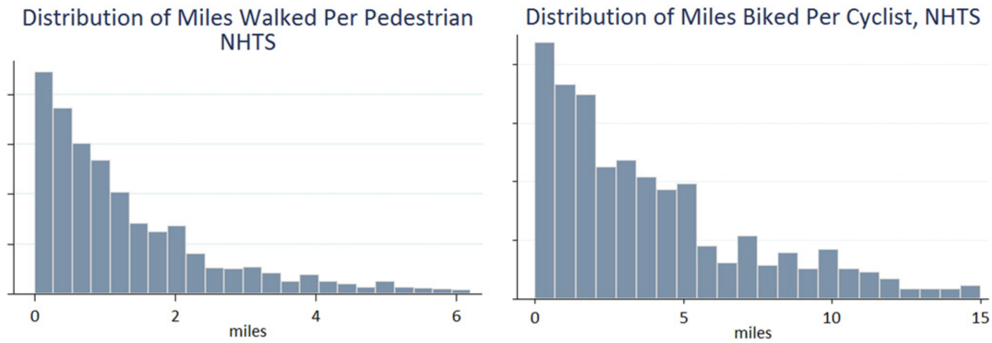


**Fig. 2.** NHTS distributions of distances walked and biked (excluding non-walkers and non-bikers).

respondents, it appears that the likelihood of biking at all decreases monotonically with density. Results of both surveys indicate that, among those who biked, Central City bikers travel shorter distances on average than bikers in the other three neighborhood types. Because of the extremely small number of Central City respondents who biked at all in the NHTS sample, I do not report NHTS Central City biking results in the remainder of this paper.

### *3.4. Using category averages to estimate walking and biking by census tract*

The final step in the method is to use the category averages to estimate total walking and biking at the population level. Eq. (1) estimates miles of walking and bicycling in 2010 for all tracts in California:

$$TotalMiles_{tract} = \sum_{\boldsymbol{i}=1}^{10} SurveyAvgMiles_{\boldsymbol{i}} * 2010Population_{tract,\boldsymbol{i}} \tag{1}$$

where $i$ indexes gender-age group categories, and each tract is classified as a neighborhood type.

These estimates range from a low of five miles walked and 1.5 miles biked (in the same census tract with only 20 residents) to a high of more than 7000 miles walked (in a tract with 11,500 residents) and just over 4000 miles biked (in a tract with over 37,000 residents).

## 4. Limitations

There are three limitations that bear mention. First and foremost, there is no way to ground truth the estimates. This is the nature of all small-area estimation, but it is worth emphasizing. Travel survey samples are not designed to be representative of the population at small geographies. This means that these survey data do not provide reliable estimates of biking and walking at the geographic level of the census tract, or even at the level of the city. The largest of these surveys has only very small samples (or no sample) in some geographic areas. I do compare the small area estimates with regional direct estimates based on the survey data (see Section 5.1), but it could be argued even at this level of geographic aggregation that the survey data are not reliable.

A second limitation is the quality of the travel survey data themselves. Historically, short trips and non-motorized trips have been undercounted in travel surveys (Weinstein and Schimek, 2005). Improved survey methods have increased the number of bicycle and pedestrian trips captured in travel surveys (Clifton and Krizek, 2004), but this is likely still an issue. As will become apparent in Section 5 of this paper, the NHTS and CHTS data provide substantially different estimates of walking and biking, even aggregated to the region level. In addition, the two surveys yielded different estimates of the percentage of the population that does not travel on any particular day. This difference could stem from differences in the sampling methodology, survey design, or survey implementation.

Finally, as noted earlier, this method estimates total walking and bicycling *by the residents of each tract*. These numbers are similar to, but not the same as, estimates of the total walking and bicycling *that occurs in that tract*. The numbers will be most similar in areas where those walking and cycling are also likely to be residents. There are two situations in which this is not true. The first is the exception of tracts that are activity centers, such as downtowns or university campuses. The second is the case of geographically small tracts, often seen in dense urban areas. In these areas, even short trips often traverse more than one tract. In this case, the estimates presented here would be valid only if, on average, the boundary-crossing trips went both ways.

## 5. Quality of the estimates

There is not a straightforward way to test the accuracy of the estimates produced using small area estimation techniques because valid data at the small area level simply are not available. To get an idea of how reliable the estimates are, I take two approaches. First, I compare the small area estimates (SAE) to direct estimates of the average miles walked and biked per person at the regional level for each of the surveys. Second, I calculate the correlation between NHTS- and CHTS-based independent estimates of average miles biked and walked per person in each gender-age-neighborhood type category.

