National Center for Sustainable Transportation
A USDOT University Transportation Center

Research Report

May 2025

Advancing Active Transportation Project Evaluation

Dillon T. Fitch-Polse, University of California, Davis

Hossain Mohiuddin, University of California, Davis

Dan Willett, University of California, Santa Barbara

Trisalyn Nelson, University of California, Santa Barbara

Matthew Favetti, University of California, Davis

Kari Watkins, University of California, Davis

ITS UCDAVIS Institute of Transportation Studies

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**
NCST-UCD-RR-25-15	N/A	N/A

4. Title and Subtitle Advancing Active Transportation Project Evaluation	**5. Report Date** May 2025 **6. Performing Organization Code** N/A
7. Author(s) Dillon T. Fitch-Polse, Ph.D. https://orcid.org/0000-0003-3760-322X Hossain Mohiuddin, Ph.D. https://orcid.org/0000-0002-6842-2424 Dan Willett, https://orcid.org/0009-0009-5526-2335 Trisalyn Nelson, Ph.D. https://orcid.org/0000-0003-2537-6971 Matthew Favetti, https://orcid.org/0000-0002-6125-6718 Kari Watkins, Ph.D. https://orcid.org/0000-0002-3824-2027	**8. Performing Organization Report No.** UCD-ITS-RR-25-07
9. Performing Organization Name and Address University of California, Davis Institute of Transportation Studies 1605 Tilia Street, Suite 100, Davis, CA 95616	**10. Work Unit No.** N/A **11. Contract or Grant No.** Caltrans 65A0686 Task Order 083 USDOT Grant 69A3552348319 and 69A3552344814
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology 1200 New Jersey Avenue, SE, Washington, DC 20590 California Department of Transportation Division of Research, Innovation and System Information, MS-83 1727 30th Street, Sacramento, CA 95816	**13. Type of Report and Period Covered** Final Research Report (April 2024 – December 2024) **14. Sponsoring Agency Code** USDOT OST-R Caltrans DRISI

15. Supplementary Notes
DOI: https://doi.org/10.7922/G2GM85PX
Dataset DOI: https://doi.org/10.5061/dryad.j9kd51cpm

16. Abstract
The purpose of this study is to examine the effectiveness of active transportation projects in increasing active transportation in California. It also serves to validate the current methods of the California Active Transportation Benefit-Cost Tool. Using count and infrastructure data from the cities of Santa Barbara and Santa Cruz, California, with updated models from the California Active Transportation Benefit-Cost Tool, the authors estimated project level changes in active transportation using two methods. The first method uses a direct demand modeled before and after bicycling and pedestrian volumes. The second method is an expected increase in bicycling and pedestrian volumes based on the project parameters and their effect sizes from the academic literature. Results show that, in general, both estimates are closely aligned. However, the results also indicate that for some projects, particularly those projects with greater change in walking and bicycling, the California Active Transportation Benefit-Cost Tool can diverge from the before-after estimate substantially at the project-level. Several suggestions for future research and improvements to the tool are made.

17. Key Words	**18. Distribution Statement**
Active transportation, infrastructure evaluation, direct demand models, before and after, benefit-cost tool	No restrictions.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	38	N/A

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

About the National Center for Sustainable Transportation

The National Center for Sustainable Transportation is a consortium of leading universities committed to advancing an environmentally sustainable transportation system through cutting-edge research, direct policy engagement, and education of our future leaders. Consortium members include: the University of California, Davis; California State University, Long Beach; Georgia Institute of Technology; Texas Southern University; the University of California, Riverside; the University of Southern California; and the University of Vermont. More information can be found at: ncst.ucdavis.edu.

Disclaimer

Acknowledgments

This study was funded, partially or entirely, by a grant from the National Center for Sustainable Transportation (NCST), supported by the U.S. Department of Transportation (USDOT) and the California Department of Transportation (Caltrans) through the University Transportation Centers program. The authors would like to thank the NCST, the USDOT, and Caltrans for their support of university-based research in transportation, and especially for the funding provided in support of this project. The authors would also like to thank the city of Santa Barbara, City of Santa Cruz, and the Santa Cruz County Regional Transportation Commission for data and project information.

Advancing Active Transportation Project Evaluation

A National Center for Sustainable Transportation Research Report

Dillon T. Fitch-Polse, Hossain Mohiuddin, and Matthew Favetti, Institute of Transportation Studies, University of California, Davis

Dan Willett and Trisalyn Nelson, Department of Geography, University of California, Santa Barbara

Kari Watkins, Department of Civil and Environmental Engineering; Institute of Transportation Studies, University of California, Davis

Table of Contents

Executive Summary iv
Introduction 1
Methods 3
  Study Area (Local Data) 3
  Data Collection 5
  Updated Method of Project Benefit Estimations 10
  Validation Procedure 15
Results 17
  Bike and Walk Model Results 17
  Estimates of Local Project Activity 17
Discussion and Conclusions 23
  Suggestions for BC Tool Improvements 24
References 26
Data Summary 28

List of Tables

Table 1. Santa Cruz Active Transportation Counts, 2016 to 2023 8

Table 2. Santa Barbara Active Transportation Counts, 2016 to 2023 8

Table 3. Number of infrastructure elements analyzed for the study 10

Table 4. List of predictors used in the bike model and in the pedestrian model 13

Table 5. Summary of Random Forest Model Performance with error margins in parentheses (Bike and Pedestrian Models) 17

List of Figures

Figure 1. Study area map for Santa Barbara 4

Figure 2. Study area map for Santa Cruz 5

Figure 3. Study area and statewide data collection sites for building the bike and pedestrian models. 6

Figure 4. Modeling process of the bike and pedestrian volume 14

Figure 5. Method of assigning the expansion factors based on the location of the permanent and short-term counters in the pedestrian model. 15

Figure 6. Estimated change in bicycling volume annually due to project. Gold diamonds indicate estimate from direct demand modeled maximum after – minimum before years. Point and error bars represent BC tool projection based on elasticities from academic literature. 19

Figure 7. Estimated percent error in change in bicycling volume annually from using the BC tool projection based on elasticities from the academic literature in comparison to the model-based before/after estimate. 20

Figure 8. Estimated change in pedestrian volume annually due to project. Gold diamonds indicate estimate from direct demand modeling maximum after – minimum before years. Point and error bars represent BC tool projection based on elasticities from academic literature. 21

Figure 9. Estimated percent error in change in pedestrian volume annually from using the BC tool projection based on elasticities from the academic literature in comparison to the model-based before/after estimate. 22

Executive Summary

California is investing substantially in active transportation (AT), including the statewide Active Transportation Program (ATP). Evaluation is critical for understanding the return on these investments. Studies show mounting evidence for how AT projects achieve benefits, yet limited tools exist to evaluate benefits.

In this report, we examine the change in walking and bicycling due to the installation of several infrastructure changes in the cities of Santa Barbara and Santa Cruz. These cities were chosen because of the available data on infrastructure projects and local traffic counts were more readily available than in any other place in California per the guidance from the California Transportation Commission.

This examination of changes to walking and bicycling is directly connected to the development of the California Active Transportation Benefit-Cost Tool (https://activetravelbenefits.ucdavis.edu/). The California Active Transportation Benefit-Cost Tool is a unified calculator designed to estimate the benefits of active transportation projects from across the State of California. Unlike other benefit-cost tools, it does not rely on user estimates of change in walking and bicycling behavior and instead estimates baseline walking and bicycling through a statewide direct demand model (developed using a Random Forest approach). The errors of the existing models supporting the California Active Transportation Benefit-Cost Tool are known to be large (see Kamalapuram, 2022 and Fitch et al., 2022), yet besides the recent work by Miah et al. (2024), there have been no other alternatives to estimating walking and bicycling volumes statewide. This project updates those models based on the work by Miah et al. (2024) and incorporates local data specific to this project.

Besides the errors associated with the few statewide direct demand models, the errors associated with other facets of the California Active Transportation Benefit-Cost Tool have not been measured. Most important are the potential errors from estimating the expected change in walking and bicycling from elasticities gathered from the scientific literature specific to each active transportation intervention. It is well documented that nearly all active transportation interventions can have widely varying effects on active transportation (Fitch-Polse and Agarwal, 2025). This project was designed to compare the estimates of the California Active Transportation Benefit-Cost Tool and before-after estimates from direct demand models of annual travel at the project level.

Results indicate that on the median, the California Active Transportation Benefit-Cost Tool estimates are closely aligned with those of the before-after analysis. This result is most likely because both approaches estimate only small changes in active transportation for most projects. However, the results also indicate that for some projects, particularly those

with greater change in walking and bicycling, the California Active Transportation Benefit-Cost Tool can diverge from the before-after estimate substantially at the project level by overestimating or underestimating the benefits.

The relatively harmonious estimate on the median suggests that the California Active Transportation Benefit-Cost Tool may already be suitable for program-level benefit calculations. However, the bias between the mean and median estimates summarized in this report should be used to adjust future versions of the tool when calculating program-level benefits.

The large outlying differences between the California Active Transportation Benefit-Cost Tool estimated demand and the before-after estimated demand suggests caution should be used when estimating the benefits of specific projects from the tool alone. More research is needed to determine the source of the bias in these projects. When feasible, the methods of this study also help guide how to integrate local count data within the framework of the California Active Transportation Benefit-Cost Tool to estimate actual accrued benefits at a project level.

Besides the guidance to improve project-level estimates with local count data and bias-correct program-level sums of benefits, two additional improvements to the California Active Transportation Benefit-Cost Tool are proposed. The first is to improve the treatment of off-street interventions in the tool which is a known limitation of the tool and may be one source of the existing bias. Second, because the direct demand models did not improve accuracy over past work, and because it is unlikely that any additional available data or model form will substantially improve predictions, there is great need for more walking and bicycling count data and in more diverse locations to improve our understanding of the benefits of active transportation projects.

