Original Paper

Estimating the Health Effects of Adding Bicycle and Pedestrian Paths at the Census Tract Level: Multiple Model Comparison

Ross Gore[1], BSc, MSc, PhD; Christopher J Lynch[1], BSc, MSc, PhD; Craig A Jordan[1], BSc, MSc; Andrew Collins[2], BSc, MSc, PhD; R Michael Robinson[1], BSc, MSc, PhD; Gabrielle Fuller[3], BSc; Pearson Ames[4], BA; Prateek Keerthi[5], BTech, MSc; Yash Kandukuri[1], BTech, MSc

[1]Virginia Modeling Analysis and Simulation Center, Old Dominion University, Suffolk, VA, United States
[2]Engineering Management & Systems Engineering, Old Dominion University, Norfolk, VA, United States
[3]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, United States
[4]Department of Economics, Chapman University, Orange, CA, United States
[5]Hampton Roads Biomedical Research Consortium, Norfolk, VA, United States

Corresponding Author:
Ross Gore, BSc, MSc, PhD
Virginia Modeling Analysis and Simulation Center
Old Dominion University
1030 University Dr
Suffolk, VA, 23435
United States
Phone: 1 757 686 6200
Fax: 1 757 686 6214
Email: ross.gore@gmail.com

Abstract

Background: Adding additional bicycle and pedestrian paths to an area can lead to improved health outcomes for residents over time. However, quantitatively determining which areas benefit more from bicycle and pedestrian paths, how many miles of bicycle and pedestrian paths are needed, and the health outcomes that may be most improved remain open questions.

Objective: Our work provides and evaluates a methodology that offers actionable insight for city-level planners, public health officials, and decision makers tasked with the question "To what extent will adding specified bicycle and pedestrian path mileage to a census tract improve residents' health outcomes over time?"

Methods: We conducted a factor analysis of data from the American Community Survey, Center for Disease Control 500 Cities project, Strava, and bicycle and pedestrian path location and use data from two different cities (Norfolk, Virginia, and San Francisco, California). We constructed 2 city-specific factor models and used an algorithm to predict the expected mean improvement that a specified number of bicycle and pedestrian path miles contributes to the identified health outcomes.

Results: We show that given a factor model constructed from data from 2011 to 2015, the number of additional bicycle and pedestrian path miles in 2016, and a specific census tract, our models forecast health outcome improvements in 2020 more accurately than 2 alternative approaches for both Norfolk, Virginia, and San Francisco, California. Furthermore, for each city, we show that the additional accuracy is a statistically significant improvement ($P<.001$ in every case) when compared with the alternate approaches. For Norfolk, Virginia (n=31 census tracts), our approach estimated, on average, the percentage of individuals with high blood pressure in the census tract within 1.49% (SD 0.85%), the percentage of individuals with diabetes in the census tract within 1.63% (SD 0.59%), and the percentage of individuals who had >2 weeks of poor physical health days in the census tract within 1.83% (SD 0.57%). For San Francisco (n=49 census tracts), our approach estimates, on average, that the percentage of individuals who had a stroke in the census tract is within 1.81% (SD 0.52%), and the percentage of individuals with diabetes in the census tract is within 1.26% (SD 0.91%).

Conclusions: We propose and evaluate a methodology to enable decision makers to weigh the extent to which 2 bicycle and pedestrian paths of equal cost, which were proposed in different census tracts, improve residents' health outcomes; identify areas where bicycle and pedestrian paths are unlikely to be effective interventions and other strategies should be used; and quantify the minimum amount of additional bicycle path miles needed to maximize health outcome improvements. Our methodology shows statistically significant improvements, compared with alternative approaches, in historical accuracy for 2 large cities (for 2016) within different geographic areas and with different demographics.

(JMIR Public Health Surveill 2022;8(8):e37379) doi: 10.2196/37379

KEYWORDS

bicycle paths; pedestrian paths; bicycling; walking; diabetes; high blood pressure; physical health; factor analysis; digital neighborhoods; data analysis

Introduction

The addition of bicycle and pedestrian paths to an area is a theoretically valuable resource for city-level planners, public health officials, and decision makers to increase physical activity and improve health outcomes. Most existing research has found a negative association between the prevalence of bicycle and pedestrian paths and poor health outcomes (ie, diabetes, stroke, obesity, heart disease, high blood pressure, and ailments to physical and mental health) [1-10].

Objectives

Our objective is to provide and evaluate a methodology for officials addressing the question "To what extent will adding specified bicycle and pedestrian path mileage to a census tract improve residents' health outcomes over time?" The methodology we propose uses factor analysis to filter and organize variables from publicly available data sets at the census tract level within a given city. The data sets included (1) the US Census [11], (2) the American Communities Survey (ACS) [12], (3) Centers for Disease Control and Prevention (CDC) 500 Cities project data [13], (4) municipality data [14,15], and (5) the GPS walking, running, and cycling tracking social network app, Strava [16,17].

The result of this analysis is a city-specific factor model describing the relationship among variables related to individuals, bicycling and walking behaviors, and health outcomes. Then, the factor model, built using past data, is used in an algorithm to predict the extent to which adding a future specified number of bicycle and pedestrian path miles to a certain location in the city quantitatively impacts certain health outcomes.

Background

We are not aware of any other applications of factor analysis to develop predictive algorithms related to the placement and efficacy of bicycle and pedestrian paths with respect to health outcomes. However, there are researchers who approach bicycle and pedestrian path planning from a similar perspective. Smith and Haghani [18] proposed an approach that adds bicycle and pedestrian paths within a city such that the length of the average trip within the bicycle and pedestrian path network is minimized, and the level of service of the bicycle and pedestrian paths is maximized. Mesbah et al [19] explored the addition of bicycle and pedestrian paths within a city by identifying locations that minimized the total travel time of automobiles within the city. Researchers assume that bicycle and pedestrian paths take road space from cars. Although this assumption may occasionally be true, in most instances, bicycle and pedestrian paths narrow car lanes but do not reduce the total number available. Duthie and Unnikrishnan [20] identified instances within a city where the addition of bicycle and pedestrian paths maximized the connectivity of the existing bicycle and pedestrian path network. This approach ignores the use of the current bicycle and pedestrian path network and aims to "open up" as many new routes as possible regardless of current demand [21].

Although they are not prevalent in identifying bicycle and pedestrian path placement, optimization techniques have also been explored for choosing existing routes rather than developing new ones. Allen-Munley et al [22] developed a model that rates bicycle routes based on predictions of injury severity [18]. Other researchers have proposed allowing users to select multiple criteria and then eliminate certain routes (ie, steep slopes and heavy traffic) before providing a set of suggestions [23,24]. More recently, researchers have explored the use of multiobjective optimization as a means of retrofitting the existing cycling infrastructure for commuter cyclists. The objective of the formulation is to maximize the network for a number of different criteria, including accessibility, minimization of the number of intersections, maximization of bicycle level of service, and minimization of total construction cost subject to space-time constraints and monetary budget [25-27].

Ospina et al [28] addressed a similar problem but framed it as a maximal covering bicycle network design problem. The maximal covering bicycle network design problem involves making investment decisions to build a cycling network aimed at maximizing the coverage of cyclists while maintaining a minimum total network cost. The derived network is subject to budget and accounts for the entire connectivity and directness as fundamental bicycle network design criteria. This approach focuses only on the network and not on the health outcomes. There is no consideration of the extent to which each path in the network improves any health outcome within an area.

It is important to note that there are arguments against defining the placement of bicycle and pedestrian paths as a systems engineering problem. Szimba and Rothengatter [29] demonstrated that interdependencies between infrastructure projects can create cost incentives to place bicycle and pedestrian paths in certain areas, even if the payoff of the addition is not optimal with respect to the use, connectivity, or health benefits of the bicycle and pedestrian path. In addition, in areas where congestion and the propagation of congestion along bicycle and pedestrian paths occur, researchers have demonstrated that optimizing the use and distance of bicycle and pedestrian paths would only exacerbate traffic within the network and not produce effective results [30-32].

Furthermore, significant work has been conducted to estimate demand [33,34] and understand why people choose to use bicycle and pedestrian paths [35-40]. Our work also considers motivation related to bicycle and pedestrian path use but does not directly attempt to optimize bicycle and pedestrian path use. We made this design choice because adding bicycle and pedestrian paths based only on the existing demand can lead to a chicken-and-egg problem. Here, areas with advanced bicycle and pedestrian path infrastructure improve, and areas without bicycle and pedestrian path infrastructure are neglected. These dynamics can create inequitable living conditions and produce enormous health and environmental disparities within a city [41].

In summary, the algorithm used in this study is unique from previous approaches used for estimating demand, evaluating network efficacy, and optimizing the placement of bicycle and pedestrian paths. The problem examined here focuses on understanding what health outcomes can be improved by adding bicycle and pedestrian paths, in which census tracts will adding bicycle and pedestrian paths improve health outcomes the most, and finally, how many miles of bicycle and pedestrian paths within a given census tract need to be added to have an impact on the residents' health outcomes.

The remainder of this paper is organized as follows. First, we review the data and methods used in our approach to construct city-specific models. Next, we apply the approach to two different cities: Norfolk, Virginia, and San Francisco, California. We then evaluate our approach for the 2 different cities. In the evaluation, our approach was tested against 2 alternate approaches for predicting improvements in health outcomes by adding bicycle and pedestrian paths. The evaluation shows that our approach offers more accurate predictions than both alternatives and that the superior difference in accuracy is statistically significant ($P<.001$ in all cases). Finally, we identify several limitations to our work and threats to its validity and review other avenues of related research.