### *5.1. Comparison of SAE to direct estimates for California regions*

The SAE uses statewide individual reports of residential neighborhood type, age, and gender to estimate average miles walked and biked per person in each category. Region-level estimates based on these numbers simply reflect how many people are in each category in the region. The direct regional estimates reported here are the average miles walked and biked by survey respondents in the region, weighted so that the survey respondents in each region are representative of the actual distribution of age and gender within the regional population. Here, I describe these calculations more precisely, present the results in Figs. 3 and 4, and discuss them.

#### *5.1.1. Regional average of small area estimates*

Calculating the SAE for each region is straightforward. Eq. (1) provides estimates of miles walked and miles biked for each tract in the state. The SAE-based regional estimates are calculated by adding up these miles by region and dividing by the total population over four years of age. Eq. (2) specifies this calculation, where there are $J$ tracts in each region and $TotalMiles_j$ is given by Eq. (1).

$$SAEMilesPerCapita_{region} = \sum_{\boldsymbol{j}=1}^{J} TotalMiles_{\boldsymbol{j}} / \sum_{\boldsymbol{j}=1}^{J} TotalPopulation_{\boldsymbol{j}} \tag{2}$$

#### *5.1.2. Regional average of direct estimates, NHTS*

Calculating the direct regional estimates from the NHTS data is straightforward as well. Using the same raw survey data that was

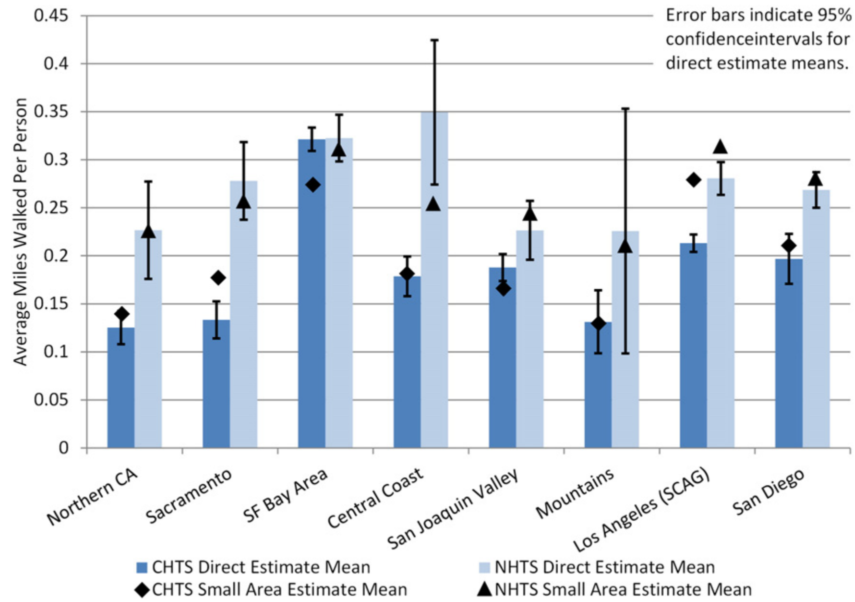


**Fig. 3.** Comparison between direct regional estimates of average miles walked and regionalized small area estimates.

used to create the small area estimates, I calculated the category-weighted average miles walked and biked per person by region. The weights in Eq. (3) adjust the respondent sample to be representative of the actual regional population in terms of the age-gender-neighborhood type categories.

$$DirectMilesPerCapita_{region} = \frac{\sum_{n=1}^{N} w_n \times TotalMiles_n}{N} \tag{3}$$

where $n$ indexes survey respondents and $w_n$ = region percent in category/sample percent in category, and there are $N$ survey respondents from the region.

### 5.1.3. Regional average of direct estimates, CHTS

The direct regional estimates from the CHTS data are also calculated using Eq. (3), but the calculation is complicated by the need to adjust the data for respondents who made zero trips on the travel diary day. This adjustment affects the number of respondents for whom *TotalMiles* is zero and, therefore, also affects both the total number of respondents $N$ and the "sample percent in category" part of the weight.

To adjust the CHTS data to reflect the estimated number of respondents who made zero trips in each region, I made the following assumptions:

1. The NHTS regional variation in the fraction of respondents reporting zero trips is assumed;
2. The CHTS regional variation in the fraction of zero-trip respondents in each gender-age category is assumed; and
3. The total number of zero trip respondents across the state is the same as in the small area estimates.

