Modeling The Health Effects Of Adding Bicycle Paths At The Census Tract-Level

Digital Neighborhoods Team

Hampton Roads Biomedical Research Consortium (HRBRC), Old Dominion University, Norfolk, Virginia, USA

* rgore@odu.edu

Abstract

Studies have shown that numerous factors (i.e. education, family structure, income, etc.) influence the extent to which adding a bicycle path improves the health of residents in a given geographic area. The complexity of the interactions among these factors prevents policy makers and practitioners (i.e. city planners) from a simple answer to the question, "How much will the proposed bicycle path improve the health of the residents in the census tract where it will be constructed?" We propose a methodology that uses factor analysis to assess data from the American Community Survey (ACS), CDC 500 Cities Project along with bike/pedestrian path location and usage data to construct a city-specific model with insight at the census tract-level. In an evaluation using data from 2010-2012, we show that given a specified amount of additional bicycle path mileage and a specific census tract, on average our model forecasts the 2016: (1) % of individuals with high blood pressure in the census tract within 1.85%, (2) % of individuals with diabetes in the census tract within 1.21% and (3) % of individuals who suffer more than two weeks worth of poor physical health days in the census tract within 1.26%. Furthermore, we show that these results are more accurate and that additional accuracy is statistically significant difference than two straight forward alternative approaches to estimating health improvements in census tracts in Norfolk, VA.

Introduction

The addition of bicycle paths to an area is a valuable resource for policy makers, public health professionals and urban planners to promote physical activity and improve health outcomes. The majority of existing research finds a negative association between the prevalence of bicycle paths and negative health outcomes (i.e. diabetes, stroke, obesity, heart disease, high blood pressure, poor physical/mental health, etc.) [1–9]. However, other studies show that bicycling can be a health hazard due to the increased risk of consuming air pollution [10, 11], being struck by a motor vehicle [7–15], or falling off the bicycle [16, 17]. Our overall aim in this paper is to investigate at a granular geographic level what health outcomes are affected by variables related to bicycling and bicycle paths. We ask: *"How much will a proposed bicycle path improve the health of the residents in the census tract (small contiguous county subdivision) where it will be constructed?"* Our methodology uses factor analysis to filter out and aggregate variables from datasets at the census tracts in Norfolk, VA for the year 2016. The data is taken from the American Community Survey (ACS) [18], and CDC 500 Cities Project [19] along with bike/pedestrian path location and usage data provided by the city of Norfolk.

Related Research

Our work builds on a significant amount of previous research. Numerous researchers have employed statistical analyses: (1) to explore the health effects of commuting via bicycle and (2) to assess the benefits of bicycle paths for communities. This work has captured data related to bicycling using an array of instruments including bicycle related longitudinal data [13, 14], telephone and web-based surveys [15, 20], weather data [21], GPS and accelerometer data [22], and information related to bicycle infrastructure [23]. The data has been rigorously analyzed to explore additional variables related to bicycling and health effects [24, 25]. However, none of these analyses construct a city-specific model of the health effects of added bicycle paths at the census tract-level intended to inform policy makers and city planners.

Additional efforts to explore the health and safety impacts of planned bicycle routes are related to our work. Specifically, work related to: (1) the development of a Bicycle Compatibility Index to determine how compatible a roadway is for allowing efficient operation of both bicycles and motor vehicles [26] and (2) existing assessments of the importance in quantifying health impacts during transportation planning [8, 27, 28]. These studies have had significant impact and demonstrate the need for granular analysis. However, they are more focused on the risk of catastrophic injury caused by a bicycling accident as opposed to improving health outcomes.

Finally, our work is influenced by previous models that enable insight into the subtleties of the effects of bicycling and bicycle paths. These efforts include a joint model of choice of residential neighborhood and bicycle ownership [29], an analysis of the environmental influences on route selection for bike and car travel [30, 31], a conceptual model to explain turning points in bicycle travel behavior [32] and improved understanding of the factors that create successful shared bicycle systems [33, 34]. In combinations, these efforts provide a platform for our work. In the next section, we describe the data sets, analysis and modeling methods in our study. Then, we evaluate the accuracy of our model by quantifying the extent to which it can predict improvements in health outcomes of individuals in a given census block when bicycle path mileage is added.

Data & Methodology

Datasets

Here, we examine relationships among demographics, bicycle paths, bicycling habits and health outcomes of 51 census tracts in Norfolk, VA during 2016. The data includes: (1) census tract boundaries used in the 2010 U. S. Census [35], (2) demographic variables surveyed by the American Communities Survey (ACS) during 2016 [18], (3) 2016 census tract-level estimates for chronic disease risk factors, health outcomes and clinical preventive service from the Center for Disease Control and Prevention (CDC) [19], and (4) 2016 bicycle path location and usage data from Norfolk, VA.