Methods

Ethical Considerations

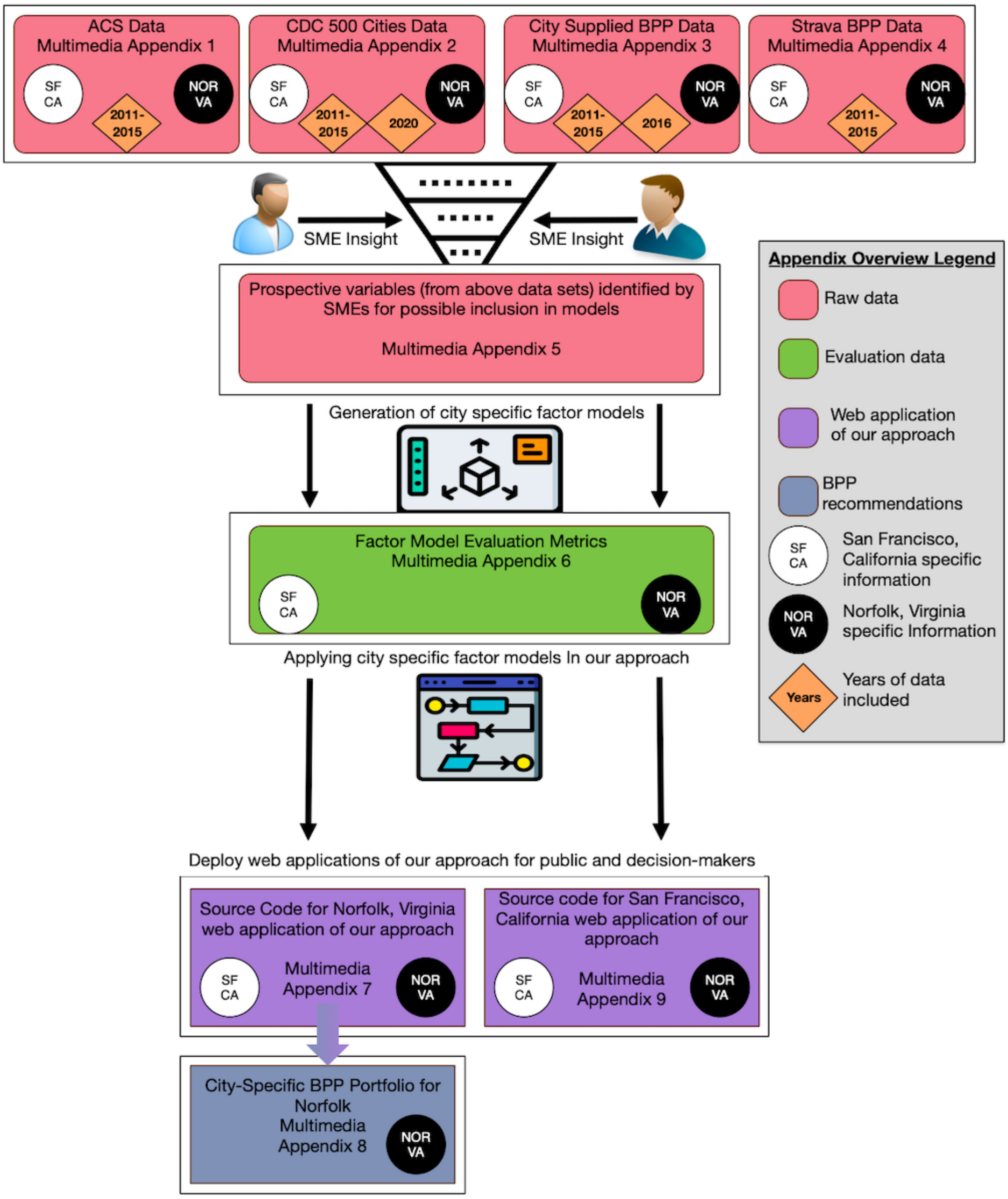
Our work uses publicly-available data related to urban infrastructure and resident demographics and health outcomes. The data sets reflect aggregate variables measured at the census tract level of a city and do not contain any personally identifiable information. Therefore, they do not involve human subjects as defined by federal regulations and their use does not require ethics board review or approval [42].

Data Sets

Overview

Our approach to modeling the health effects of adding bicycle and pedestrian paths at the census tract level uses data from (1) census tract boundaries used in the US Census [11]; (2) demographic variables from the ACS [12]; (3) census tract–level estimates for health outcomes, health statuses, healthy behaviors, and disease prevention from the CDC [13]; (4) bicycle and pedestrian path location and use data from Norfolk, Virginia, and San Francisco, California [14,15]; and (5) bicycle and pedestrian path use data from the GPS walking, running, and cycling tracking social network app, Strava. Combining these data sets resulted in >400 variables for each census tract in Norfolk, Virginia, and San Francisco, California [16,17]. An overview of all the data sets and other supplementary materials supplied in the multimedia appendices of this paper is shown in Figure 1.

Figure 1. An overview of the data sets and other supplementary materials supplied in the multimedia appendices. ACS: American Communities Survey; BPP: bicycle and pedestrian path; CDC: Centers for Disease Control and Prevention; NOR: Norfolk; SF: San Francisco; SME: subject matter expert.



US Census and ACS

Census tracts are small, contiguous, and relatively permanent statistical subdivisions of a county or an equivalent entity. The populations in census tracts vary from 1200 to 8000. Census tracts provide a stable geographic unit for statistical analysis in the US Census and ACS [43].

The ACS is an ongoing national survey that samples a subset of individuals within the same geographic areas in the US Census. Using the same questions, data were collected each month throughout the year. In contrast, the US Census provides a more comprehensive sample of individuals in the United States, collecting data from more individuals during a particular period (March to August) but administered only once every 10 years. A metaphor helps elucidate the differences between the 2 surveys. The US Census serves as a high-resolution photograph of the US population once every 10 years, whereas the ACS serves as many low-resolution continually updated videos over the same period [43]. Multimedia Appendix 1 provides the data included in the ACS for this study.

CDC 500 Cities Project

The census tract–level estimates and methodology for estimating health outcomes, health statuses, healthy behaviors, and disease prevention are provided by the CDC 500 Cities project. The 500 Cities project is a collaboration between the CDC and the Robert Wood Johnson Foundation. The small area estimates provided by the project allow policymakers and local health departments to better understand the burden and geographic distribution of health-related variables in their jurisdictions and assist them in planning public health interventions [13]. The data included in the CDC 500 Cities project for this study are provided in Multimedia Appendix 2.

City-Supplied Bicycle and Pedestrian Path Data

The bicycle and pedestrian path data for Norfolk, Virginia, and San Francisco, California include the latitude and longitude location of bicycle lanes, routes, and paths built and maintained in each city. Bicycle use data were taken from bicycle counters used in each city [14,15]. The data included from Norfolk, Virginia, and San Francisco, California, for this study are provided in Multimedia Appendix 3.

Strava Data

We used the Strava Metro rollup data set for Norfolk, Virginia, and San Francisco, California. This data set contains walking, running, and bicycling activity counts per road segment for a given year. These counts can then be aggregated at the census tract level. The road count segment is referred to as edge within Strava. Each edge is associated with a latitude and longitude bounding box using the Strava application programming interface [16,17]. The Strava data for Norfolk, Virginia, and San Francisco, California for this study are provided in Multimedia Appendix 4. There are limitations to using the Strava data, which we describe in the *Discussion* section.

Data Selection

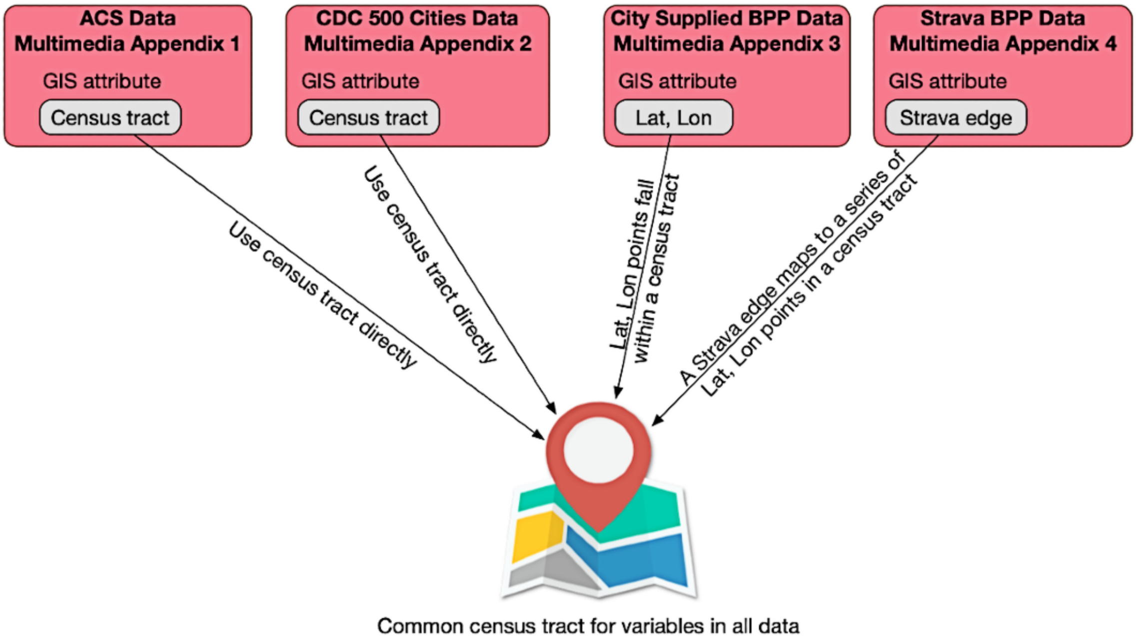
Our data set included a wide range of variables collected from multiple sources. From this data set, we selected a subset of the variables that individuals with domain expertise identified as possibly contributing to the use of bicycle and pedestrian paths and the impact of bicycle and pedestrian paths on health outcomes when additional mileage was added to a geographic area (ie, census tract). The expertise of these individuals spanned social work, health science and nutrition, community health, public health, and transportation. Textbox 1 shows the categories of variables identified by domain experts for each census tract in Norfolk, Virginia, and San Francisco, California. Multimedia Appendix 5 provides the list of observed variables for each category. These variables can be combined using common Geographical Information System attributes to align data at the census tract level. The approach for joining these data together at the census travel level is shown in Figure 2.

Textbox 1. The categories of variables from our data sets that are included in our factor analysis for Norfolk, Virginia, and San Francisco, California.

Data set and variable category

- American Communities Survey
  - Race
  - Educational attainment
  - Employment status
  - Income and benefits
  - Marital status
  - Sex and age
  - Commuting to work
  - Citizenship
  - Health insurance
  - Occupation
  - Household by type
  - Relationship
- Centers for Disease Control and Prevention 500 Cities project
  - Health outcomes
  - Health risk behaviors
  - Prevention
  - Health status
- City Bicycle and Pedestrian Path data
  - Bicycle and Pedestrian Path use data
  - Bicycle and Pedestrian Path mileage data
- Strava Bicycle and Pedestrian Path data
  - Bicycle and Pedestrian Path use data

Figure 2. The approach to joining together the data sets at the census tract level. ACS: American Communities Survey; BPP: bicycle and pedestrian path; CDC: Centers for Disease Control and Prevention; GIS: Geographical Information System.