Figs. 3 and 4 depict the comparison between SAE-based regional results and direct regional estimates from each of the two surveys. The height of the bars represents the direct regional estimate of average miles walked and biked per person for each survey, with error bars

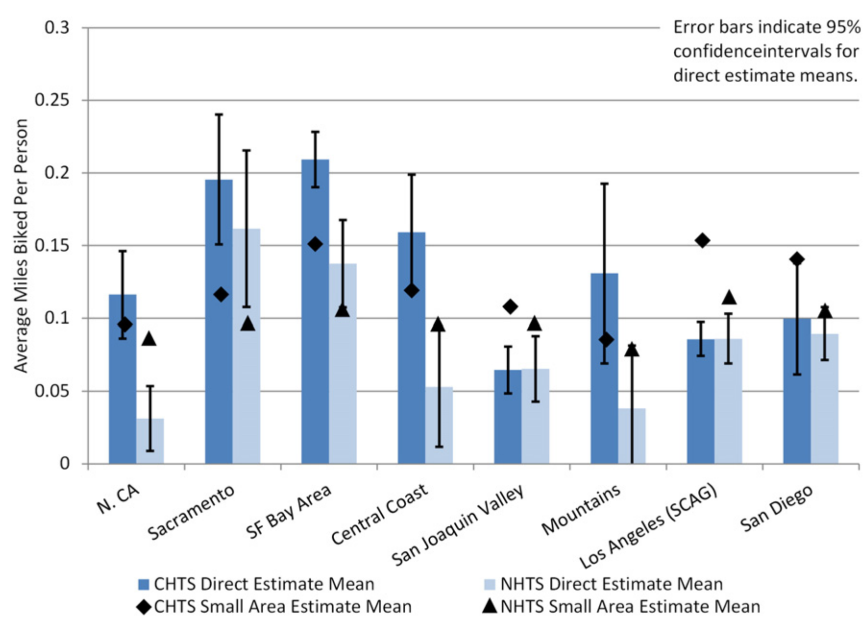


**Fig. 4.** Comparison between direct regional estimates of average miles biked and regionalized small area estimates.

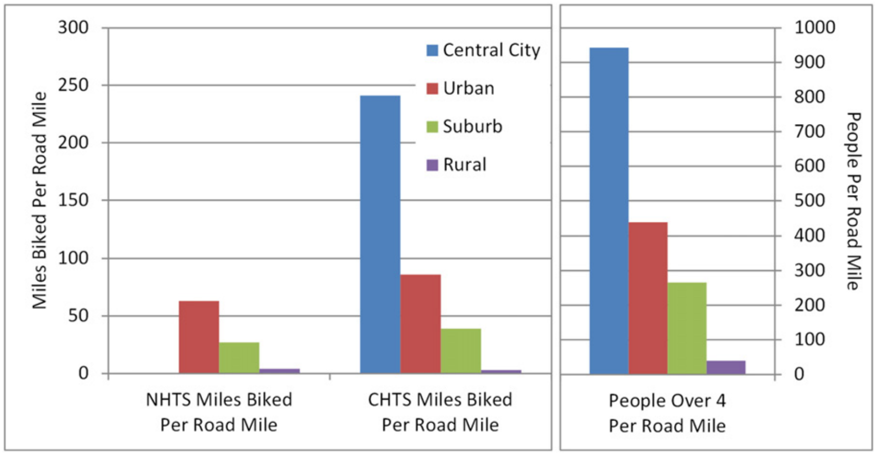


**Fig. 5.** Pedestrian road use intensity by neighborhood type and population per road mile.

providing a 95% confidence interval for this average value. The diamond and triangle symbols represent the SAE means for the CHTS- and NHTS-based estimates, respectively.

Looking at Fig. 3, the first point of note is that the CHTS estimates of walking are lower across the board than those from the NHTS, and their 95% confidence intervals do not overlap for most regions. Fig. 4 indicates that, for biking, the CHTS estimates are higher than the NHTS estimates in most regions, and confidence intervals overlap for only half of the regions. These comparisons provide evidence that even direct mean estimates from travel surveys are not reliable at the regional level.

Comparing the SAE and direct regional estimates of walking for each survey, I find good agreement in six out of eight California regions using the NHTS data; these SAEs are within the 95% confidence intervals for the direct mean estimates. This agreement occurs in half of the regions for the CHTS-based estimates. Interestingly, the SAE from both surveys suggests more walking in the Los Angeles region than the direct regional estimates indicate.

Comparing the regional estimates of cycling in Fig. 4 tells a different story. The SAE-based regional estimates from the NHTS data are approximately 0.10 miles for most regions, whereas there is substantial variation in the direct regional estimates. The CHTS-based SAE and direct regional estimates both exhibit more variation, but are not necessarily consistent with one another. One reason for the apparent unreliability of these estimates is that the number of survey respondents who biked at all is small (570 in the NHTS versus 1700 in the CHTS). Another explanation could be that variation in cycling infrastructure between regions is substantial, and is not incorporated in the SAE values. This is consistent with the finding that the CHTS-based SAE values are lower than the regional direct estimates for places with good bike infrastructure (Bay Area, Sacramento), and higher than the regional direct estimates for places with worse-than-average bike infrastructure (Los Angeles, San Joaquin).