Combined these data sets result in more than 400 variables for each of the 51 census tracts. Census tracts are small, contiguous, relatively permanent statistical subdivisions of a county or equivalent entity. Nationally, the populations in census tracts vary from 1,200 to 8,000. Census tracts provide a stable geographic unit for statistical analysis of the U.S. Census and ACS [35]. The ACS is an ongoing national survey that samples a subset of individuals within the same geographic areas considered in the U.S. Census. Using the same questions, data is collected each month throughout the course of a year. In contrast, the U.S. census provides a more comprehensive sample of individuals in the U.S., collecting data from more individuals during a particular time period (March to August), but administered only once every ten years. A metaphor helps elucidate the difference between the two surveys. The U.S. census serves as a high-resolution photograph of the U.S. population once every 10-years while the ACS serves as many low-resolution continually updating videos over the same period of time [18].

The census tract-level estimates and methodology for estimating chronic disease risk factors, health outcomes and clinical preventive service are provided by the CDC′s 500 cities project. The 500 Cities project is a collaboration among the CDC and the Robert Wood Johnson Foundation. The small area estimates provided by the project allow policy makers 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 [19].

The bicycle path data provided by the city of Norfolk includes the latitude and longitude location of bike lanes, routes and paths built and maintained in Norfolk, VA through the end of 2016 [36]. The bicycle usage data is taken from the 25 automated bike counters permanently installed in the city and the results of manual counts at 80 different locations throughout the city. The automated bike counters monitor bicycling volume 24 hours a day; manual counts provide a subset of total users. Manual counts were taken between September 14 and 20 from 4:30 p.m. – 6:30 p.m. as part of the National Bicycle and Pedestrian Documentation Project.

Latitude and longitude coordinates identify each automated and manual-counting location [12]. While the geographic resolution of this data ranges from very high (latitude/longitude) to high (census-tract) we are still able to assemble a fine grained data set by mapping each latitude/longitude pair in the bicycle path and usage data set to its respective census tract. The result is our combined data set. In what follows we discuss the statistical analysis we perform on this data set to: (1) filter out unrelated variables, (2) aggregate similar related variables together and (3) identify the factors related to census tract-level bicycling habits and health and their interrelations in quantitative terms.

Factor Analysis

Our dataset includes a wide range of variables collected from multiple sources. From this dataset we selected a subset of the variables that individuals with domain expertise identified as possibly contributing to the use of bicycle paths and the impact that bicycle paths have on health outcomes when additional mileage is added to a geographic area (i.e. census tract). These include the following variables for each census tract in Norfolk City.

Table 1. Included variables for each census tract in Norfolk, VA.

Variable	Description	Originating Data Source
Strava Trip Rate	Description	Strava
% of Individuals Who Report More Than 2 Weeks of Poor Mental Health	Description	500 Cities
% of Individuals with Diabetes	Description	500 Cities
% of Individuals With High Blood Pressure	Description	500 Cities
% of Individuals Who Are Obese	Description	500 Cities
% of Individuals Who Are White	Description	ACS
% of Individuals Who Are Black	Description	ACS
% of Individuals Who Are Hispanic	Description	ACS
% of Individuals With A High School Education	Description	ACS
% of Individuals With A College Education	Description	ACS
Median Income	Description	ACS
% of Individuals Who Work From Home	Description	ACS
% of Individuals Who Own 2+ Vehicles	Description	ACS
% of Individuals Whose Parents Are Married	Description	ACS
% of Individuals Whose Mother Is Not Married	Description	ACS
% of Individuals Who Bike To Work	Description	ACS
# of Bike Path Miles	Description	City of Norfolk, VA
# of Individuals Using The Bike Paths	Description	City of Norfolk, VA

Next, we apply factor analysis to filter this list of variables and aggregate them into a conceptual model composed of a much lower number of meaningful unobserved variables (factors). Factor analysis is a multiple step process. First, half (n = 26 census tracts) of the data (selected randomly) is used in an exploratory factor analysis (EFA). The purpose of the exploratory analysis is to hypothesize the structure of a conceptual model. Then, a confirmatory factor analysis (CFA) is performed using the other half (n = 25 census tracts) of the data to confirm (or reject) the hypothesized model structure [37].

Exploratory factor analysis is an iterative process with three stages. First variables with low communality and/or weak cross-loadings are discarded. Next, the optimal number of factors is estimated based on the correlation of the remaining variables. Finally, the variables are organized into the estimated number of factors. Since the public health intervention we are evaluating is bicycle and pedestrian paths we required at least one of the factors to be composed entirely of variables directly related to bicycling habits and/or bicycle paths. Furthermore, if any of the factors have low communality and/or weak cross loadings the variables in the factor causing the issue are discarded and the process is repeated. Commonality is a measure that reflects the percent of variance in a given variable that can be explained by an aggregated factor.

Factors made up of variables with high communality accurately reflect each of the variables included in the factor. Cross loadings reflect the communality of a given variable with respect to a different factor. A variable should only have high communality with the factor it is included in. A variable with *strong* cross loadings has: (1) high communality with the factor it is included in and (2) low communality with all other factors. If this is not true, the cross loadings for the variable are considered weak because it means that the variable is not unique to a single factor [37].