Factor Analysis

Overview

Next, we applied factor analysis to reduce these observed variables into latent variables (ie, factors). Factor analysis generates a model that measures how changes in one factor predict changes in another by reducing a large number of observed variables to a handful of comprehensible underlying factors. The result is an interpretable and actionable model of concepts that are otherwise difficult to measure [44].

The Honesty-Humility (H), Emotionality (E), Extraversion (X), Agreeableness (A), Conscientiousness (C), and Openness to Experience (O) 6D model of the human personality structure is a widely known result of the application of factor analysis. The ability of factor analysis to reduce the many observed variables related to personality into 6 distinct factors has pushed the state of the art in psychological research [45]. Our goal of applying factor analysis was similar.

We applied exploratory factor analysis (EFA) to filter the observed variables from the data described in Textbox 1 and reduced them into a model composed of factors that include residents' (1) demographics and background characteristics (DBC), (2) health, and (3) bicycling and pedestrian habits (BPH). Using this model, we can understand how changes in one factor predict changes in others.

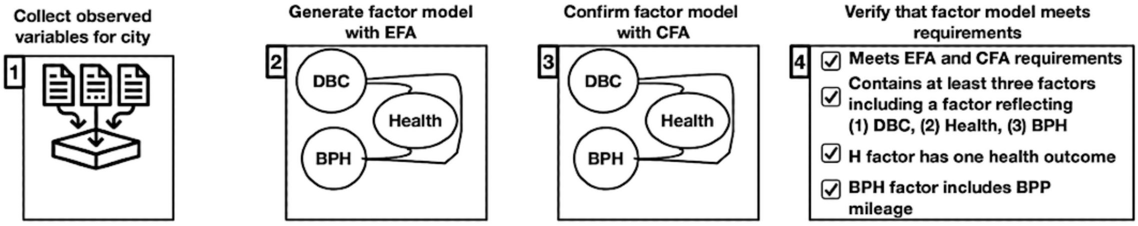
EFA Summary

In our approach, EFA was used to fit a factor model. Before the EFA began, data corresponding to half of a given city's census tracts were selected at random. In the application of our approach, data from 2011 to 2015 were used. Then, using these data, an EFA model was fitted.

Figure 3 shows the fitting of the model using EFA. The process is iterative, and each iteration comprises 3 stages. Figure 3A shows the observed variables that underwent analysis for a given iteration. These observed variables are organized into a number of factors that optimize the fit of the model in Figure 3B. The optimization constructs a model with the minimum number of factors such that the observed variables associated with each factor have maximum commonality with one another and minimal commonality with the observed variables in all other factors. Commonality reflects the amount of variance an observed variable shares with other variables in a factor [44,46].

Finally, the model was assessed. The assessment tests if all factors are composed of variables with high communality (>0.5) with respect to the factor they are associated with and low communality (<0.5) with all other factors. If this is true, the process terminates. Otherwise, variables that do not meet the communality requirement are discarded and the process is repeated for another iteration. Figure 3C shows the assessment stage of the iteration. The requirements imposed in this stage are consistent with the established factor analysis guidelines [46].

Figure 3. The process of generating a factor model for a city and verifying that it meets our defined restrictions. BPH: bicycling and pedestrian habits; BPP: bicycle and pedestrian path; CFA: confirmatory factor analysis; DBC: demographics and background characteristics; EFA: exploratory factor analysis.



Confirmatory Factor Analysis Summary

Next, the fit of the hypothesized model was confirmed or rejected by applying confirmatory factor analysis (CFA) using the other half of the data from 2011 to 2015. The goal of CFA is to confirm or reject the hypothesized model. As a result, (1) only observed variables were included, (2) the variables were loaded onto the same factors as in the CFA, and (3) the communality of the variables in the model was assessed. The model was confirmed if it satisfied the same requirements as specified for EFA [46].

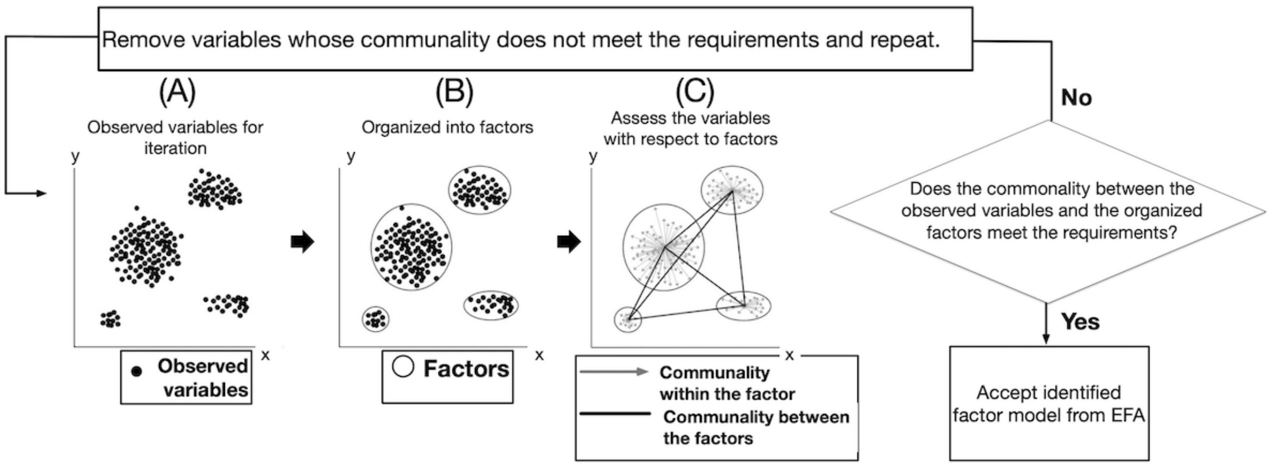
Factor Restrictions and Limitations

The application of factor analysis imposes several limitations on our approach for estimating the health effects of adding bicycle and pedestrian paths to the city-specific factor model. First, a model that meets our requirements must be generated using EFA and confirmed using CFA. Furthermore, to apply our algorithm, the model must consist of at least three factors reflecting residents' (1) DBC, (2) health, and (3) BPH. Finally, the health factor must include at least one observed variable related to a health outcome, and the BPH must include an observed variable related to the amount of bicycle and pedestrian path mileage in the census tract. The process of generating a factor model and determining whether it meets these restrictions is illustrated in Figure 4.

We imposed these restrictions because our health outcome prediction algorithm computes the factor scores for each census tract in a city based on these factors. Factor scores are continuous numbers reflecting the extent to which each census tract manifests each factor. For each factor, the scores were distributed normally, with a mean of 0 and an SD of 1. Large positive values reflect census tracts where the factor is heavily present, and large negative values reflect census tracts where the factor is not present at all [47].

Without these factors, the proposed algorithm could not be applied. It does not have sufficient data or structure to produce estimates of the health effects of adding bicycle and pedestrian paths. This is a limitation of the proposed approach. This limitation is discussed in more detail in the *Discussion* section.

Figure 4. The three stages of an EFA iteration—(A) observed variable identification, (B) organization of variables into factors, and (C) assessment of the communality of variables within and between each of the identified factors. EFA: exploratory factor analysis.



Estimating the Health Effects of Adding Bicycle Paths at the Census Tract Level

Overview

Given a factor model hypothesized by EFA and confirmed by CFA, we proposed an algorithm to predict the health effects of adding bicycle and pedestrian paths at the census tract level. For this purpose, we defined the input as an observed variable identified from the factor model. The variable then progressed through a sequence of steps that were applied to each census track and resulted in a predicted health outcome change for each identified health factor. The steps of this algorithm are enumerated in the following sections. Finally, the output from the algorithm was a list of hypothesized health improvement outcomes.

Input

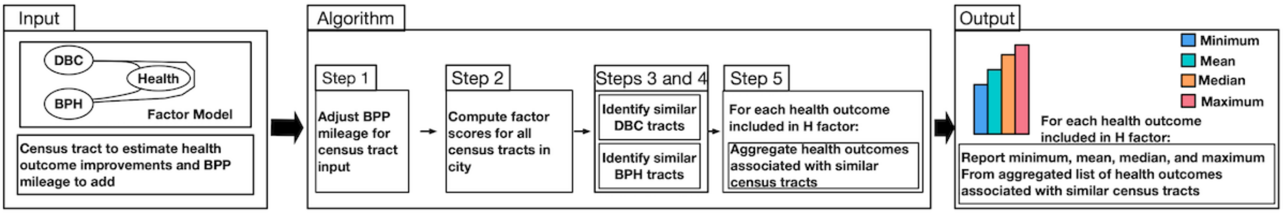
In our problem statement, there was only one observed variable in the model that could be changed directly by a city-level planner, public health official, or decision maker. This variable represented the additional bicycle and pedestrian path mileage for a census tract within a city. This was the input to our algorithm, along with the factor model generated for the city.

Algorithm

The algorithm proceeded as follows, as conveyed visually in Figure 5.

1. The algorithm adds the bicycle and pedestrian path mileage to the specified census tract in the data set for the city.
2. Factor scores are computed for the following three factors: DBC, health, and BPH.
3. Given the DBC factor score for the input census tract, the algorithm identifies all other census tracts in the city with a DBC factor score within the threshold value—x. This list of census tracts reflects those that are similar to the input census tract with respect to the DBC factor. Recall that the factor scores are normally distributed, with an SD of 1. Thus, a census tract within a factor score x of the tract being analyzed reflects a census tract within SDs of the input tract [47].
4. Given the BPH factor score for the input census tract (which includes the newly added bicycle and pedestrian path mileage), the algorithm identifies all other census tracts in the city with BPH factor scores within x. This list of census tracts reflects those that are similar to the input census tract with respect to the BPH factor.
5. For each observed health outcome within the health factor, the algorithm creates a list that stores the difference between the value of the health outcome for each census tract identified in steps 3 and 4 and the value of the health outcome for the input census tract. This list of differences is a distribution of hypothesized improvements in a health outcome by adding a specified amount of bicycle and pedestrian path mileage to a census tract. Any differences that are <0 are discarded because these differences indicate that adding bicycle and pedestrian path mileage to the census tract will degrade health outcomes.

Figure 5. Instantiation of the algorithm for predicting how much additional BPP mileage in a census tract will improve health outcomes. BPH: bicycling and pedestrian habits; BPP: bicycle and pedestrian path; DBC: demographics and background characteristics.



Output

For each list of hypothesized improvements for health outcomes generated in step 5, the algorithm output the minimum, mean, median, and maximum values of the improvements to the user. The algorithm could also report the entire distribution of possible improvements and SD of the distribution for each health outcome.