Introduction

California is making substantial investments in active transportation (AT), including the statewide Active Transportation Program (ATP). Evaluation is critical for understanding the return on these investments. A recent literature review shows mounting evidence for the ways AT projects achieve benefits (Fitch-Polse and Agarwal, 2025), yet there are limited tools to evaluate benefits. For example, the Cal-B/C tool for AT projects from the Caltrans Economics Branch[1], while used for active transportation projects from some state programs, is not used for the California ATP project evaluation. This is because the tool requires users to input expected change in travel behavior, which limits the ability to standardize project-level effects across the State. Through collaboration with Caltrans, UC Davis developed the California AT Benefit-Cost Tool. From hereafter, this tool will be referred to as "the BC tool"—which differs from a conventional benefit-cost tool in that the costs are given by the user, and the benefits are calculated from uniformly projected changes in travel behavior to improve AT benefit estimation. California Transportation Commission (CTC) has used the BC tool for ATP program-level estimates of changes in safety, physical activity, local pollutants, and emissions. The BC tool is also an available framework and model for uniform project-level ATP assessments across the State. However, the tool currently does not consider local data on walking and bicycling counts, limiting its usefulness for estimating the realized benefits of completed projects. In this project, we conducted data collection and analysis of local data to supplement the BC tool.

This project explores the incorporation of local data on walking and bicycling levels into the BC framework and underlying model to make it more useful as a planning tool and improve local-scale AT project evaluation. This project uses local data to validate the estimates of bicycle and/or pedestrian activity which are critical for accurate predictions of project benefits in the BC tool framework.

The BC tool relies on the effect sizes (or elasticities) from previous studies found in different academic literature and described by Fitch-Polse and Agarwal (2025). The tool consists of direct demand models of bicyclists and pedestrians using data from statewide bike and pedestrian counts on different sites, built environment data, accessibility data developed by *PeopleforBikes*, census data, safety data, Strava Metro, and weather data. The details of the models can be found in Fitch et al. (2022). The tool has several limitations. As the tool relies on the statewide network and intersections data, it is limited in estimating and evaluating the actual benefits of local active-transportation-related completed projects in the local area. The tool also does not incorporate the variability of

[1] https://dot.ca.gov/programs/transportation-planning/division-of-transportation-planning/state-planning/transportation-economics

data collected from permanent and non-permanent counters while producing the statewide combined dataset for model building.

Since the development of the models for the BC tool, Miah et al. (2024) developed a method to infuse permanent and short-term bike counter data using expansion factors. Different expansion factors were developed based on the data from the permanent counter sites and applied to transform the short-duration count sites. Their method transforms hourly count data of short-term counters to daily traffic data for a site based on an hour of day (HOD). Then, the daily traffic of the short-term site is converted to weekly traffic based on the day of the week (DOF) expansion factor. The weekly traffic is averaged to get the daily average traffic and then converted to monthly total traffic by multiplying the days of the month. Finally, using the month of the year (MOY) factor, the daily average traffic is converted into yearly total traffic and then into average annual daily bike traffic (AADBT) by dividing the total traffic by the number of days. The expansion method allows for a much broader range of data to be considered in modeling bicycling and pedestrian demand, since so many short-term counts are conducted across California. The expansion factors for the bike model are based on fine grained Strava Metro activity for bicycling. The expansion factors for the pedestrian model are developed based on permanent pedestrian counter (i.e., Eco-Counter) data in the region.

In this study, we leverage the work by Miah et al. (2024) and Fitch et al. (2022) to improve active travel volume estimates within the study area and estimate changes in volumes due to active transportation infrastructure investments. This study specifically focuses on (1) advancing and validating the most critical component of the existing BC tool, the estimates of walking and bicycling activity, and (2) estimating project-specific benefits using local data and the BC tool. In this project, we achieve these goals by using bicyclist and pedestrian count data in Santa Cruz and Santa Barbara. These cities have a historical record of collecting accurate bicycling and pedestrian count data before and after implementing active transportation infrastructure and programs.

Methods

Study Area (Local Data)

This study focuses on data collected from the cities of Santa Cruz and Santa Barbara in California. These cities make good study sites because they both have made commitments and taken action to support the accessibility, safety, and participation of active transportation through infrastructure changes, among other programs. Both cities have also adopted comprehensive plans to bolster and improve active transportation (City of Santa Cruz, 2017; City of Santa Barbara, 2006; City of Santa Barbara, 2016). In conjunction with these plans, both cities have adopted a Vision Zero policy to eliminate all traffic fatalities and serious injuries by 2030.

Santa Cruz and Santa Barbara have also deployed bicycle and pedestrian count programs. These count programs inform planning and decision-making for active transportation projects. In addition, both cities have recently installed numerous roadway upgrades and new infrastructure designed for pedestrians and bicyclists, such as bike paths, sidewalks, and crosswalks. Figure 1 and Figure 2 display the location of infrastructure changes and pedestrian and bicycle count sites between 2018-2024 in Santa Barbara and Santa Cruz, respectively.