Specifically, we required each variable to have: (1) a communality > 0.40 and (2) no cross loading with any other factor > 0.32. These requirements are consistent with established factor analysis guidelines [37]. The results of iteratively filtering and aggregating the variables previously listed through the EFA yield the three-factor conceptual model shown in Figure 1(a). It is important to note the extent to which the requirements associated with the EFA filter and organize variables. Of the 18 variables previously listed, only 11 are included in Figure 1 and they are organized into 3 factors.

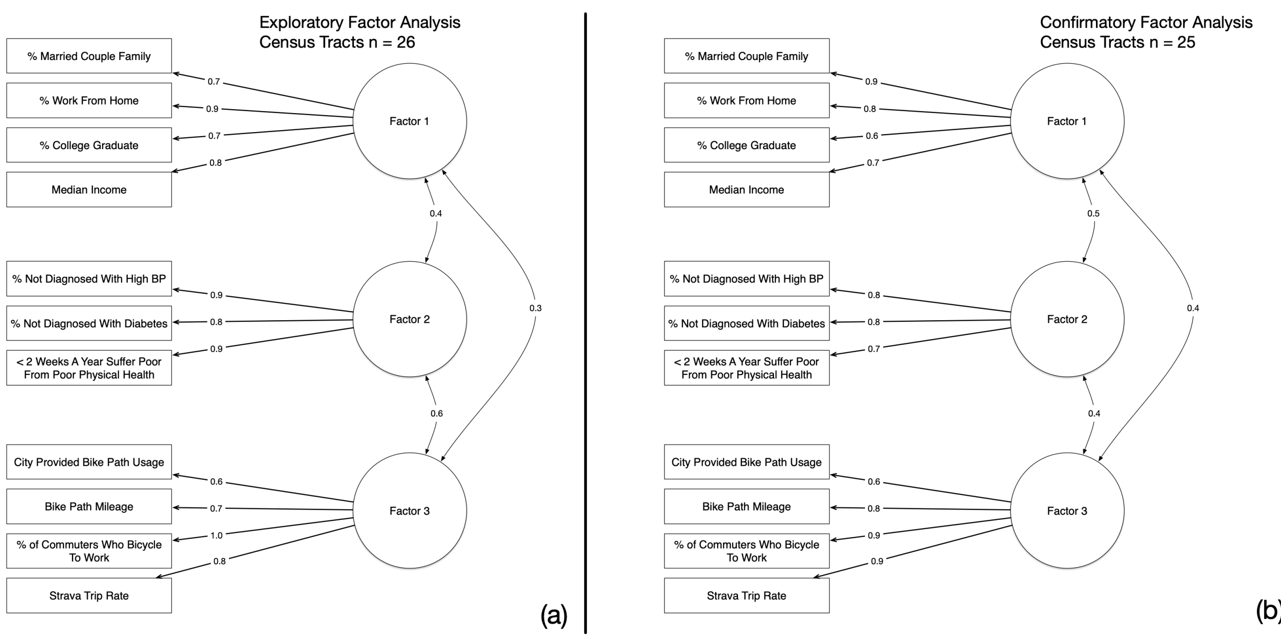

Fig 1. Exploratory and Confirmatory Factor Analysis.

The CFA was completed using the variables and number of factors identified in the EFA. However, because the goal of this analysis is to confirm or reject the hypothesized conceptual model, no cross loadings were permitted. In other words variables were forced to load on a single factor. We used a multi-faceted criteria to confirm the hypothesized model following established guidelines: (1) variable loadings ≥ 0.50; (2) average variance extracted ≥ 0.50; (3) reliability of factors ≥ 0.70. Factor reliability is a measure of how well the variables that make up a factor correlate with one another [37]. The variable loadings for the CFA are shown in Figure 1(b). The reliability values of the factors (1-3) are: 0.892, 0.903 and 0.774. The average variance explained values of the factors (1-3) are: 0.734, 0.781 and 0.703. These all exceed threshold the minimum threshold (0.700) established for CFA [37]. Next, we describe how this conceptual model enables us to design an algorithm to estimate the health effects of adding bicycle paths at the census tract-level.

Model & Estimation

Discussion of Factors

Before proceeding to explaining our estimation algorithm it is useful to name the factors and discuss the variables include within them. We name Factor 1 - *Socio-Economic Characteristics*, Factor 2 - *Health Outcomes* and Factor 3 – *Cycling Enthusiasm.*

While these names are subjective we feel they are straightforward given the variables that make up each factor. Within the *Socio-Economic Characteristics* factor the variables *% College Graduate*, *Median Income*, *% Work From Home* and *% Married Couple Family* are measures derived from questions in the ACS collected during 2016. While the inclusions of these variables in this factor makes intuitive sense, it was surprising that other variables from the ACS related to demographics often correlated with college and married couple families (race, number of vehicles owned, etc) were excluded by the analysis.

