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 xxxx

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 zero observations in only 550 out of 8057 total tracts in the state (USDOT, 2011; CDT, 2013). However, even this large sample only

http://dx.doi.org/10.1016/j.jtrangeo.2016.06.023
0966-6923/

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.