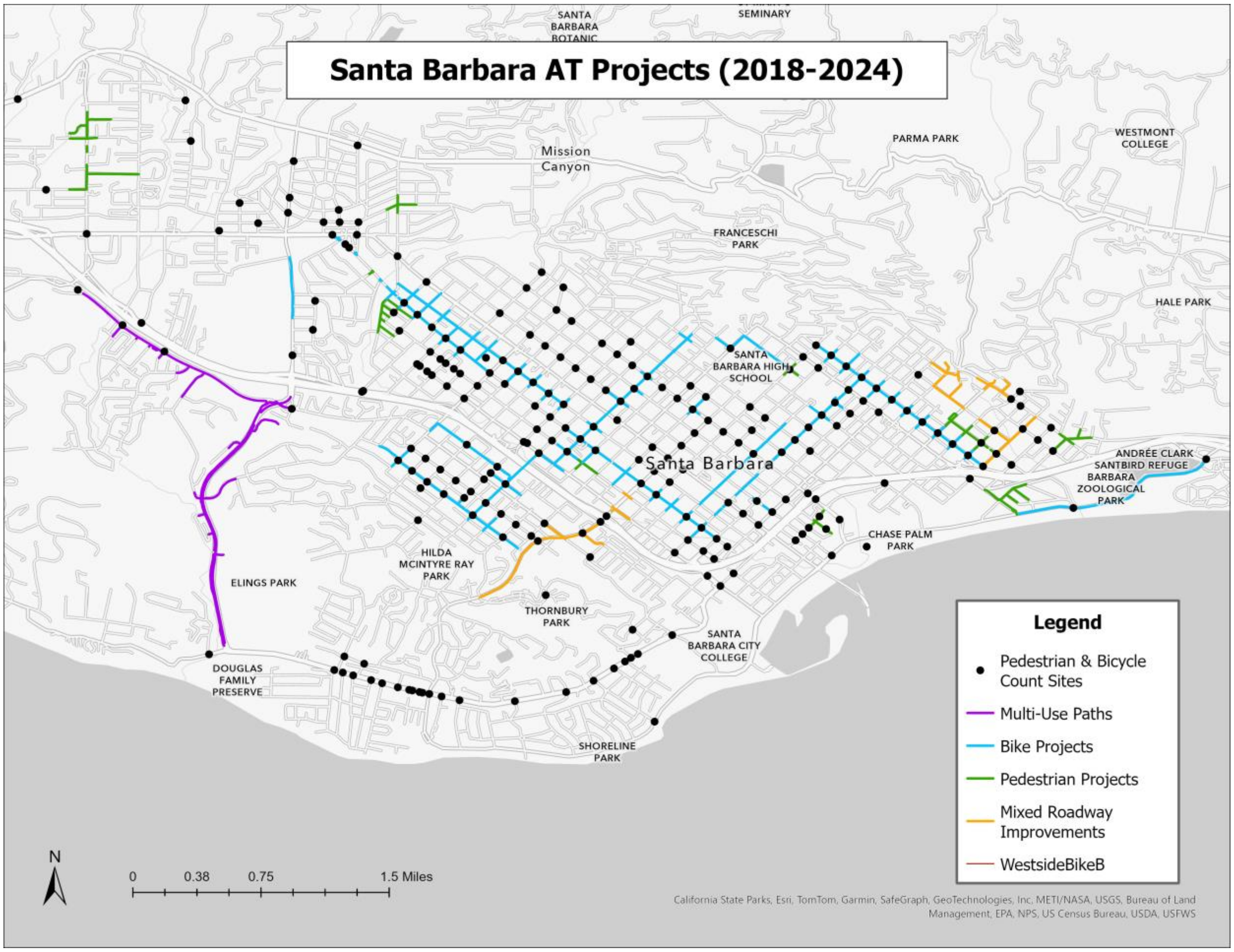

Figure 1. Study area map for Santa Barbara

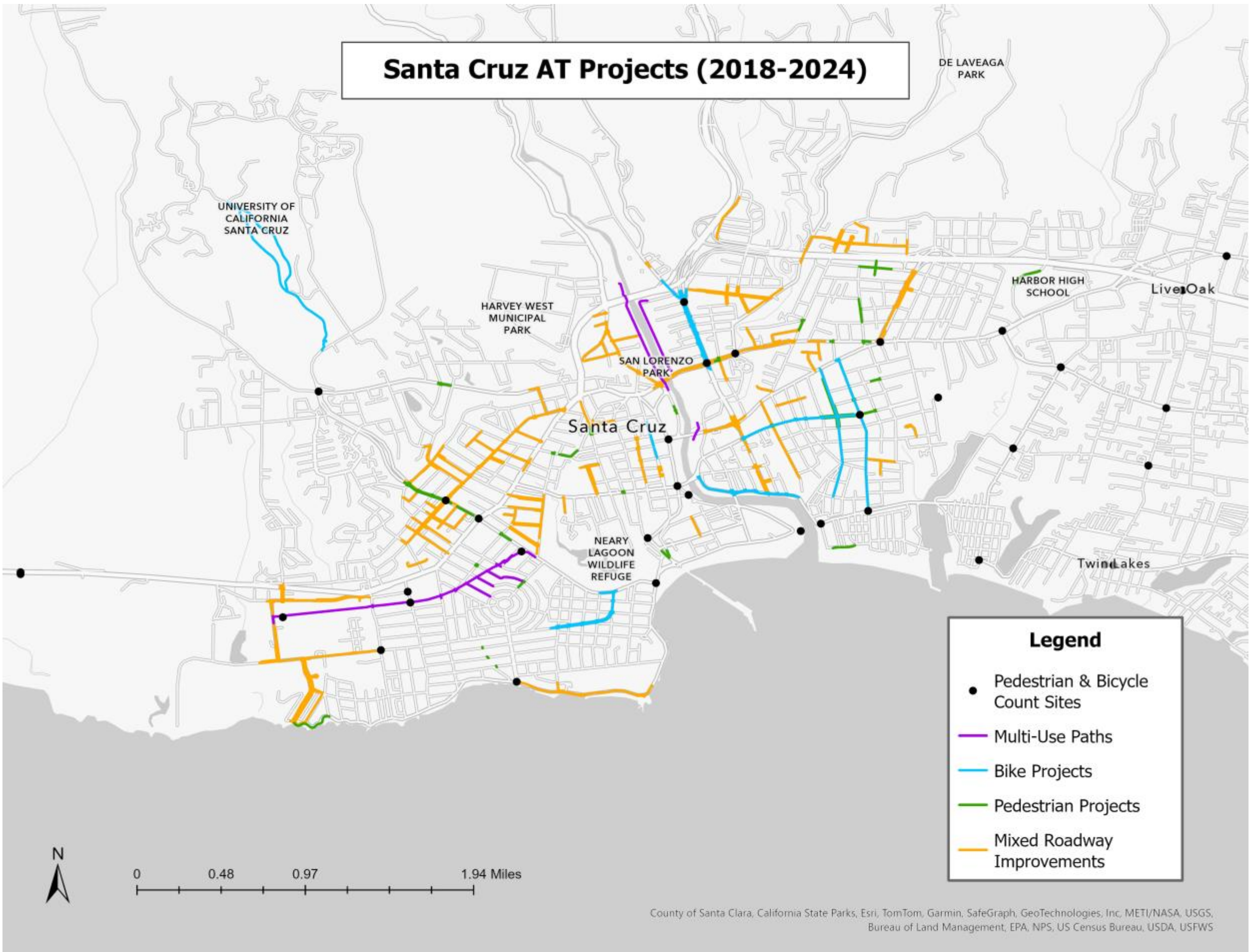

Figure 2. Study area map for Santa Cruz

Data Collection

In this study, we included both pedestrian and bicycle counts between 2018 and 2024 from the area of Santa Barbara and Santa Cruz. Although we had count data in 2018 and 2024, we removed projects installed in those years because we could not estimate “before” volumes before 2018 or “after” volumes after 2024. Along with the counts from the local area, we also included other count data from permanent and short-term bike count locations in California from Miah et al. (2024) to produce a count dataset for the bike model. We included pedestrian count data from permanent and short-term pedestrian count locations in California from Kamalapuram (2022) to produce a count dataset for the pedestrian model. The locations of the combined bike and pedestrian counts used to build the models are shown in Figure 3.

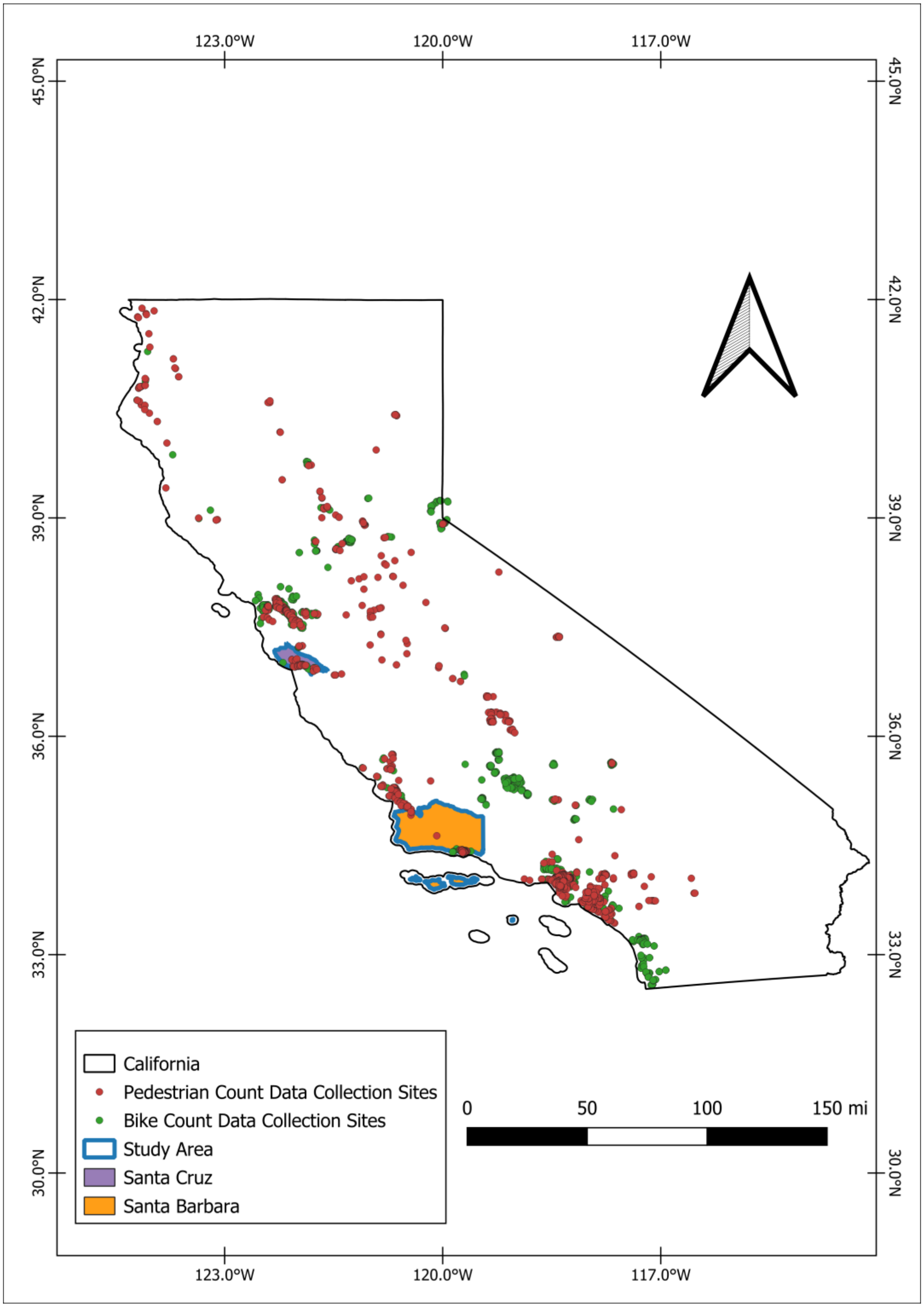

Figure 3. Study area and statewide data collection sites for building the bike and pedestrian models.

The following section describes the bike and pedestrian data collection process from Santa Barabara and Santa Cruz. These data were then combined with the data acquired from Mintu et al. (2024) and Kamalapuram (2022) to build the dataset for modeling.

Counts from Santa Barbara and Santa Cruz

This study incorporates bicycle and pedestrian count data from Santa Barbara and Santa Cruz. Both cities deploy counters from Eco-Counter along roads and paths, which use sensors and tubes to capture screenline counts of passing bicyclists and pedestrians. The City of Santa Cruz has four permanent Eco-Counters installed throughout the city, the oldest of which has been active since 2016. In Santa Barbara, two mobile Eco-Counters have been operated by the University of California, Santa Barbara since 2023. These mobile Eco-Counters were installed at nine different sites for durations of at least two weeks in 2023 and 2024. The Santa Barbara mobile Eco-Counters and two of the Santa Cruz permanent Eco-Counters report counts on 15-minute intervals. The other two permanent Eco-Counters in Santa Cruz report counts on hourly intervals. This study also incorporates short-term manual count data into the models. The team received manual pedestrian and bicycle intersection and screenline count data from the Santa Cruz County Regional Transportation Commission collected during 2016, 2018, and 2021. The research team organized data from over 20 sites located within the city of Santa Cruz. The 2016 counts took place in October on weekdays between 4PM to 6PM. The 2018 and 2021 counts took place in May and occurred on both weekdays and weekends, with the majority of counts occurring between 4PM to 6PM and 11AM to 1PM. Other counts occurred at 1-hour intervals throughout the day. Manual count data from two count programs in Santa Barbara were also incorporated into this study. The first program was organized by the City of Santa Barbara, which counted bicyclists' and pedestrians' turning movements at over 30 sites from 2016 to 2023. The turning movement counts were deployed on weekdays and weekends for durations between 1-5 days and recorded counts at 15-minute intervals during daylight hours. The second program is a bike count program run by the University of California, Santa Barbara, covering eight screenline sites in Santa Barbara in 2023. These bike counts were recorded at two-hour intervals on weekdays.