In the *Health Outcomes* factor the exclusion of chronic diseases that are similar to diabetes and high blood pressure (obesity, stroke and coronary heart disease) is unexpected, but not unprecedented. Previous research has shown that despite burning more calories per week there is not a statistically significant difference in the BMI of cyclists and non-cyclists [38]. Similarly, other researchers found that bicycling is a more effective at reducing the insulin resistance (a precursor of diabetes) of individuals than reducing the body fat percentage of individuals [38] and that it is effective at reducing insulin resistance even in non-obese individuals [39]. This insight alone is valuable. It demonstrates which health outcomes in Norfolk City policy makers and city planners can expect to be improved by the addition of bicycle paths. Finally, the variables included in the *Cycling Enthusiasm* factor are expected. These are the only four variables in our data set related to bicycling habits.

All three factors in our conceptual model co-vary with one another. This means that given an increase in a variable in one of the factors, one would predict an increase in the other two factors. However, a policy maker or city planner can directly control only one of the variables in one of the factors. This variable is *Bike Path Mileage* in the *Cycling Enthusiasm* factor. As a result, this variable is the only input in our model. The other eight variables among the three factors serve as internal components used in its analysis.

Estimation Algorithm (EA)

Our algorithm to estimate the effects of adding bicycle paths at the census-tract level proceeds as follows:

1. The policy maker, public health professional or urban planner specifies the census tract and the number of bike path miles s/he is considering.

2. The algorithm adds the bike path mileage to the appropriate census tract in the data set.

3. The algorithm computes factor scores for each census tract in the data. Factor scores are continuous numbers, which reflect the extent to which each census tracts manifests each factor. For each factor, scores are distributed normally with a mean of 0 and a standard deviation of 1. Thus, 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 [40].

4. The algorithm identifies all census tracts with a: (1) *Socio-Economic Characteristics* factor score and (2) *Cycling Enthusiasm* factor score within x of the respective factor for the census tract being analyzed. This list of census tracts reflects those census tracts that are similar to the census tract being analyzed in two dimensions: (1) *Cycling Enthusiasm* (based on the added bicycle paths) and (2) its existing *Socio-Economic Characteristics.* Recall, factor scores are normally distributed with a standard deviation of one. Thus, a census tract within a factor score of x of the tract being analyzed reflects a census tract with a factor score within x standard deviations.

5. The algorithm computes the difference between the existing diabetes rate for the census tract being analyzed and the diabetes rate for the tracts identified in Step 4. These diabetes rate

improvements are kept in two lists according to the factor with which they are associated. These two lists are named: (1) Cycling Enthusiasm Diabetes Improvement List and (2) Socio-Economic Characteristics Diabetes Improvement List.

6. The algorithm computes the difference between the existing high blood pressure rate for the census tract being analyzed and the high blood pressure rate for the tracts identified in Step 4. These high blood pressure rate improvements are kept in two lists according to the factor with which they are associated. These two lists are named: (1) Cycling Enthusiasm High Blood Pressure Improvement List and (2) Socio-Economic Characteristics High Blood Pressure Improvement List.

7. The algorithm computes the difference between the existing poor physical health rate for the census tract being analyzed and the poor physical health rate for the tracts identified in Step 4. These high blood pressure rate improvements are kept in two lists according to the factor with which they are associated. These two lists are named: (1) Cycling Enthusiasm High Poor Physical Health Improvement List and (2) Socio-Economic Characteristics Cycling Enthusiasm High Poor Physical Health Improvement List.

8. The algorithm removes each element of the diabetes improvement lists, high blood pressure improvement lists and poor physical health improvements lists that is < 0. This reflects the assumption that adding bike paths to a census tract will not be detrimental to the health of residents in the census tract with respect to diabetes and stroke outcomes.

9. The algorithm reports the average of the median values of the diabetes improvement lists, the high blood pressure improvement lists and poor physical health improvements lists to the policy maker or city planner. These values reflect expected improvement in each health outcome of adding the bicycle path mileage to the census tract. The algorithm can also report the minimum and maximum values of each improvement list.

In the next section, we fit the estimation algorithm parameter x (Step 4) and then evaluate the algorithm's accuracy against alternative approaches.

Evaluation

The accuracy of our estimation algorithm is elucidated through empirical evaluation against alternative approaches. Specifically, for a given census tract that added a specified amount of bike path miles in a given year, we evaluate how accurately our algorithm predicts, in terms of percentages of people in 2016, that the CDC 500 Cities Project estimates: (1) have high blood pressure, (2) are diagnosed with diabetes and (3) report more than two weeks of poor physical health days.

Evaluation Test Suite

To evaluate our algorithm, we created a new data set that provides the same variables as the first data set, but using the data for each census tract in Norfolk City from ACS, CDC 500 Cities Project and Virginia Dept. of Transportation (VDOT) for 2010, 2011 and 2012. For each year, the miles of bike paths that were added to the different census tracts in Norfolk City were assessed: (2010) 21.27 miles of bicycle paths added in 29 different census tracts, (2011) 24.05 miles bicycle paths added in 22 different census tracts and (2012) 32.36 bicycle paths added in 37 different census tracts. Combined these, ≈ 75 miles of bike path added to more than 40 census tracts, form the test cases that enable us to evaluate the accuracy our algorithm.