### *5.2. Correlation between NHTS- and CHTS-based small area estimates*

Looking at correlations between the two independent NHTS- and CHTS-based estimates of average miles walked and biked per person in each category produces encouraging results. Overall correlations are 0.81 for miles walked and 0.85 for miles biked. The correlations are higher in Urban and Suburb neighborhood types (0.9 or higher in all cases), and substantially lower in Rural neighborhood types in spite of relatively large sample sizes (0.6 or lower). Because of small sample sizes in Central City neighborhoods, there are effectively only two categories for miles biked and three for miles walked, so correlations are not particularly meaningful. Taken together, this evidence indicates that the estimates in Urban and Suburb areas are the most robust.

It is important to note that, although the correlation is high between the estimates developed with the two surveys, the actual numbers differ quite substantially. This suggests that the estimates are best used to compare relative differences between tracts, rather than as absolute numbers.

## 6. Application examples

The estimates of miles walked and biked by census tract can be combined with other data to produce metrics useful for planning. The first

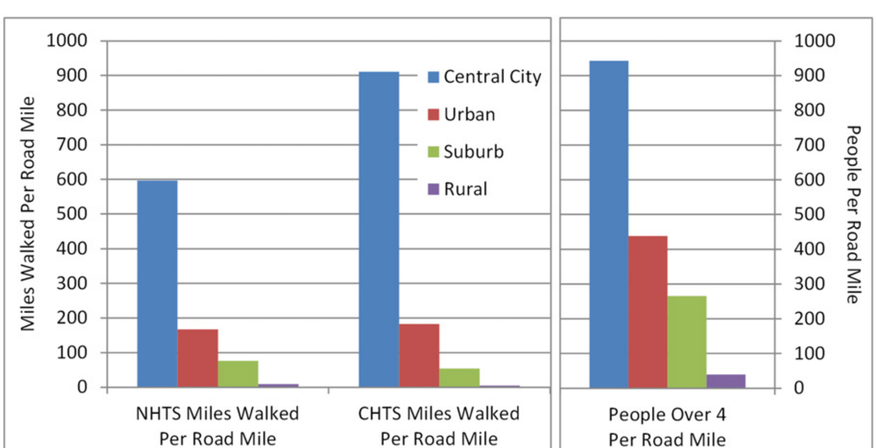


**Fig. 6.** Cyclist road use intensity by neighborhood type and population per road mile.

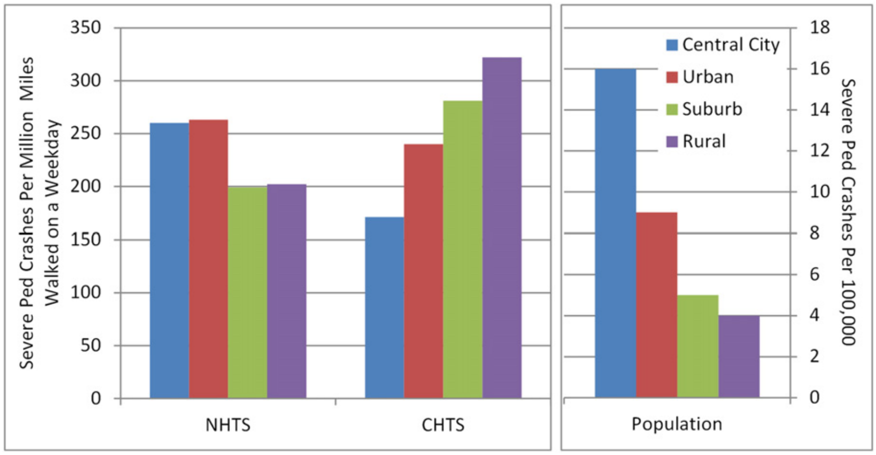


**Fig. 7.** Annual severe pedestrian crash rates by neighborhood type.

metric divides estimated miles of walking and biking by walkable (i.e., non-highway) road miles. The results provide information about where roads are being used most intensively by pedestrians and cyclists, which can then be used to prioritize non-motorized infrastructure needs. The second metric divides pedestrian and cyclist crash data by estimates of walking and biking miles to calculate a crash rate for each census tract, which will shed light on which parts of the state are relatively safe or especially dangerous for these activities. In each case, the metrics developed here are compared to a default population-based metric.

### 6.1. Intensity of roadway use by cyclists and pedestrians

To prioritize bicycle and pedestrian infrastructure needs in different census tracts, it is useful to know where the roads are most heavily used by cyclists and pedestrians. This is often accomplished by manual counts at specific intersections, but manual counts are labor intensive and, therefore, cannot provide comprehensive spatial coverage. The ratio of the estimates of total miles walked and biked to the total miles of non-highway roads in each tract provides an indicator for pedestrian and cyclist intensity of road use in each tract.