Counts were compiled and organized following the Federal Highway Administration's Traffic Monitoring Guide for Non-Motorized Traffic. Count site locations were associated with the Strava Metro simplified Open Street Map network so they could be associated with all the other data sources used to model active transportation (see below).

Table 1 and Table 2 include a summary of pedestrian and bicycle counts in Santa Barbara and Santa Cruz.

Table 1. Santa Cruz Active Transportation Counts, 2016 to 2023

Neighborhood	Pedestrian Traffic	Pedestrian Hours of Data Collection	Bike Counts	Bike Hours of Data Collection
Boardwalk	1,770,996	48,748	1,366,931	48,748
Circles	1,348	10	704	10
Downtown	2,273	18	781	18
King St	96	2	156	2
Lower Seabright	996	14	1014	14
Mission Hill	358,812	28,704	440,636	28,704
Mission St	539	6	258	6
Soquel Ave	233	8	248	8
UC Santa Cruz	712	8	642	8
Upper Seabright	5,083,542	64,420	1,139,152	64,420
Water St	212	4	167	4
Westside	305,408	31,012	421,936	31,012

Table 2. Santa Barbara Active Transportation Counts, 2016 to 2023

Neighborhood	Pedestrian Traffic	Pedestrian Hours of Data Collection	Bike Counts	Bike Hours of Data Collection
Bel Air	988	91	1,331	273
Downtown	132,508	347	48,281	1,432
East Beach	1,479	54	0	0
East Mesa	1,406	78	0	0
Eastside	15,963	442	3,401	962
Hidden Valley	810	172	2,858	536
Laguna	7,913	78	247	65
Lower East	12,689	221	1,075	130
Lower State	110,269	416	52,505	1,924
Lower West	925	39	179	65
Oak Park	20,030	482	2,550	853
Samarkand	151	78	291	312

Neighborhood	Pedestrian Traffic	Pedestrian Hours of Data Collection	Bike Counts	Bike Hours of Data Collection
San Roque	63	13	188	65
Upper East	9,909	156	659	65
Upper State	5,356	196	8,285	786
Waterfront	18,233	52	11,006	262
West Beach	2,626	44	420	39
West Mesa	1,523	84	0	0
Westside	17,096	598	1,974	494

Infrastructure Data / Active Transportation Project-Specific Data

The research team compiled an infrastructure dataset from projects completed in or after 2018 in Santa Barbara and Santa Cruz. The infrastructure dataset tracks attributes found to increase pedestrian and bicycle volumes and improve safety. These attributes include roadway features such as bike lanes, off-road paths, road diets, and intersection features like crosswalks, curb extensions, and ADA ramps. These attributes were selected based on the listed attributes required for benefit estimation using the BC tool developed by Fitch et al. (2022).

The team received a list of grant-funded projects and infrastructure improvements from city planners in Santa Cruz and Santa Barbara. Examples of project funding include the Caltrans Active Transportation Program, Highway Safety Improvement Program, and Community Development Block Program. Documentation for infrastructure projects was found in each city's public web archives. These documents were used to determine project locations, infrastructure improvements, and completion dates.

Some infrastructure projects spanned locations throughout the cities. To track each project's infrastructure improvements and where they occurred, the team used Google Maps Street View to toggle between imagery taken before and after a project's completion date. If an infrastructure change occurred following the project's completion date, the length or amount of upgraded infrastructure was recorded along with the Strava network links associated with the infrastructure location. Table 3 outlines the new infrastructure tracked in Santa Cruz and Santa Barbara for this study.

Table 3. Number of infrastructure elements analyzed for the study

Infrastructure Element	Santa Cruz*	Santa Barbara*
ADA Ramp	343	230
Bike Boulevard	0 ft	22,402 ft
Bike Box	12	0
Buffered Bike Lane	14,893 ft	11,261 ft
Conventional Bike Lane	8,575 ft	4,716 ft
Crossing island	0	16
Crosswalk	225	77
Curb Extension	56	111
Flashing Beacon	73	51
Lighting	60	235
Off-Street Path	7,557 ft	13,750 ft
Protected Bike Lane	436 ft	3069 ft
Road Diet	0 ft	14,628 ft
Sidewalk	4,360 ft	4,333 ft
Traffic Signal	1	19

* Units in counts unless specified.

Ancillary Data

In addition to the permanent and short-term count site data, and existing variables from Kamalapuram (2022), we collected bike and pedestrian crash data for California from 2019 to 2023 from the Transportation Injury Mapping System (TIMS). We also collected the Strava bike volume data for California from 2019 to 2023 from Strava Metro. We collected block-level census data for California for each year from 2019 to 2023 from the American Community Survey's 5-year estimates. These data were used as variables within the direct demand models.

Updated Method of Project Benefit Estimations

We estimated two direct demand models: the bike model and the pedestrian model. The dependent variable for the bike model was the annual average daily bike traffic (AADBT), and for the pedestrian model it was the annual average daily pedestrian traffic (AADPT). The use of direct demand models for active transportation prediction has been a common approach to large-scale volume estimation due to the lack of treatment for these modes in traditional travel demand models (Broach et al, 2024; Kaiser et al., 2025, Nelson et al, 2021, Nordback et al., 2017, Schneider et al., 2021). Yet, applications of direct demand models continue to show large errors whether applied at the city or statewide level. Nonetheless, it is the current state of the practice, so this project employs a similar approach.

The models in this report were developed based on the AADPT and AADBT data for each count site for each year from 2019 to 2023. However, many of the sites did not have multi-year data. We considered data for each site for each year as one observation.

We estimate the existing active travel volumes in the form of AADBT and AADPT using the Random Forest Machine Learning algorithm. The Random Forest is an ensemble learning method that uses many decision tree models with feature subsetting to increase predictive accuracy. Based on bicycle and pedestrian count data availability, these models estimate bicycle travel on the links (bi-directional travel) and pedestrian travel at the intersections (sum of all crossing movements). Both models are built based on the previously developed model of Kamalapuram (2022) and Miah et al. (2024). Based on the available bike data from Miah et al. (2024), pedestrian data from Kamalapuram (2022), and the locally collected permanent and short-term new bike and pedestrian count data from the study area, we developed an initial database for the model. The following section describes the different categories of predictor variables used in bike and pedestrian models.

Population Data

Block-level census data, such as population, race, gender, income, and commute to work (i.e., percentage of bike share for commute to work, percentage of share of walk, and percentage share of transit to commute work), were included as explanatory variables in the model.

Strava Metro Data

In processing the count data we found many of the bike and pedestrian count sites were not aligned with the Strava network. We decided to select the closest Strava link to the latitude and longitude of the counter locations to join Strava trip counts from the Strava Metro website. This approach might have some limitations, which can be addressed in future studies by taking the averages of the multiple Strava links nearest to bicycle and pedestrian data collection sites. Strava data for a specific bike or pedestrian count site were selected by extracting the Strava yearly count value of the nearest Strava link to the bike count site.

Crash data

We used the bike and pedestrian crash incidence locations from TIMS spanning 2019 to 2023 in California for the crash data. In the bike model, we produced a 10-meter buffer around the road link and counted the number of bike and pedestrian crashes within that buffer for each year to construct the safety/crash variable. In the pedestrian model, we used a 100-meter buffer around the intersection location and counted the number of bike and pedestrian crashes for each year to produce the safety/crash variable. We included both pedestrian and bicyclist crashes due to the limited amount of count data for each mode. While the crashes for each mode likely follow different profiles, we assumed that some of the crash risk for bikes would translate to crash risk for pedestrians.

Using crash data to predict volume is subject to one important limitation. Because crashes are usually considered an outcome of traffic volume, by including them in the model we likely bias the model parameters (we are essentially conditioning on a post-outcome variable) and so we cannot infer causal meaning to the parameters. In this case, we are using the model solely for predictive purposes, and including crashes increased out-of-sample predictive performance, so we proceeded with the causally dubious, yet better predictive model. However, future research should consider the implications for this decision in the use of the predicted volumes for benefit calculations, specifically calculations of safety benefits.

Accessibility Data

The network accessibility metrics used in this study are the low and high-stress connections of census blocks from the *PeopleForBikes*' Bicycle Network Analysis (BNA) tool. The BNA tool uses a computationally intensive routing algorithm that calculates census block-level accessibility metrics for several activity types. The details of this method can be found on the *PeopleForBikes* website.[2] The accessibility percentages for the count location were aggregated by Kamalapuram (2022), and those accessibility metrics are used as explanatory variables of this study.

Other features

We included the study area-specific dummies in the model so that our model can capture some of the variability of the study area. Before using the Strava bike counts as a predictor in the pedestrian count estimation model, we tested how the Strava bike counts are related to pedestrian counts using Santa Barbara as a study area. We found a moderate correlation (R = 0.3 to 0.58 depending on the location) between Strava bike count and pedestrian count data from Eco-Counters in Santa Barbara. Strava running data showed no clear correlation with pedestrian counts and so were not considered in the models. Table 4 lists the variables used for the bike and pedestrian Random Forest models.

The bicycle and pedestrian models also include annual average precipitation and minimum and maximum temperature, all obtained from the National Oceanic and Atmospheric Administration (NOAA), and built environment and road characteristics data from the Smart Location Dataset provided by the Environmental Protection Agency (EPA).

---

[2] https://cityratings.peopleforbikes.org/about/methodology

Table 4. List of predictors used in the bike model and in the pedestrian model

Bike Model		**Pedestrian Model**	
Variable	*Source*	*Variable*	*Source*
Strava Bike volume of the nearest Strava link to the site	*Strava Metro*	Strava Bike volume of the nearest Strava link to the site	*Strava Metro*
Network Accessibility Metric from the Bicycle Network Analysis (BNA)	*PeopleforBike* and Kamalapuram (2022)	Network Accessibility Matric the Bicycle Network Analysis (BNA)	*PeopleforBike* and Kamalapuram (2022)
Roadway characteristics	*PeopleforBike* and Kamalapuram (2022)	Roadway characteristics	*PeopleforBike* and Kamalapuram (2022)
Census block group-level variables	American Community Survey 5-year estimate	Census block group-level variables	American Community Survey 5-year estimate
Weather data	National Oceanic and Atmospheric Administration (NOAA)	Weather data	National Oceanic and Atmospheric Administration (NOAA)
Safety Data/ Crash Data	Statewide Transportation Injury Mapping Systems (TIMS)	Safety data	Statewide Transportation Injury Mapping Systems (TIMS)
Built environment data	Smart Location Data by Environmental Protection Agency (2021)	Built environment data	Smart Location Data by Environmental Protection Agency (2021)
Study Area Dummy (Santa Barbara and Santa Cruz)	-	Study Area Dummy (Santa Barbara and Santa Cruz)	-
Interaction between study area dummy and year of data	-	Interaction between study area dummy and year of data	-
Internation between the study area dummy and Strava bike volume	-	Internation between the study area dummy and Strava bike volume	-

We developed the bike and pedestrian model using the groups of predictors listed in Table 4. Permanent and short-term bike counts were used to calculate the AADBT variable, which is the dependent variable of the model. One challenge in using both short-term and permanent counters is combining those data, as those data are likely to follow different distributional patterns. It has been shown that brief duration counts result in biased estimates of annual averages (Laustsen et al., 2016). Because of this potential bias, we transformed the short-duration bike count data using the expansion factors developed by Miah et al. (2024) to make them consistent with the permanent counter data for the model. The steps to reduce bias, scale short-term counts to AADBT, and predict volumes are illustrated in Figure 4.

The first step of the process is to combine the count data, including the transformed data from Miah et al. (2024). The second step was to estimate and cross validate the Random Forest model. The third step was to predict volumes for each of the years with data available (2018-2023). Finally, the last step was to compare the demand model-based estimates with the estimates assumed from the new infrastructure data (see Validation Procedure below).

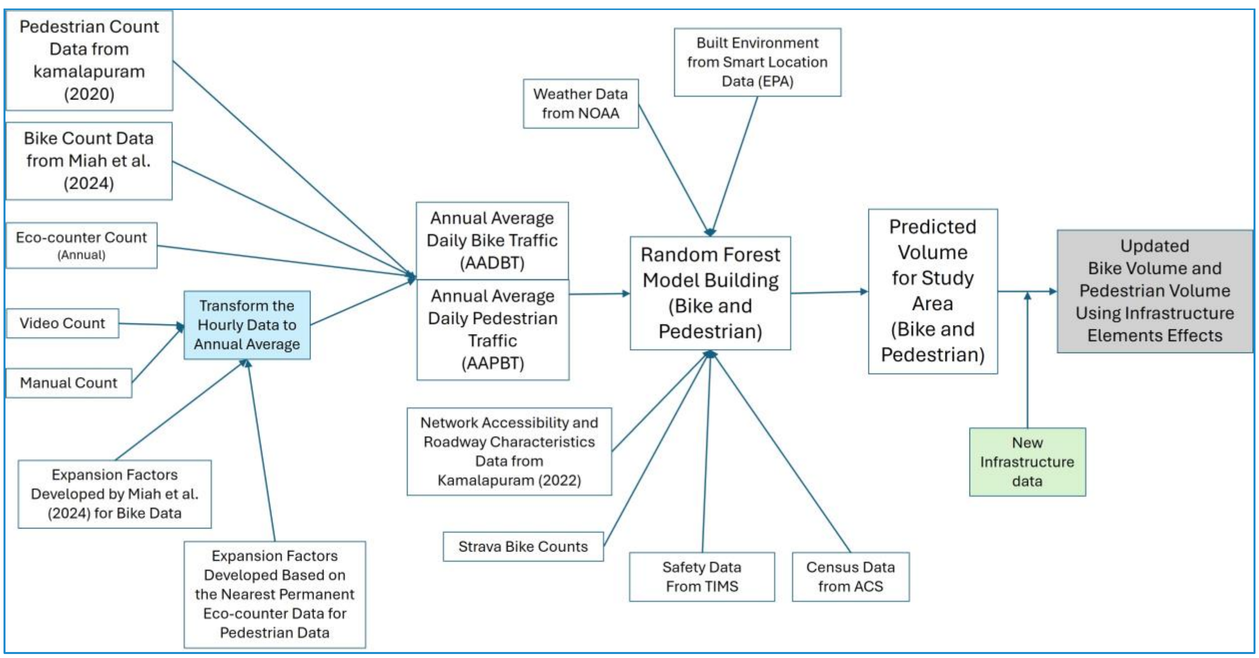

Figure 4. Modeling process of the bike and pedestrian volume

We also transformed the short-term pedestrian count data using newly developed expansion factors for this study. We plotted the short-term and permanent counter locations on a map and assigned the nearest permanent counter site to each short-term counter site (see Figure 5). Then, we developed the hour of the day, day of the week, and month of the year expansion factors for each permanent count site and used those factors to transform the short-term count data of the sites nearest the permanent counter site (see Figure 5). The steps in Figure 4 for the pedestrian data workflow worked similarly to the steps for the bicyclist data workflow. The only difference is that the expansion factoring in step one was newly developed by us based on the nearest permanent counter, not Strava volumes.

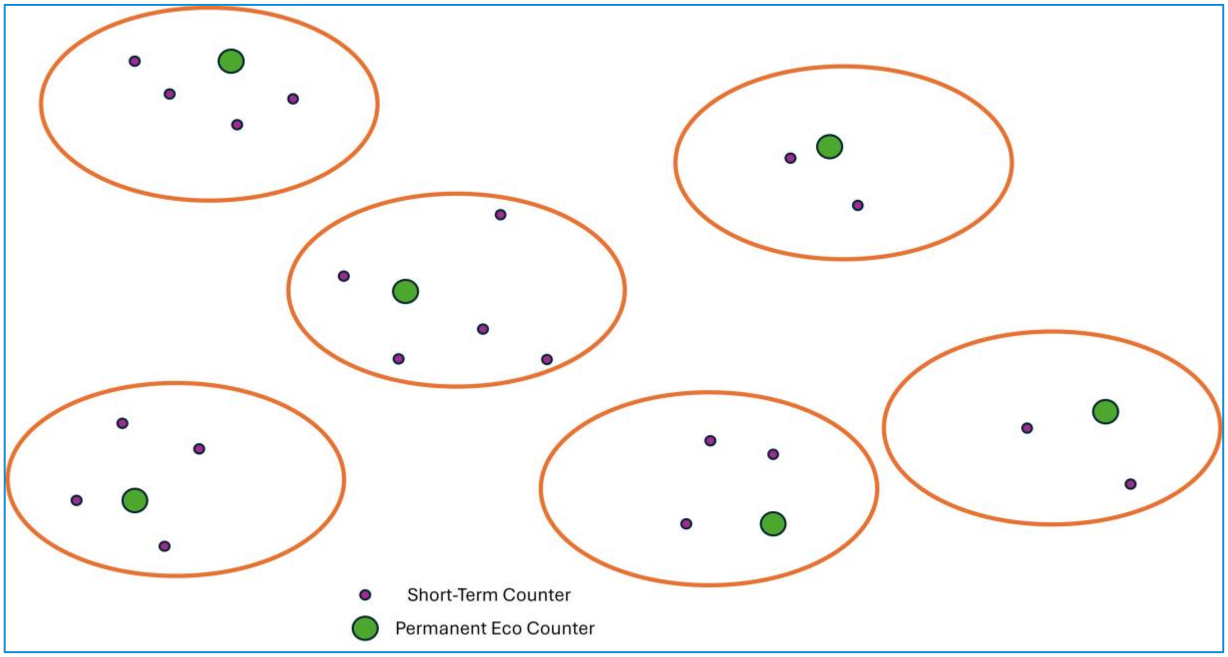

Figure 5. Method of assigning the expansion factors based on the location of the permanent and short-term counters in the pedestrian model.

Validation Procedure

After estimating the models and ensuring errors were comparable to Miah et al. (2024) and Kamalapuram (2022), we applied the bike model to predict bicycling volumes on all the links of the study area. The links were produced using the Strava network data collected from the Strava Metro of Santa Barbara and Santa Cruz. We predicted the average daily bike traffic (AADBT) for 2019, 2020, 2021, 2022, and 2023 for every link in the study area. We also produced intersections/nodes using the Strava link data of the study area. We applied the pedestrian model to those intersections/nodes to estimate the average daily pedestrian traffic (AADPT) for 2019, 2020, 2021, 2022, and 2023 for every intersection of the study area. Each link or intersection where the model was applied has the associated predictor data for each year from the 2019 to 2023 period.

The AADBT and AADPT were then converted to project-level active travel by summing the volume estimates for the links and nodes in each project. To ensure that the estimates for each project did not include the time during construction, we chose to summarize the "before" project activity as the minimum estimate in a year prior to project completion, and the "after" project activity as the maximum estimate in a year following project completion. The calculated "after" minus "before" formed the before/after estimate of project-level walking and bicycling. To validate the BC tool estimates (based on baseline direct demand model estimates plus an expected change from elasticities), we compared the BC tool estimated project-level walking and bicycling to the before/after estimate. The BC tool estimates are based on the direct demand model prediction during construction years. Using the outputs from the direct demand model, the equations in the BC tool were used to forecast the expected change in active transportation following project completion. The BC tool equations use the academic literature elasticities for specific infrastructure interventions and consider the project element elasticities as a function of their length or count in the context of the entire project reach (See Fitch et al., 2022). For example, in a Santa Cruz Project (MB Sanctuary Scenic Trail Segment 7, Phase 1), the model-predicted before and after volumes for the nearly 4,000 ft project were ~12,000 and ~19,900 AADBT, respectively (an increase in ~7,900 AADBT). In comparison, the BC Tool estimated an increase of nearly double that amount (~13,800).

Results

Bike and Walk Model Results

We conducted tenfold cross-validation for both the bike and the pedestrian models. This cross-validation technique allows for the estimate of multiple validation metrics and an estimate of their error (Table 5). The bike model had a mean absolute error of 67 while the pedestrian model's mean absolute error was 929. The errors are like those found by Miah et al. (2024) and Kamalapuram (2022). The lack of improvement in model metrics from the earlier models is difficult to explain. It may be that by adding the local data, which was predominantly short-term counts, that needed to be factored to annual averages, we increased the source of error in the response variables compared to past work. Although this model did not improve on the predictive ability of past models, it has several added benefits for this and future research. Unlike the past direct demand models in California, this model was designed to predict specifically in Santa Barbara and Santa Cruz (through the inclusion of indicator variables and interactions specific to those geographies) for this project. However, the removal of those terms would facilitate a more general predictive model. Further, these models were the first to be developed across multiple years to evaluate actual change in walking and bicycling which can be used to estimate the actual effects of projects on walking and bicycling demand.

Table 5. Summary of Random Forest Model Performance with error margins in parentheses (Bike and Pedestrian Models)

Model Metrics	Bike Model	Pedestrian Model
RMSE	123.24 (+/- 13.13)	2331.22 (+/- 476.92)
MAE	67.09 (+/- 3.25)	929.09 (+/- 152.87)
R2	0.72 (+/- 0.07)	0.62 (+/- 0.16)

Estimates of Local Project Activity

Bike Validation

The comparison between the direct demand modeled (max after – min before years) and the BC tool projected estimate of project level bicycling volume were heavily skewed (Figure 6 and Figure 7). The BC tool had a median error of only -2.7 AADBT, but an average error of 2,909 AADBT at the project footprint level. The large difference between the mean and median error are primarily driven by the BC tool predicting many zeros which are much closer to the small increases estimated through the direct demand modeling (max after - min before) approach. When either method produced a large value away from zero, the errors are very large. This suggests that the method may be very inaccurate for any given project, but the median estimate is likely to be accurate. To visualize this effect, Figure 6 plots the direct demand modeled max after – min before estimate against the mean and range of estimates from the BC tool approach. Four projects have exceptionally large errors, which is the cause for the large mean/median discrepancy. It is not clear from the data why those projects have such large errors. Follow-up with local agencies may be needed to help determine the source of error.

Not only are there several outlier projects, but the BC tool method also seems to overpredict bicyclist volumes by a much larger margin when projects have caused a greater change in bike volume (Figure 6). In Figure 7, the differences between the two approaches are plotted and sorted as percentages. The errors are much larger on the positive side of the scale than on the negative. It is not clear why the BC tool approach would be more biased for larger volume projects, but it suggests that when there are a lot of bicyclists in a project, the variation in the effect of a project may be greater.

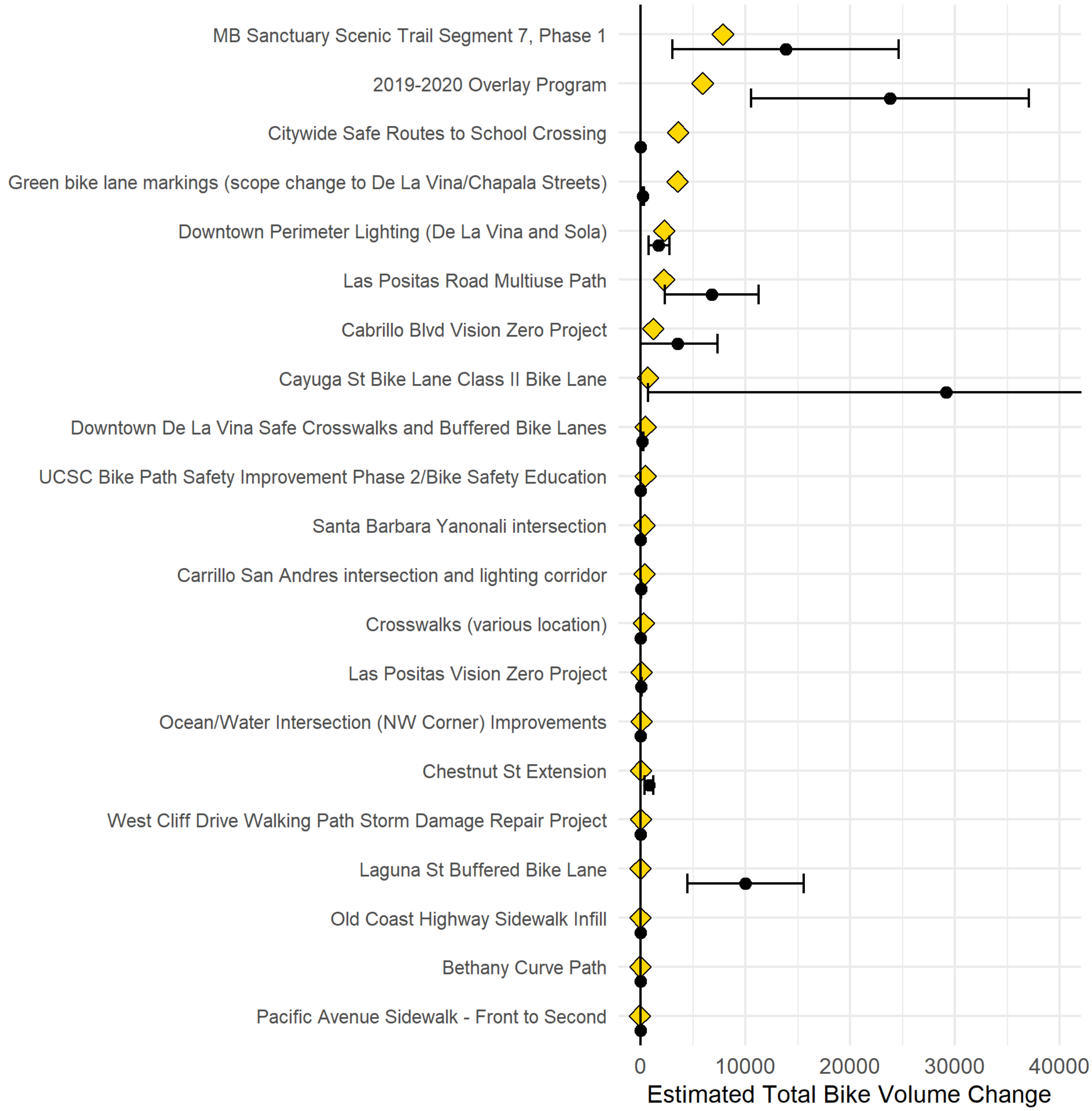

Figure 6. Estimated change in bicycling volume annually due to project. Gold diamonds indicate estimate from direct demand modeled maximum after – minimum before years. Point and error bars represent BC tool projection based on elasticities from academic literature.

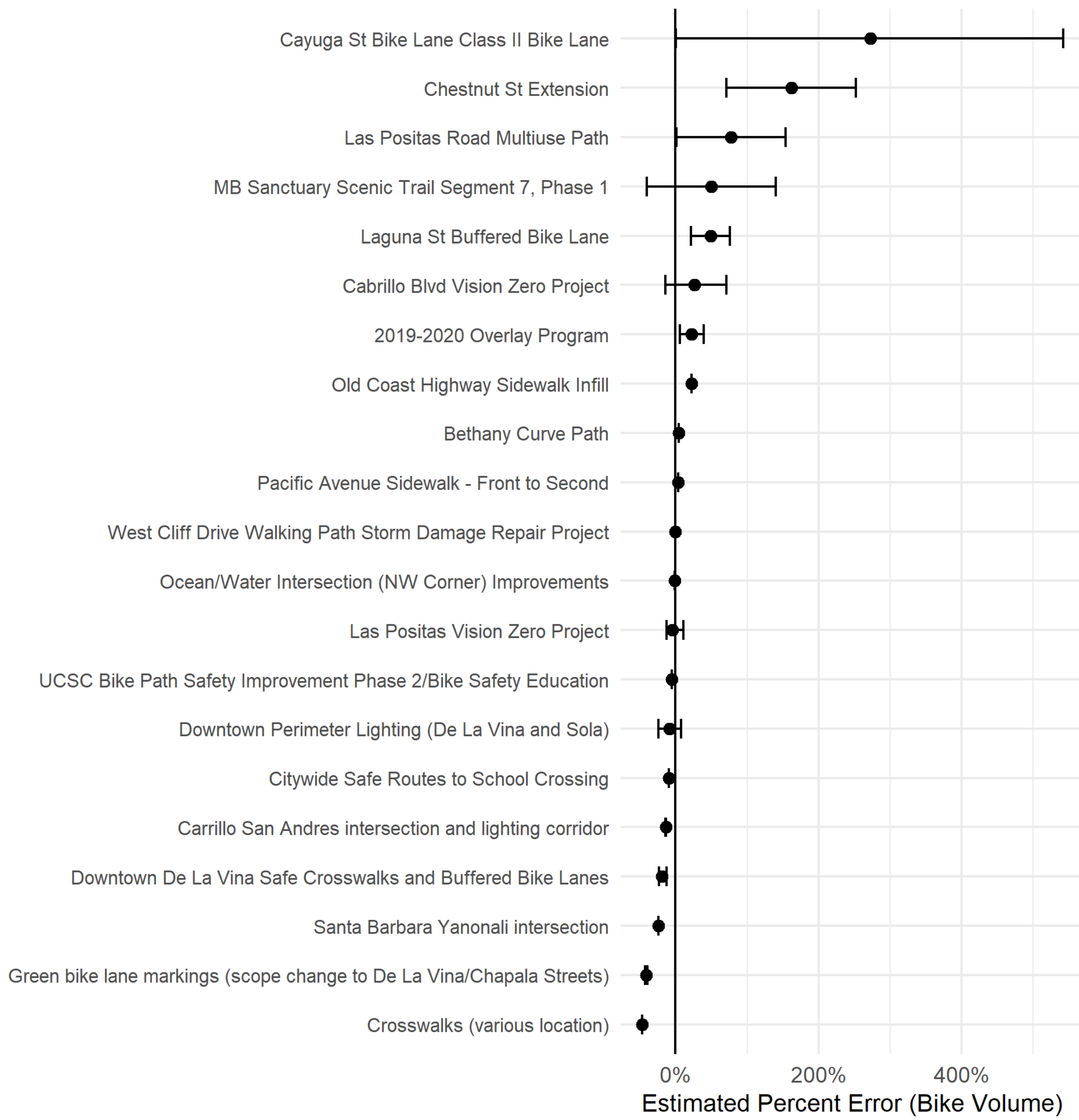

Figure 7. Estimated percent error in change in bicycling volume annually from using the BC tool projection based on elasticities from the academic literature in comparison to the model-based before/after estimate.

Pedestrian Validation

Like the bicyclist model, the comparison between the direct demand modeled (max after – min before years) and the BC tool projected estimate of project level pedestrian volume were heavily skewed (Figure 8 and Figure 9). The BC tool had a median error of only -8.5 AADPT, and an average error of -40.4 AADPT at the project level. The difference between the mean and median error for pedestrians is much smaller than that for bicyclists. This is likely because the outliers in the pedestrian projects are in both the positive and negative direction (Figure 8 and Figure 9). Additionally, since most pedestrian projects are focused on the safety of existing pedestrians, changes in pedestrian demand are rare, and both the direct demand modeling (max after - min before) approach and BC tool approach estimate change little for most projects.

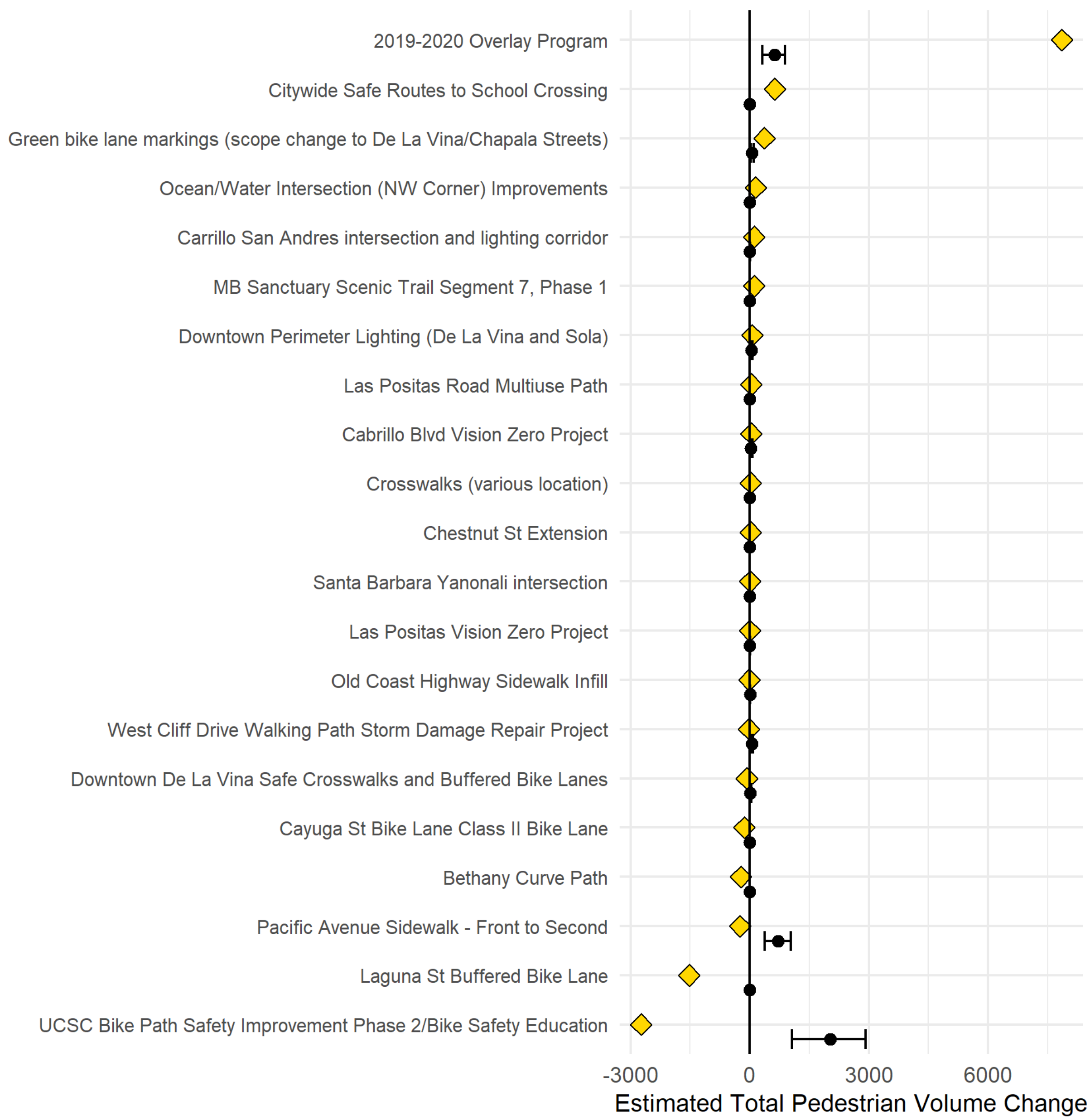

Figure 8. Estimated change in pedestrian volume annually due to project. Gold diamonds indicate estimate from direct demand modeling maximum after – minimum before years. Point and error bars represent BC tool projection based on elasticities from academic literature.

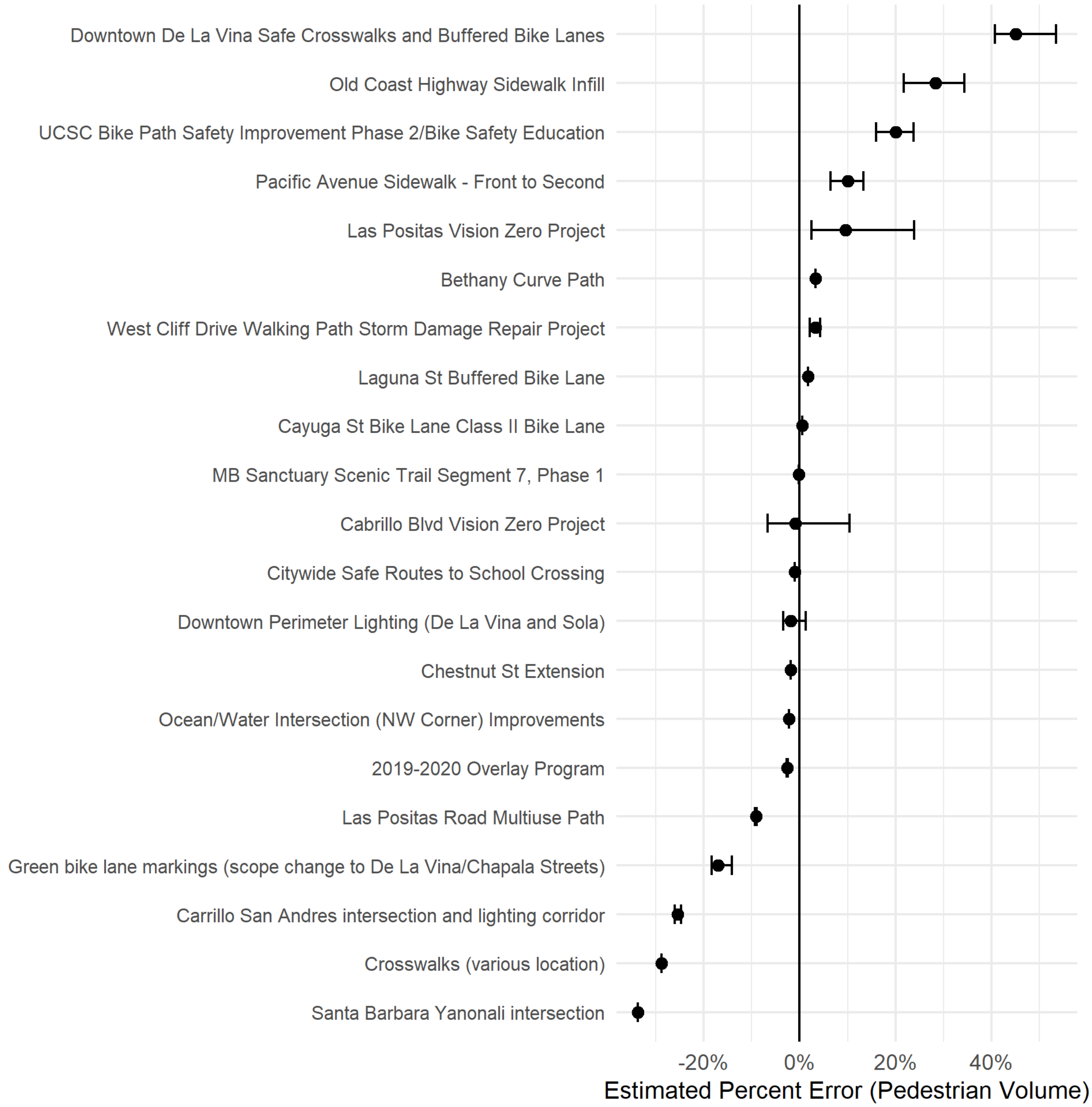

Figure 9. Estimated percent error in change in pedestrian volume annually from using the BC tool projection based on elasticities from the academic literature in comparison to the model-based before/after estimate.