We choose to use data from this three-year period for two reasons. First, studies have shown that similar passive physical activity related interventions require several years to realize any health benefits within a large-scale population (i.e. thousands) [2], [41], [37]. Second this period reflects a nontrivial investment from policy makers in Norfolk, VA into the installation of bike paths. The almost 75 miles of bike paths added during this period account for ≈25% of the total bike path miles in the city installed in the last 20 years.

Given this test suite of bike paths additions, we evaluate two versions of the estimation algorithm. The first version (EA) reflects exactly the algorithm described in the Methodology Section. The second version (EA′) reflects, running EA in a given census tract for every tenth-mile increment of bike path addition up to and including the specified amount of mileage, collecting the results for each run and then reporting the largest median improvement. By operating in this manner, EA′ implements the assumption that adding more bike path mileage (i.e. 1.0 miles as opposed to 0.5 miles) to a given census tract will not be detrimental to the expected improvement in the diabetes, high blood pressure and poor physical health rate.

Evaluation of Algorithms EA and EA′

For each census tract that added bike path mileage in 2010, 2011 and 2012 we run EA and EA′ to estimate the improvement in the high blood pressure, poor physical health rate and diabetes rate of residents in the census tract realized in 2016. The improvement estimate reflects the median value (i.e. Step 9) reported by the algorithm. The distribution of the algorithms' error for each health outcome and the summary statistics for the distribution are shown in Figure 2 and Table 2 respectively. The versions of EA and EA′ shown in the figure and table use the value of x that resulted in the smallest mean absolute error from the evaluation data. Recall, this parameter reflects the similarity threshold for the factor scores of a different census tract to be included in the analysis.

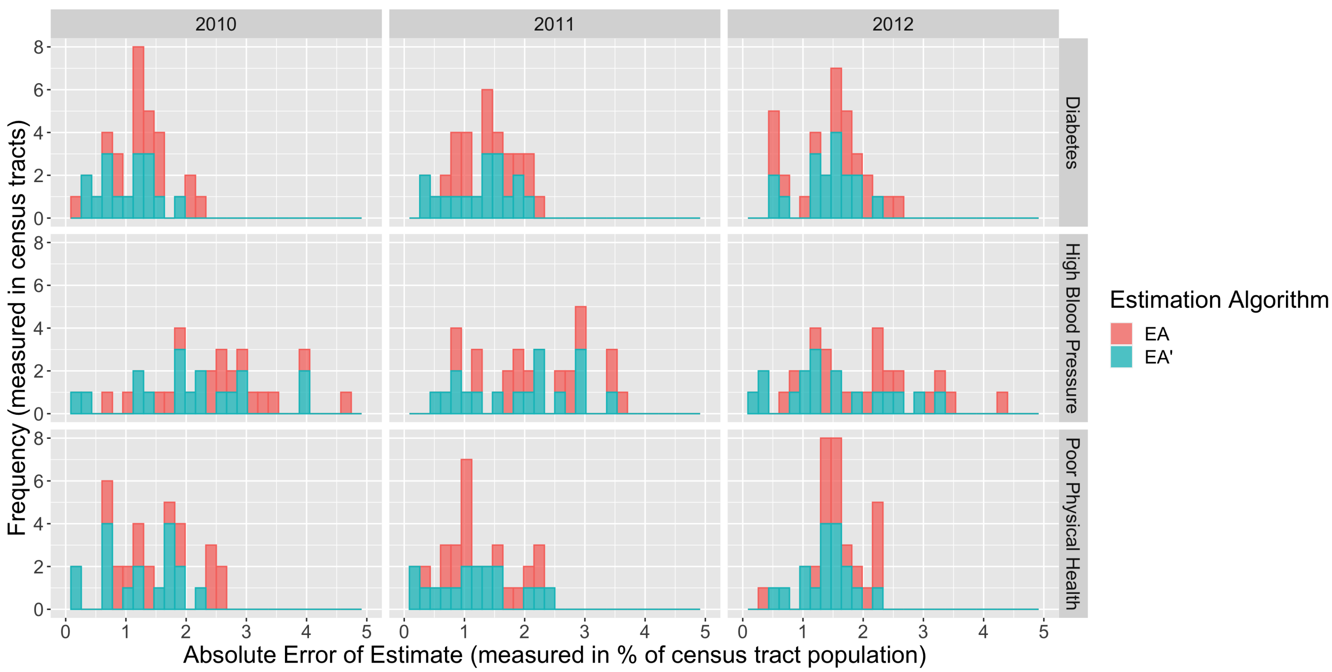

Fig 2. Distribution of Absolute Error for Improvement Predictions from Best Algorithm Parameterizations for Health Outcomes by Year .