Fig. 5 depicts both NHTS- and CHTS-based small area estimates of the mean miles walked per non-highway road mile for each neighborhood type, with the distribution of population per road mile for comparison. Fig. 6 depicts this information for miles biked. Overall, roads and sidewalks are most heavily used by pedestrians and cyclists in the most densely-populated neighborhoods of the state. This overall pattern is expected, but the particulars of the estimated rates of use are informative. Specifically, roads in Urban neighborhoods are used more than twice as heavily by both pedestrians and cyclists as roads in Suburb neighborhoods. Results based on both surveys indicate that Central City road use by pedestrians is approximately an order of magnitude (10 times) greater than Suburban pedestrian road use. Both cyclists and pedestrians use Rural roads at an extremely low level.

The walking results contrast sharply with the estimates of population per road mile, which indicate only a three-fold difference between Central City and Suburb neighborhood types. This is because people in more urban neighborhood types walk much more than people in rural areas. For cycling, we see the same pattern, but it is somewhat less pronounced because there is not as much of a difference between neighborhood types for cycling.

### 6.2. Cyclist and pedestrian crash rates

A second important application of estimates of total miles walked and biked is to gain a better understanding of the relative safety of walking and cycling at a detailed geographic resolution. To accomplish this, activity estimates are merged with California's Statewide Integrated Traffic Records System (SWITRS) data on pedestrian and cyclist crashes by census tract for the years 2003–2012 (TIMS, 2014). With this information, crash *rates* can be calculated by census tract: the number of pedestrian crashes that occurred in each tract per mile walked, and the number of cyclist crashes that occurred in each tract per mile biked. Crashes are divided into two categories: severe crashes are those that resulted in severe injury or death for the pedestrian or cyclist, and non-severe crashes are those that resulted in a non-severe visible injury

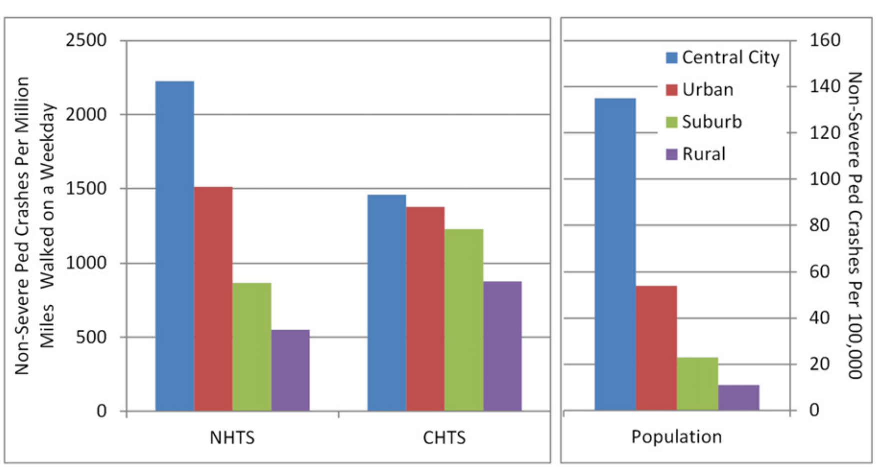


**Fig. 8.** Annual non-severe pedestrian crash rates by neighborhood type.

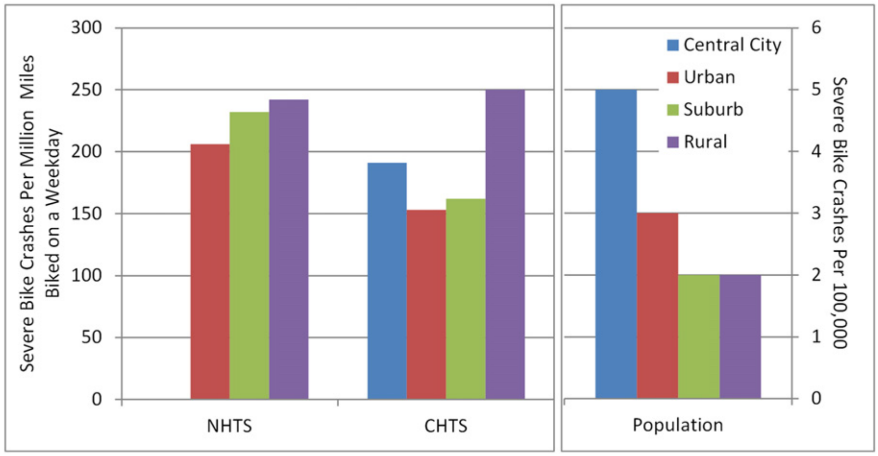


**Fig. 9.** Annual severe bike crash rates by neighborhood type.

or a complaint of pain by the pedestrian or cyclist. These crash severity classifications are provided in the SWITRS dataset.