Results

Overview

The accuracy of our algorithm was elucidated through an empirical evaluation of alternative approaches for two different cities (Norfolk, Virginia, and San Francisco, California). In our evaluation, we computed how accurately each approach predicted the health outcome improvements of the bicycle and pedestrian paths added in each city in 2016. Specifically, for a given census tract, in each city that added bicycle and pedestrian paths miles in 2016, we evaluated how accurately our algorithm estimated an improvement in health outcomes in 2020. We chose to use a 5-year time-lapse period for our evaluation because research has shown that is the expected amount of time for a fully realized change in health outcomes given outdoor exercise infrastructure interventions [48,49].

Factor Analyses

Applying the process described in the *Methods* section and shown in Figures 3 and 4 with the data from half the census tracts in each city for each year from 2011 to 2015 yields the EFA models shown in Figure 6A (n=195) and Figure 7A (n=490). Confirmation of these models using the remaining half of the census in each city for each year from 2011 to 2015 is shown in Figure 6B (n=190) and Figure 7B (n=485). Within the figures, the numbers labeled with single-headed arrows reflect the commonality of an observed variable with the associated factor. The double-headed arrows reflect the shared variance between factors [44,46]. The goodness-of-fit statistics corresponding to the CFA for each model are provided in Multimedia Appendix 6 (Norfolk, Virginia) and Multimedia Appendix 6 (San Francisco, California) along with guidelines on how to interpret the goodness-of-fit statistics.

Figures 6 and 7 show that the factor models for each city met our requirements. These models served as inputs for our estimation algorithm in the evaluation. It is important to note that although each model had the three required factors (DBC, health, and BPH), there were differences in the observed variables that form the factors. The factor analysis showed that changes in high blood pressure, diabetes, and poor physical health were predicted by changes in DBC and BPH in Norfolk, Virginia, whereas changes in stroke and diabetes were predicted by changes in DBC and BPH in San Francisco, California. This was not unexpected or a violation of the requirements of our approach. Although we required the 3 factors to be present, we anticipated that different observed variables would form these 3 factors for different cities.

Figure 6. Exploratory factor analysis and confirmation factor analysis models for Norfolk, Virginia, using data sets from 2011 to 2015. Single-headed arrows reflect the commonality of an observed variable with a factor. Double-headed arrows reflect the value of the shared variance between factors. BPH: bicycling/pedestrian habits; BPP: bicycle and pedestrian path; DBC: demographics and background characteristics.

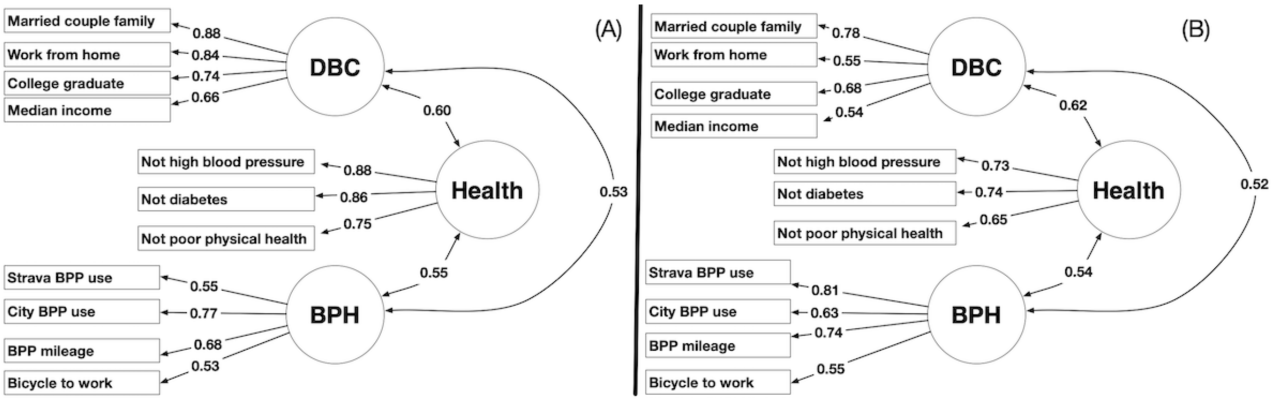
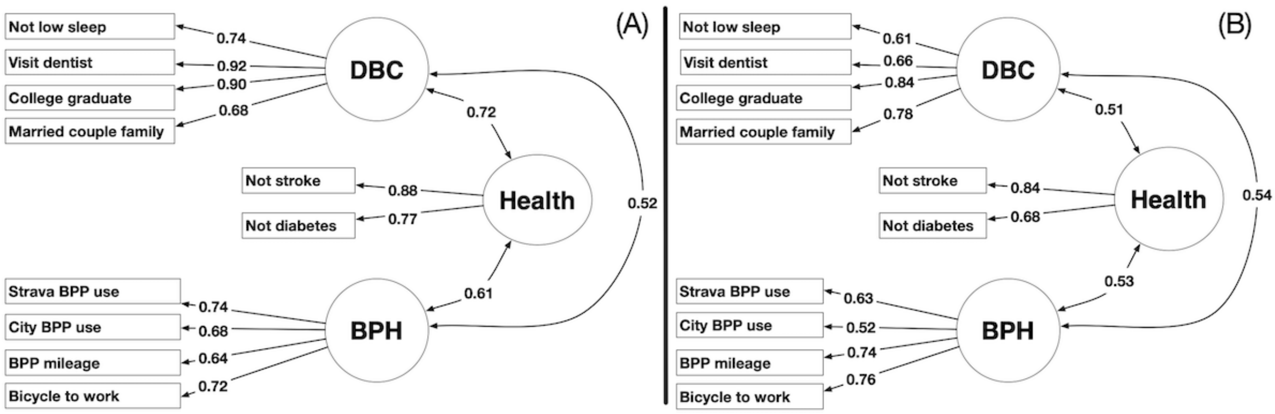


Figure 7. Exploratory factor analysis and confirmation factor analysis models for San Francisco, CA, using data sets from 2011 to 2015. Single-headed arrows reflect the commonality of an observed variable with a factor. Double-headed arrows reflect the value of the shared variance between factors. BPH: bicycling and pedestrian habits; BPP: bicycle and pedestrian path; DBC: demographics and background characteristics.



Evaluation

Recall that our algorithm took an input: (1) the factor model for a given city and (2) the census tract and amount of bicycle and pedestrian path mileage to be added. It then output the minimum, mean, median, and maximum estimated improvements by adding the bicycle and pedestrian path mileage to the input census tract. In the evaluation, we only used the median improvement estimate from the algorithm.

In our evaluation, we used our factor model constructed using data from 2011 to 2015 to estimate the accuracy of our approach and 2 alternative approaches with respect to the improvements in health outcomes provided by bicycle and pedestrian paths installed in 2016. The evaluation included 31.58 miles (50.81 km) of bicycle and pedestrian paths added in Norfolk, Virginia, across 31 census tracts and 52.36 miles (84.25 km) of bicycle and pedestrian paths added tracts in San Francisco, California, across 49 census tracts. Table 1 provides additional details regarding the setup of the evaluation.

Table 1. Evaluation setup metadata for Norfolk, Virginia, and San Francisco, California, in 2016.

	Norfolk, Virginia	San Francisco, California
BPP[a] miles (km) added	31.58 (50.81)	52.36 (84.25)
Census tracts with paths added, n	31	49
Census tracts in city, n	77	195
Health outcomes evaluated	Diabetes %; poor physical health %; high blood pressure %	Diabetes %; stroke %

[a]BPP: bicycle and pedestrian path.

Alternative Approaches

We evaluated our algorithm using 2 alternative approaches. The first alternative assumed that each health outcome within a census tract in the future would be same as the average value for that health outcome for the census tract from 2011 to 2015. This approach mirrored the prediction that the temperature tomorrow would be the same as the average temperature of the previous 5 days.

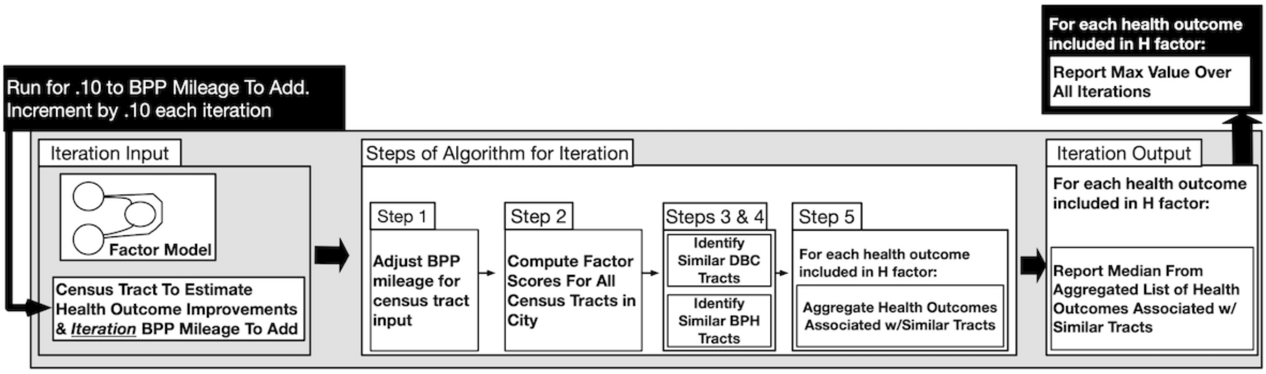
The second alternative used linear regression modeling [50]. This approach used regression to predict future changes in each health outcome using a weighted linear combination of the (1) DBC factor and (2) BPH factor scores of the census tract based on the constructed factor model using data from 2011 to 2015, after the specified increase in mileage.

Approach

We evaluated our approach by using x=0.50. Recall that x is the threshold used to identify similar census tracts in terms of the (1) DBC factor and (2) BPH factor scores. In addition, our evaluation approach is an extension of the algorithm described in the *Methods* section. For our evaluation, given a specified number of bicycle and pedestrian path miles to be added and a census tract, we ran the algorithm for every 0.10-mile increment of bicycle and pedestrian paths up to the specified number of miles.

Each time the algorithm was executed, the median improvement from the algorithm was collected. The largest improvement over all the runs was reported. A version of our approach is shown in Figure 8. It implemented the assumption that adding more bicycle and pedestrian path mileage (ie, 1.0 miles as opposed to 0.5 miles) to a given census tract would not be detrimental to the expected improvement in a health outcome.

Figure 8. The specific version of our algorithm included in the applied evaluation. BPP: bicycle and pedestrian path; BPH: bicycling and pedestrian habits; DBC: demographics and background characteristics.