Figure 2 and Table 2 elucidate several trends in the accuracy of the EA and EA′ algorithm. The most noticeable trend is that the algorithms predict the diabetes and poor physical health rates more accurately than they predict the high blood pressure rate independent of the year. This result is expected. The CDC 500 cities data shows that there is two times more variation in the rate of high blood pressure in Norfolk, VA than the diabetes and poor physical health rates [19]. As a result, the high blood pressure rate is more difficult to predict resulting in less accuracy from EA and EA′.

Table 2. Summary Statistics of Absolute Error for Best Algorithm Parameterizations

Algorithm	Health Outcome	x-value	Mean Absolute Error (MAE)	Standard Deviation
EA	Diabetes	0.50	1.40	0.55
EA′	**Diabetes**	**0.25**	**1.21**	**0.52**
EA	High Blood Pressure	0.50	2.27	1.00
EA′	**High Blood Pressure**	**0.25**	**1.85**	**0.98**
EA	Poor Physical Health	0.50	1.48	0.64
EA′	**Poor Physical Health**	**0.25**	**1.26**	**0.58**

The results also show that the most accurate versions of the EA and EA′ algorithms use different parameterizations of x. Our revised algorithm, EA′, which assumes that adding more bike path mileage (i.e. 1.0 miles as opposed to 0.5 miles) to a given census tract will not be detrimental to the expected improvement in health outcomes is most accurate with a lower value of x (0.25 vs. 0.50) than our original algorithm, EA, which does not make this assumption. The difference in the parameterization reflects the need for a mechanism to smooth out the variation in the data used by the algorithm. Our revised algorithm, EA′, implements this smoothing by: (1) running the original EA algorithm in a given census tract for every 0.10 mile increment of bike path addition up to and including the specified amount of mileage, (2) collecting the results for each run and (3) then reporting the largest median improvement. By aggregating all of the predictions from each run of the EA algorithm, EA′ produces an output that is less sensitive to variations in the data used by the algorithm. As a result, it is most accurate when using a narrow similarity threshold reflected in a small value of x ($x = 0.25$). However, our original EA does not implement this assumption. As a result it is most accurate when an additional mechanism is used to smooth out the variation in the data. This smoothing is achieved in EA by a wider similarity threshold reflected in a larger value of x (0.50).

The final trend apparent in the results data is that the accuracy of each of the algorithms is independent of the year of the test suite data. This provides evidence that improvement in the high blood pressure, poor physical health and diabetes rates is visible after as few as four years and remains visible for at least six years after bicycle paths have been added to a census tract. These findings are consistent with studies related to interventions and the time period required to realize and maintain health benefits within a large-scale population [2, 37, 41].

Evaluation Against Alternative Approaches

The effectiveness of our estimation algorithm is elucidated through empirical evaluation against alternative approaches. Specifically, we compare our approach to: (1) a baseline approach (*Baseline*) based entirely on historical data and (2) a statistical approach that uses linear regression modeling (*LR*) [42]. For a given year, a given census tract and a certain number of bike path miles added, each approach predicts the improvement in the diabetes, high blood pressure and poor physical health rates in 2016.

The baseline approach assumes there will be no improvement from the current year′s data. This approach mirrors predicting that the weather tomorrow will be the same as the weather today. The statistical approach uses regression to predict future changes in each health outcome using a weighted linear combination of the *Socio-Economic Characteristics* factor and the increase the in *Cycling Enthusiasm* score of the census tract. We fit the LR model using 50% of the data selected at random from the test suite. Then we evaluate it on the remaining 50% of the data. The other approaches (Baseline, EA and EA′) are evaluated using all the data.

Discussion

Table 2 shows that both of our approaches are more effective than the alternatives. We expected EA and EA′ to outperform the baseline approach. The CDC 500 Cities and Norfolk, VA data shows that the

Table 3. Summary Statistics of Absolute Error for Best Algorithm Parameterizations

Algorithm	Health Outcome	x-value	Mean Absolute Error (MAE) All Years	Standard Deviation All Years
EA	Diabetes	0.50	1.40	0.55
EA′	**Diabetes**	**0.25**	**1.21**	**0.52**
LR	Diabetes	NA	2.13	0.81
Baseline	Diabetes	NA	2.68	0.92
EA	High Blood Pressure	0.50	2.27	1.00
EA′	**High Blood Pressure**	**0.25**	**1.85**	**0.98**
LR	High Blood Pressure	NA	3.40	1.37
Baseline	High Blood Pressure	NA	3.86	1.72
EA	Poor Physical Health	0.50	1.48	0.64
EA′	**Poor Physical Health**	**0.25**	**1.26**	**0.58**
LR	Poor Physical Health	NA	2.43	0.84
Baseline	Poor Physical Health	NA	2.96	1.19