Discussion and Conclusions

The results from this research showed that by using a direct demand model and incorporating local data, project level estimates of walking and bicycling demand are possible with sparse count data. Currently there are no reliable gold standards to compare these estimates to. Because the count data were not generally collected to specifically evaluate the projects, it did not align in space and time to be useful for direct before/after estimates at the project level. Because of this limitation, we treated the model-based estimates as the “ground truth”, but additional research with much more fine-scale count data designed to evaluate specific projects over time is needed to validate the project-level estimates. The cross validation of the demand models suggests there still remain large sources of uncertainty in predicting count data, which indicates that project-level estimates are likely to have errors of similar magnitude. Other potential errors include the aggregation of count data at the project level, where proximity to other projects, or the general context of the project within the broader network is likely to bias results. Active transportation infrastructure interventions do not occur in a vacuum, and the summing of link level volume from direct demand models is unlikely to account for the complexity of how people travel on transportation networks. In this way, direct demand models are not only limited in their data inputs, but also in their ability to represent travel patterns.

When comparing the elasticity approach of the BC tool to the before/after model-based estimates, the medians are reasonably similar for both bike and pedestrian models. This suggests that the elasticity approach is appropriate in the aggregate (when comparing a group of many projects). However, the errors at the project level were found to be substantial in several project cases, suggesting care must be taken when using the BC tool in its current form for project-level estimates of walking and bicycling.

Several new research questions have emerged from these results. First, *what is the cause for the large outlying prediction errors of the BC tool?* We suggest that qualitative research with practitioners familiar with the projects could help generate hypotheses for future evaluation. Second, *why are the bike model errors more biased than the pedestrian model errors?* And *why is the bias on the positive side of the scale?* While interviews with local practitioners might help form new hypotheses, we also suggest that more model improvements are needed. Future research on more variables or alternative techniques that account for the context of projects (e.g., connectivity and accessibility) is likely to help. Third, *what should be done in the BC tool to improve estimation of benefits at the project level?* Below we offer some suggestions for this question.

Suggestions for BC Tool Improvements

Improve the communication of the tool to inform users that reliability is an issue at the project level, but not program level. This communication should be aimed as a warning but should not attempt to dissuade project-level comparisons. Because there are no other tools available to estimate project-level benefits in a systematic manner, and because the BC tool will be undergoing improvements, it is still a valuable approach to project comparison and prioritization. Encouraging users to examine not only the mean effect but the range in effects estimated by the tool will help users become more aware of the uncertainty and can help inform planning decisions.

Improve the treatment of off-street interventions in the tool. Currently, the BC tool only estimates the base demand of off-street interventions based on the average demand of the other project roads and intersections. This is likely to be a poor estimate of a well-used infrastructure type. Without a full travel demand model, making estimates of new facilities is a challenge. However, finding an approach that incorporates the added connectivity and accessibility gained from new facilities is likely to improve the accuracy of project activity estimation when off-street facilities are included.

Improve the models of active transportation through more count data collection. The lack of count data for walking and bicycling is a known problem for estimating exposure for safety, and it is a clear problem for quantifying the other benefits of active transportation projects. There is great need to collect more data in more diverse locations to make project-level evaluation of benefits possible.

Improve the models of active transportation through more direct inclusion of local context. These improvements could be the addition of new variables in the direct-demand models, or alternative model structures that can account for local context. However, the efficacy of this work is not likely to compare to that of simply collecting more walking and bicycling data (as suggested above) given the general dearth of data available in California.

Link local and generalized models to improve estimates. To balance local and generalized models it will be helpful to define regions in California where we expect effects of interventions to be similar. By identifying which places are similar, in characteristics that matter for walking and bicycling, we will be able to combine local data from multiple areas to reduce error. We will also be able to make better predictions in places with limited data, relying on patterns observed in other similar places.

Consider allowing for submission of local count data to improve estimates. Because many users may have local count data that are not included in the development of the current models, an ability to update the models or adjust their estimates based on local count data could be one way to improve estimates of active transportation. This approach should not replace the current approach of the tool, but instead form an alternative use where users estimate the "achieved" benefits (through before and after data) of a project, rather than the forecasted benefits.

Consider monetizing the benefit calculations for easy cost-effectiveness comparisons. The current Cal-B/C tool is an example of the potential outcomes that could be included as a part of the BC tool. By using assumptions which monetize benefits, project and program level comparison could be more easily achieved.

Create a program-level interface for the BC tool. Because the tool is more accurate in the aggregate, consider a secondary interface to sum the benefits of groups of projects. This could be used by grant agencies for reporting or local implementors looking to prioritize groups of projects. When creating this summary, ensure that the reported bias is corrected. One way to correct the estimates is to take the program-level sum of benefits and subtract the difference between the program mean and program median.

References

Broach, J., Kothuri, S., Miah, M. Md., McNeil, N., Hyun, K., Mattingly, S., Nordback, K., Proulx, F. (2024). Evaluating the potential of crowdsourced data to estimate network-wide bicycle volumes. *Transportation research record* 2678, 3: 573-589.

City of Santa Barbara (2006). Pedestrian Master Plan. Retrieved from https://santabarbaraca.gov/sites/default/files/documents/Public%20Works/Pedestrian%20Master%20Plan/01-%20Cover%20and%20Table%20of%20Contents.pdf

City of Santa Barbara (2016). Bicycle Master Plan. Retrieved from https://santabarbaraca.gov/government/departments/public-works/public-works-downtown-team/transportation-policy

City of Santa Cruz (2017). Active Transportation Plan. Retrieved from https://www.cityofsantacruz.com/home/showpublisheddocument/60966/636353003776970000

Fitch, D. T., Kamalapuram, S., Favetti, M., & Handy, S. L. (2022). Caltrans Active Transportation Benefit–Cost Tool: Technical Documentation. https://activetravelbenefits.ucdavis.edu/technicaldocs

Fitch-Polse, D. T., & Agarwal, S. (2025). The benefits of active transportation interventions: A review of the evidence. *Journal of Transport and Land Use,* 18(1), 77–122. https://doi.org/10.5198/jtlu.2025.2468

Kaiser, S. K., Klein N., and Kaack, L. R. (2025). "From counting stations to city-wide estimates: data-driven bicycle volume extrapolation." *Environmental Data Science* 4: e13.

Kamalapuram S. (2022). Estimating bicycle and pedestrian ridership using the Random Forest algorithm. UC Davis. ProQuest ID: Kamalapuram_ucdavis_0029M_21235. Merritt ID: ark:/13030/m56t7vgs. Retrieved from https://escholarship.org/uc/item/1qj4f71g

Laustsen, K., Mah, S., Semler, C., Nordback, K., Sandt, L., Sundstrom, C., Raw, J., Jessberger, S., 2016. Coding Nonmotorized Station Location Information in the 2016 Traffic Monitoring Guide Format. USDOT FHWA

Miah, M. Md., Griswold, J.B., Proulx, F., Bigham, J., & Grembek, O. (2024). Methodology of a Large-Scale Annual Average Bicycle Traffic (AADBT) Estimation. Poster presentation at the 102nd Annual Meeting of the Transportation Research Board, Washington, DC.

Nordback, K., Sellinger, M., and Phillips T. (2017). Estimating Walking and Bicycling at the State Level. Portland, OR: Transportation Research and Education Center

(TREC). http://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1129&context=trec_reports.

Schneider, R. J., Schmitz, A., and Qin X. (2021). Development and validation of a seven-county regional pedestrian volume model. *Transportation research record* 2675, 6: 352-368.

Data Summary

Products of Research

Two sets of data were collected for this study. The first is a list of infrastructure projects and their attributes. This data was collected through document review, discussions with practitioners, and review of historical imagery on Google Street View. The second data set includes the walking and bicycling counts for the two study cities. This data was added to existing data to update direct demand models and estimate expected project walking and bicycling.

Other data sets that were used for this study include all the data from the California Active Transportation Benefit/Cost tool (URL), data from Miah et al., (2024), data from Strava Metro, American Community Survey, and data from the Smart Location Database. This data was used to update new direct demand models.

Data Format and Content

The infrastructure data is stored in two comma delimited (csv) files (import-sc-infrasutrcture.csv holds the data from Santa Cruz, and import-sb-infrastructure.csv hold data from Santa Barbara). Data and metadata can be found on the UC Davis Dryad Data Repository (https://doi.org/10.5061/dryad.j9kd51cpm).

Data Access and Sharing

All supplemental data can be accessed at its original source location. People interested in the intermediate data processed as a part of this project can contact the authors. Because of use restrictions, the data from Strava Metro cannot be made available without permission from Strava.

Reuse and Redistribution

Please find the license information for restrictions on the use and distribution of this data on the Dryad dataset page.