Measures of Effectiveness

For a given city and a given approach to estimating the improvement in a health outcome for bicycle and pedestrian paths added in 2016, we computed the following two measures of effectiveness (MOEs): (1) the root mean squared error (RMSE) and (2) the mean absolute error (MAE). These are 2 established metrics used to measure the accuracy of continuous variables. MAE measures the average magnitude of the errors in a set of predictions without considering their direction. It reflects the average over the evaluation of the absolute differences between the prediction and actual observation where all individual differences have equal weight. RMSE also measures the average magnitude of the error. However, it reflects the square root of the average squared differences between the predicted and actual observations. Within the RMSE, the errors were squared before they were averaged. As a result, the RMSE gives a relatively high weight to large errors [51]. By using both metrics as MOEs, we could capture the accuracy of each approach for decision makers who (1) view all errors equally (MAE) and (2) view large errors as particularly undesirable (RMSE).

Measures of Success

We deem our approach successful if, for each city included in our evaluation, our approach is more accurate across every MOE than the best alternative approach, and these differences are all statistically significant at $P<.01$, when a 1-tailed paired sample t test is applied. We used a 1-tailed paired sample t test to determine whether the mean paired difference between the MOEs of our approach and an alternate approach was <0 (ie, our approach was more accurate). In this procedure, paired observations reflected the MOEs for a given census tract. Within the pair, one observation corresponded to our approach, and the other corresponded to an alternative approach [52].

Discussion

Principal Findings

In our evaluation, we compare the accuracy of our factor model approach, a linear regression approach, and predict no change approach. Each approach estimates the improvements in health outcomes provided by bicycle and pedestrian paths installed in 2016 in 31 census tracts in Norfolk, Virginia and 49 census tracts in San Francisco, California. The results of the evaluation are shown in Table 2.

Table 3 shows that our approach is more accurate than the alternatives, and Table 4 shows that those improvements in accuracy over the best alternative are statistically significant because $P<.001$ for every health outcome in each city when the 1-tailed paired t test is applied.

We expected our approach to outperform the "predict no change approach" because the CDC 500 Cities project and bicycle and pedestrian path data for both cities show that most of the time when a bicycle path of any length is added, the health outcomes identified by the factor analysis improve within 5 years. However, we did not know whether our approach outperformed the linear regression approach.

The results of the evaluation showed that our approach outperformed the linear regression models because it assumed that critical thresholds within the DBC and BPH factors existed (parameter x in steps 3 and 4 of the algorithm). The linear regression approach did not make this assumption [50]. By accounting for this threshold, our approach ensured that it did not overpredict the improvement offered by additional bicycle path miles when the DBC or BPH factor for the census tract indicated that the additional path miles would be ineffective.

By not accounting for this threshold, the linear regression approach could overpredict the expected improvement in health outcomes within a census tract. This was because the linear regression approach assumed that some amount of bicycle and pedestrian paths in each census tract would yield a population without any negative health outcomes. This is unrealistic. Our evaluation results in Tables 3 and 4 demonstrate that linear regression yields statistically significant inferior accuracy, as measured by our 1-tailed paired *t* test.

Table 2. Evaluation of approaches for bicycle and pedestrian paths added in Norfolk, Virginia, in 2016.

Health outcome and MOE[a] (% of individuals who experience a negative health outcome)	Predict no change (census tract: n=31), mean (SD)	Linear regression (census tract: n=31), mean (SD)	Our approach (census tract: n=31), mean (SD)
Diabetes			
MAE[b]	2.33 (0.66)	2.14 (0.67)	1.63 (0.59)
RMSE[c]	2.41 (0.62)	2.29 (0.61)	1.67 (0.55)
Poor physical health			
MAE	2.69 (0.72)	2.21 (0.69)	1.83 (0.57)
RMSE	2.64 (0.69)	2.27 (0.66)	1.94 (0.56)
High blood pressure			
MAE	2.95 (1.17)	2.27 (1.07)	1.49 (0.85)
RMSE	3.18 (1.13)	2.38 (0.92)	1.55 (0.82)

[a]MOE: measure of effectiveness.

[b]MAE: mean absolute error.

[c]RMSE: root mean squared error.

Table 3. Evaluation of approaches for bicycle and pedestrian paths added in San Francisco, California, in 2016.

Health outcome and MOE[a] (% of individuals who experience a negative health outcome)	Predict no change (census tract: n=49), mean (SD)	Linear regression (census tract: n=49), mean (SD)	Our approach (census tract: n=49), mean (SD)
Diabetes			
MAE[b]	2.32 (1.19)	2.18 (1.18)	1.24 (0.91)
RMSE[c]	2.44 (1.11)	2.41 (1.11)	1.35 (0.90)
Stroke			
MAE	2.68 (0.58)	2.78 (0.68)	1.81 (0.52)
RMSE	3.19 (0.52)	2.97 (0.64)	1.88 (0.49)

[a]MOE: measure of effectiveness.

[b]MAE: mean absolute error.

[c]RMSE: root mean squared error.

Table 4. Assessment of whether the improved accuracy of bicycle and pedestrian paths added in 2016 is statistically significant.

City, health outcome, and MOE[a]	Statistical significance of our approach MOE versus best alternative MOE, *P* value
Norfolk, Virginia (census tract: n=31)	
Diabetes	
MAE[b]	<.001
RMSE[c]	<.001
Poor physical health	
MAE	<.001
RMSE	<.001
High blood pressure	
MAE	<.001
RMSE	<.001
San Francisco, California (census tract: n=49)	
Diabetes	
MAE	<.001
RMSE	<.001
Stroke	
MAE	<.001
RMSE	<.001

[a]MOE: measure of effectiveness.

[b]MAE: mean absolute error.

[c]RMSE: root mean squared error.

Comparison With Prior Work

Our study builds on a significant amount of previous research. Numerous researchers have used statistical analyses to (1) explore the health effects of commuting via bicycle or by foot [4,53-62] and (2) assess the health benefits of bicycling and bicycle and pedestrian paths versus the risk of injury or death [63-67]. This study captured data related to walking and bicycling using telephone and web-based surveys [53,54,68], GPS, accelerometers, heart rate monitors [6,58,69-77], bicycling shares [78-80], and social media [17,81].

Predicting which bicycle and pedestrian paths residents will choose is also related to our work. Within this arena, researchers have found different results with respect to the extent to which bicycle and pedestrian path users prefer to take paths that minimize the total travel distance. For example, Broach et al [71,82] used data from Portland, Oregon, to formulate a model that estimated that preferred routes were <10% longer than the shortest path distance. Similarly, Winters et al [39] found that 75% of trips in Vancouver, British Columbia, Canada, were within 10% of the shortest path distance. However, Aultman-Hall et al [83] found no clear relationship between the shortest path distance and percent route deviation in Ontario, Canada, and Krizek et al [84] looked at data in Minneapolis, Minnesota, and found that the average path traveled was roughly twice as long as the shortest path available.

There is also significant research focused on understanding the rate at which future use of bicycle and pedestrian paths will change, as commuters who currently do not use bicycle and pedestrian paths start to transition into commuting by foot or bicycle. Waldykowski et al [85] developed a simulation that explored the conditions under which motor vehicle commuters switch over to commute by bicycle and pedestrian path [85]. Similarly, Mahfouz et al [86] combined distance decay, route calculation, and network analysis methods to examine (1) where future bicycle and pedestrian path commuter demand is within a city, (2) if it is likely to rise, and (3) how such demand could be accommodated within existing bicycle and pedestrian path networks. Finally, Liu et al [87] proposed a connectivity measure that captures the importance of a link in connecting the origins of cyclists and nearby subway stations and incorporated it into a statistical model.

In addition, researchers have attempted to better understand the impact of bicycle and pedestrian paths on health outcomes. This work includes (1) cost-benefit analysis of bicycle and pedestrian paths with respect to health improvements [10,88]; (2) lessons learned from cities with especially enthusiastic cycling culture such as Amsterdam, the Netherlands; Barcelona, Spain; and Chicago, Illinois [49,89,90]; and (3) understanding what type of bicycle and pedestrian paths cyclists and pedestrians prefer [69].

These studies demonstrate the need for granular analysis with actionable outcomes with respect to bicycle and pedestrian paths. Furthermore, although the studies have had a significant impact on the research community, none of them constructed a city-specific model to advise decision makers about the extent to which adding bicycle and pedestrian paths to a census tract would improve residents' health outcomes. Our study addresses this problem within a larger bicycle and pedestrian path research area.

Limitations

Data Limitations

Strava has emerged as a tool of interest for collecting data on bicycling, running, and walking, understanding the effects of new interventions for users, and promoting safety among riders. However, this crowdsourced data are biased toward recreational riders, who are frequent users of GPS-enabled fitness apps. Thus, there is a need to quantify and correct the inherent bias in crowdsourced data to better represent all residents across various demographics. Strava users tend to be more frequently identified as male, be older, and have more income than the general population [17]. In addition, there are limitations to how well the data counted by municipalities reflect the actual volume of bicycle and pedestrian traffic on bicycle and pedestrian paths [91,92]. Research has shown that accounting for biases in placement, time, and day of the week needs to be performed to address these issues [93,94].

Controlling for these biases in the Strava and municipal count data is beyond the scope of our work. However, it is important to note that there were biases in the data. Ultimately, these limitations mean that the Strava data sets that informed our study are nonuniform subsamples of the traffic of cyclists, walkers, and runners in Norfolk, Virginia, and San Francisco, California.

It is also important to note that the use of e-bikes has changed significantly during the period of our study [6]. e-Bikes present a potential opportunity to encourage active transportation while reducing personal barriers to active transportation [95,96]. Survey results suggest that e-bikes may reduce some personal barriers to traditional cycling and allow riders to travel greater distances [97,98]. In addition, e-bikes may have the added benefit of promoting health among individuals who are otherwise reluctant to engage in physical activity [99] and improve metabolic fitness [100] and enjoyment [101]. Exploring how the increased use of e-bikes affects our approach is an opportunity for future work.

Approach Limitations

Recall that our approach uses 5 years of past data to fit a factor model and requires the factor model to consist of at least three factors where unique factors reflect residents' (1) DBC, (2) health, and (3) BPH. In addition, the health factor must include at least one observed variable related to a health outcome, and the BPH factor must include an observed variable related to the amount of bicycle and pedestrian path mileage in the census tract. For cities in which these requirements cannot be met, our approach cannot be applied. This limits its utility and geographic area of applicability. However, related research has shown that these factors are important to account for and often present when understanding who chooses to use bicycle and pedestrian path and how effective bicycle and pedestrian paths are in improving health outcomes [2,56,78,102-104]. Furthermore, these factors provide a structure that enables our approach to predict improvements in health outcomes more accurately than the alternative approaches.