majority of the time when a bike path of any length is added the high blood pressure, poor physical health rate and diabetes rates of residents in a census tract improve within a five years [18, 19, 36]. However, we did not know if our approach would outperform the LR approach. Examining the results of the evaluation shows that both versions of the EA algorithm outperform the LR algorithm due to the LR approach's assumption that critical thresholds within either factor do not exist. Due to this assumption, the LR approach can drastically over predict the improvement in the high blood pressure, poor physical health rate and diabetes rate within a census tract. This occurs because LR allows any increase in the *Cycling Enthusiasm* factor score to compensate for any reduced *Socio-Economic Characteristics* factor score, even when such a substitution is invalid [42, 43]. The result is inferior accuracy. Our EA and EA′ algorithms do not suffer from this pitfall. Both algorithms employ a similarity threshold (parameter x in Step 4) for each factor to identify similar census tracts. This threshold ensures that the algorithms do not over predict the improvement offered by additional bike path miles in circumstances where the *Socio-Economic Characteristics* of the residents in the tract indicate that the additional path miles will be an ineffective intervention.

Applicability to Practitioners and Policy Makers

The practical significance of our work is that the findings are directly actionable for policy makers, public health professionals and urban planners within Norfolk, VA by providing concrete insight into the question, "How much will the proposed bicycle path improve the health of the residents in the census tract where it will be constructed?" Specifically, it enables them to: (1) weigh the extent to which two bicycle paths of equal cost proposed in two different census tracts improve health outcomes of the residents, (2) identify areas where bicycle paths are unlikely to be effective public health interventions and other strategies should be employed 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.

Validity Threats

Despite these capabilities, internal and external validity threats affect our study. Threats to internal validity arise when factors affect the dependent variables without the evaluators' knowledge. It is possible that some flaws in the implementation of our model could have affected the results of the evaluation. However, our model used established libraries to conduct the factor analysis and the source code passed numerous internal reviews. Threats to external validity occur when the results of the evaluation cannot be generalized. Although the evaluation was performed for 3 years of data using 75 miles of added bike paths in over 40 census tracts, the results cannot be generalized to other cities or other time periods.

Conclusion

We proposed a methodology that considers demographics and habits of individuals using publicly available data at the census tract level for Norfolk, VA. Our methodology used factor analysis to assess data from the American Community Survey (ACS), and CDC 500 Cities Project along with bike/pedestrian path location and usage data to construct a city-specific model with insight at the census tract-level. In an evaluation using data from 2010-2012, we showed that given a specified amount of additional bicycle path mileage and a specific census tract, on average our model forecasts the 2016: (1) % of individuals with high blood pressure in the census tract within 1.85%, (2) % of individuals with diabetes in the census tract within 1.21% and (3) % of individuals who suffer more than two weeks worth of poor physical health days in the census tract within 1.26%. Furthermore, we showed that these results are superior to two straightforward approaches for estimating health improvements.

References

1. Ding D, Gebel K. Built environment, physical activity, and obesity: what have we learned from reviewing the literature? Health & place. 2012;18(1):100–105.

2. Pucher J, Dijkstra L. Promoting safe walking and cycling to improve public health: lessons from the Netherlands and Germany. American journal of public health. 2003;93(9):1509–1516.

3. Oja P, Titze S, Bauman A, De Geus B, Krenn P, Reger-Nash B, et al. Health benefits of cycling: a systematic review. Scandinavian journal of medicine & science in sports. 2011;21(4):496–509.

4. Moritz WE. Survey of North American bicycle commuters: design and aggregate results. Transportation Research Record. 1997;1578(1):91–101.

5. Khreis H, May AD, Nieuwenhuijsen MJ. Health impacts of urban transport policy measures: A guidance note for practice. Journal of Transport & Health. 2017;6:209–227.

6. Gao J, Helbich M, Dijst M, Kamphuis CB. Socioeconomic and demographic differences in walking and cycling in the Netherlands: How do these translate into differences in health benefits? Journal of Transport & Health. 2017;6:358–365.

7. Buekers J, Dons E, Elen B, Panis LI. Health impact model for modal shift from car use to cycling or walking in Flanders: application to two bicycle highways. Journal of Transport & Health. 2015;2(4):549–562.

8. Tsunoda K, Kitano N, Kai Y, Tsuji T, Soma Y, Jindo T, et al. Transportation mode usage and physical, mental and social functions in older Japanese adults. Journal of Transport & Health. 2015;2(1):44–49.

9. Smith O. Commute well-being differences by mode: Evidence from Portland, Oregon, USA. Journal of Transport & Health. 2017;4:246–254.

10. Bigazzi AY, Broach J, Dill J. Bicycle route preference and pollution inhalation dose: Comparing exposure and distance trade-offs. Journal of Transport & Health. 2016;3(1):107–113.

11. 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. Journal of Transport & Health. 2015;2(4):460–473.

12. Hollingworth MA, Harper AJ, Hamer M. Risk factors for cycling accident related injury: The UK Cycling for Health Survey. Journal of Transport & Health. 2015;2(2):189–194.

13. Hamann CJ, Peek-Asa C, Lynch CF, Ramirez M, Hanley P. Epidemiology and spatial examination of bicycle-motor vehicle crashes in Iowa, 2001–2011. Journal of Transport & Health. 2015;2(2):178–188.

14. Ohlin M, Berg HY, Lie A, Algurén B. Long-term problems influencing health-related quality of life after road traffic injury–Differences between bicyclists and car occupants. Journal of Transport & Health. 2017;4:180–190.

15. Bhatia D, Richmond SA, Loo CJ, Rothman L, Macarthur C, Howard A. Examining the impact of cycle lanes on cyclist-motor vehicle collisions in the city of Toronto. Journal of Transport & Health. 2016;3(4):523–528.

16. De Hartog JJ, Boogaard H, Nijland H, Hoek G. Do the health benefits of cycling outweigh the risks? Environmental health perspectives. 2010;118(8):1109–1116.

17. Noland RB, Quddus MA. Analysis of pedestrian and bicycle casualties with regional panel data. Transportation research record. 2004;1897(1):28–33.

18. Teschke K, Reynolds CC, Ries FJ, Gouge B, Winters M. Bicycling: Health risk or benefit. UBC Med J. 2012;3(2):6–11.

19. 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;343.

20. Methorst R, Schepers P, Christie N, Dijst M, Risser R, Sauter D, et al. 'Pedestrian falls' as necessary addition to the current definition of traffic crashes for improved public health policies. Journal of Transport & Health. 2017;6:10–12.

21. Hu W, Dong Q, Huang B. Correlations between road crash mortality rate and distance to trauma centers. Journal of Transport & Health. 2017;6:50–59.

22. Council NR, et al. Using the American Community Survey: benefits and challenges. National Academies Press; 2007.

23. Scally CP, Pettit KL, Arena O. 500 Cities Project. 2017;.

24. Area NCU. Bicycle/Pedestrian Plan;.

25. City N. Norfolk City bicycle plan. 2011;.

26. VDOT P, Committee BA. Virginia Dept. of Transportation;.

27. Heinen E, Maat K, Van Wee B. Day-to-day choice to commute or not by bicycle. Transportation Research Record. 2011;2230(1):9–18.

28. Crane M, Rissel C, Standen C, Ellison A, Ellison R, Wen LM, et al. Longitudinal evaluation of travel and health outcomes in relation to new bicycle infrastructure, Sydney, Australia. Journal of Transport & Health. 2017;6:386–395.

29. Stinson MA, Bhat CR. Frequency of bicycle commuting: internet-based survey analysis. Transportation Research Record. 2004;1878(1):122–130.

30. Liao Y, Chang SH, Ku PW, Park JH. Associations of public bicycle use with transport-related and leisure-time physical activity in Taiwanese adults. Journal of Transport & Health. 2017;6:433–438.

31. Fairnie GA, Wilby DJ, Saunders LE. Active travel in London: the role of travel survey data in describing population physical activity. Journal of Transport & Health. 2016;3(2):161–172.

32. Sears J, Flynn BS, Aultman-Hall L, Dana GS. To bike or not to bike: Seasonal factors for bicycle commuting. Transportation research record. 2012;2314(1):105–111.

33. Brown BB, Smith KR. Complex active travel bout motivations: Gender, place, and social context associations. Journal of transport & health. 2017;6:335–346.

34. Akar G, Clifton KJ. Influence of individual perceptions and bicycle infrastructure on decision to bike. Transportation research record. 2009;2140(1):165–172.

35. Baltes MR. Factors influencing nondiscretionary work trips by bicycle determined from 1990 US census metropolitan statistical area data. Transportation research record. 1996;1538(1):96–101.

36. Chieng M, Lai H, Woodward A. How dangerous is cycling in New Zealand? Journal of Transport & Health. 2017;6:23–28.

37. Zientek LR. Exploratory and confirmatory factor analysis: understanding concepts and applications. Taylor & Francis; 2008.

38. Dannenberg AL, Ricklin A, Ross CL, Schwartz M, West J, White S, et al. Use of health impact assessment for transportation planning: importance of transportation agency involvement in the process. Transportation Research Record. 2014;2452(1):71–80.

39. 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. Transportation Research Record. 2008;2082(1):17–26.

40. Winters M, Teschke K, Grant M, Setton EM, Brauer M. How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel. Transportation Research Record. 2010;2190(1):1–10.

41. Goodman A, Green J, Woodcock J. The role of bicycle sharing systems in normalising the image of cycling: An observational study of London cyclists. Journal of transport & health. 2014;1(1):5–8.

42. Marik PE, Varon J. Nitroglycerin, stroke and hypertension: a word of caution. Crit Care. 2015;18(3):59.

43. Brown BB, Tharp D, Tribby CP, Smith KR, Miller HJ, Werner CM. Changes in bicycling over time associated with a new bike lane: relations with kilocalories energy expenditure and body mass index. Journal of transport & health. 2016;3(3):357–365.