Tract-level estimates for the entire state are provided online (Salon and Handy, 2014). Figs. 7–10 depict average crash rates by neighborhood type, and include a comparison to what the crash rates are on a straight per capita basis. Note that, due to gaps in the crash data, it was not possible to assign approximately 10% of the state's crashes to specific census tracts. The analysis shown here does not include these crashes, and therefore the estimates for crash rates are slightly underestimated.

Using the SAE from the two surveys as a basis for activity levels, there is no robust pattern by neighborhood type in severe crash rates per mile of activity for pedestrians (Fig. 7) and cyclists (Fig. 9) in California. Crash rates per capita do appear to show a pattern, but because people bike and walk more or less depending partially on their neighborhood type, these population-based patterns are likely misleading. Note that the finding of no pattern does not mean that there are no hot spots for severe crashes; problematic intersections and road segments have been clearly identified in specific locations (see, e.g., Schuurman et al., 2009). This finding implies, however, that these especially dangerous locations are spread across neighborhood types in the state.

At first glance, the lack of a spatial pattern in severe crash rates might appear counterintuitive, since the more urban areas have much more vehicle traffic in them. However, urban areas also have many more miles walked by pedestrians and miles cycled by bicyclists. In addition, and likely critical for this finding, the average speed of vehicles in the different neighborhood types is different; vehicles travel faster in less dense areas, making it more likely that those crashes that do occur will be severe. The following comparison is striking. Approximately 36% of the 10 years of reported pedestrian crashes in Rural areas were classified as severe, while this figure in Central City neighborhoods is only 11%. For reported bicycle crashes, 14% were severe in Rural areas versus only 5% in the Central City. In both cases, Urban and Suburb percentages of crashes that were severe are in between.

There is a clear pattern in non-severe crash rates per mile of activity (see Figs. 8 and 10), with non-severe crashes higher in more urban neighborhood types. Some portion of this pattern in non-severe crashes may result from a higher likelihood of reporting such crashes to the police in urbanized areas (the density of police is higher, so reporting is easier). However, much of the pattern is likely a truth about the relative severity of crashes in more and less urban areas. These data indicate that the likelihood of experiencing a crash at all is higher for pedestrians and cyclists in more urban environments, but the likelihood of that crash being severe is lower.

## 7. Conclusion

An average of 835 million dollars were invested in bicycle and pedestrian facilities in each of the past five years in the U.S. (FHWA, 2015a). Approximately 5000 pedestrians and bicyclists die and more than 100,000 are injured on U.S. streets in a typical year (NHTSA, 2010–2015). Knowing how much and where bicycle and pedestrian

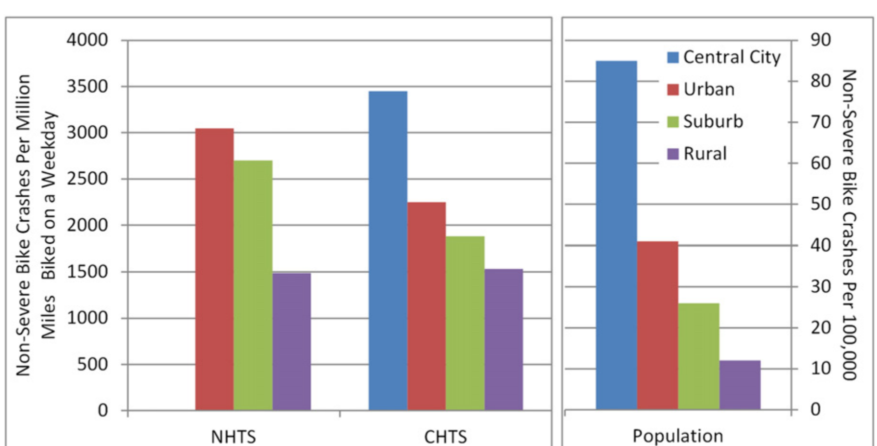


**Fig. 10.** Annual non-severe bike crash rates by neighborhood type.

activity occurs is critical as cities prioritize future investments and consider new safety measures.

This paper has introduced a simple small-area estimation technique to estimate pedestrian and cyclist activity levels at a fine geographic scale. Activity levels were estimated for all census tracts in California using two separate travel survey datasets. These activity estimates were used to produce two measures useful to planners: the intensity of infrastructure use by pedestrians and cyclists, and walking and bicycling crash rates. In the absence of more comprehensive data on bicycle and pedestrian activity at small geographic scales, this method could become a valuable tool for planners. It is simple to execute and it represents a clear improvement over activity estimates that are based only on population.