Validity Threats

Threats to internal and external validity affected our study. Threats to internal validity arose when factors affected the dependent variables without evaluators' knowledge. It is possible that some flaws in the implementation of our model could have affected the evaluation results. However, our approach used established libraries to conduct factor analysis, and the source code passed internal reviews [105,106].

Threats to external validity occur when evaluation results cannot be generalized. Although the evaluation was performed using more than 83 miles of added bicycle paths in 80 census tracts across the 2 cities, the factor models and accuracy results cannot necessarily be generalized to other areas. In addition, the factor analysis that generates our models assumes that each pair of variables follows a bivariate normal distribution. Although we verified that this assumption was true in our data, it may not be generalizable to other data sets and other cities where the approach is applied. However, it is very important to note that our approach, which yielded models producing these results, can be applied to other cities assuming that factor models that meet our requirements exist [105,106].

Conclusions

Our work is directly actionable for policy makers, public health professionals, and urban planners in Norfolk, Virginia, and San Francisco, California, by providing concrete insight into the question "To what extent will adding specified bicycle and pedestrian path mileage to a census tract improve residents' health outcomes over time?" Specifically, it enables them to (1) weigh the extent to which 2 bicycle and pedestrian paths of equal cost proposed in 2 different census tracts improve the health outcomes of the residents, (2) identify areas where bicycle and pedestrian paths are unlikely to be effective public health interventions and other strategies should be used to help residents, and (3) quantify the minimum amount of bicycle path miles that need to be added in a given census tract to maximize the improvement in health outcomes for residents. Our results demonstrate that for 2 different cities, our approach estimates improvements in health outcomes more accurately than alternate approaches, and these improvements are statistically significant.

A web application that implements our algorithm and summarizes its findings in an actionable manner is available [107]. Multimedia Appendix 7 provides the source code for the web application. This application was used to identify a recommended set of bicycle and pedestrian paths across census tracts in Norfolk, Virginia. A time series forecast of the expected improvements in health outcomes for these recommendations was also conducted. These artifacts, which are examples of the types of analyses enabled by our approach, are available in Multimedia Appendix 8. A similar web application that implements our algorithm for San Francisco, California, is available [108]. The source code for it is provided in Multimedia Appendix 9.

Acknowledgments

This study was funded by the Hampton Roads Biomedical Research Consortium (300675-010 IRAD). The Hampton Roads Biomedical Research Consortium reviewed and approved the submission of the manuscript.

Data Availability

The contents of the multimedia appendices are specified below and are supplied in the paper. They are also available on the web as Mendeley Data [109].

Authors' Contributions

RG, CAJ, RMR, AC, and CJL conceived and designed the experiments. RG performed the experiments. RG, GF, YK, and PK analyzed the data. RG, PA, PK, and YK contributed the reagents, materials, and analysis tools. RG, CJL, CAJ, GF, AC, and RMR wrote the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

American Community Survey data used for factor analysis.
[ZIP File (Zip Archive), 4528 KB-Multimedia Appendix 1]

Multimedia Appendix 2

The Centers for Disease Control and Prevention 500 Cities project data were used for factor analysis and evaluation.
[ZIP File (Zip Archive), 2144 KB-Multimedia Appendix 2]

Multimedia Appendix 3

City-supplied bicycle and pedestrian count and path length data used for factor analysis.
[ZIP File (Zip Archive), 488 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Strava-supplied bicycle and pedestrian trip count data used for factor analysis.
[ZIP File (Zip Archive), 49327 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Prospective variables identified by subject matter experts for factor analysis.
[ZIP File (Zip Archive), 10 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Goodness-of-fit measure descriptions and statistics for generated factor models.
[ZIP File (Zip Archive), 5 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Data and source code for web application deployment of model for Norfolk, Virginia.
[ZIP File (Zip Archive), 23680 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Data and source code for research artifacts produced from the model for Norfolk, Virginia, including recommended portfolios of bicycle and pedestrian paths and time series of expected health outcome improvements.
[ZIP File (Zip Archive), 381 KB-Multimedia Appendix 8]

Multimedia Appendix 9

Data and source code for web application deployment of the model for San Francisco, California.
[ZIP File (Zip Archive), 20857 KB-Multimedia Appendix 9]

References

1. Calise TV, Dumith SC, Dejong W, Kohl 3rd HW. The effect of a neighborhood built environment on physical activity behaviors. J Phys Act Health 2012 Nov;9(8):1089-1097. [doi: 10.1123/jpah.9.8.1089] [Medline: 22207103]
2. Cohen JM, Boniface S, Watkins S. Health implications of transport planning, development and operations. J Transport Health 2014 Mar;1(1):63-72. [doi: 10.1016/j.jth.2013.12.004]
3. Dill J. Bicycling for transportation and health: the role of infrastructure. J Public Health Policy 2009;30 Suppl 1:S95-110. [doi: 10.1057/jphp.2008.56] [Medline: 19190585]
4. Ding D, Gebel K. Built environment, physical activity, and obesity: what have we learned from reviewing the literature? Health Place 2012 Jan;18(1):100-105 [FREE Full text] [doi: 10.1016/j.healthplace.2011.08.021] [Medline: 21983062]
5. Garrard J. Health benefits of cycling. In: Pucher J, Buehler R, editors. City Cycling. Cambridge, MA, USA: MIT Press; 2012:11.
6. Hoj TH, Bramwell JJ, Lister C, Grant E, Crookston BT, Hall C, et al. Increasing active transportation through e-bike use: pilot study comparing the health benefits, attitudes, and beliefs surrounding e-bikes and conventional bikes. JMIR Public Health Surveill 2018 Nov 29;4(4):e10461 [FREE Full text] [doi: 10.2196/10461] [Medline: 30497998]
7. Khreis H, May AD, Nieuwenhuijsen MJ. Health impacts of urban transport policy measures: a guidance note for practice. J Transport Health 2017 Sep;6(C):209-227. [doi: 10.1016/j.jth.2017.06.003]
8. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
9. Oja P, Titze S, Bauman A, de Geus B, Krenn P, Reger-Nash B, et al. Health benefits of cycling: a systematic review. Scand J Med Sci Sports 2011 Aug;21(4):496-509. [doi: 10.1111/j.1600-0838.2011.01299.x] [Medline: 21496106]
10. Wang G, Macera CA, Scudder-Soucie B, Schmid T, Pratt M, Buchner D. Cost effectiveness of a bicycle/pedestrian trail development in health promotion. Prev Med 2004 Feb;38(2):237-242. [doi: 10.1016/j.ypmed.2003.10.002] [Medline: 14715217]
11. Smith J. Non-Premium Medical Expenditures for Families and Children: 2010 and 2011. United States Census Bureau. 2014 Mar. URL: https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=393640797E042C0AFA92FA26C6D21B45?doi=10.1.1.639.5426&rep=rep1&type=pdf [accessed 202-08-17]
12. National Research Council. Using the American Community Survey: Benefits and Challenges. Washington, DC, USA: The National Academies Press; 2007.
13. National Center for Chronic Disease Prevention and Health Promotion (U.S.), Division of Population Health, Epidemiology and Surveillance Branch, Centers for Disease Control and Prevention (U.S.), Robert Wood Johnson Foundation, CDC Foundation. 500 Cities Project : local data for better health, 2016 : Allen, TX. Centers for Disease Control and Prevention. 2018 Sep. URL: https://stacks.cdc.gov/view/cdc/61637 [accessed 2022-08-17]
14. San Francisco Open Data Portal. DataSF. 2021. URL: https://datasf.org/opendata/ [accessed 2022-02-03]
15. Norfolk Open Data Portal. City of Norfolk. 2022. URL: https://data.norfolk.gov/ [accessed 2022-02-03]
16. Albergotti R. Strava, popular with cyclists and runners, wants to sell its data to urban planners. The Wall Street Journal Blog. 2014. URL: https://www.wsj.com/digital-print-edition?mod=wsjheader [accessed 2022-02-03]
17. Lee K, Sener IN. Strava Metro data for bicycle monitoring: a literature review. Transport Rev 2020 Jul 30;41(1):27-47. [doi: 10.1080/01441647.2020.1798558]
18. Smith H, Haghani L. A mathematical optimization model for a bicycle network design considering bicycle level of service. In: Transportation Research Board 91st Annual Meeting. 2012 Presented at: TRID '12; January 22-26, 2012; Washington, DC, USA.
19. Mesbah M, Thompson R, Moridpour S. Bilevel optimization approach to design of network of bike lanes. Transport Res Record 2012 Jan 01;2284(1):21-28. [doi: 10.3141/2284-03]
20. Duthie J, Unnikrishnan A. Optimization framework for bicycle network design. J Transp Eng 2014 Jul;140(7):04014028. [doi: 10.1061/(ASCE)TE.1943-5436.0000690]
21. Larsen J, El-Geneidy A, Yasmin F. Beyond the quarter mile: re-examining travel distances by active transportation. Can J Urban Res 2010;19(1):70-88.
22. Allen-Munley C, Daniel J, Dhar S. Logistic model for rating urban bicycle route safety. Transp Res Record 2004 Jan 01;1878(1):107-115. [doi: 10.3141/1878-13]
23. Hochmair H. Towards a classification of route selection criteria for route planning tools. In: Proceedings of the 11th International Symposium on Developments in Spatial Data Handling. 2004 Presented at: SDH '04; August 23-25, 2004; Leicester, UK. [doi: 10.1007/3-540-26772-7_37]
24. Su JG, Winters M, Nunes M, Brauer M. Designing a route planner to facilitate and promote cycling in Metro Vancouver, Canada. Transp Res Part A Policy Pract 2010 Aug;44(7):495-505. [doi: 10.1016/j.tra.2010.03.015]
25. Zhu S, Zhu F. Multi-objective bike-way network design problem with space–time accessibility constraint. Transportation 2019 Jun 19;47(5):2479-2503. [doi: 10.1007/s11116-019-10025-7]