One substantial caveat that deserves repeating here is that the method assigns all pedestrian and cyclist activity to the census tract where people reside, rather than where the activity actually occurs. Most non-motorized trips are short, however, and begin and/or end at home. However, there are some places where many nonresidents might walk or bike (i.e., downtowns, university campuses, etc.). An important implication is that this method should *not* be used to calculate crash risk within activity centers. This is because pedestrian or bicyclist exposure in these types of locations would be systematically underestimated, meaning that crash risk could be overestimated.

Despite this and other limitations, there is strong potential for this simple method to be both used and extended. Although the analysis here uses actual population numbers from the Census, the method could also be applied to forecasts of population by category in future years. In addition, the impact of land use policies could be captured by changing the neighborhood type classifications for zones. Geographic units other than census tracts could also be used, though it is important that the units are internally homogeneous with respect to neighborhood type.

Improvements to the method itself could be made as well. This paper uses one of the simplest techniques among small-area estimation methods. Future research could investigate the utility of incorporating more advanced regression-based small-area estimation methods, taking into account individual built environment characteristics rather than representing built environment variation with only four neighborhood types. Studies show, for example, that bicycle activity is strongly correlated with bicycle infrastructure (e.g., Dill and Carr, 2003), but there is (at this writing) no state-wide database on such infrastructure that would enable the inclusion of this variable in the method. Conducting this analysis at the region rather than the state level might produce improved estimates as well, due to regional variation in infrastructure and weather.

Efforts are underway to improve and expand systematic bicycle and pedestrian counts (FHWA, 2015b) and to incorporate bicycle and pedestrian modes into travel demand forecasting models in a more robust way. Increasing use of GPS-enabled travel surveys will also provide more accurate data on non-motorized travel. The method presented here is a useful option to estimate activity in the meantime, and itself might be improved over time.

## Acknowledgements

This research was part of a project entitled *Non-Motorized Travel: Analysis of the National Household Travel Survey California Add-On Data*, completed with funding from the California Department of Transportation (Contract No. 65A0404). The author would also like to thank Susan Handy, who was the Principal Investigator on the project and provided useful comments on earlier drafts of this manuscript. The contents of this report reflect the views of the author who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration.

## References

ABS (Australian Bureau of Statistics), 2006. A Guide to Small Area Estimation – Version 1.1. ABS, Canberra, Australia Accessed June 2016 at: http://nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/3a60738d0abdf98cca2571ab00242664/$FILE/May%2006.pdf.

Beck, L.F., Dellinger, A.M., O'Neil, M.E., 2007. Motor vehicle crash injury rates by mode of travel, United States: using exposure-based methods to quantify differences. Am. J. Epidemiol. 166 (2), 212–218.

CDT (California Department of Transportation), 2013. 2010–2012 California Household Travel Survey Final Report, Version 1.0. CDT, prepared by NuStats Research Solutions, LLC, Sacramento, CA (Available online at dot.ca.gov).

Clifton, K.J., Krizek, K.J., 2004. The utility of the NHTS in understanding bicycle and pedestrian travel. Proceedings of the National Household Travel Survey Conference: Understanding our nation's travel. Transport Research Board, Washington, D.C. Accessed June 2016 online at: http://onlinepubs.trb.org/onlinepubs/archive/conferences/nhts/Krizek.pdf.

Dhondt, S., Macharis, C., Terryn, N., Van Malderen, F., Putman, K., 2013. Health burden of road traffic crashes, an analysis of clinical data on disability and mortality exposure rates in Flanders and Brussels. Crash Analysis and Prevention 50, 659–666.

Dill, J., Carr, T., 2003. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. Transp. Res. Rec. 1828, 116–123.

FHWA (Federal Highway Administration), 2015a. Federal-Aid Highway Program Funding for Pedestrian and Bicycle Facilities and Programs. FHWA, Washington, D.C. Accessed June 9, 2015, at: www.fhwa.dot.gov/environment/bicycle_pedestrian/funding/bipedfund.cfm.

FHWA (Federal Highway Administration), 2015b. Federal Highway Administration Pedestrian and Bicycle Activities, January. FHWA, Washington, D.C. Accessed June 10, 2015, at: www.fhwa.dot.gov/environment/bicycle_pedestrian/overview/trb_summaries/trb2015.cfm.

Griswold, J., Medury, A., Schneider, R., 2011. Pilot models for estimating bicycle intersection volumes. Transp. Res. Rec. 2247, 1–7.

Handy, S.L., Boarnet, M.G., Ewing, R., Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. Am. J. Prev. Med. 23 (2), 64–73.

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., Xu, Z., 2012. Estimating use of non-motorized infrastructure: models of bicycle and pedestrian traffic in Minneapolis, MN. Landsc. Urban Plan. 107 (3), 307–316.

Jestico, B., Nelson, T., Winters, M., 2016. Mapping ridership using crowdsourced cycling data. J. Transp. Geogr. 52, 90–97.

Krizek, K.J., Johnson, P.J., Tilahun, N., 2005. Gender differences in bicycling behavior and facility preferences. In: Rosenbloom, S. (Ed.), Research on Women's Issues in Transportation. Washington, D.C, Transportation Research Board, pp. 31–40.

Lee, C., Moudon, A.V., 2006a. Correlates of walking for transportation or recreation purposes. J. Phys. Act. Health 3 (1 Suppl), S77–S98.

Lee, C., Moudon, A.V., 2006b. The 3Ds + R: quantifying land use and urban form correlates of walking. Transp. Res. Part D: Transp. Environ. 11 (3), 204–215.

Madre, J.L., Axhausen, K.W., Brög, W., 2007. Immobility in travel diary surveys. Transportation 34 (1), 107–128.

McAndrews, C., 2011. Traffic risks by travel mode in the metropolitan regions of Stockholm and San Francisco: a comparison of safety indicators. Inj. Prev. 17 (3), 204–207.

McAndrews, C., Beyer, K., Guse, C.E., Layde, P., 2013. Revisiting exposure: fatal and non-fatal traffic injury risk across different populations of travelers in Wisconsin, 2001–2009. Crash Analysis and Prevention 60, 103–112.

Mindell, J.S., Leslie, D., Wardlaw, M., 2012. Exposure-based, 'like-for-like' assessment of road safety by travel mode using routine health data. PLoS One 7 (12). http://dx.doi.org/10.1371/journal.pone.0050606 (e50606).

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. Crash Analysis and Prevention 43 (5), 1624–1634.

Molino, J.A., Kennedy, J.F., Inge, P.J., Bertola, M.A., Beuse, P.A., Fowler, N.L., Emo, A.K., Do, A., 2012. A Distance-Based Method to Estimate Annual Pedestrian and Bicyclist Exposure in an Urban Environment. F2HA (FHWA-HRT-11-043), Washington, D.C. Accessed May 16, 2016 at: http://www.fhwa.dot.gov/publications/research/safety/pedbike/11043/11043.pdf.

NHTSA (National Highway Traffic Safety Administration), 2015. Traffic Safety Fact Sheets for Pedestrians and Bicyclists (2010–2015). NHTSA, Washington, D.C. Accessed May 16, 2016, at: http://www-nrd.nhtsa.dot.gov/CATS/listpublications.aspx?id=A&ShowBy=DocType.

Pulugurtha, S.S., Repaka, S.R., 2008. Assessment of models to measure pedestrian activity at signalized intersections. Transp. Res. Rec. 2073, 39–48.

Saelens, B.E., Handy, S.L., 2008. Built environment correlates of walking: a review. Med. Sci. Sports Exerc. 40 (7 Suppl), S550–S556.

Salon, D., Handy, S.L., 2014. Estimates of Miles Walked and Biked in California by Census Tract. University of California at Davis, Davis, CA Accessed May 16, 2016, at: http://ultrans.its.ucdavis.edu/projects/estimates-miles-walked-and-biked-california-census-tract.

Schneider, R., Henry, T., Mitman, M., Stonehill, L., Koehler, J., 2012. Development and Application of a Pedestrian Volume Model in San Francisco, California. Transp. Res. Rec. 2299, 65–78.

Schuurman, N., Cinnamon, J., Crooks, V.A., Hameed, S.M., 2009. Pedestrian injury and the built environment: an environmental scan of hotspots. BioMed Central Public Health 9 (233), 1–10 Online at: http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-9-233.

Teschke, K., Harris, M.A., Reynolds, C.C., Shen, H., Cripton, P.A., Winters, M., 2013. Exposure-based Traffic Crash Injury Rates by Mode of Travel in British Columbia. Can. J. Public Health 104 (1), e75–e79.

TIMS (Transportation Injury Mapping System), 2014. SWITRS data source. University of California at Berkeley, Safe Transportation Research and Education Center, Berkeley, CA Online at: http://tims.berkeley.edu.

Turner, S.M., Shunk, G., Hottenstein, A., 1998. Development of a methodology to estimate bicycle and pedestrian travel demand. Texas A & M University, Texas Transportation Institute, College Station, TX (FHWA/TX-99/1723-S).

USDOT (U.S. Department of Transportation), 2011. 2009 National Household Travel Survey User's Guide, Version 2. USDOT, Washington, D.C. Available online at: http://nhts.ornl.gov.

Weinstein, A., Schimek, P., 2005. How much do americans walk? An analysis of the 2001 NHTS. Paper presented at the Transportation Research Board Annual Meeting, January 9–13, Washington, D.C.

Zhu, M., Cummings, P., Chu, H., Xiang, H., 2008. Urban and rural variation in walking patterns and pedestrian crashes. Injury Prevention 14 (6), 377–380.