26. Zuo T, Wei H. Bikeway prioritization to increase bicycle network connectivity and bicycle-transit connection: a multi-criteria decision analysis approach. Transp Res Part A Policy Pract 2019 Nov;129:52-71. [doi: 10.1016/j.tra.2019.08.003]
27. Buehler R, Dill J. Bikeway networks: a review of effects on cycling. Transp Rev 2016;36(1):9-27. [doi: 10.1080/01441647.2015.1069908]
28. Ospina JP, Duque JC, Botero-Fernández V, Montoya A. The maximal covering bicycle network design problem. Transp Res Part A Policy Pract 2022 May;159(C):222-236. [doi: 10.1016/j.tra.2022.02.004]
29. Szimba E, Rothengatter W. Spending scarce funds more efficiently - including the pattern of interdependence in cost-benefit analysis. J Infrastruct Syst 2012 Dec;18(4):242-251. [doi: 10.1061/(asce)is.1943-555x.0000102]
30. Yang H, Bell MG. Models and algorithms for road network design: a review and some new developments. Transp Rev 1998 Jul;18(3):257-278. [doi: 10.1080/01441649808717016]
31. Farahani RZ, Miandoabchi E, Szeto WY, Rashidi H. A review of urban transportation network design problems. Eur J Oper Res 2013 Sep;229(2):281-302. [doi: 10.1016/j.ejor.2013.01.001]
32. Kepaptsoglou K, Karlaftis M. Transit route network design problem: review. J Transp Eng 2009 Aug;135(8):491-505.
33. An M, Chen M. Estimating nonmotorized travel demand. Transp Res Record 2007 Jan 01;2002(1):18-25. [doi: 10.3141/2002-03]
34. Turner S, Shunk G, Hottenstein A. Development of a methodology to estimate bicycle and pedestrian travel demand. Federal Highway Administration. 1998 Sep. URL: https://static.tti.tamu.edu/tti.tamu.edu/documents/1723-S.pdf [accessed 2022-02-03]
35. Handy SL, Xing Y, Buehler TJ. Factors associated with bicycle ownership and use: a study of six small U.S. cities. Transportation 2010 Apr 21;37(6):967-985. [doi: 10.1007/s11116-010-9269-x]
36. Park H, Lee YJ, Shin HC, Sohn K. Analyzing the time frame for the transition from leisure-cyclist to commuter-cyclist. Transportation 2011;38(2):305-319. [doi: 10.1007/s11116-010-9299-4]
37. Parkin J, Wardman M, Page M. Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation 2007 Aug 3;35(1):93-109. [doi: 10.1007/s11116-007-9137-5]
38. Pucher J, Dill J, Handy S. Infrastructure, programs, and policies to increase bicycling: an international review. Prev Med 2010 Jan;50 Suppl 1:S106-S125. [doi: 10.1016/j.ypmed.2009.07.028] [Medline: 19765610]
39. Winters M, Brauer M, Setton EM, Teschke K. Built environment influences on healthy transportation choices: bicycling versus driving. J Urban Health 2010 Dec;87(6):969-993 [FREE Full text] [doi: 10.1007/s11524-010-9509-6] [Medline: 21174189]
40. Winters M, Davidson G, Kao D, Teschke K. Motivators and deterrents of bicycling: comparing influences on decisions to ride. Transportation 2010 Jun 13;38(1):153-168. [doi: 10.1007/s11116-010-9284-y]
41. Zuniga-Teran AA, Gerlak AK, Elder AD, Tam A. The unjust distribution of urban green infrastructure is just the tip of the iceberg: a systematic review of place-based studies. Environ Sci Policy 2021 Dec;126:234-245. [doi: 10.1016/j.envsci.2021.10.001]
42. 45 CFR 46. Code of Federal Regulations. Title 45. Public Welfare. Department of Health and Human Services. Part 46. Office for Human Research Protections, US Department of Health and Human Services. 2017 Jan 19. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/revised-common-rule-regulatory-text/index.html [accessed 2022-08-17]
43. Berkley J. Using American community survey estimates and margins of error. United States Census Bureau. 2017. URL: https://www.census.gov/content/dam/Census/programs-surveys/acs/guidance/training-presentations/20170419_MOE_Transcript.pdf [accessed 2022-02-03]
44. Rummel RJ. Applied Factor Analysis. Evanston, IL, USA: Northwestern University Press; 1988.
45. Lee K, Ashton MC. Psychometric properties of the HEXACO personality inventory. Multivariate Behav Res 2004 Apr 01;39(2):329-358. [doi: 10.1207/s15327906mbr3902_8] [Medline: 26804579]
46. Thompson B. Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications. Washington, DC, USA: American Psychological Association; 2004.
47. DiStefano C, Zhu M, Mindrila D. Understanding and using factor scores: considerations for the applied researcher. Pract Assess Res Eval 2009;14(1):20. [doi: 10.7275/da8t-4g52]
48. Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. Am J Prev Med 1998 Nov;15(4):379-397. [doi: 10.1016/s0749-3797(98)00076-2] [Medline: 9838979]
49. van Goeverden K, Nielsen TS, Harder H, van Nes R. Interventions in bicycle infrastructure, lessons from Dutch and Danish cases. Transp Res Procedia 2015;10:403-412. [doi: 10.1016/j.trpro.2015.09.090]
50. Kutner MH, Nachtsheim CJ, Neter J, Li W. Applied Linear Statistical Models. 5th edition. Boston, MA, USA: McGraw-Hill Irwin; 2005.
51. Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci Model Dev 2014 Jun 30;7(3):1247-1250. [doi: 10.5194/gmd-7-1247-2014]
52. Lowry R. Concepts and Applications of Inferential Statistics. Poughkeepsie, NY, USA: Richard Lowry; 2014.
53. Moritz WE. Survey of North American bicycle commuters: design and aggregate results. Transp Res Record 1997 Jan 01;1578(1):91-101. [doi: 10.3141/1578-12]

54. McKenzie B. Modes Less Traveled—Bicycling and Walking to Work in the United States: 2008–2012. United States Census Bureau. 2014 May 8. URL: https://www2.census.gov/library/publications/2014/acs/acs-25.pdf [accessed 2022-02-03]
55. Buehler R, Pucher J. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. Transportation 2011 Jul 6;39(2):409-432. [doi: 10.1007/s11116-011-9355-8]
56. Avila-Palencia I, de Nazelle A, Cole-Hunter T, Donaire-Gonzalez D, Jerrett M, Rodriguez DA, et al. The relationship between bicycle commuting and perceived stress: a cross-sectional study. BMJ Open 2017 Jun 23;7(6):e013542 [FREE Full text] [doi: 10.1136/bmjopen-2016-013542] [Medline: 28645948]
57. Pucher J, Dijkstra L. Promoting safe walking and cycling to improve public health: lessons from The Netherlands and Germany. Am J Public Health 2003 Sep;93(9):1509-1516. [doi: 10.2105/ajph.93.9.1509] [Medline: 12948971]
58. Ek A, Alexandrou C, Söderström E, Bergman P, Delisle Nyström C, Direito A, et al. Effectiveness of a 3-month mobile phone-based behavior change program on active transportation and physical activity in adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 08;8(6):e18531 [FREE Full text] [doi: 10.2196/18531] [Medline: 32510462]
59. Chatterjee K, Sherwin H, Jain J, Christensen J, Marsh S. Conceptual model to explain turning points in travel behavior: application to bicycle use. Transp Res Record 2012 Jan 01;2322(1):82-90. [doi: 10.3141/2322-09]
60. Lindqvist AK, Rutberg S. One step forward: development of a program promoting active school transportation. JMIR Res Protoc 2018 May 08;7(5):e123 [FREE Full text] [doi: 10.2196/resprot.9505] [Medline: 29739733]
61. Robroek SJ, Brouwer W, Lindeboom D, Oenema A, Burdorf A. Demographic, behavioral, and psychosocial correlates of using the website component of a worksite physical activity and healthy nutrition promotion program: a longitudinal study. J Med Internet Res 2010 Sep 30;12(3):e44 [FREE Full text] [doi: 10.2196/jmir.1402] [Medline: 20921001]
62. van den Berg MH, Schoones JW, Vliet Vlieland TP. Internet-based physical activity interventions: a systematic review of the literature. J Med Internet Res 2007 Sep 30;9(3):e26 [FREE Full text] [doi: 10.2196/jmir.9.3.e26] [Medline: 17942388]
63. Schepers P, Fishman E, Beelen R, Heinen E, Wijnen W, Parkin J. The mortality impact of bicycle paths and lanes related to physical activity, air pollution exposure and road safety. J Transp Health 2015 Dec;2(4):460-473. [doi: 10.1016/j.jth.2015.09.004]
64. Teschke K, Reynolds CC, Ries FJ, Gouge B, Winters M. Bicycling: health risk or benefit? UBC Med J 2012;3(2):6-11 [FREE Full text]
65. Johan de Hartog J, Boogaard H, Nijland H, Hoek G. Do the health benefits of cycling outweigh the risks? Environ Health Perspect 2010 Aug;118(8):1109-1116 [FREE Full text] [doi: 10.1289/ehp.0901747] [Medline: 20587380]
66. Rojas-Rueda D, de Nazelle A, Tainio M, Nieuwenhuijsen MJ. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. BMJ 2011 Aug 04;343:d4521 [FREE Full text] [doi: 10.1136/bmj.d4521] [Medline: 21816732]
67. Zhang C, Soliman-Hamad M, Robijns R, Verberkmoes N, Verstappen F, IJsselsteijn WA. Promoting physical activity with self-tracking and mobile-based coaching for cardiac surgery patients during the discharge-rehabilitation gap: protocol for a randomized controlled trial. JMIR Res Protoc 2020 Aug 19;9(8):e16737 [FREE Full text] [doi: 10.2196/16737] [Medline: 32812886]
68. Krykewycz GR, Pollard C, Canzoneri N, He E. Web-based "crowdsourcing" approach to improve areawide "bikeability" scoring. Transp Res Record 2011 Jan 01;2245(1):1-7. [doi: 10.3141/2245-01]
69. Hood J, Sall E, Charlton B. A GPS-based bicycle route choice model for San Francisco, California. Transp Lett 2011;3(1):63-75. [doi: 10.3328/tl.2011.03.01.63-75]
70. Casello JM, Usyukov V. Modeling cyclists' route choice based on GPS data. Transp Res Record 2014 Jan 01;2430(1):155-161. [doi: 10.3141/2430-16]
71. Broach J, Dill J, Gliebe J. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp Res Part A Policy Pract 2012 Dec;46(10):1730-1740. [doi: 10.1016/j.tra.2012.07.005]
72. Bricka S, Zmud J, Wolf J, Freedman J. Household travel surveys with GPS: an experiment. Transp Res Record 2009 Jan 01;2105(1):51-56. [doi: 10.3141/2105-07]
73. Elwood S, Goodchild MF, Sui DZ. Researching volunteered geographic information: spatial data, geographic research, and new social practice. Ann Assoc Am Geograph 2012 May;102(3):571-590. [doi: 10.1080/00045608.2011.595657]
74. Dorn D, Gorzelitz J, Gangnon R, Bell D, Koltyn K, Cadmus-Bertram L. Automatic identification of physical activity type and duration by wearable activity trackers: a validation study. JMIR Mhealth Uhealth 2019 May 23;7(5):e13547 [FREE Full text] [doi: 10.2196/13547] [Medline: 31124470]
75. Muggeridge DJ, Hickson K, Davies AV, Giggins OM, Megson IL, Gorely T, et al. Measurement of heart rate using the polar OH1 and Fitbit charge 3 wearable devices in healthy adults during light, moderate, vigorous, and sprint-based exercise: validation study. JMIR Mhealth Uhealth 2021 Mar 25;9(3):e25313 [FREE Full text] [doi: 10.2196/25313] [Medline: 33764310]
76. Lindqvist A, Rutberg S, Söderström E, Ek A, Alexandrou C, Maddison R, et al. User perception of a smartphone app to promote physical activity through active transportation: inductive qualitative content analysis within the smart city active mobile phone intervention (SCAMPI) study. JMIR Mhealth Uhealth 2020 Aug 05;8(8):e19380 [FREE Full text] [doi: 10.2196/19380] [Medline: 32755889]

77. Suminski Jr RR, Dominick G, Saponaro P. Assessing physical activities occurring on sidewalks and streets: protocol for a cross-sectional study. JMIR Res Protoc 2019 Jul 30;8(7):e12976 [FREE Full text] [doi: 10.2196/12976] [Medline: 31364605]
78. Bachand-Marleau J, Lee BH, El-Geneidy AM. Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. Transp Res Record 2012 Jan 01;2314(1):66-71. [doi: 10.3141/2314-09]
79. Goodman A, Green J, Woodcock J. The role of bicycle sharing systems in normalising the image of cycling: an observational study of London cyclists. J Transp Health 2014 Mar;1(1):5-8 [FREE Full text] [doi: 10.1016/j.jth.2013.07.001] [Medline: 25568838]
80. DeMaio P. Bike-sharing: history, impacts, models of provision, and future. J Public Transp 2009 Dec;12(4):41-56. [doi: 10.5038/2375-0901.12.4.3]
81. Evans-Cowley JS, Griffin G. Microparticipation with social media for community engagement in transportation planning. Transp Res Record 2012 Jan 01;2307(1):90-98. [doi: 10.3141/2307-10]
82. Broach J, Gliebe J, Dill J. Bicycle route choice model developed using revealed preference GPS data. In: Proceedings of the 90th Annual Meeting of the Transportation Research Board. 2011 Presented at: TRB '11; January 23-27, 2011; Washington, DC, USA.
83. Aultman-Hall L, Hall FL, Baetz BB. Analysis of bicycle commuter routes using geographic information systems: implications for bicycle planning. Transp Res Record 1997 Jan 01;1578(1):102-110. [doi: 10.3141/1578-13]
84. Krizek KJ, El-Geneidy A, Thompson K. A detailed analysis of how an urban trail system affects cyclists' travel. Transportation 2007 Jul 18;34(5):611-624. [doi: 10.1007/s11116-007-9130-z]
85. Wałdykowski P, Adamczyk J, Dorotkiewicz M. Sustainable urban transport—why a fast investment in a complete cycling network is most profitable for a city. Sustainability 2021 Dec 23;14(1):119. [doi: 10.3390/su14010119]
86. Mahfouz H, Arcaute E, Lovelace R. A road segment prioritization approach for cycling infrastructure. arXiv 2021 May 8.
87. Liu B, Bade D, Chow JY. Bike count forecast model with multimodal network connectivity measures. Transp Res Record 2021 Jul 16;2675(7):320-334. [doi: 10.1177/03611981211021849]
88. Edwards RD, Mason CN. Spinning the wheels and rolling the dice: life-cycle risks and benefits of bicycle commuting in the U.S. Prev Med 2014 Jul;64:8-13. [doi: 10.1016/j.ypmed.2014.03.015] [Medline: 24657549]
89. Collins AJ, Jordan CA, Robinson RM, Cornelius C, Gore R. Exploring good cycling cities using multivariate statistics. Environ Syst Decis 2020;40(4):526-543. [doi: 10.1007/s10669-019-09753-z]
90. Harkey DL, Reinfurt DW, Knuiman M. Development of the bicycle compatibility index. Transp Res Record 1998 Jan 01;1636(1):13-20. [doi: 10.3141/1636-03]
91. Hyde-Wright A, Graham B, Nordnack K. Counting bicyclists with pneumatic tube counters on shared roadways. ITE J 2014 Feb;84(2):32-37.
92. Griffin G. The National Academies of Sciences, Engineering, and Medicine. 2011 Sep 6. URL: https://trid.trb.org/view/1141140 [accessed 2022-02-03]
93. Figliozzi M, Johnson P, Monsere C, Nordback K. Methodology to characterize ideal short-term counting conditions and improve AADT estimation accuracy using a regression-based correcting function. J Transp Eng 2014 May 1;140(5):404-414. [doi: 10.1061/(asce)te.1943-5436.0000663]
94. Nordback K, Marshall WE, Janson BN. Development of estimation methodology for bicycle and pedestrian volumes based on existing counts. Colorado Department of Transportation. 2013 Oct. URL: https://rosap.ntl.bts.gov/view/dot/26644 [accessed 2022-02-03]
95. Berntsen S, Malnes L, Langåker A, Bere E. Physical activity when riding an electric assisted bicycle. Int J Behav Nutr Phys Act 2017 Apr 26;14(1):55 [FREE Full text] [doi: 10.1186/s12966-017-0513-z] [Medline: 28446180]
96. Louis J, Brisswalter J, Morio C, Barla C, Temprado JJ. The electrically assisted bicycle: an alternative way to promote physical activity. Am J Phys Med Rehabil 2012 Nov;91(11):931-940. [doi: 10.1097/PHM.0b013e318269d9bb] [Medline: 23085705]
97. MacArthur J, Dill J, Person M. Electric bikes in North America: results of an online survey. Transp Res Record 2014 Jan 01;2468(1):123-130. [doi: 10.3141/2468-14]
98. Haustein S, Møller M. Age and attitude: changes in cycling patterns of different e-bike user segments. Int J Sustain Transp 2016 Mar 23;10(9):836-846. [doi: 10.1080/15568318.2016.1162881]
99. Gloekler S, Wenaweser P, Lanz J, Stoller M. How e-biking can boost cardiovascular health. Eur Heart J 2015 Aug 14;36(31):2033. [doi: 10.1093/eurheartj/ehv154] [Medline: 25935878]
100. Peterman JE, Morris KL, Kram R, Byrnes WC. Pedelecs as a physically active transportation mode. Eur J Appl Physiol 2016 Aug;116(8):1565-1573. [doi: 10.1007/s00421-016-3408-9] [Medline: 27299435]
101. Sperlich B, Zinner C, Hébert-Losier K, Born DP, Holmberg HC. Biomechanical, cardiorespiratory, metabolic and perceived responses to electrically assisted cycling. Eur J Appl Physiol 2012 Dec;112(12):4015-4025. [doi: 10.1007/s00421-012-2382-0] [Medline: 22446956]
102. Brown BB, Smith KR. Complex active travel bout motivations: gender, place, and social context associations. J Transp Health 2017 Sep;6:335-346 [FREE Full text] [doi: 10.1016/j.jth.2017.01.014] [Medline: 29104857]

103. Pinjari AR, Eluru N, Bhat CR, Pendyala RM, Spissu E. Joint model of choice of residential neighborhood and bicycle ownership: accounting for self-selection and unobserved heterogeneity. Transp Res Record 2008 Jan 01;2082(1):17-26. [doi: 10.3141/2082-03]
104. Griffin G, Nordback K, Götschi T, Stolz E, Kothuri S. Monitoring bicyclist and pedestrian travel and behavior: current research and practice. Transportation Research Circular(E-C183). 2014 Mar. URL: https://onlinepubs.trb.org/onlinepubs/circulars/ec183.pdf [accessed 2022-02-03]
105. Spencer CD. Two types of role playing: threats to internal and external validity. Am Psychol 1978 Mar;33(3):265-268. [doi: 10.1037/0003-066x.33.3.265]
106. Yu CH. Threats to validity of research design. Creative Wisdom!. 2010. URL: http://www.creative-wisdom.com/teaching/WBI/threat.shtml [accessed 2022-02-03]
107. Modeling Bike and Pedestrian Paths at the Census Tract Level In Norfolk, VA. Modeling The Health Effects Of Adding Bicycle Paths At The Census Tract-Level. 2022. URL: https://rgore-vmasc.shinyapps.io/norfolk-bike-ped/ [accessed 2022-02-03]
108. San Francisco Census Tract Stroke and Diabetes Rate Improvement Estimation Based On Adding Bicycle and Pedestrian Paths. Modeling The Health Effects Of Adding Bicycle Paths At The Census Tract-Level. 2022. URL: https://rgore-vmasc.shinyapps.io/sf-bike-ped/ [accessed 2022-02-03]
109. Gore R. Modeling The Health Effects of Adding Bicycle and Pedestrian Paths At The Census Tract-Level Supplementary Information. Mendeley Data. 2022. URL: https://www.narcis.nl/dataset/RecordID/oai%3Aeasy.dans.knaw.nl%3Aeasy-dataset%3A242304 [accessed 2022-07-21]

Abbreviations

ACS: American Communities Survey
BPH: bicycling and pedestrian habits
CDC: Centers for Disease Control and Prevention
CFA: confirmatory factor analysis
DBC: demographics and background characteristics
EFA: exploratory factor analysis
MAE: mean absolute error
MOE: measure of effectiveness
RMSE: root mean squared error

Edited by A Mavragani, Y Khader; submitted 11.04.22; peer-reviewed by J Delgado-Ron, C Jeong; comments to author 14.05.22; revised version received 13.07.22; accepted 29.07.22; published 24.08.22

Please cite as:
Gore R, Lynch CJ, Jordan CA, Collins A, Robinson RM, Fuller G, Ames P, Keerthi P, Kandukuri Y
Estimating the Health Effects of Adding Bicycle and Pedestrian Paths at the Census Tract Level: Multiple Model Comparison
JMIR Public Health Surveill 2022;8(8):e37379
URL: https://publichealth.jmir.org/2022/8/e37379
doi: 10.2196/37379
PMID: