4-2018

"Safety in Numbers" and Bicycle Safety: A Detailed Analysis of the Denver Metropolitan Area

Rachael Thompson Panik
Clemson University, rtpanik@gmail.com

Follow this and additional works at: https://tigerprints.clemson.edu/all_theses

Part of the Urban, Community and Regional Planning Commons

"SAFETY IN NUMBERS" AND BICYCLE SAFETY:
A DETAILED ANALYSIS OF THE DENVER METROPOLITAN AREA

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of City and Regional Planning

by
Rachael Thompson Panik
April 2018

Accepted by:
Dr. Eric Morris, Committee Chair
Dr. John Gaber
Dr. Patrick Gerard

ABSTRACT

Recently across the US, there has been a push to accommodate and encourage the viability of alternative modes of transportation—especially bicycling. Leaders across all levels of government, trade groups, advocacy and policy groups, and others are promoting different methods to make urban areas more bikeable. Now, as planning practice is moving towards implementing a transportation system that serves different types of travelers, the US faces challenges involved with retrofitting existing automobile-oriented streets.

While implementing bicycle safety initiatives is becoming a popular movement among municipalities, there have been differing opinions on the best way to make cities more bikable in academic literature (Pucher & Buehler, 2012). There is an ongoing debate about what types of improvements will be the most effective at reducing crash rates and/or decreasing individual risk for cyclists. Since 2003, one of the key factors in this debate has been the phenomenon of "safety in numbers."

"Safety in numbers," or SiN, describes the observed inverse correlation between bicycle ridership and cyclist risk (Jacobsen, 2003). As ridership numbers increase, the relative risk per cyclist is said to decrease (all else being equal). When examining large-scale datasets, such as national ridership counts and crash statistics, research suggests there is a significant negative, non-linear correlation (exponentially decreasing) between ridership and crashes per rider. This means that while the total number of crashes increases with ridership, the rate of crashes per rider decreases.

While bicycle safety and SiN are well-researched topics, there are still many questions about the SiN effect that are still unclear. First, the full character of the SiN effect is not explicit in the existing literature. Nearly all studies of the phenomenon have been conducted with large units of analysis (cities, countries, etc.). No study to the researcher's knowledge has considered the SiN effect at the individual street level with real data. Second, because SiN has not been studied with small units, there has not been a way to control for road conditions that also effect bicycle crash rates. And third, it is not clear how all of the factors that determine cyclist injury and fatalities—including SiN, bicycle infrastructure, speed limit, road design, congestion, etc.—interact with one another.

These gaps in collective understanding about safety in numbers has led to disagreements among scholars about its nature and implications for practice. One of the major debates surrounding SiN and policy has been its use as an argument to dissuade investment in separated bicycle infrastructure. Some think that separated infrastructure may undermine some of the safety benefits that may affect cyclists because of SiN; the goal of this type of infrastructure is to limit motorists' conflict points with cyclists, and because of this, separated infrastructure may actually endanger other cyclists on the road because fewer cyclists are interacting with drivers in mixed traffic, lessening drivers' incentives to adjust their behavior (assuming that behavior modification underlies the SiN effect) (Thompson et al., 2017).

Despite limited understanding about this topic, SiN is has been used to make policy justifications, specifically pitting policy-only solutions against infrastructure improvement ones (Bhatia & Wier, 2011; City of Berkeley, 2010). It is crucial, then to understand the SiN effect more fully. My research addresses these gaps in the literature and provides recommendation for practice.

My research reports several major findings. First, the safety in numbers effect is reflected on the individual road segment level; using a Cragg double hurdle model, I showed that numbers are a significant predictor of crashes, even when other control variables—infrastructure, congestion measures, speed limit, functional class, median household income, and road length—are added to the model. Second, my research shows that the SiN effect is best characterized by a non-linear, exponentially decreasing mathematical model, even on the segment level. Third, my research created detailed predictions that quantify how the SiN effect changes under different conditions. The most notable of these findings was twofold. First, there was no significant difference in the predicted number of crashes for segments with or without bike lanes as the number of trips increased. And second, facilities with separated bike lanes also receive a safety benefit from increased exposure, but the benefit is not as strong as on segments without separated bike lanes.

In summary, my research verified existence of SiN on the road segment level as well as characterizes the effect mathematically. I also suggest that practicing planners

should encourage more biking to improve overall road user safety, but that this should

be done in tandem with other measures such as bicycle infrastructure.

ACKNOWLEDGEMENTS

My thesis in its entirety is the sum of an entire support system that I have been blessed with for the last several years. First, I would like to express my deepest and most earnest thanks to my committee members—Dr. Eric Morris, Dr. John Gaber, and Dr. Patrick Gerard—with special thanks to Dr. Eric Morris for chairing my committee and guiding my thesis process. Their expertise and encouragement were invaluable to creating this research.

Second, besides my committee members, I would also like to thank Jared Draper, Rebecca Sanders, Spencer Gardner, and other employees at Toole Design Group for their willingness to share expertise, provide data for my analyses, and encourage me regularly. Third, I would like to thank my fellow classmates at Clemson University for supporting me with kind words and laughter for the last two years. Special thanks to my classmates Zemin Dhanani and Providence Adu, for sharing their GIS skills so freely with me during endless hours running analyses at Clemson, and to Katherine Bernier for being my sounding board during all of the challenges one faces when writing a thesis.

Finally, and most importantly, I would like to thank my family and friends for supporting me throughout my education, with especially warm thanks to my husband William Panik, who cooked innumerable meals for me, encouraged me when I was tired, believed in my abilities without fail, and who always listened to me while I shared my passions for research and urban planning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.   LITERATURE REVIEW

1.1   Introduction: Why cycle in cities?

Recently across the US, there has been a push to accommodate and encourage the viability of alternative modes of transportation—especially bicycling. Leaders across all levels of government, trade groups like the American Planning Association, advocacy and policy groups, and others are promoting different methods to make urban areas more bikeable. For example, cities all over the US have been creating or updating active transportation plans to implement a systematic approach to policy making and infrastructure retrofitting. Municipal leaders and government agencies have identified the benefits of having a transportation system that allows more biking (Guide, 2011). Many of the studies on urban transportation's relationship to quality of life have pointed towards similar conclusions: urban spaces that support convenient biking and walking can improve the quality of life of their residents in many respects, including improved health, increased economic prosperity, and a stronger sense of place (Campbell & Wittgens, 2004; Elvik, 2009).

From an economic perspective, dedicating resources to urban designs and land use plans that promote active transportation can bring about lucrative returns on investment. It has been shown that through increasing investment in public amenities,

communities can boost private investment, tourism, and the surrounding property values (Richards, 2014). For example, Lancaster, California turned an arterial with five lanes into a "Main Street" of sorts by investing in streetscaping and traffic calming, and lowering the speed limit. The total investment cost about $12 million (in 2014 dollars), and in return, the city attracted more than $300 million in private investment (Richards, 2014). Similar investments in other cities may lower the price and increase the appeal of alternative modes of transportation (Sorensen, Wachs, Min, Kofner, & Ecola, 2008)

In terms of urban design and community, biking can promote a sense of place among residents (Richards, 2014). As it becomes more prevalent in the US, biking can alleviate congestions issues that are economically wasteful and detract from quality of life (Sorensen et al., 2008), and, arguably, happiness (Montgomery, 2013; Morris & Guerra, 2015).

There are also social benefits from biking and forms other active transportation; biking can help combat some of the more urgent public health problems in the US. Evidence has shown that desirable individual behaviors must be supported with environmental factors—such as places that encourage biking and walking—to lessen risk for major health issues, such as diabetes and obesity (Botchwey, Trowbridge, & Fisher, 2014; Elvik, 2009; Johan de Hartog, Boogaard, Nijland, & Hoek, 2010). The built environment can also support and encourage biking. A study of Dutch cycling habits revealed that Dutch people have a half of a year longer life expectancy than comparable

countries worldwide, most likely due to cycling (Fishman, Schepers, & Kamphuis, 2015; P. Schepers, Twisk, Fishman, Fyhri, & Jensen, 2017).

Biking also improves quality of life for those who do not have or cannot afford their own vehicles. Biking is faster and more efficient than walking and thus can improve mobility options for those who cannot drive (Glaeser, Kahn, & Rappaport, 2008; Wegman, Zhang, & Dijkstra, 2012). Ownership of a bicycle can also notably increase mobility for those living in poverty (Wegman, et. al, 2012).

1.2   Challenges to cycling in the United States

There are many challenges to encouraging more cycling in the US. Road infrastructure design practices in the United States over the last 100 years have caused many problems that urban planners and municipal officials are still trying to mitigate. After the mass production of personal cars, our transportation system became increasingly more automobile-centric to the detriment of bicycling safety (Botchwey et al., 2014; Wegman et al., 2012) Now, as planning practice is moving towards implementing a transportation system that serves different types of travelers, the US faces challenges involved with retrofitting existing automobile-oriented streets. One study identifies travel characteristics of cyclists that are not catered to by our existing road system (Wegman et al., 2012). They include:

- *Causing vulnerability* in a crash due to cyclists' lack of physical protection and speed differentials between cyclists and vehicles;

- *Not accommodating flexibility* in behavior, meaning that cyclists cannot adapt their riding habits (including trip route, type of facility used, or travel lane on the road) in response to other factors like weather, debris or damaged paving, or heavy traffic (Twaddle, Schendzielorz, Fakler, & Amini, 2014);

- *Increasing the propensity to fall* off of the bicycle due to uneven pavement, poorly designed roads, or very narrow travel lanes;

- *Causing inconspicuousness of bicyclists to drivers* when in mixed traffic;

- *Not catering to cyclists' extra energy expenditures* that are required to cycle in hilly areas or over long distances through providing extra right of way for cyclists in topographically challenging situations;

- And *ignoring differential ability* among a variety of riders and trip types.

In most cases, our present-day transportation infrastructure system and development patterns have been designed to optimize vehicular travel (Wegman et al., 2012). Up until recent decades, these design practices have not been questioned. Now, best practice manuals produced by all levels of government and trade organizations have begun to reassess road design standards to be more equitable for all users. Major examples include the Federal Highway Administration's Separated Bike Lane Planning and Design Guidelines (2015), the Massachusetts DOT Separated Bike Lane Planning and

Design Guide (*Separated Bike Lane Planning and Design Guide, 2015)*, and the National

Association of City Transportation Officials (NACTO) *Urban Bikeway Design Guide*

(2011). All of these are compilations of bicycle facility and streetscape design guidelines

at varying levels of detail. They are all driven by a forwarding-thinking combination of

engineering/human factors research melded with an urban design perspective. Each

guide recognizes bicycles as not just other vehicles in the stream of mixed traffic, but

instead as their own distinct travel mode with needs that are different than motor

vehicles'. NACTO's guides in particular recognize cyclist (and pedestrian) activity as a

driving force in creating lively streets and aesthetic spaces.

Bicycle safety has been a popular topic for municipalities in recent years. Presently,

cycling is associated with a much higher risk of injury than driving a personal automobile

or taking transit per kilometer travelled—7.5 times higher, to be exact (Elvik, 2009). One

of the more innovative ways that cities are trying to combat cycling crashes is through

"vision-zero" plans. Vision-zero plans outline a series of steps that would theoretically

eliminate cyclist fatalities by a given year. This is done by using crash statistics from a

municipality to identify street and intersection characteristics that correlate significantly

with higher crash risk to identify dangerous intersections. Once those characteristics are

determined, other roads and intersections with similar characteristics are targeted for

improvements, sometimes before there are even fatalities.

The types of challenges facing municipalities as they try to become more bicycle-

friendly vary in differing US regions. In the southern portion of the US, for example,

there are unique challenges for biking and other modes of active transportation. Many

places in the South experienced the post-World War II boom of development consisting

mostly of lower density residential and commercial developments—or "sprawl." While

other places in the US also experienced this type of development, the degree to which

the Southern states adopted sprawling development was much more intense in terms of

development density and land use due to rapid periods of growth in the era of the

automobile (Ewing & Hamidi, 2014). Interestingly, the 9 out of the top 10 states for

number of bicycle fatalities are all in the southeastern corner of the US (Alabama,

Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and

Tennessee) (Price, 2016).

Another major issue for biking in the American South is that there has been a

dearth of funding for improving cycling opportunities. Southern states have spent

notably less of the federal transportation funding given to them on bicycle

infrastructure (only 1.7%) compared with the national average (2.1%), and Southern

states also spend less per capita on bicycle- and pedestrian-only projects (Price, 2016).

As has been noted, development patterns that discourage bicycling have also been

prevalent in a majority of Southern states (Price, 2016). As has been noted,

development patterns that discourage bicycling have also been prevalent in a majority

of Southern states (Price, 2016). However, in other places in the US, such as the Denver,

Colorado region (and other Midwestern areas), Portland, Oregon, and many cities across

California, bicycling is more common and supported by local governments through policies and infrastructure.

In Southern states and across the US—including the most bicycle-friendly cities—bicycle crashes and their resulting injuries and fatalities are a far-reaching and not well understood issue. To reduce or even eliminate these fatalities, more measures should be taken to ensure that bicycling becomes safer (Bhatia & Wier, 2011).

While implementing bicycle safety initiatives is becoming a popular movement among municipalities, there have been differing opinions on the best way to make cities more bikable in academic literature (Pucher & Buehler, 2012). There is an ongoing debate about what types of improvements will be the most effective at reducing crash rates and/or decreasing individual risk for cyclists. Since 2003, one of the key factors in this debate has been the phenomenon of "safety in numbers."

1.3   History of "safety in numbers" research

"Safety in numbers," or SiN, describes the observed inverse correlation between bicycle ridership and cyclist risk (Jacobsen, 2003). As ridership numbers increase, the relative risk per cyclist is said to decrease (all else being equal). When examining large-scale datasets, such as national ridership counts and crash statistics, research suggests there is a significant negative, non-linear correlation (exponentially decreasing) between

ridership and crashes per rider. This means that while the total number of crashes increases with ridership, the rate of crashes per rider decreases.

This phenomenon is generally called the non-linearity of risk, and it applies to other vulnerable road users as well, including pedestrians and motorcyclists (Elvik, 2009; Wegman et al., 2012). The theory can be generally described as follows: as more cyclists enter the system, they face lower risks on an individual basis (per capita). The natural implication of this phenomenon is to promote policies to encourage more people to cycle to improve overall safety for cyclists (Jacobsen, 2003).

It is important to differentiate between the theory of "safety in numbers" and cycling in groups. It is probably true that, when a person rides a bicycle with other cyclists for trips, that person's risk of having a crash with an automobile decreases (possibility because of increased visibility of a group of cyclists, more people watching for a potential crash, etc.). The SiN theory, however, is different. SiN does not refer to singular crash scenario risks. Instead, SiN refers to a system-wide phenomenon; as more and more cyclists enter the transportation system, relative risk (risk per rider or risk per trip) decreases.

Jacobsen (2003) published the first study on the relationship between the number of cyclists and the number of crashes per capita. He found that the likelihood of cyclists being struck by an automobile decreases in a nonlinear, inverse fashion as the number of people biking increases.

Jacobsen used five ridership and crash data sets in his study. Three of the data sets allowed him to measure injuries/capita for walking and biking across several scales—across multiple cities within the same state (California), across multiple Danish cities, and across multiple European countries. The other data sets were time series, meaning that injuries/fatalities were measured annually. To specify the relationship between injuries per capita versus the number of reported riders across cross-sectional data sets, he used least squares analysis, operationalized as follows:

$$I = aE^b$$

$where \; I = injury \; measure \; (injuries \; or \; deaths \; per \; capita),$
$E = measure \; of \; biking \; or \; walking \; (kilometers \; biked \; per \; capita \; per \; day)$
$a = scaler \; (to \; be \; calculated)$
$b = type \; of \; relaitonship \; (to \; be \; calculated; typical \; characerizations \; below)$

Equation 1: Safety in numbers

$b = 1: linear\ relationship\ between$
$increase\ of\ exoposure\ and\ crash$

$I = aE^b$
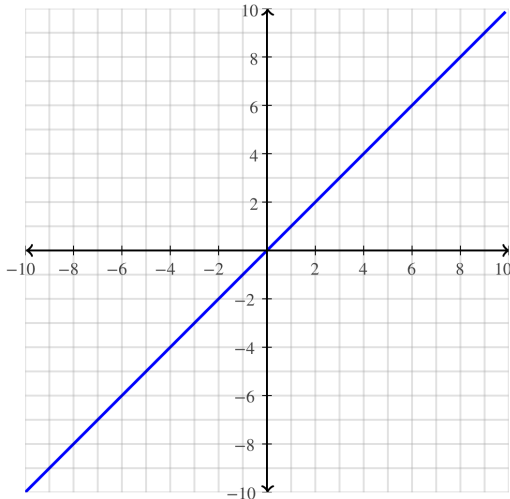
Figure 1: Linear increase of risk
with more exposure

$b < 1 : less\ than\ linear$
$relationship$

$I = aE^b$
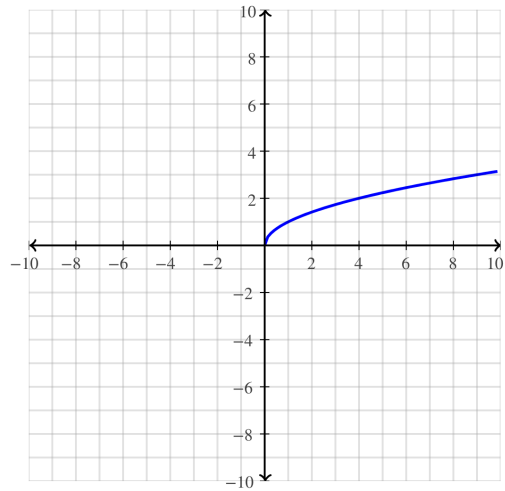
Figure 2: "Less than linear" increase
of risk with more exposure

$b < 0$
$=\ negatively\ exponentially\ related;$
$of\ cyclists\ to\ motorists\ and\ crashes$

$I = aE^b$
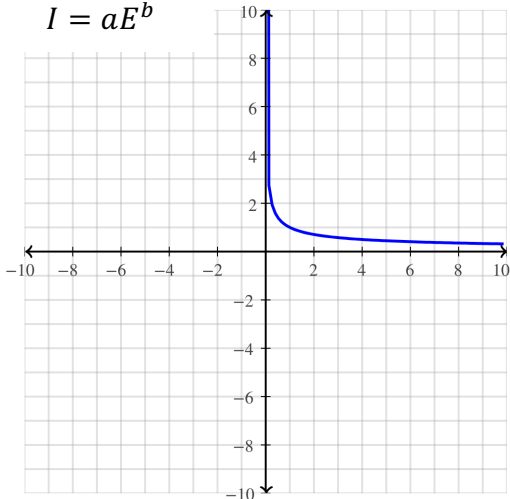
Figure 3: Decay of risk with more exposure

$b > 1$
$=\ more\ than\ linear\ relationship,$
$or\ positively\ exponentially\ related$

$I = aE^b$
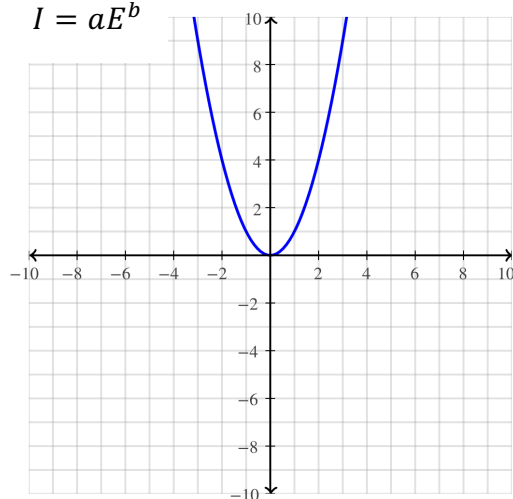
Figure 4: Exponential growth of risk with
more exposure

Graphing Tool Source: https://www.mathpapa.com/calc/tutorial/graphing-equations/

10

When calculating the injury measure in terms of units of biking (injuries per kilometer of biking), modeled as:

$$\ln\left(\frac{I}{E}\right) = \ln\left(aE^{b-1}\right)$$

$$\ln(I) - \ln(E) = \ln(a) + (b-1)\ln(E)$$

$$\frac{I}{E} = aE^{b-1},$$

Equation 2: Safety in numbers (in terms of units of biking)

Jacobson found that the injury measure decreased significantly in places where there are higher kilometers of biking yearly (Jacobsen, 2003). His empirically derived values for *b* fell below 0, meaning that the relationship between crash rates and exposure was less than linear, or exponentially decreasing (similar to Figure 3). He attributed this 1) to changes in driver behavior in response to seeing more cyclists as drivers become more attuned to having cyclists on the road, and 2) to drivers in places with many cyclists being more likely to bike themselves, thus being more aware of cyclists on the road (Jacobsen, 2003). His models have been used to predict the injury rates due to a given increase in cycling.

The range of the safety-in-numbers effect for cyclists

Upper curve is based on the highest coefficient found in the literature (0.669)

Lower curve is based on the lowest coefficient found in the literature (0.085)
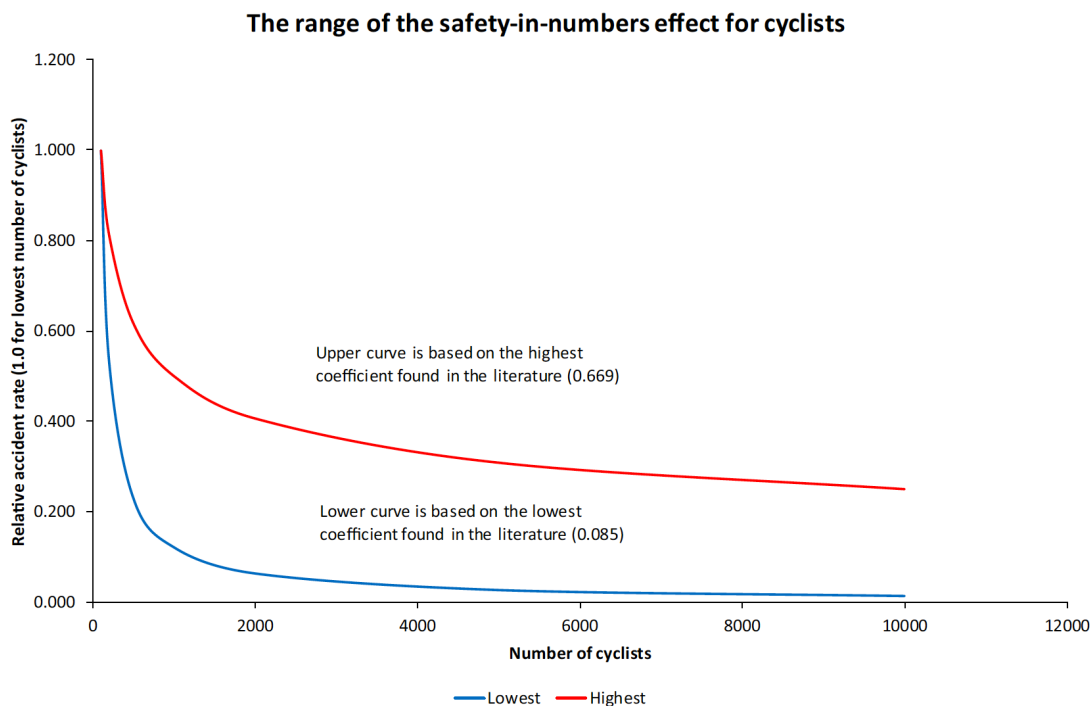
Figure 5: Safety in numbers summarized from literature

Thus there are multiple constructs proposed as to why this phenomenon occurs, the most popular of which is that SiN is caused by something similar to the theory from the psychology field known as conditioning; humans learn to expect a certain outcome in response to the same event. Applied to bicycling and traffic safety, the theory is that drivers adjust their behavior as they see and interact with cyclists on the road, ultimately becoming safer drivers (Wagner, 1972). In regard to SiN, it is supposed that drivers who see cyclists regularly become conditioned to their presence and therefore adjust their driving behavior to become safer around cyclists (Jacobsen, 2003).

Other studies have supported this finding. Wegman et al. (2012) found a similar non-linear relationship when comparing bicycle fatality rates per kilometer in European

countries, finding that countries with higher kilometers of bicycle ridership per year

have a relatively low fatality rate, and countries with very low traffic levels (less than 20

km per person per year) have relatively high rates (Wegman et al., 2012). However, in

countries with around 200-300 kilometers per year (which is approximately the median)

there is a large variance of crash rates. This means that there are likely other factors

affecting crash rates other than the SiN effect, such as investment mechanisms, bicycle

facilities and bicycle usage (Wegman et al., 2012). People may bike because it is safe due

to factors other than "numbers," like bicycle infrastructure or bike-friendly topography.

Thus, there may be a spurious relationship between exposure measures and the

number of crashes; intrinsically safe places may attract large numbers of bicyclists.

Interestingly, accident rates in very low-income countries are much higher, even though

the large majority of people walk and bike as their mode of transportation (Elvik, 2009).

This supports the hypothesis that the number of kilometers travelled on a bicycle is not

the only factor determining cyclists' safety. Neither this study nor Jacobsen's study

controlled for these factors, such as the presence or absence of bicycle lanes.

Another way SiN has been researched has been through agent-based models

(ABMs). In one study, step-based models (Wagner, 1972) were used to simulate drivers

adapting their behavior over time after interacting with cyclists on the road. Agent-

based models are useful for modeling micro-level "disaggregate populations that give

rise to macro-level phenomena" (Thompson et al., 2017). Said another way, ABMs

model individual interactions between cyclists and cars to try to explain the larger-scale

phenomena of SiN. In Thompson's research, the model shows how drivers' ability to adapt their driving habits around cyclists may change in response to physically separated infrastructure, like a cycle track or side path (Thompson et al., 2017). Results from the Thompson ABM study support the existence the SiN phenomenon and somewhat characterize the observed nature of SiN. This research created a simulated transportation system with randomly assigned sections of separated bicycle infrastructure and bicyclists in mixed traffic conditions. A rendering of the system is shown below. Buildings are blue, cars are white, cars interacting with cyclists are red, cyclists are black, roads are grey, and separated bicycle infrastructure is green.
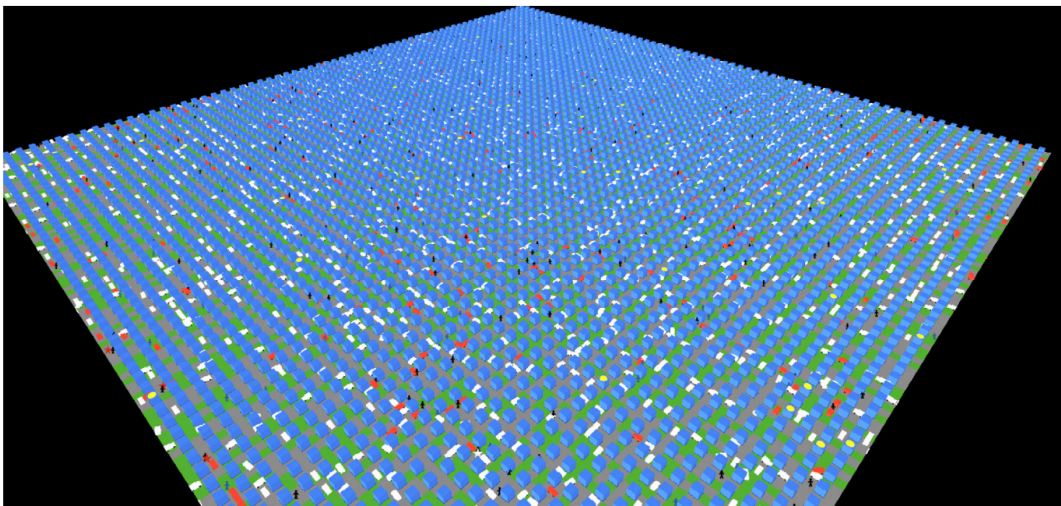
Figure 6: Simulated transportation system from Thompson et al. (2017)

Driver and cyclist behavior was governed by a step function that modeled classical conditioning (Wagner, 1972), as shown below:

$$V(t_k) = \frac{\sum_{i=1}^{n} V_i(t_k)}{n}$$

where

$$V_i(t_k) = V_i(t_{k-1}) + \Delta V_i(t_k)$$

and the change in association $\Delta V_i(t_k)$ is calculated as

$$\Delta V_i(t_k) = \begin{cases} \alpha\beta(\lambda - V_i(t_{k-1})) & \text{if driver } i \text{ encounters at least one cyclist at time } t_k \\ 0 & \text{if driver } i \text{ last encountered a cyclist after } t_{k-m}, \text{ and} \\ -\alpha\beta V_i(t_{kc-1}) & \text{otherwise.} \end{cases}$$

where $V(t_k)$ = mean level of driver's ability to associatiate road they are on

with the presence of cyclists ,

t_k = given time in simulation

n = population of drivers

α = cyclist saliency (or the cyclists' dectability by drivers)

β = saliency of the road (or the cylists occupying a positon on the road

shared by adriver)

λ = association value (or the ability of drivers to associate

certain segments of road with cyclist presence)

Figure 7: Step function governing motorists' behavior from Thompson et al. (2017)

This model utilized three variables—(1) the saliency values (the ability for cyclists

and road segments (α and β), (2) the association value (the ability of drivers to associate

certain segments of road with cyclists), and (3) the amount of interactions cyclists have

with vehicles—to study the behavioral adaptation around cyclists which likely underpins

the SiN effect (Thompson, et al. 2017). The scope of this research was to investigate

how cyclists' risk increased/decreased from using physically separated facilities as

drivers changed their behavior from interacting with cyclists. The virtual drivers were

programmed to alter their behavior to drive safer around cyclists incrementally as they

intermingled with cyclists in the network. When cyclists used separated infrastructure,

however, they were essentially invisible to drivers, meaning that drivers did not alter

their behavior incrementally in response to those using separated infrastructure.

The study found that, as more cyclists began using separated infrastructure, the

remaining cyclists that were interacting with cars had a higher risk per capita (at high

driver association values). See results below. Each line indicates relative risk, or "RR," at

a given "BA," or the association level, λ.
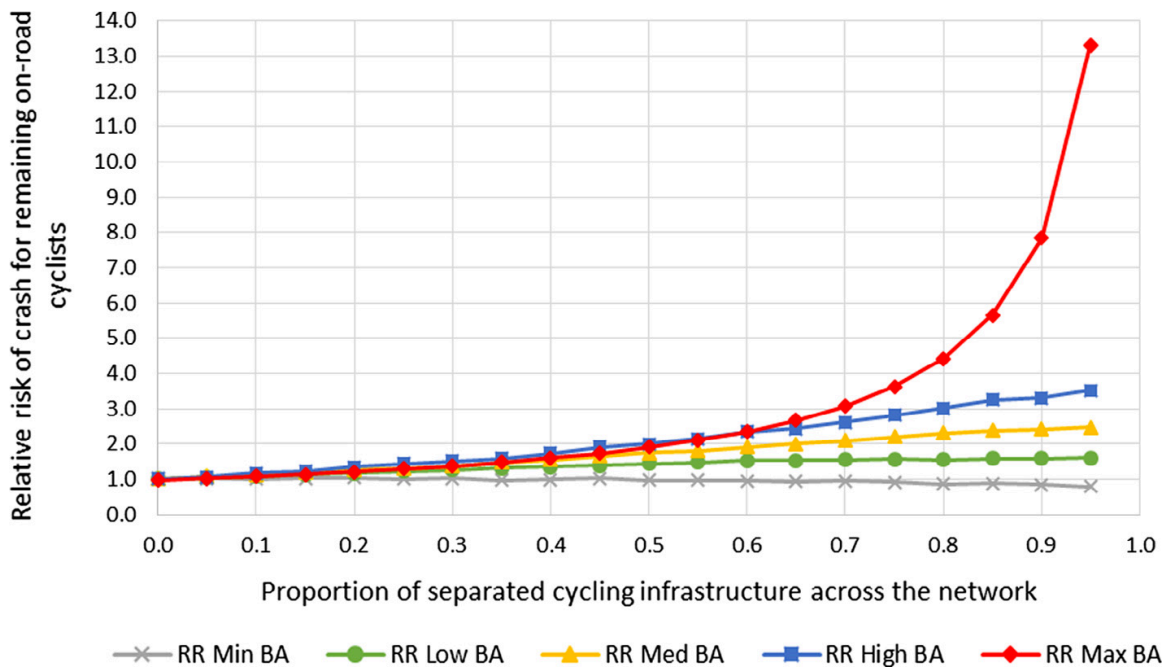
Figure 8: Relative risk per cyclist with varying association levels and proportions of separated cycling infrastructure from Thompson et al. (2017)

As more cyclists left the "cycle track"—or segments of road within a system in which cyclists do not interact at all with drivers (green segments in Figure 6)—and entered into mixed traffic, relative risk per cyclist increased marginally. As previously explained, this model built in the assumption that separated infrastructure keeps motorists from interacting directly with cyclists, the driver's ability to expect a cyclist (and therefore change their behavior) did not improve unless cyclists shared the street with cars. Instead, cyclists using the "cycle track" were essentially invisible to the drivers. Results showed that only when the simulated system had greater than 80%

separated infrastructure and the drivers had high association strength (λ) was there

significantly higher relative risk of crashes for the cyclists remaining on the street. This

means that, according to agent-based models, cyclists would only be at increased risk by

using separated infrastructure if 80% of their travel is on physically separated

infrastructure with no interaction with vehicles (Thompson et al., 2017). In this scenario,

the few riders on the non-separated road would be experience much higher risk per

capita than in a scenario with less separated infrastructure. In the most extreme case

(where over 70% of the system was separated infrastructure and drivers had the

maximum association level [the ability for drivers to associate streets with cyclists]) the

total number of crashes within the system—not just risk per cyclist—also increased. In

summary, this study proposed that separated bicycle infrastructure could potentially

cause cyclists higher risk if we assume that they do not benefit from the "numbers"

effect when using it.

There are several notable issues with this approach. First, the assumption that

bicyclists using separated cycling infrastructure are "invisible" to drivers ignores one of

the major questions in this research area: *how* does SiN accrue to cyclists? This has not

been clearly proven in existing literature, but Thompson et. al's research built the

assumption into the virtual drivers' behavior; the invisible infrastructure makes no

distinction about whether or not drivers must interact with cyclists or just *see* them on

the road, but instead just assumes. This major and potentially unfounded assumption

may inappropriately ascribe SiN effects (or lack thereof) to separated infrastructure. For

example, if SiN only affects cyclists who interact with motorists in mixed traffic, then assuming bicyclists using separated infrastructure are "invisible" to drivers is logical. However, this has yet to be determined in the literature. It may be that motorists only have to see cyclists for the effect to take hold, in which case only very limited infrastructure (greenways and trails, for example) would be totally out of sight for the drivers. Even outside this assumption, the scenario in which cyclists complete 80% of miles travelled on separated infrastructure is highly unlikely in the US within the foreseeable future due to very few complete networks of bicycle infrastructure throughout the country.

1.4   Arguments within SiN research

Jacobsen's (and others') findings have been challenged in the literature due to methodological issues, conceptual validity, and general usefulness. It has been pointed out that finding a correlation between ridership and crashes is not necessarily indicative of causation; what could be nothing more than a statistical relationship may not be borne out in reality (Bhatia & Wier, 2011; Elvik, 2009; Wegman et al., 2012). Making any sort of policy recommendations for promoting cycling based on these correlations alone could be overstepping and preemptive—and potentially unethical if doing so causes more crashes and fatalities due to the lack of supporting infrastructure, law enforcement, etc. (Bhatia & Wier, 2011).

First, it has been pointed out that Jacobson's analysis was cross sectional, meaning that there could be issues with reverse causation. Most importantly, as has been pointed out, "safety may actually cause numbers" (Bhatia & Wier, 2011). It is plausible that more people may bike in places where it is safer to do so; if that is the case, decrease in relative risk is actually due to safe infrastructure rather than the "numbers," which clearly has major implications for the body of literature on bicycle safety and for urban planning practice. Jacobson's results do not overcome this temporal issue—whether the safety or the numbers comes first.

Second, Jacobson's research did not account for the built environment influencing driver behavior through design speed and traffic volumes, both of which have been identified as important and spurious factors in cyclist and pedestrian safety. And other factors, such as traffic law enforcement or traffic laws (Berg, 2006; Lavetti & McComb, 2014), topography, and weather (Wegman et al., 2012), may also be confounding variables that challenge the validity of SiN as a theory. For example, Wegman, et al. (2012) found that, in countries with high levels of ridership and safety, there are correlations between the number of cyclists and a higher density of bicycle facilities.

Third, others have pointed out that SiN may be caused by altered driver behavior *and* may also be related to respect. In places where there is not a strong culture of bicycling, cyclists may not be given the same level of respect on the road as other mode users, which could ultimately contribute to less courteous (and less safe) treatment

from other users. In a study of bicycle-car accidents in Finland it was found that there were differences between expected and actual rights (for example, who *believes* they have the right of way in a turning scenario vs. who *actually* has the right of way) on the road (Räsänen & Summala, 1998).

 It is possible that cyclists may be more respected in cities with higher cycling rates because drivers are also more likely to bike for some of their trips, creating in them a sense of empathy for cyclists (Wegman et al., 2012). Similarly, an Australian study found that drivers who were also cyclists were 1.5 times more likely to self-report safe driving behaviors around cyclists than drivers who never cycle themselves (Johnson, Oxley, Newstead, & Charlton, 2014). Drivers who also cycle also report more positive attitudes and a better knowledge of road rules that pertain to cycling (Johnson et al., 2014). This could be a part of what underpins the SiN effect. While this still suggests that more cycling may lead to increased safety overall, the existing literature (to my knowledge) has not measured this directly in any analysis, so the actual effect on user safety is unclear.

 SiN's usefulness as a theory has also been challenged. Some disagree about whether or not it provides useful information for practice, making the argument the measure used for safety in SiN studies (relative risk for individual cyclists) is not a valid measure of safety. In terms of bicycle safety research, "safety" has been described from two somewhat opposed perspectives: (1) in terms of individual risk, and (2) in terms of aggregate number of deaths or fatalities.

All SiN research has used this first construct of safety, most likely because of the available data. SiN has been measured in terms of risk per individual; as the number of riders increase, the risk *per individual* decreases. Said another way, the likelihood of serious injury/death per rider decreases as the number of riders in the system and as the ratio of cyclists to motorists increases. However, this conceptualization of safety says little about the actual number of crashes in the system.

Bhatia and Weir (2012) point this out in their report. In spite of the decreased risk per person with increased exposure to motor vehicles, the actual number of crashes has continued to increase as more cyclists enter the system. They argue that because of this, there actually is *not* safety in numbers. If "safety" is defined in terms of crashes within a system over a given period of time (Hauer, 1982), there is merit to their argument. While pure research is certainly useful for the sake of furthering knowledge, they argue that safety research is different in its ethical impetus. Arguably, the point of transportation safety research is to reduce the number of preventable crashes, injuries, and fatalities in the transportation system (Bhatia & Wier, 2011). If safety is conceptualized the way Bhatia and Weir (2011) and Hauer (1982) have defined it, then SiN measured in terms of individual risk may not have much value as a topic of research as it does not contribute much towards improving "safety" because many people still die each year from cycling/vehicle crashes.

In light of this important distinction, some have argued that SiN as a topic area may distract from the overall problem of cyclist fatalities in the US (Bhatia & Wier,

22

2011). When researched under this conceptualization of "safety," SiN research may just

shift the blame to victims of a transportation system that is (arguably) not designed for

their safety (Bhatia & Wier, 2011). The conversations about SiN shift the cause of death

and injury to those biking and away from policies and road designs that could protect

them. Cyclists' accidents often see more fatalities than automobile accidents due to

cyclists' vulnerability on the road and their lack of protection, as well as the disparities in

speed between cyclists and cars (Wegman et al. 2012). To assign responsibility to

bicyclists to fix a transportation system by numbers alone may be poor research and

even unethical (Bhatia & Wier, 2011). As Bhatia and Weir (2011) pointed out:

> "Some transportation agencies appear to use higher prevalence of walking as the
>
> primary explanation of high pedestrian injury frequencies. For example, on a
>
> website describing pedestrian safety in their community, the City of Berkeley,
>
> California states: "Compared to other cities, Berkeley has a high number of
>
> bicycle and pedestrian injuries. The main reason for this is because so many
>
> people walk and bike in Berkeley, not because it is a dangerous place" (City of
>
> Berkeley, 2010). Such statements appear to readily discount both the burden of
>
> injury and the contribution of transportation system design, speed, and other
>
> environmental factors" (p.238).

Bhatia and Wier went as far as to say that this measure of SiN may not be real

safety research because it is not applicable to the practice of making all road users safer

because it says nothing about reducing total crashes and therefore, in their opinion, is

ineffectual.

1.5   Debate around SiN as it applies to practice

Jacobsen's widely quoted response to the "safety in numbers" phenomenon is

that "policies that increase walking and biking appear to be an effective route improving

the SiN of walking and biking" (p. 209, 2003) . However, this statement has been

controversial among researchers.

Jacobsen was not specific about the types of policies that would be best for

promoting cycling. Certain policies could focus on implementing dedicated

infrastructure to separate bicyclists from auto traffic such as bike lanes. While there has

been a decades-long debate about protected cycling infrastructure, or infrastructure

that has some physical barrier between cyclists and motorists, versus mixed-traffic

cycling, SiN is has been used to make policy justifications, specifically pitting policy-only

solutions against infrastructure improvement ones (Bhatia & Wier, 2011; City of

Berkeley, 2010). For example, the City of Berkley has used Jacobsen's report and

statistics to encourage more biking and walking in the area (City of Berkeley, 2010).

Berkeley's website used his results to justify policies that may encourage biking and

walking regardless of the presence of bicycling infrastructure or improved road design standards (2019). Also, some countries have begun to measure "cycling success"—as in how much better cycling is becoming in a place—in terms of increasing the number of bicycle travel miles travelled each year, as opposed to reducing the numbers of crashes or another comparable safety measure. It is crucial to remedy the public safety issue of high bicycle crash, injury, and fatality rates, but it is debatable what types of policies or other improvements will be truly effective. It is unclear whether or not the SiN effect is strong enough to actually protect riders.

One of the major debates surrounding SiN and policy has been its use as an argument to dissuade investment in separated bicycle infrastructure. Some think that separated infrastructure may undermine some of the safety benefits that may affect cyclists because of SiN; the goal of this type of infrastructure is to limit motorists' conflict points with cyclists, and because of this, separated infrastructure may actually endanger other cyclists on the road because fewer cyclists are interacting with drivers in mixed traffic, lessening drivers' incentives to adjust their behavior (assuming that behavior modification underlies the SiN effect) (Thompson et al., 2017). Some believe that the benefit of SIN may only accrue to cyclists who are riding in traffic without any exclusive infrastructure (bike lanes, bike paths, etc.) dedicated to them (Thompson et al., 2017).

The counter-argument to that perspective is that protected facilities may attract new ridership—will mixed-traffic cycling attract enough cyclists to accrue the benefits of

SIN? It may be that some of the deterrents to cycling, such as high crash rates (or perceived high crash rates) and lack of confidence in their cycling abilities, may keep more road users from switching modes to cycling, meaning that the SiN effect could not take (J. Schepers & Heinen, 2013). Mixed-traffic cycling may be too stressful for new cyclists; would-be cyclists may not feel comfortable enough with their cycling skills to ride alongside automobile traffic, which could deter cyclists and again keep SiN from coming to fruition. A health-focused study of bicycling in Portland, Oregon, a city with a strong bicycle culture, found that a disproportionately large share of bicycling, both for recreation and utilitarian purposes, occurred on streets with existing bicycle infrastructure, such as a bike path or bike lane (Dill, 2009). When tested on the micro level (at an intersection or on a single strip of separated infrastructure), there is also evidence that physically separated infrastructure can reduce the risk for cyclists (Wegman et al., 2012).

Some research has determined that both drivers and cyclists are more comfortable with their travel with physically separated bicycling facilities as opposed to mixed-traffic cycling (R. Sanders & Cooper, 2013; R. L. Sanders, 2014; R. L. Sanders, 2015). If this is the case, then it is another reason for supporting separated infrastructure, and this could be used as another counter-argument against proponents of mixed-traffic cycling only.

As has been shown, the exact nature of SiN is not clear in the existing literature, but it is affecting policies. If some cyclists and practitioners believe that cyclists would

be safer overall (because of the SIN effect) if they were treated as just another vehicle on the road without any special protections, their stances may influence how infrastructure—or lack thereof—is implemented in the future. In the US, where there is not a strong culture for cycling and cycling infrastructure planning is a newer practice, SiN may not be a strong enough force in itself to change cycling culture (or lack thereof) (Richards, 2014), or to mitigate copious amounts of cyclist injuries and/or deaths in our transportation system.

SiN may also change the way the future road safety is predicted. Accident prevention models can be used to predict the amount of future accidents based on a change in the level of bicycle ridership. However, these models will need to account for a SiN effect if it can be quantified (Elvik, 2009):

$$BCM = \propto V_m^{\beta 1} V_b^{\beta 2}$$

$$where \ BCM = predicted \ annual \ number \ of \ bicycle \ crashes,$$
$$\propto = scaler$$
$$V_m \ = volume \ of \ motoroists$$
$$V_b \ = volume \ of \ bicyclists$$
$$\beta = defines \ relationship \ between \ volumes \ and \ the \ number \ of \ accidents$$
$$0.4 \ < \beta \ < 0.9 \ (Elvik, 2009)$$

Equation 3: Accident prediction model example

However, in a review of bicycling habits and infrastructure patterns in the Netherlands, Jacobsen's prediction model did not match changes in cyclist fatality rates in response to increased cycling (P. Schepers et al., 2017). This study found that, in a given timeframe, the distance of cycling per capita increased by 20% and the fatality rate decreased by 80%, where Jacobsen's model would only predict a decrease of 10% (P. Schepers et al., 2017). It is clear that there were other factors influencing crash rates in these results, like road safety measures, enforcement (Berg, 2006), or bicycle infrastructure.

Wegman, et al. (2012) also found that, in countries with high levels of ridership and low crash rates, there are also correlations present between the number of cyclists and a higher density of bicycle facilities. This study pushes back against Jacobsen's belief that policies should encourage more cycling *alone* to improve cyclists' safety; instead, policies should be wrapped up in a "package" of policy changes, infrastructure retrofitting, and increased education (P. Schepers et al., 2017; Wegman et al., 2012). It is not evident from existing literature which of those three approaches is the most effective in reducing cyclist injuries and fatalities.

While SiN may not be a conclusive argument for discounting separated infrastructure, it should encourage reflection about estimating the safety benefits from a given extent of separated infrastructure in the context of its overall connectivity of a given bicycle network (Thompson et al., 2017). To better inform practice, to contribute to the overall understanding of bicycling safety in cities, and to further the academic

literature on SiN, it is important to more fully understand SIN as it applies to different

scales of analysis and how the effect varies over space and how the phenomenon

changes in the presence of bicycle facilities.

1.6   Gaps within SiN and road safety research

It is not clear how all of the factors that determine cyclist injury and fatalities

(including SiN) interact with one another. In a review of literature, the following factors

have been identified as potentially influencing bicycle injuries and fatalities (as shown in

Figure 9 below) (Wegman et al., 2012):

- *Travel behavior*: This factor consists of several other subcategories: (1) locations of

    attractions, (2) needs, opportunities, and abilities, or NOA, and (3) travel

    resistance. A summary of these points is as follows:

    (1) First, the locations of attractions and destinations can be thought of as land

    use patterns. Certain land uses may attract or deter cycling (Richards, 2014). The

    transportation system that supports and influences land use can also cater to or

    endanger cyclists. An example would be that a denser network of roads

    supported by denser land might be more suitable for bicycle travel because this

road structure shortens distances between destinations (P. Schepers, Hagenzieker, Methorst, Van Wee, & Wegman, 2014).

(2) Second, the needs and opportunities (like the need for utilitarian travel to work or the desire to travel for recreation, for example), and abilities of cyclists (depending on age, disability, confidence of cyclists, and level of fitness) can affect how and where cyclists travel.

(3) Third, and finally, rider discomfort in certain road conditions, incurred cost (in time or money), and perceived risk (Aldred & Crosweller, 2015; R. L. Sanders, 2014; R. L. Sanders, 2015) can influence travel behavior and mode choice.

- *Exposure to motor vehicles*: Travel behavior results in varying amounts and types of exposure to motor vehicles.

- *Risk of crash*: As cyclists interact with motor vehicles, they risk crashing and experiencing injuries or death. Multiple factors affecting this risk have been identified, including (1) infrastructure or road design, (2) road users and their behaviors, (3) vehicles and their design properties, which can be more or less harmful to cyclists in a crash scenario, and (4) vehicle speed.

These factors have been used to create a framework for determining the rate of bicycle injuries and fatalities (P. Schepers et al., 2014):
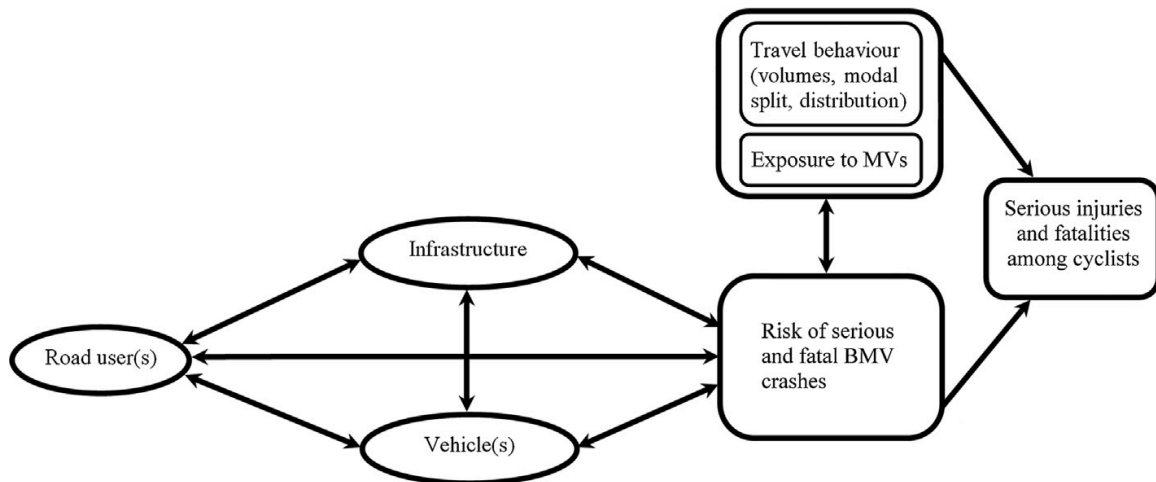
Figure 9: Conceptual framework of factors that influence cyclists' injuries and fatalities

(P. Schepers et al., 2014; Wegman et al., 2012)

In studying the effects of each of the components of this framework, Schepers et al., (2012) found the most influential factors in travel behavior and exposure to motor vehicles (the top part of the conceptual framework) is what is called "network level separation," or the degree to which cyclists are exposed to high-speed motor vehicles. Cyclists may not be able to use high-speed roads, such as highways or divided arterials, meaning that they are not exposed to high-speed vehicular traffic. This shifts a notable

number of cyclists to roads where vehicular traffic is lower. The other important factors were low cycling speed, the use of one-way bicycle paths, and intersection treatments that protect cyclists through reducing driver speeds and increasing cyclists' visibility to drivers (Richards, 2014; P. Schepers et al., 2014). This study postulates that this nonlinearity of risk could be due funneling drivers on to higher-speed roadways where there are less cyclists and less crashes in general (Wegman et al., 2012).

SiN likely comes into play in this conceptual framework of exposure to motor vehicles and risk of crashes. While the reality of the statistical relationship may have been verified through meta-analysis (Elvik, 2009; Elvik & Bjørnskau, 2017) the actual nature of SiN has been harder to determine. Factors that may strengthen or weaken the effect of SiN include (Wegman, et al. 2012; Elvik, 2009; Elvik & Bjørnskau, 2017):

- *The number of pedestrian or cyclists*: There may be a stronger effect from more riders when there are few cyclists than when there are many. This could be for a number of reasons, including marginal changes in drivers' behavior being more obvious when more cyclists are added in places with few cyclists, cyclists' feeling less confident riding alone versus riding when ride alongside others, or even better law enforcement in communities where there are larger numbers of cyclists versus smaller numbers of cyclists, leading to all mode users obeying traffic laws more carefully.

- *The number of motor vehicles as compared to the number of cyclists*: The total
  number of cyclists on the road may matter less than the ratio of cyclists to
  automobiles.

- *Skill level of pedestrians or cyclists*: Cycling crash rates vary across age groups
  (Wegman et al., 2012). In places where there is a largely elderly population, for
  example, there may be a difference in the effect of SiN compared with a place
  where there are varied ages. Elderly travelers have less crashes in cars than on
  bicycles, so in this scenario there may be more crashes (Wegman, et.al ,2012).

- *Nature of the transportation system*: Existing conditions of the transportation
  infrastructure system—including bicycle facilities—may also affect the strength of
  SiN.

Outside of understanding how these factors interact with one another, there are
other notable holes in the literature. As noted previously, the most apparent one is
addressing the temporal issues in SiN theory. It is still unclear if numbers cause safety or
if safety causes numbers (Bhatia & Wier, 2011). SiN has also not been closely studied at
regional levels. Most of the studies consider either municipal-scale or nationwide-scales
of bicycling and crash rates.

SiN has also not been studied spatially. Factors that may strengthen or weaken the effect of SiN—traffic volumes, road designs, prevalence of bicycle infrastructure, and general attitudes towards biking—likely vary spatially. For example, one municipality may have stronger bicycle and pedestrian programming than its neighboring municipality, leading towards more awareness of and respect for cyclists, while the neighboring municipality may not. It is unclear how this would affect crash rates in relation to SiN when comparing the two municipalities.

To date, there is not a statistical model that describes the spatial nature of SiN. It is conceivable, however, that spatial analyses may help generate hypotheses about the nature of SiN over a given area. Further studies may even generate a spatial model that would predict how the SiN effect may be used to predict changes in crash counts or identify areas of concern. All SiN studies thus far have used linear regression to determine correlation. One of the underlying assumptions in a regression analysis, however, is that there is no spatial autocorrelation. In determining the spatial relationship among ridership and crash data sets, it is important to determine if there are real spatial effects that would alter existing assumptions underlying the SiN effect.

It is likely that road segments that are closer to one another are more similar than those that are further apart, but the extent to which they are similar is unknown. Multiple factors may contribute to potential spatial autocorrelation, but the most obvious and important is the fact that the segments do not exist alone in space but rather are a part of an entire network. Bicycle mobility and safety are related to how

much road network is in a given space. For example, a square mile of a road network that has 10 miles of road segments (total) inherently provides less connectivity (and ultimately mobility) than a square mile that has 25 total miles of road. This cannot be controlled on an individual segment level, but instead must take into to consideration spatial units within the entire network.

Similarly, nearly all SiN analyses in the literature have been conducted at either large scales (crashes per country or municipality) or on a theoretical basis (agent-based models). Key studies are summarized in the following table. There is little understanding of how this relationship would change by studying bicycle ridership data on the segment level from cities and corresponding crash rates. Testing SiN in this way can ground-truth underlying assumptions, as well as test the assumption that SiN is best characterized by exponential relationships.

Summarized generally, the main gaps in the understanding of safety in numbers are three-fold. First, studies have not been disaggregated to smaller units of analysis on a large scale (like individual streets within an entire road network). Second, because of SiN has not been studied with small units, there has not been a way to control for road conditions that also effect bicycle crash rates. And finally, because of the previous two gaps in knowledge, it is still unclear how the SiN effect interacts with other important factors to affect bicycle safety as a whole. The research presented in this study contributes to the literature by addressing these issues. The following section describes the research methods and data used to address these major gaps in understanding.

Table 1: Summary of key SiN studies

Title	Author(s)	Year	Units of Analysis	Exposure Measure	Injury Measure
"Safety in numbers: more walkers and bicyclists, safer walking and biking"	P. Jacobsen	2003	(1) Walking/biking in 68 California cities. (2) Walking/biking in 47 Danish towns (3) Bicycling in 14 European Countries (4) Walking in 8 European countries (5) United Kingdom (6) Netherlands	(1) Portion journey to work trips on foot/bike (2) km walked/biked a day (3) km biked/capita/day (4) trips on foot/capita/day (5 – (6) billion km biked annually	(1) – (2) Injuries per capita (3) – (4) Fatalities per capita (5) – (6) Fatalities
"Estimating the safety benefit of separated cycling infrastructure adjusted for behavioral adaptation among drivers; an application of agent-based modelling"	J. Thompson, J. Wijnands, G. Savino, B. Lawrence	2013	Virtual road network and individual streets/bicycle facilities	Virtual cyclists interacting with virtual cars	Collisions of virtual cyclists and cars
"The Dutch road to a high level of cycling safety"	P. Schepers, D. Twisk, E. Fishman, A. Fyhri, A. jenson	2015	(1) Netherlands (2) European countries	(1), (2) Billions of km	(1) Fatalities per billion bicycle km, (2) Road Fatalities per 100,000 population

2.   RESEARCH DESIGN AND DATA SOURCES

2.1   Design Overview

To better inform practice, to contribute to the overall understanding of bicycling safety in cities, and to further the academic literature on SiN, it is important to more fully understand SIN as it applies to different scales of analysis. The following research questions have been developed to address some of the gaps in understanding within the literature:

Research Questions:

Is the "safety in numbers" phenomenon reflected in analyses of individual streets within a road network? If so: (1) how do crash rates with exposure when other variables are controlled, and (2) how does SiN vary under different road conditions?

As previously stated, SiN research has not been conducted with small units of analysis, with the exception of a the previously mentioned agent-based models. It is therefore unclear whether or not the relationship between the number of crashes and the number of cyclists will be the same when investigating individual road segments within a given area as it is when considering ridership and crashes when aggregated to a city or country level. Conducting a SiN analysis on a segment scale allows for the

consideration of other road condition variables that may also affect cyclist safety like congestion, speed limits, road functional class, presence of bicycle infrastructure, etc. This study attempts to fill this gap in understanding by conducting an analysis on an individual road segment level. The following research design has been developed to analyze SiN in this way for a subsection of the ten-county Denver-Arora metropolitan (governed by the Denver Region Council of Governments, or DRCOG).

Results from researching bicycle SiN at the unit of analysis of the individual street segment will inform existing literature in several ways. First, conducting the research on individual streets will identify whether or not the large-scale phenomenon of SiN is actually the same across all units of analyses. While studies of this type have been done with theoretical models (see Thompson et al., 2017), research of this type has not been done with data from an actual municipality. Furthering this understanding with data reflecting reality may begin to inform us about what the SiN phenomenon actually means for riders' experiences in different conditions.

Second, testing the effect of the number of trips in a system on the likelihood of a crash while controlling for other factors that are known to affect bicycle safety will shed light on how the SiN phenomenon really affects cyclist safety. Specifically, my research will help answer the question "Do numbers cause safety, or does safety cause numbers?" This new understanding will inform the framework previously discussed by prioritizing what may matter most for protecting cyclists. Understanding this component of SiN could potentially inform bicycle facility design as well as the existing literature.

This research approach also allows us to better understand one of the most important issues in the literature—whether "numbers" causes safety or safety causes "numbers." Results from this research shed light on whether lower crash rates are predicted by the number of trips even when other factors that are known to affect cyclists' safety are controlled, such as road functional class, presence of bicycle infrastructure, speed limit, etc.

The following design is used in this research:

- **Part 1:** Determine if road segments in the Denver Metropolitan Region experience a SiN effect (a correlation between the number of trips made by cyclists on a given road segment and bicycle crashes per rider on that segment).

- **Part 2:** Characterize the SiN effect mathematically, investigating whether linear or exponential models best describe the phenomenon.

- **Part 3:** Determine how crash rates vary based on trips when other variables are controlled

- **Part 4:** Measure the variation of the SiN effect under different road conditions.

The following sections further explain the proposed design, define the study areas, describe the data sources, and outline the limits of the research methods.

2.2   Study area summary

The study area selected for research is a 10-county metropolitan area, Denver,
Colorado. The ten counties included in the study area are Arapahoe, Boulder,
Broomfield, Clear Creek, Denver, Weld, Adams, Jefferson, Douglas, and Gilpin Counties
(shown in Figure 10). The following table provides several summary statistics that help
characterize the area. To summarize the table, the selected study area is large,
consisting of 3 million people and more than 1.2 million households. This is also a
wealthier area than the rest of Colorado.

Table 2: Study area summary

Denver-Arora Metropolitan Statistical Area Characteristics	
Population Total	3,303,417
Number of Households	1,262,786
Average Household Income in MSA	$92,956
Average Household Income in CO	$68,813
Average Home Value	$294,850

Source: US Census Data[1]

---

[1] Obtained from *American Fact Finder*
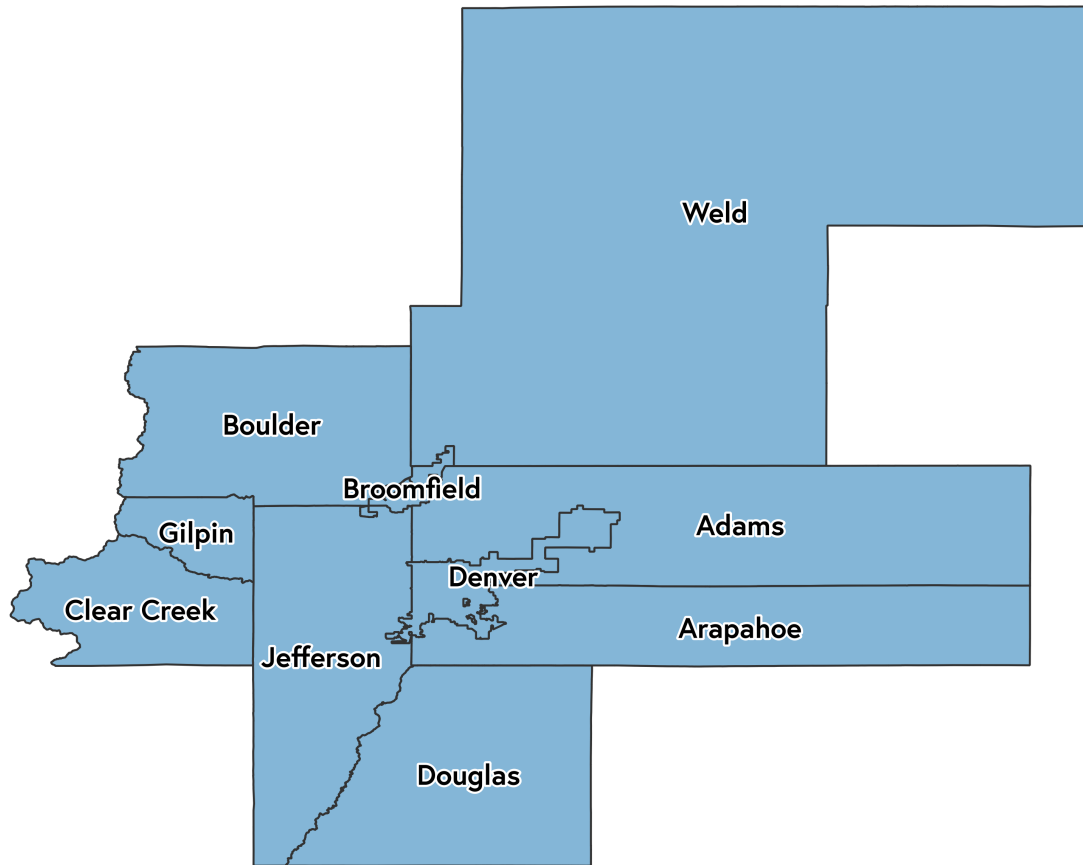
Figure 10: Denver Region (Counties)

Source: Birmingham Business Alliance, 2017 [1]

Many metro areas across the US have had an increase, albeit a small one, in the number of people commuting to work via bicycle in the last decade. This includes Denver. In 2014, a little over 2% of commuters in Denver traveled by bicycle, which was 30% above the national average at the time (Hyer, 2014). This is due in part to very bicycle-friendly topography and climate (Denver Region Council of Governments, 2017)

and consistent investment in bicycle infrastructure for over a decade (Douglas, 2017).

Around 1.4 million trips each day are made by foot or bicycle each day in the DRCOG

jurisdiction, and about 162,000 of them are on a bicycle (Denver Region Council of

Governments, 2017). Bicycle commuting in Denver has risen 32% over the previous

decade (Denver Region Council of Governments, 2017).

Denver is suitable for continued growth in bicycle ridership. However, bicycle

commuting has been dropping slightly (in Denver and around the US) in more recent

years, likely due to the decrease in gas prices starting in 2016. Higher levels of driving,

rising household incomes, and, arguably, declining road safety have been identified as

other reasons for less biking (Anderson, 2017). Despite these general trends downward,

parts of the study area, including downtown Denver, have seen some upticks in

ridership. In 2017, around 35,000 commuted by bike on National Bike to Work day (a

notable increase from the previous years), which has created momentum for expanding

the existing bicycle culture to more users (Douglas, 2017).

In terms of traffic fatalities, from 2005 – 2016 there were 9632 reported bicycle

crashes (excluding data from 2009, for reasons explained below). A majority of the

reported crashes resulted in injuries, and nearly three-quarters of these incidents

happened at intersections (Denver Region Council of Governments, 2017). To combat

these issues, the DRCOG has taken steps towards implementing safety initiatives,

including Vision Zero Planning and implementing new bicycle infrastructure (Denver

Region Council of Governments, 2017).

This study area was chosen for several reasons. First, data required for these analysis is readily available through partnerships with the Toole Design Group, a national leader in bicycle and pedestrian planning and engineering, and with the Denver Regional Council of Governments. Second, this region has a strong bicycle culture with established record keeping practices for ridership counting and crash reporting. The Denver Region Council of Governments (DRCOG) has kept detailed crash data that is publicly available dating back to 2005. The DRCOG also has a large database of publicly available GIS data that has been meticulously catalogued for over a decade. Presently, the Denver Metro area, like the rest of the US, would likely fall into the ridership category described by Wegman et al. (2012), in which there is a relatively high median number of riders (compared to surrounding areas) but still high crash and fatality rates; in these scenarios, it is not well understood how more exposure affects bicycle safety. If this is the case, this study may help to articulate the nature of SiN in these ambiguous cases.

2.3   Data Sources

Several types of data are used in this study:

- *Strava Data for the entire DRCOG jurisdiction:* Strava is a self-described social media network for cyclists and runners ("Colorado Strava MetroTraining," 2017). It

is available to users through web-based or phone applications. The purpose of

Strava is to foster community among the running and cycling populations around

the world. The application allows athletes to log the location of their runs and

bicycling trips, and it then provides performance data for each activity (such as

average speed, total distance travelled, and descriptive statistics about other

athletes who have completed the same route). A part of the Strava mission is

realized through a department of the company called Strava Metro. The mission of

Strava Metro is to create high quality spatial data to make active transportation

more viable in cities across the world. This program provides aggregated, spatially-

referenced datasets of cycling information from users' data to local planning

organizations to better inform bicycle and pedestrian planning practices. The

resulting data set has very detailed ridership counts (including the number of

unique riders and total trips) per segment of road for the entire transportation

system (all roads) within a state of city over a given time period. The DRCOG study

area has purchased Strava datasets for the 2016 calendar year. In the image shown

below, the lines collectively represent the entire transportation system in a subset

of my study region. Each line contains many attributes, including number of trips

per segment, number of unique riders per segment, etc. For my entire study area,

there are over 536,000 segments.

o   It should be noted that Strava data may not be representative of all cyclists

because it does not represent all riders for two reasons. First, the data shows

number of trips per segment for those who use the app, but this obviously does

not capture all riders' trips. Second, it represents routes of riders who may be

enthusiasts an stronger cyclists than the average rider. This is discussed in

detail in the limitations section of this document (Section 4.3).
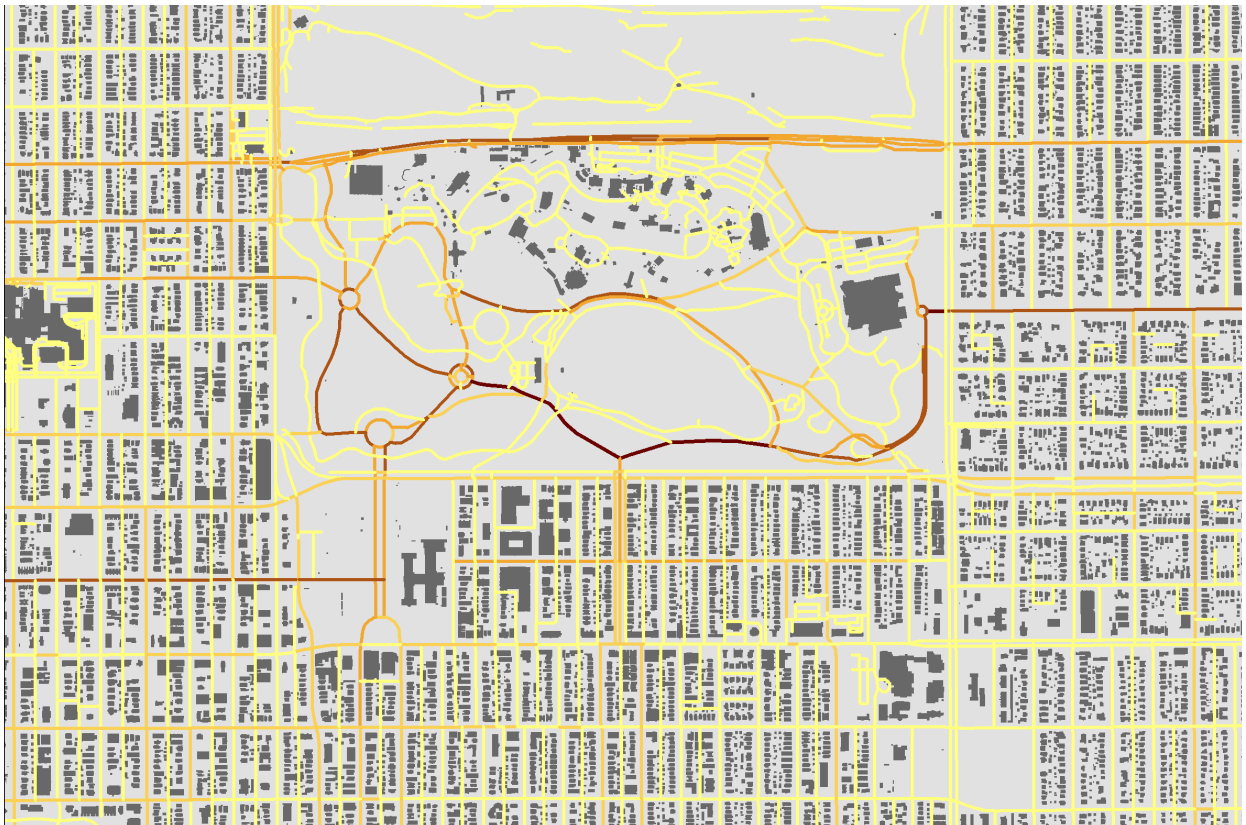
Figure 11: Strava data in study area--downtown Denver zoom

Note: Lighter segments represent streets with lower numbers of trips, and darker segments

represent streets with higher numbers of trips

- *DRCOG Crash Data:* DRCOG has collected detailed, geospatially-referenced crash data since 2005. These datasets are publically available through the DRCOG Regional Data Catalog, the goal of which is to encourage data-driven community and municipal planning within its jurisdiction (Denver Region Council of Governments, 2017). The data used in this study are all bicycle crashes and fatalities from 2005 to 2015. Data from 2016 is not available as it has not yet been processed by the DRCOG. Upon in-depth review of the data, crashes from 2009 are not included in the analysis, as there are many incorrectly georeferenced crashes, compromising the validity of the entire dataset for that year.

- *American Fact Finder:* Demographic data from the US Census Bureau has been obtained at the block group (2,145 groups total) and county (10 counties total) levels. Specifically, median household income at the block group level has been added to models for each segment within the study area.

- *Other shapefiles and data from DRCOG:* Other shapefiles from each of the study areas were obtained from the DRCOG, including municipal boundaries, functional classification for each road, volume to capacity ratios on roads, bicycle infrastructure, and speed limits.

- *2017 TIGER/Line shapefiles*: The US Census Bureau provides shapefiles for GIS

  software that show all legal boundaries. These files are to be paired with

  demographic data from the American Fact Finder to identify block group

  boundaries of the median household income values.

2.4   Geoprocessing Methods

As noted above, multiple different geospatial datasets were in these analyses. In

order to accurately combine each of these datasets into a single dataset, several

geoprocessing methods were executed using ESRI ArcMap, a geographical information

software (GIS) program. The following sections summarize the geoprocessing tools used

to create the final dataset.

Clipping

The original Strava Metro dataset contained all road segments for the entire

state of Colorado. Since the study area only contained the 10 counties under the DRCOG

jurisdiction, the study area was used to clip the segment dataset down to only the

segments of interest. Two counties (Clear Creek and Gilpin Counties) were cut from the

analysis because they did not contain any reported bicycle accidents for the 9 years

(2009 data is excluded) of crash data.

Spatial Joins

The spatial join functionality in ArcMap creates joins between datasets based on a common spatial reference point (as opposed to an attribute field commonality). Because most of the datasets were independent of one another entirely (meaning that they had no common attribute from which a table join could connect their attributes), they had to be joined based on their spatial location. The output of the spatial join function is a single dataset and including the road network with the attributes from both datasets contained within it. The following datasets were combined using spatial joins:

- *Strava Metro (clipped to the study area)* – The Strava dataset was treated as the "base" dataset to which the rest of the layers were joined.

- *Functional Class and Volume to Capacity Ratio* – The DRCOG and Colorado Department of Transportation (CODOT) each provided one shapefile that contained functional classification and V/C ratio (the number of cars using each segment divided by the traffic capacity of that segment). These files were joined spatially to the "base" Strava Metro shapefile.

- *Bicycle infrastructure*—The DRCOG also provided shapefiles that contained the location of bicycle infrastructure. By joining this file to the "base" Strava dataset, each segment with bicycle infrastructure was identified. "Sharrows," or road paint designed to encourage motorists to share roads with cyclists, were not included because they do not add notable safety benefit for cyclists compared

to other forms of bicycle infrastructure. Standard bike lanes and physically

separated bicycle facilities were differentiated for analysis.

- *County*–The county shapefile was joined spatially to the Strava Metro base file.

  The resulting dataset contained each segment categorized by the county in

  which it is located. Segments that stretched across county boundaries were split

  on the boundary to create two separate segments.

- *Median Household Income on the Block Group Level* – Median household

  income values were assigned to the base dataset on the block group level in the

  same way as the county shapefile.

Speed limits were included in the base dataset and therefore did not need to use

the spatial join function.

Appending

Spatial joining is a powerful tool, but it can only operate when there are perfect

overlaps of datasets spatially. In particular, this did not apply with the geospatially

referenced crash points and the road segments. Each crash was referenced as a point to

a specific geospatial coordinate (latitude and longitude points) with reasonable

accuracy, but they did not always overlap the segments exactly. To deal with this

discrepancy, the points were overlaid to the correct segment using the append

mechanism in ESRI ArcMap. The append function triangulated each point to its three

closest segments. The crash point was then appended to segment with the minimum

distance from the crash point. In order to count the number of crashes per segment, a

variation of the spatial join tool was used in which the resulting shapefile contained a

sum of all the crash points along each individual segment.

2.5   Research Design Concept—Step 1: Determine if the Denver Metropolitan Region

experiences a SiN effect

To determine whether or not there is a SiN effect in the study area, this report

uses the same non-linear characterization used in a majority of the literature (similar to

the relationship shown in Figure 3). This analysis is conducted on the individual

segment scale, utilizing all 536,519 segments. In the literature, the following model has

been used to investigate SiN:

$$I = aE^b$$

With the data sets used in this research, the key variables would be

operationalized as shown in the table below.

Table 3: Operationalization of Variables

Variable	Conceptualization	Operationalization
I = Incident Measure	Measurement of number of incidents between cyclists and automobiles	Number of crashes with a bicycle involvement per segment
E = Exposure Measure	Opportunities for bicycle-motorist interaction	Number of trips (via Strava Metro Data) per segment
I/E = Relative Risk	Relative probability for a crash to occur	Number of crashes with bicycle involvement per 10,000 trips*

For analyses, the log-linear form of I/E is used.

However, these approaches are not appropriate here. Tobit models use a latent variable that may cause the resulting crash rate predictions to be below zero. This is not easily interpreted for crash rates since they cannot be negative. Poisson distributions are also not appropriate here because they consider count data, whereas this research considers crash *rates*, not individual crashes (Ma, Yan, & Weng, 2015). To deal with this issue of the highly right-skewed dataset bounded at zero, a two-step exponential hurdle model was used to model the relationship between the number of trips per segment and the number of crashes on that segment.

Cragg's two-part hurdle model considers the "hurdle" between zero and non-zero outcomes by modeling the zero outcomes and non-zero outcomes separately (Cragg, 1971; Ma, Yan, & Weng, 2015). When applied to crash rates in this scenario, it

utilizes a probit model to analyze the relationship between the exposure measure and whether or not a crash occurs on a segment at all for the given time span. For segments that have positive crash rates (n = 5,534), a second, conditional model examines how the exposure measure affects the crash rate within the same time frame using an exponential non-linear regression analysis. (The log of the dependent variable, crashes/10,000 trips, is taken because crash frequency has rightward skew). Thus, the results show whether the number of trips increases (or decreases) the likelihood of there being at least one crash on the segment, and whether it increases (or decreases) the number of crashes per 10,000 trips, conditional on there being a crash on the segment. To render the results more interpretable, predictions for unconditional crashes per 10,000 trips can be generated using the output from each of the two models. This research showed predictions for crashes per 10,000 trips at different levels of ridership while holding all control variables at their means.

2.6   Research Design Concept—Step 2: Characterize mathematical relationship of the SiN effect

The second step of the research design involved determining whether SiN has the same exponential relationship with crashes/trip as has been identified in the literature. This step in the research process investigates whether a linear model would better fit the relationship, as judged by the model's pseudo r-squared. As has been

shown in the literature, the prevailing assumption is that SiN is best characterized by a non-linear, exponentially decreasing relationship (crashes per rider declines with more trips, but at a decelerating rate as the number of trips increases). However, it is crucial to verify that this is replicated at a smaller unit of analysis.

2.7   Research Design Concept —Step 3: Crash rates in response to exposure when other variables are controlled

The ultimate goal of the third portion of the research design is to identify how certain independent variables—including the exposure measure— affect crash rates. This was done through utilizing a hurdle model as described above while adding other independent variables which may be expected to affect bicycle safety and crash rates to the analysis. This is done to help identify whether an apparent SiN effect is really due to large numbers of trips being attracted to areas where it is safer to bicycle. As identified in the literature, the following independent variables were included in the hurdle analysis:

- *Number of trips:* This dependent variable will be used to determine if the number of trips per segment is a statistically significant factor in predicting crashes rates. This will be used as an indicator for the SiN effect; if higher ridership values covary significantly with lower injury intensity scores, this would support the SiN effect's

existence. It should be noted here that Strava Metro data is used to determine the number of trips. While Strava data is not representative of the entire cyclist population on each street, it is assumed that trips reported through the Strava data is a proportional representation of all trips on each segment. I had hypothesized that there would be a negative relationship between the number of trips and crash rates.

- *Presence of bicycle infrastructure*: The presence of any type of bicycle infrastructure (including bike lanes, greenways, trails, or any physically separated infrastructure), except for shared lane markings, or "sharrows," is included in in this analysis. One dummy variable indicates the presence of bike lanes, and another indicates bicycle facilities that are physically separated from motorized traffic (including trails/greenways, on-street separated bike lanes, and side paths). If the presence of bicycle infrastructure is associated with fewer crashes per rider, as I had hypothesized, this may inform how infrastructure and the SiN effect relate to one another temporally—there may be fewer crashes/rider in inherently safer areas not due to SiN but because safer areas attract more riders and more trips. The inclusion of this variable, and my other control variables, represent a significant step towards more fully understanding safety in numbers as a whole.

Again, I predicted that the number of crashes per 10,000 trips decreases in the presence of bicycle infrastructure (for both independent variables).

- *Volume to capacity (V/C) ratio*: More congested areas may be more dangerous for cyclists, so this may be a more powerful indicator of crash rates than ridership or presence of infrastructure. In this analysis, exact V/C ratios were available for some segments, and binary measures of congestion (where "congested" segments have V/C ratios greater than 1 and "not congested" segments have V/C ratios less than 1) were available for the remaining segments. For consistency, segments were all reduced to binary variables ("congested" or "not congested") and coded as a dummy variable. It was expected that V/C ratio will be positively associated with crash rates; more congested roads may be more dangerous for cyclists.

- *Speed limits*: Speed limits on given roads are likely to influence cycling safety. Higher speeds may contribute to crash and fatality rates, so it was assumed that there will be a positive relationship between these variables.

- *Median household income:* Lower income areas typically have higher rates of active transportation injuries and fatalities than other areas even when other variables

are controlled. This could be due to less suitable road conditions in low income

areas, or because these areas see more dangers biking/driving habits. It was

hypothesized that areas with higher median household incomes have lower crash

rates than comparatively lower incomes areas.

- *County:* In order to control for possible cultural differences across cycling

  populations (for example, riders in Boulder might ride more safely than riders in

  Denver), each of the counties was coded, and dummy variables for the counties

  were assigned to each segment. Controlling for counties may also partially control

  for other factors. For example, road network density, which is not explicitly

  measured here, varies significantly in some counties; Denver County has a very

  dense road network and intersection density, but Weld County is much larger

  geographically and has a far less dense network of roads throughout.

- *Segment length:* The Strava Metro dataset breaks up each street into segments,

  where a segment is the length between intersections. Segment length is included in

  the analysis as a control variable. I expected there would be more crashes on

  longer segments. Note that segment length may also be a loose proxy for road

  network density; we would expect shorter segments in areas with denser road

  networks.

2.8   Research Design Concept —Step 4: Variation of the SiN effect under different

road conditions.

The previous step determines if the number of trips is a significant predictor of (1)

whether or not there will be a positive crash rate and (2) the number of crashes/10,000

trips, conditional that there is at least a single crash per segment. The final step of this

research considers the other independent variables' effects on predicted crash rates. To

do this, I created predictions (extrapolated from the hurdle model) for each of my

independent variables (excluding segment length, median household income, and

county code as they are not considered crash predictors but control variables). I made

the following prediction for the number of crashes at given the number of trips

considering:

- Bike lanes (present or not present)

- Physically separated bicycle facilities (present or not present)

- Speed limit (for 25, 35, 50, 60, and 70 mph)

- Volume to capacity ratio (for congested or not congested)

- Functional class (for local roads, collectors, minor arterials, major arterials, and

  interstates)

Results from this step will quantify how the SiN effect varies under different road

conditions.

3.   RESULTS

Results from each step of the research design are shown below. Interpretation

and discussion of these results is contained in the following sections.

3.1   Step 1 Results

A histogram of the number of segments with at least one crash is shown below

in Figure 12 (n = 5,534). Even without the zero values, the data were highly right-

skewed, as is expected with count data. The figure below shows a histogram of the

distribution of the number of trips on segments, excluding those with zero trips. It was

also highly right-skewed; this somewhat fits a typical count distribution, but this

histogram is extreme in its excess of zeros when zeroes were included (this is over

500,000 segments). This is clear in the second histogram, which includes those

segments.
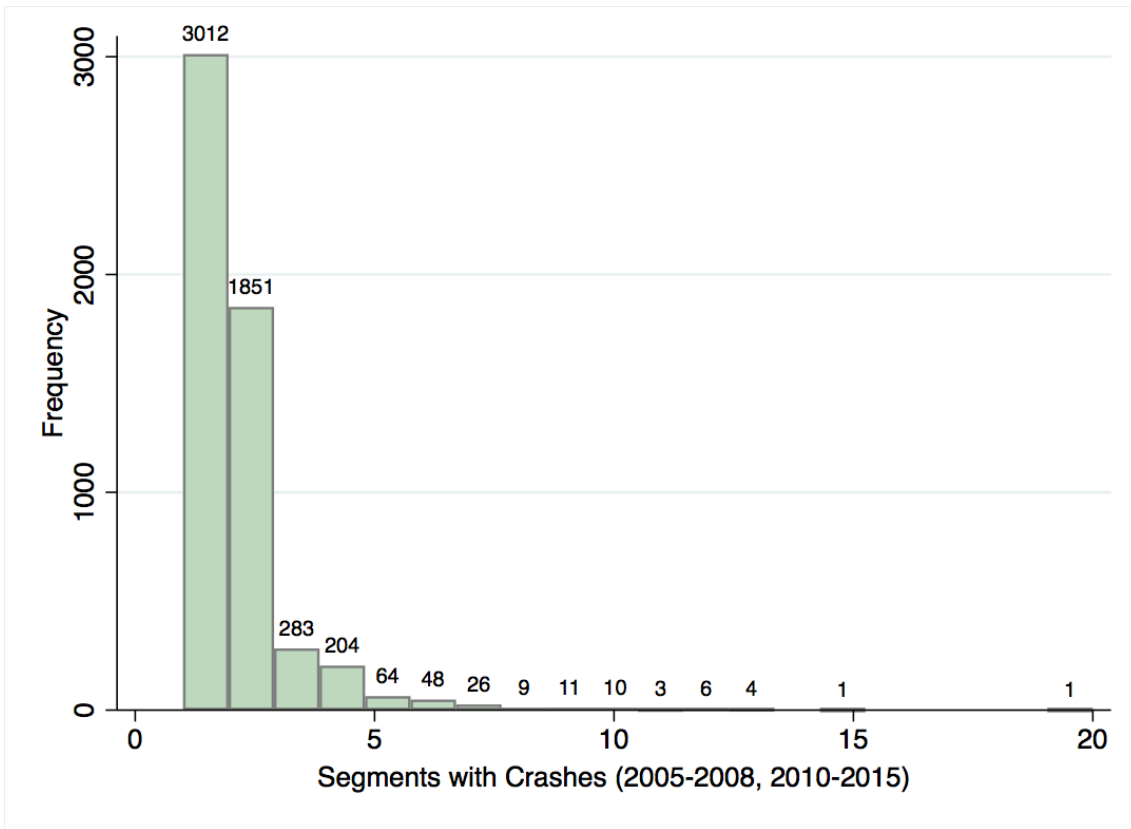
Figure 12: Histogram of segments with at least one crash

Table 4 below shows the results of the first step of the research design. The exponential hurdle model predicting crash rates per 10,000 trips based on the number of trips in the segment had a pseudo r-squared value of 0.0236. Note that this is McFadden's pseudo r-squared (explained in the equation below), which is a measure of both variability and of the goodness of fit of the model (UCLA: Statistical Consulting Group, 2011). Like r-squared values, pseudo r-squared values range between zero and one, and higher values indicate a better fit (UCLA: Statistical Consulting Group, 2011).

$$pseudo - R^2 = 1 - \frac{\ln\ L(M_{full})}{\ln\ L(M_{intercept})}$$

$$where$$
$$L = estimated\ likelihood$$
$$M_{full} = model\ with\ predictors$$
$$\boldsymbol{M_{intercept} = model\ without\ predictors}$$

Equation 4: McFadden's pseudo r-squared

For both the probit/selection model (which considers whether or not there is a positive crash rate per segment) and the conditional model (which considers effects on crashes per 10,000 trips only on segments with positive crash rates), results showed that the number of trips as an independent variable was significant. Interestingly, the variables' magnitude in both models is very high, but in opposite directions. Due to the low pseudo r-squared, this model explains very little of the variability seen in the data. Figure 13 below shows the predicted marginal effects with a 95% confidence interval.

Table 4: Hurdle model for number of trips predicting crashes per 10,000 trips

Pseudo R^2 = 0.0236

Conditional Model (exponential) Dependent variable = crashes/10,00 trips	Coef.	Std. Err.	z	P>z
Total Trips	-3.47E-4	6.57E-6	-52.8	**<0.001**
Selection (Probit) Model Dependent variable = crashes/10,00 trips				
Total Trips	3.51E-5	0.000	25.6	**<0.001**

Both the probit and the conditional models showed that the number of trips is a highly statically significant at a 95% confidence interval. For the probit model, this means that the number of trips can fairly reliably predict whether or not there will be a positive crash rate on each segment. This makes sense; it is more likely that there will be at least one crash on segments with many trips. Since the conditional model only considers crash-rate-positive segments, these results mean that the number of trips is also a reliable predictor of crash rates of segments (assuming there is at least one crash on that segment). It suggests that crashes/rider decreases with the number of trips, as the SiN hypothesis suggests.

Again, the pseudo r-squared value in this analysis is low (0.0236). This is not unexpected. A univariate analysis is not likely to explain much of the variation that we see in crash rates. It is both intuitive and proven in previous research that other factors

besides "numbers" are strong predictors of crash rates, so it makes sense that only

considering one factor that may influence crashes is a poor predictor of variability.

Because hurdle models contain probit analyses, the coefficients cannot be easily

used for direct interpretation. A more useful approach to understanding the effect size

of the independent variable in both models is considering the predictions for

unconditional crash rates generate using marginal effects. In this scenario, the marginal

effects are generated by multiplying the probability of a crash at specified intervals of

the independent of interest (number of trips) by the predicted number of crashes/rider

conditional on there being at least one crash. This generates predictions of how many

crashes will occur with increasing number of trips. Predictions of crashes per 10,000

trips as trips increase from 0 to 5,000 are shown below in Figure 13.

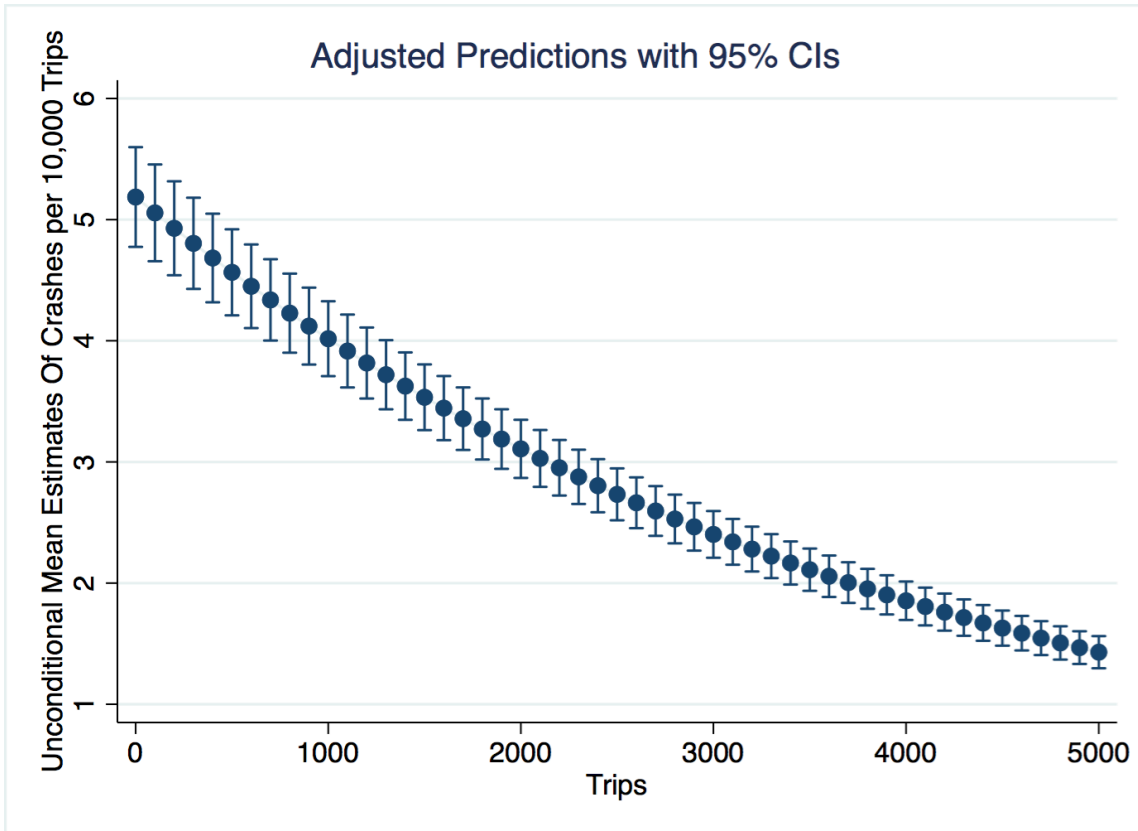**Figure 13: Safety in numbers marginal effects & predictions for unconditional crashes**

The predictions in Figure 13 show that estimated number of crashes decreases from approximately five crashes per 10,000 trips to less than two crashes per 10,000 trips as trips increase from 0 to 5,000. This fits the hypothesized relationship and mirrors what the literature has shown in analyses with larger units of analysis.

3.2 Step 2 Results

To ensure that an exponential relationship is the best fit to describe the SiN effect at this unit of analysis, results from both a normal linear regression analysis and a linearized exponential regression analysis (taking the natural log of crashes/rider) are shown below in Table 5.

Table 5: Regression analyses comparisons

Linear Regression		*Log-Linear Regression*	
R-squared =	0.0321	R-squared =	0.407
Adjusted R-squared =	0.0303	Adjusted R-squared =	0.406

The r-squared value for the linear regression and the exponential regression were 0.0321 and 0.4071, respectively. While r-squared values do not paint the entire picture for "goodness of fit," they do provide insight into how much variation in the dependent variable is explained by the independent variable. Because the r-squared values are markedly different—by nearly an entire order of magnitude—it is easier to discern what type of model is the better fit. These results support the selection of a log-linear approach to the hurdle models, and it also supports that the relationship is best characterized by an exponential relationship.

3.3 Step 3 Results

For the second step of the research design, other variables shown to affect bicycle crash rates (bike lanes, functional class, median household income, physically separated bicycle infrastructure, segment length, and V/C ratio) were considered in tandem with number of trips. These datasets' distributions are shown in the figures below.
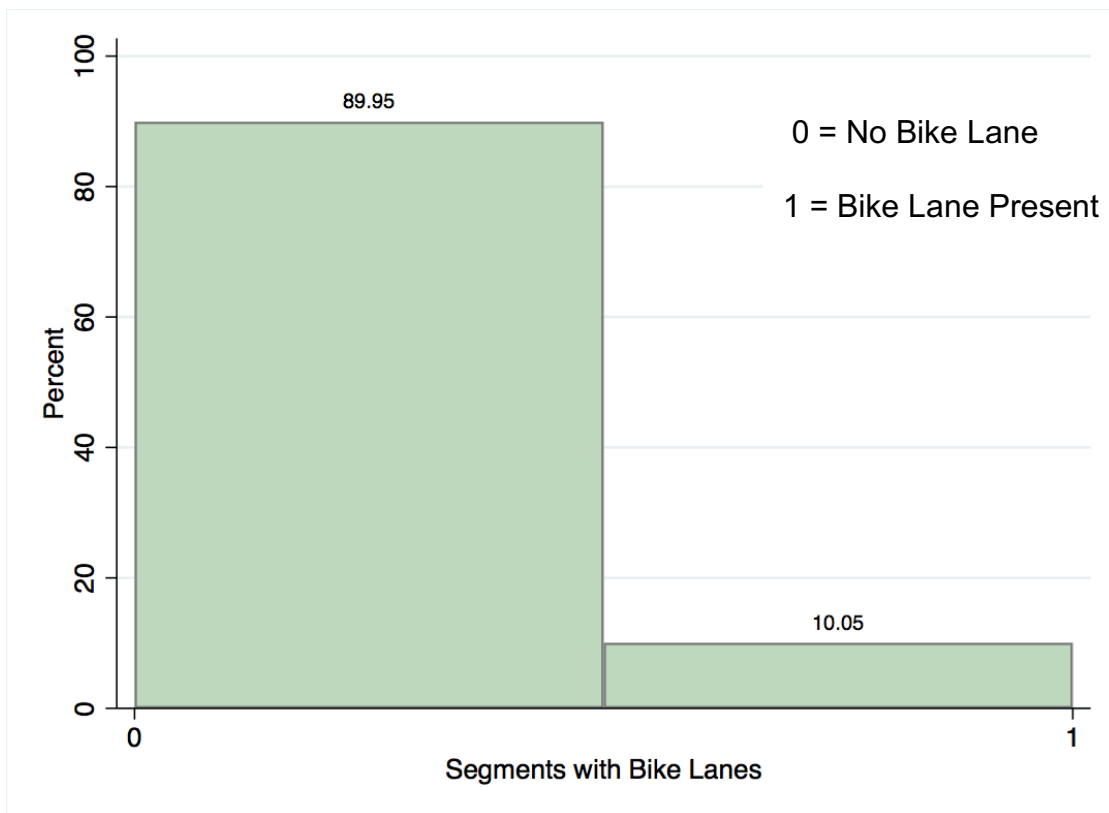
Figure 14: Distribution of segments with bike lanes
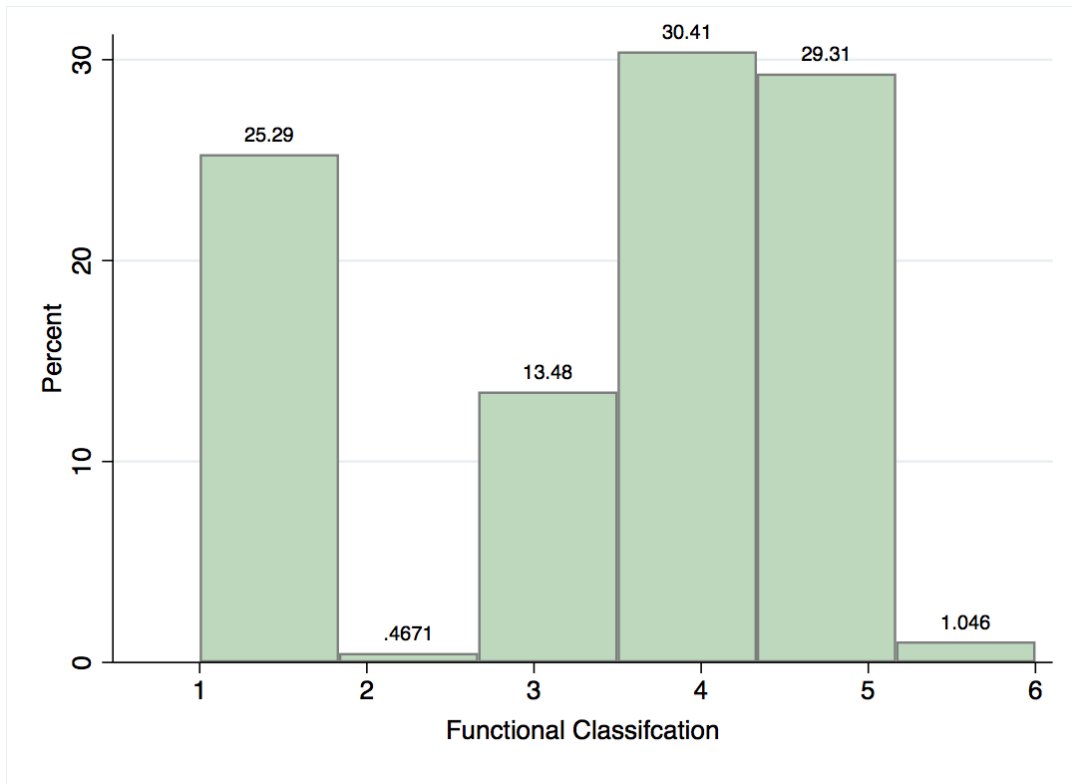
**Functional
Classification**
1 = Local Roads
2 = Minor Collectors
3 = Major Collectors
4 = Minor Arterials
5 = Major Arterials
6 = Interstate

Figure 15: Distribution of Functional Classification

Note: In analyses functional classes (FC) 2 and 3 were combined and renamed FC 3: Collectors

Data from American Fact Finder, 2016, In US Dollars
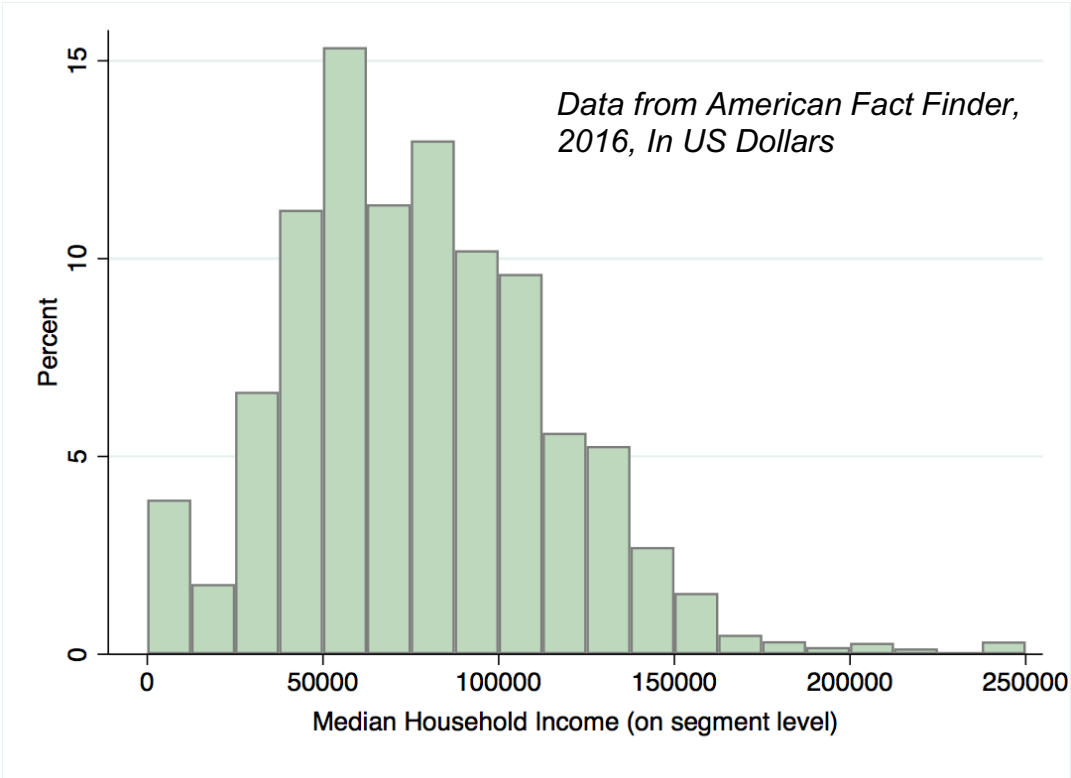
Figure 16: Median household income distribution
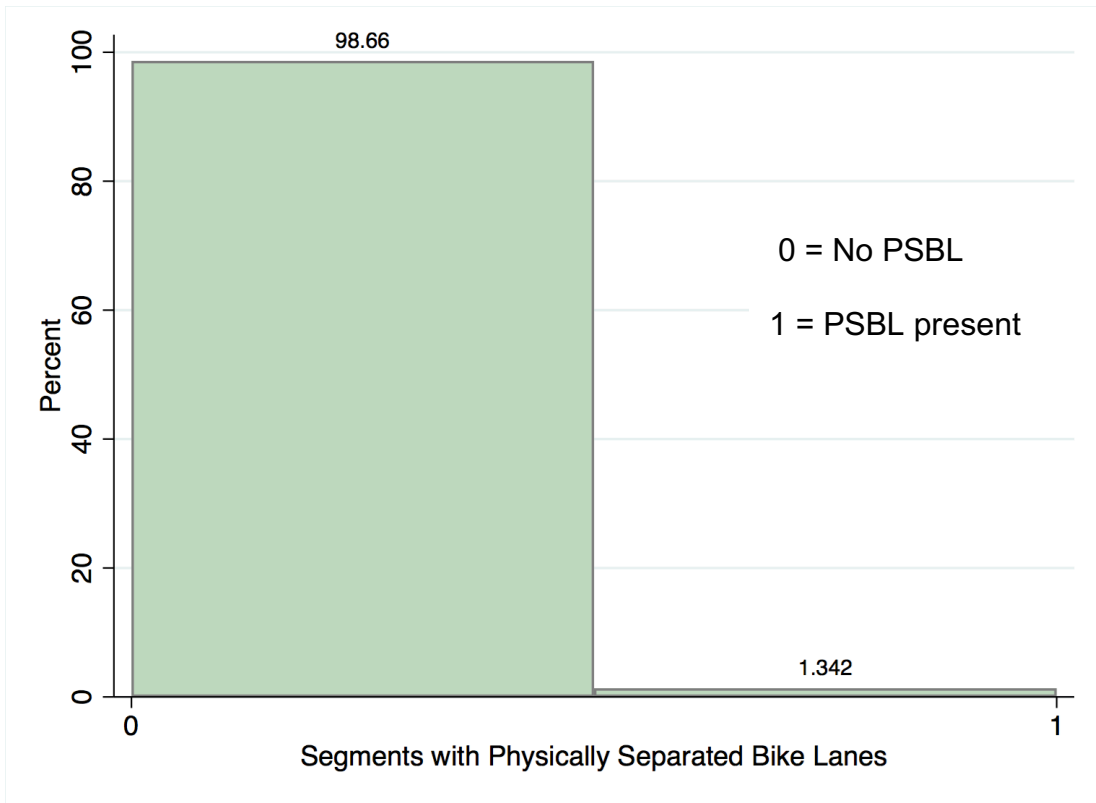
Figure 17: Distribution of segments with physically separated bike lanes (PSBL)
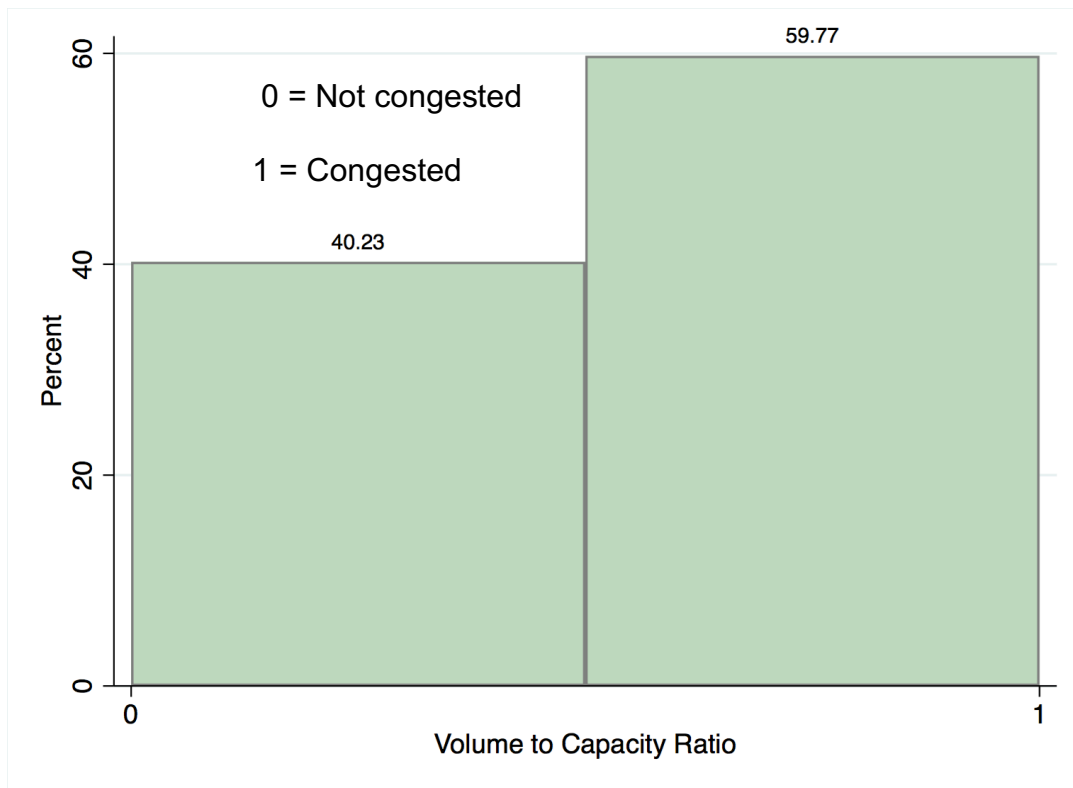
Figure 18: Distribution of congested segments

Most of the distributions of the independent variables are not surprising. Figure 14 shows that a large majority of the segments do not have bike lanes. This is to be expected in such a large area in the United States. The distribution of functional class is a little surprising, though. It is expected that there would be more "local" roads in the dataset, and there are a very small number of roads that are classified as "minor collectors." For the analysis, "minor collectors" and "major collectors" are joined into a single functional class, "collectors," due to the small number of minor collectors.

The median household income (Figure 16) is also, not surprisingly, reasonably normally distributed but with some rightward skew. Also unsurprisingly, there are even fewer separated bicycle facilities than bike lanes. Design standards for these types of bicycle facilities are new and, in some ways, only just now being formalized (*Urban Bicycle Design Guide,* 2011; *Separated Bike Lane Planning and Design Guide, 2015)* and the process of building bicycle infrastructure can be slow in the US. The distribution of volume to capacity ratio on each segment is also not surprising. Given the aging infrastructure crisis in the United States ("Making the Grade," 2017), it is somewhat impressive that only 60% of segments in this study area are considered "congested."

Table 6 and Table 7 below show the results of the second step of the research design. This two-part analysis showed (1) what factors predict whether or not there will be a positive crash rate and (2) what factors are significant predictors of crash rates (assuming crash rates are above zero). Factors considered include the number of trips (serving as the variable testing the "safety in numbers" effect), functional class of the segment, whether or not the segment has a bike lane or a physically protected bicycle facility (like a trail or side path), the median household income surrounding the segment (on a block group level), speed limit along the segment, and the volume capacity ratio. This analysis also controlled for segment length and county. It should be noted that while the study area consisted of all 10 counties, there were only crashes in 8 of the 10 counties. Statistically significant factors at a 95% confidence level are in black text with the P>z value in bold.

Table 6: Conditional hurdle model for factors affecting crashes per 10,000 trips

Conditional Model: Dependent variable = crashes/10,00 trips	Coef.	Std. Err.	z	P>z
Pseudo R^2 = 0.0788				
Number of Trips	-0.0003	6.55 E-5	-49.57	**<0.001**
3 - Collectors	-0.174	0.0720	-2.41	0.0160
4 - Minor Arterials	-0.729	0.0633	-11.5	**<0.001**
5 - Major Arterials	-0.480	0.0662	-7.25	**<0.001**
6 - Interstate	0.0213	0.392	0.050	0.957
Bike Lane	-0.846	0.0575	-14.7	**<0.001**
Physically Separated Bike Facility	-0.741	0.182	-4.07	**<0.001**
Median Household Income	-6.5E-7	6.0E-6	-1.070	0.283
Speed Limit	0.0044	0.002	2.76	**0.005**
Volume/Capacity Ratio	0.410	0.048	10.29	**<0.001**
Segment Length	0.6797	0.179	3.79	**<0.001**
County Code (compared with Adams County)				
Arapahoe (2)	-0.0172	0.0857	-0.200	0.841
Boulder (3)	-0.636	0.0902	-7.05	**<0.001**
Broomfield (4)	-0.391	0.144	-2.71	0.007
Denver (5)	-0.270	0.0802	-3.36	**0.001**
Douglas (6)	-0.1598	0.0968	-1.65	0.0990
Jefferson (7)	-0.4157	0.0832	-5.00	**<0.001**
Weld (8)	0.1579	0.1585	1.00	0.319

Table 7: Probit hurdle model for factors affecting crashes per 10,000 trips

Selection Model Dependent variable = crashes/10,00 trips	Coef.	Std. Err.	z	P>z
Pseudo R² =0.0788				
Number of Trips	2.81E-5	1.60E-6	18.02	**<0.001**
3 - Collectors	0.175	0.01867	9.39	**<0.001**
4 - Minor Arterials	0.07289	0.01612	4.52	**<0.001**
5 - Major Arterials	0.00740	0.0166	0.440	0.656
6 - Interstate	-0.441	0.0882	-5.00	**<0.001**
Bike Lane	0.30880	0.01528	20.2	**<0.001**
Physically Separated Bike Facility	0.13317	0.04729	2.82	**0.005**
Median Household Income	6.94E-7	1.60E-8	-4.47	**<0.001**
Speed Limit	0.01785	0.00028	63.9	**<0.001**
Volume/Capacity Ratio	-0.12375	0.01197	-10.3	**<0.001**
Segment Length	-0.42436	0.05102	-8.32	**<0.001**
County Code (compared with Adams County)				
Arapahoe (2)	0.0575	0.0218	2.64	0.0080
Boulder (3)	0.130	0.0232	5.60	**<0.001**
Broomfield (4)	0.0133	0.0361	0.370	0.714
Denver (5)	0.0119	0.0204	0.580	0.561
Douglas (6)	-0.0489	0.0242	-2.02	0.0430
Jefferson (7)	0.00321	0.0212	0.150	0.879
Weld (8)	0.0735	0.0404	1.82	0.0690

The probit selection model shows that the number of trips is still a significant predictor of whether or not there will be a positive crash rate on a segment even when we control for other factors. Similarly, in the conditional model, it is shown that the number of trips is a significant predictor of crash rates for crash-rate-positive segments. Both results are as expected.

For this model, the pseudo r-squared value is somewhat higher than the model in step 1, which only considered "numbers" as a predictor for crashes. This makes sense as including more variables into the model will explain more of the variation seen in the dependent variable (crashes per 10,000 trips).

To understand the effect of trips on the total number of crashes, we generate predictions again. To create these predictions using this hurdle model and its accompanying control variables, this analysis uses marginal effects at the means methods. This method generates predictions of the total number of crashes for a given number of trips while all other independent variables in the model were at their mean value. Predictions of the number of unconditional crashes for each 10,000 trips were generated using both the probit model and the conditional model, multiplying the predicted probability of having at least one crash by the predicted number of crashes/10,000 trips if there is a crash. The predictions, then, show how many crashes (per 10,000 trips) there will be on a segment that is otherwise "average in every way." Prediction results for the multivariate analyses are shown below in Figure 19.
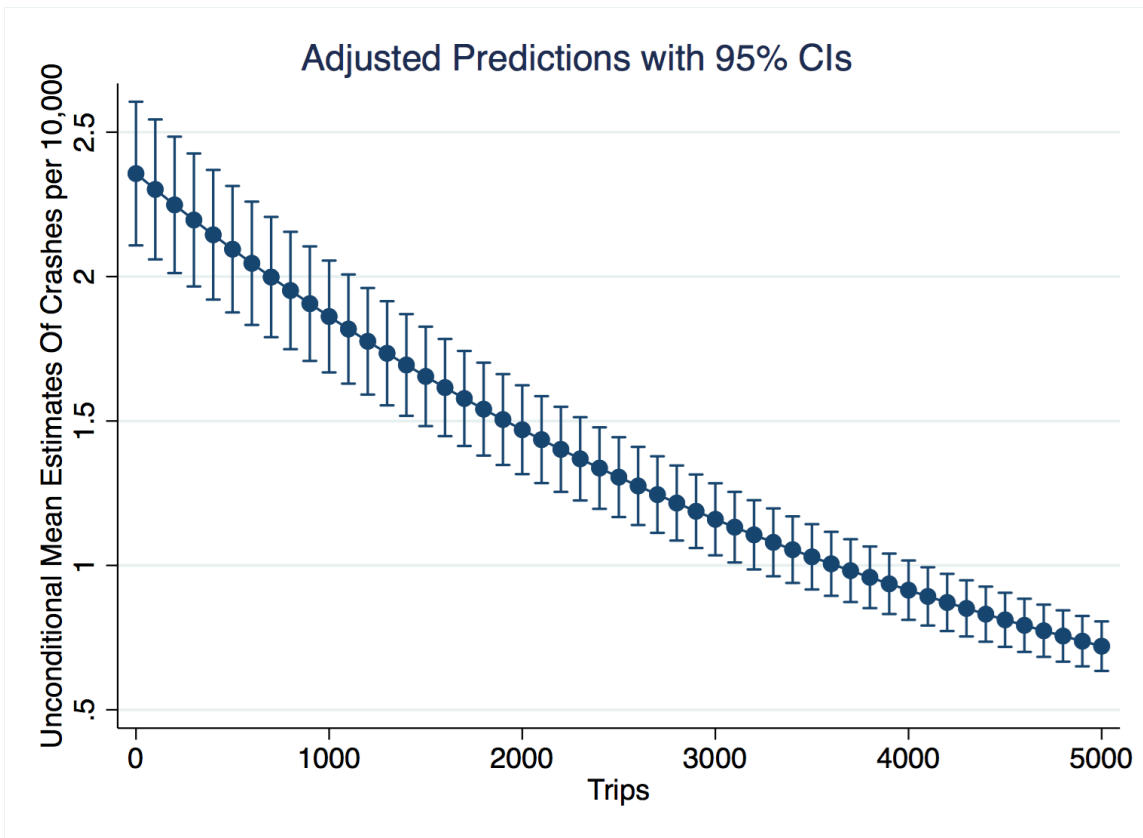
Figure 19: Marginal effects considering all covariates

These results continue to uphold the SiN effect and support the mathematic characterization made previously in the literature. The predictions show increasing the number of trips decreases the number of crashes *even when other variables were at their "average";* the number of crashes decreases from 2.5 per 10,000 trips to less than 1 crash per 10,000 trips as the number of trips increases from 0 to 5,000.

3.4   Step 4 Results

These following figures show predictions are based on the hurdle model used in previous steps, and they quantify the number of predicted crashes on segments with varying road conditions.
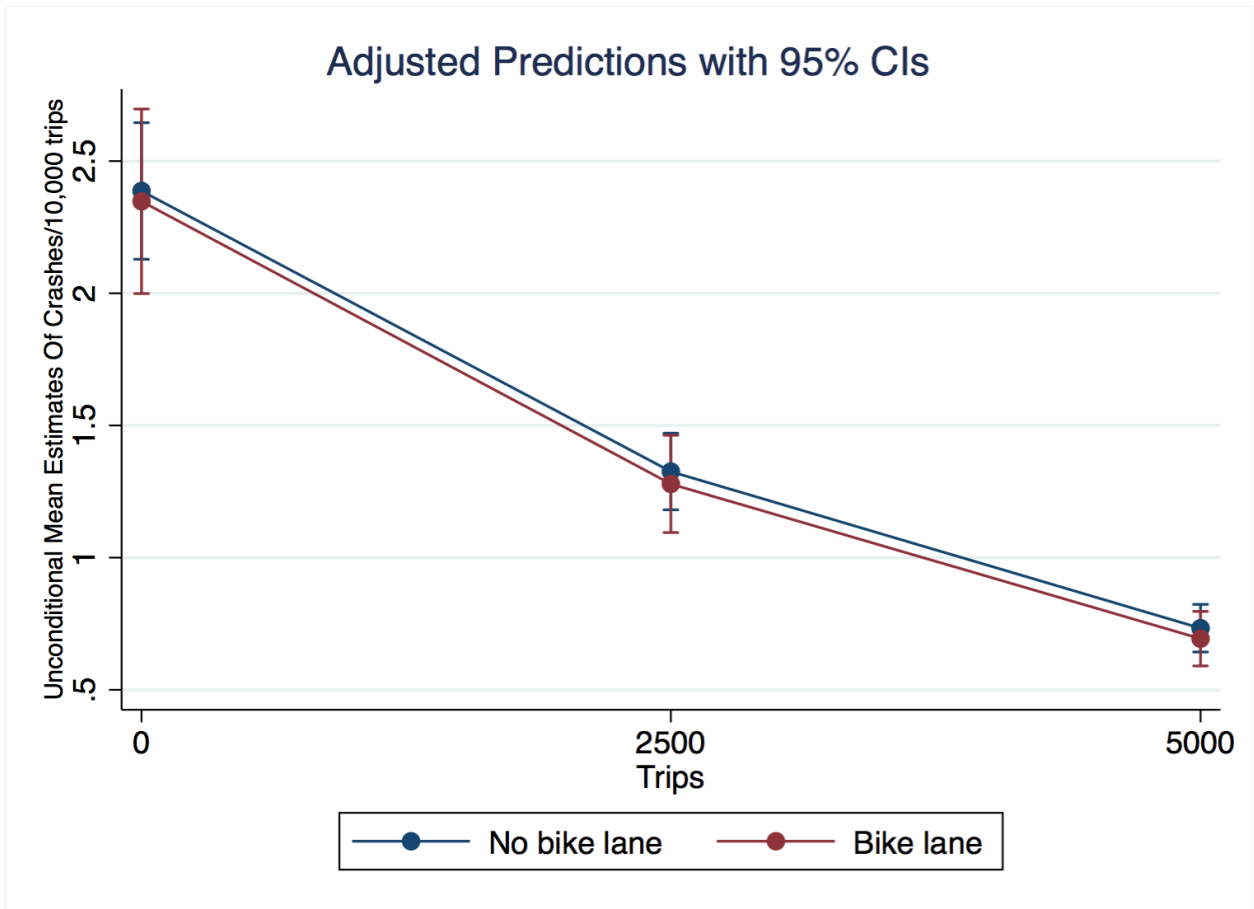
Figure 20: Crash predictions considering exposure and bike lanes

Figure 20 shows that as exposure increases from 0 to 5,000 trips the number of crashes on segments where there are bike lanes versus where there are no bike lanes

are very similar. The number of crashes on segments where there is no bike lane is very slightly smaller than where there are bike lanes (about 2.4 to 2.35, respectively).
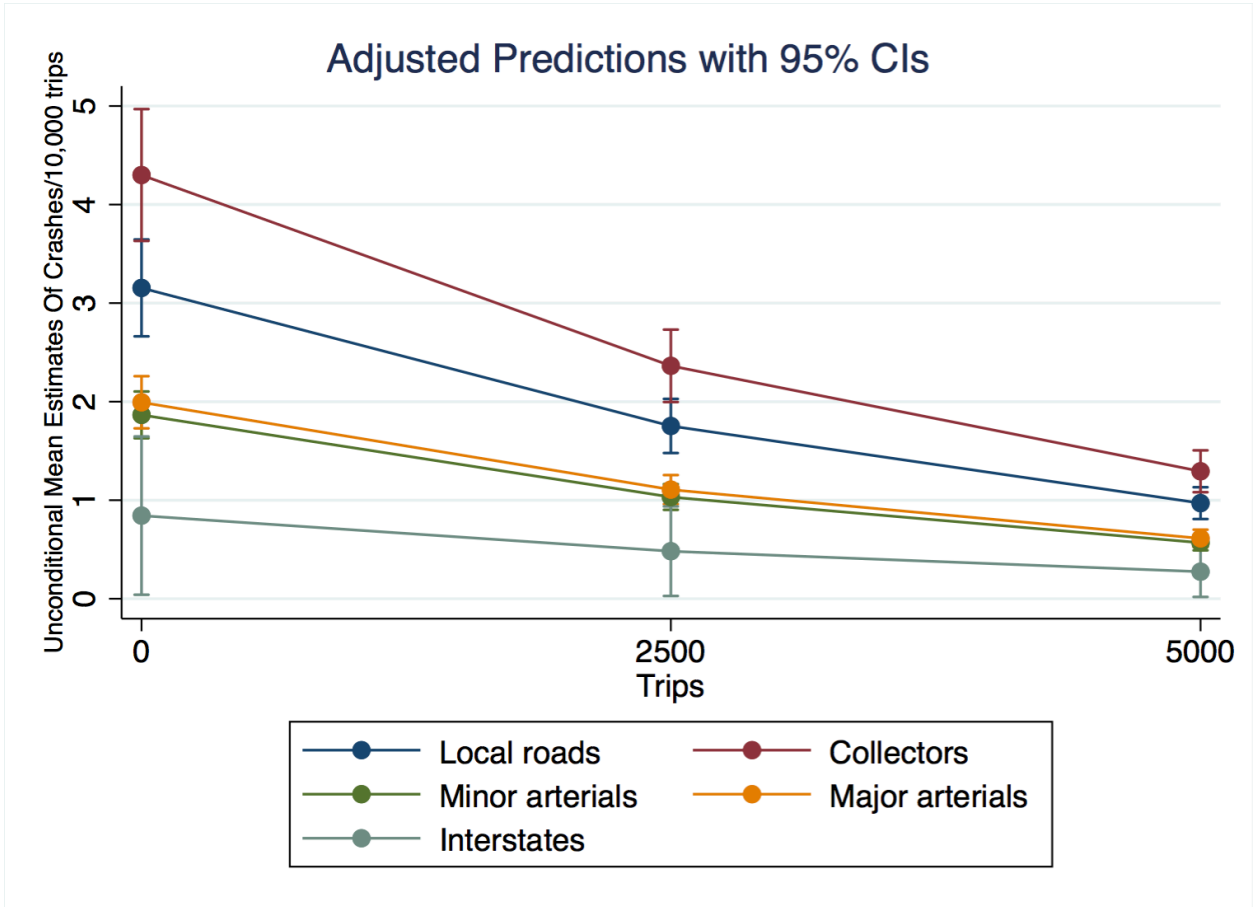
Figure 21: Crash predictions considering exposure and functional class

Results from

Figure 21 show that as the number of trips increases from 0 to 5,000, the number of predicted crashes decreases for all functional classifications, but not in the same way. The highest predicted crashes are on segments that are either collectors or

local roads, and the lowest are interstates and minor arterials. From these results, collectors see the most safety benefit from increased riders, and interstates see the least safety benefit from increased ridership.
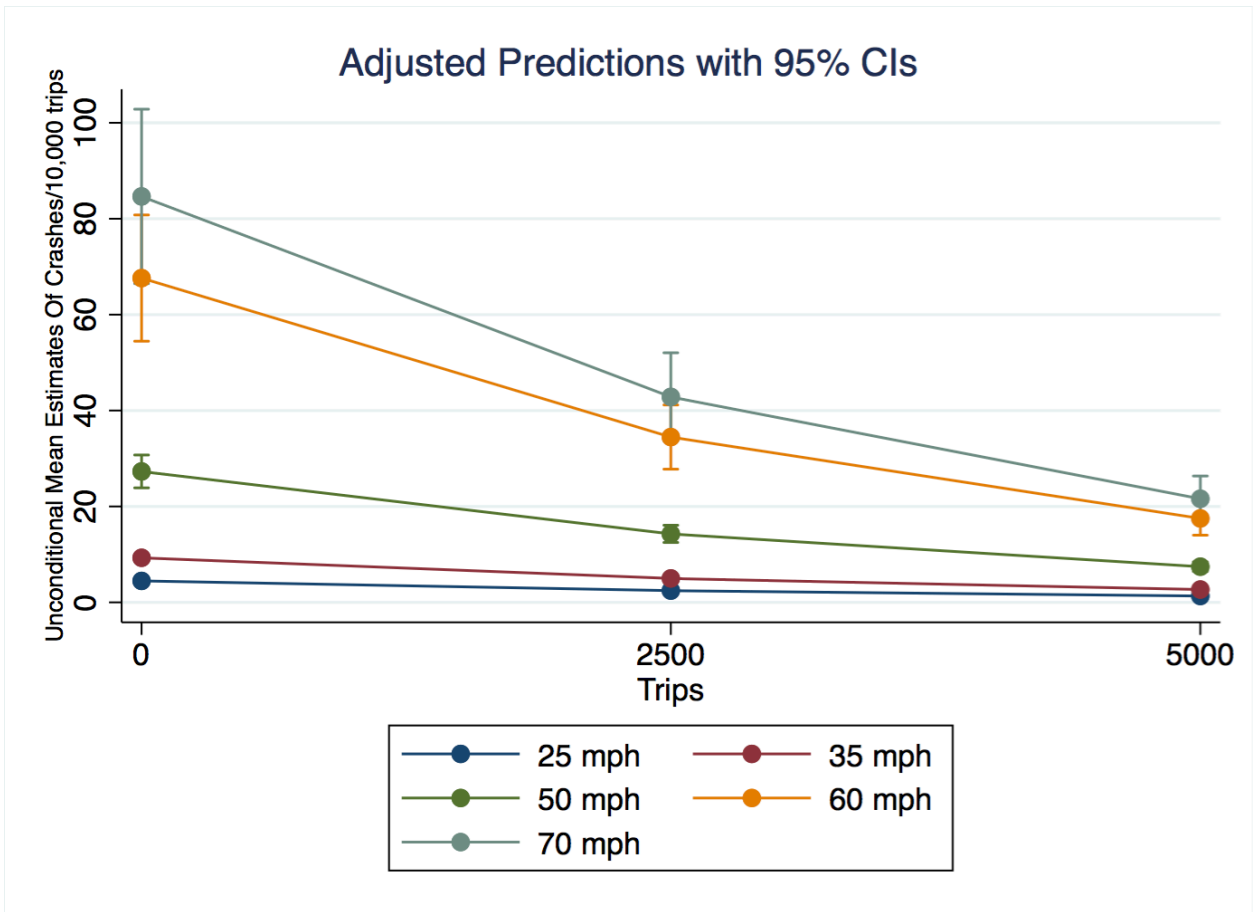
Figure 22: Crash predictions considering speed limit and functional class

Figure 22 indicates that the number of crashes decreases for all speed limits as more riders use the segments, but that they decrease from different starting points and at different rates. The highest speed roads (70 mph) sees the greatest decrease in the number of crashes (from 80 to about 20 crashes per 10,000 trips) as the number of trips

increases from 0 to 5,000. These results show that higher speed roads benefit from more riders per segment than lower speed roads.
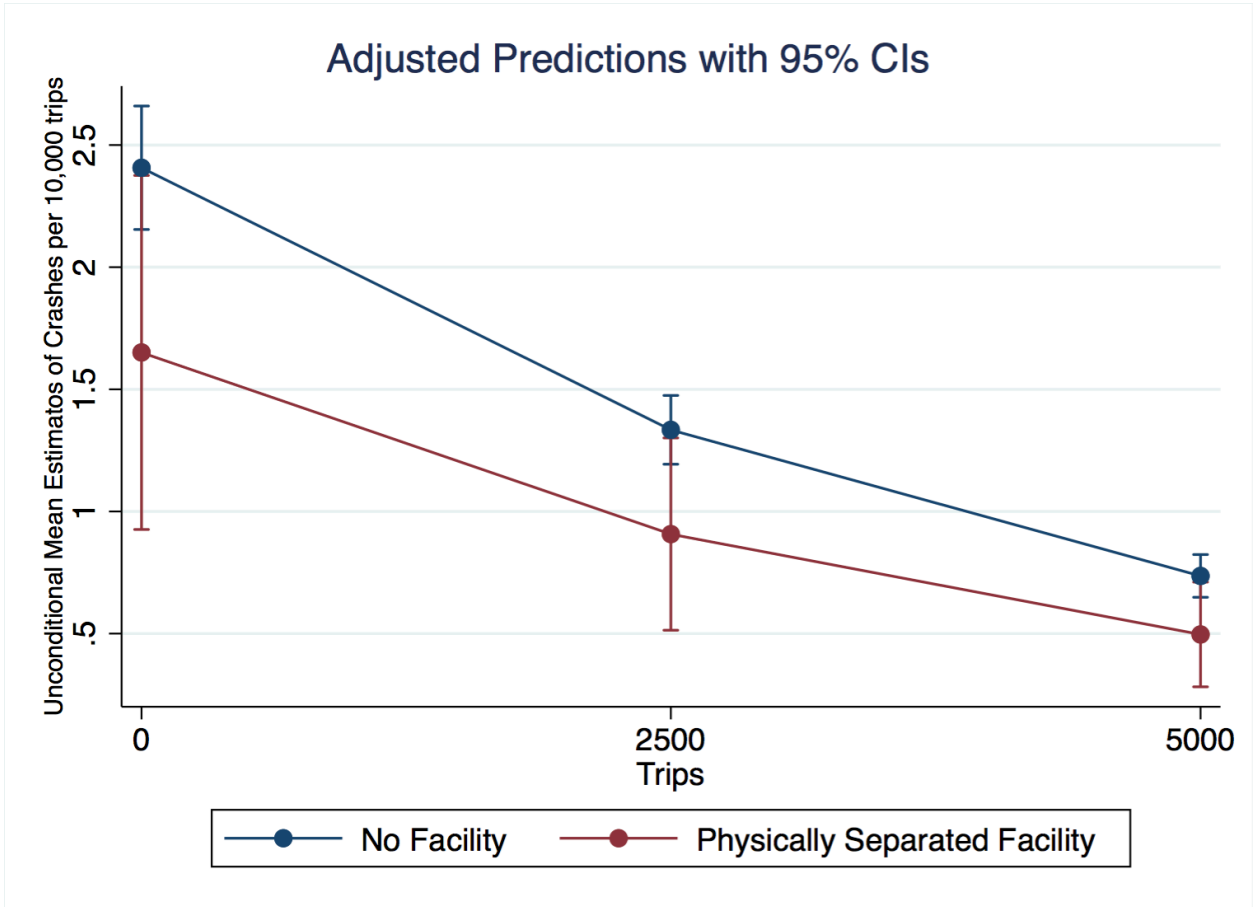
Figure 23: Crash predictions considering exposure and physically separated bicycle lanes (PSBL)

Results from Figure 23 that segments with no facility benefit from more riders than segments with physically separated bicycle infrastructure; as trips increase from 0 to 5,000, the number of predicted crashes decreases from 2.5 to less than one and 1.5 to less than 0.5 (per 10,000 trips) for segments with no facilities and PSBLs, respectively.
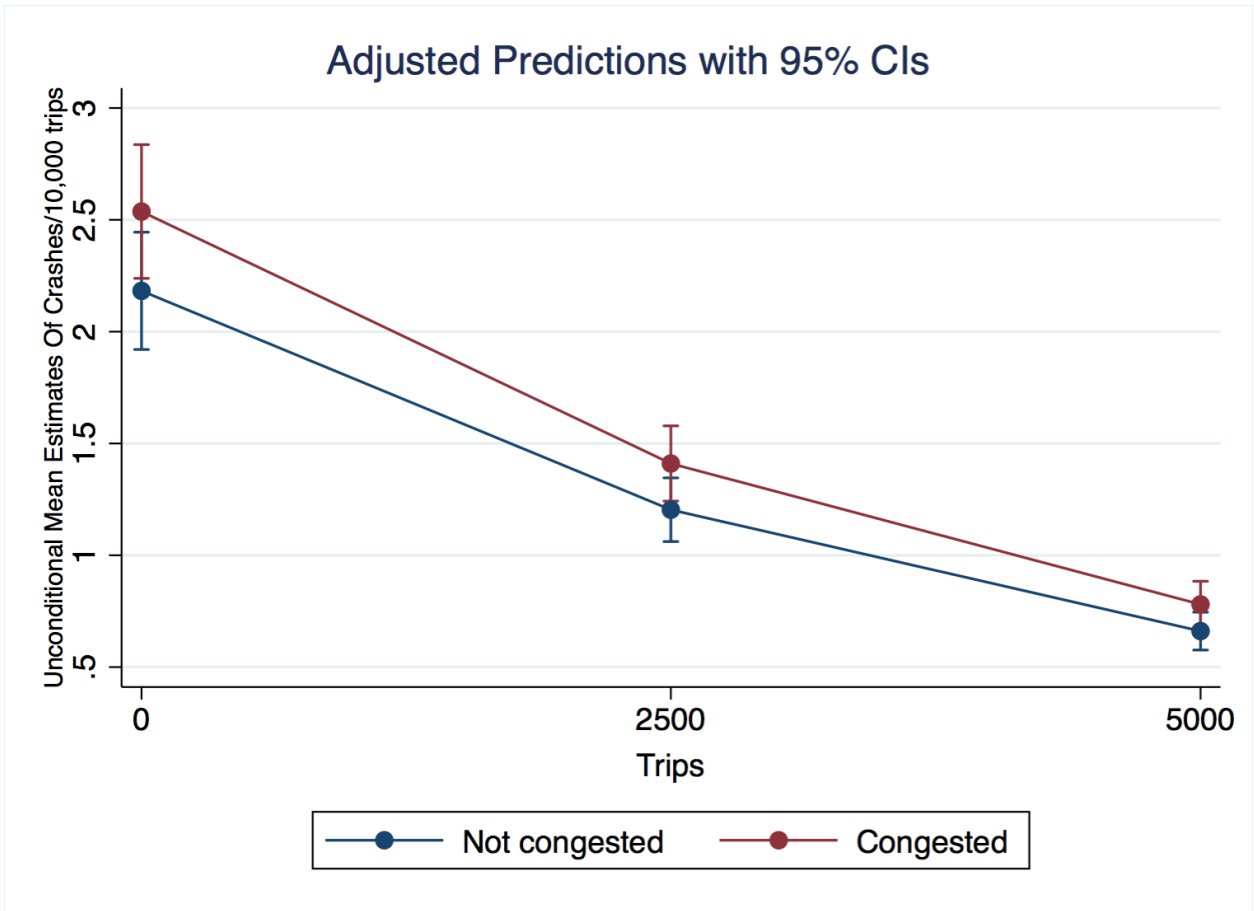
Figure 24: Crash predictions considering exposure and congestion

Figure 24 shows that the number of crashes per 10,000 trips decreases as the segments have more trips for both congested and not congested segments. Segments that are congested, however, benefit more from higher exposure than not congested segment; the number of predicted crashes per 10,000 trips decreases from 2.5 to less than one on congested segments, whereas crashes decrease from 2.25 to less than one on segments that are not congested.
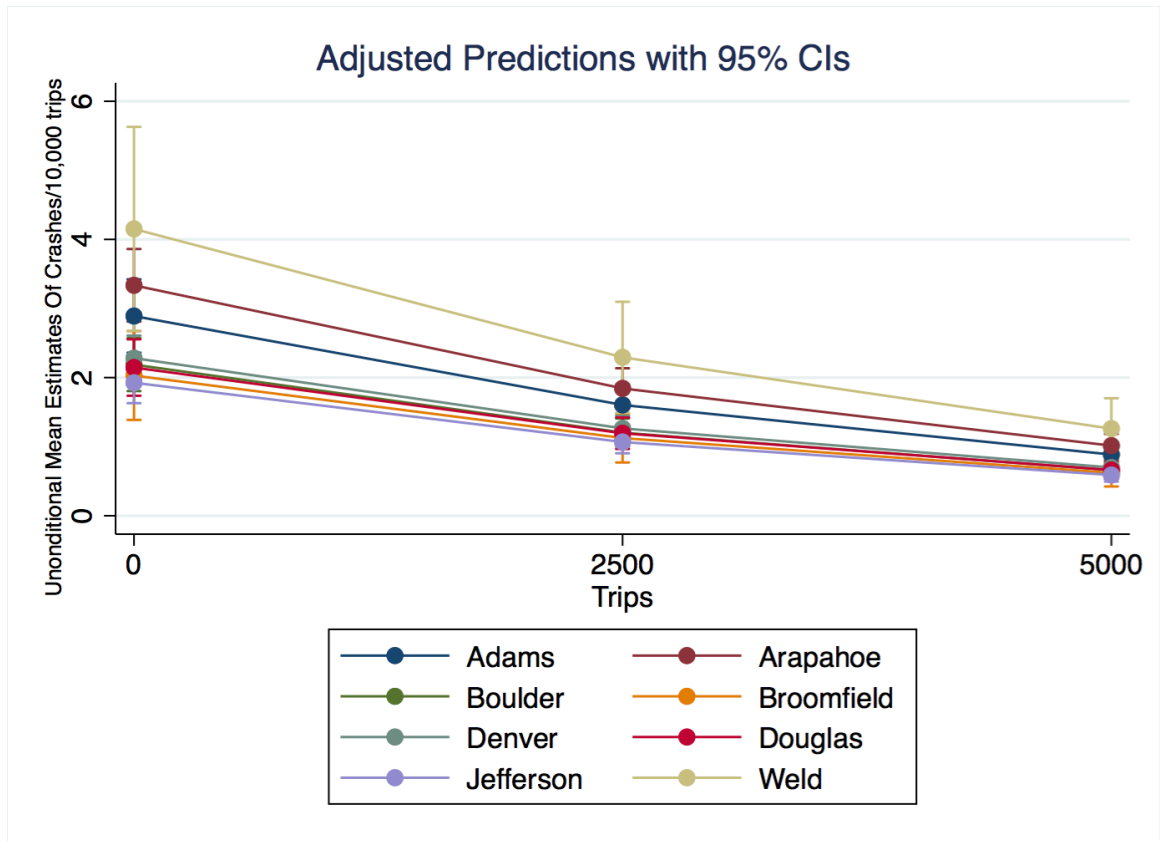
Figure 25: Crash predictions considering exposure and congestion

Figure 25 shows that the number of crashes decreases across all counties as the number of riders increases. The counties that benefit the most from increased exposure are Douglas and Weld Counties, which are both more rural, less populated, and lower density (in terms of road network) counties. The counties that benefit the least are Jefferson County and Broomfield counties.

4.   DISCUSSION OF RESULTS, LIMITATIONS, AND STEPS FOR FUTURE RESEARCH

4.1   Discussion of Results

The distribution of the number of segments with crashes is, as mentioned previously, extremely right-skewed, as is the number of trips per segment. This could be because of several reasons. First, under reporting of bicycle accidents (Elvik & Mysen 1999), which is noted in the literature as a significant issue, could lead to excess segments with zero crashes. Second, this could also be because of the nature of the data itself; each segment of road (defined in the data as the length between intersections) varies in length depending on the road network, but the majority are as short as a block in length (100-200 feet), especially in the urban settings. Because nearly every road is divided up into smaller segments, there are less crashes per unit than there would be if each segment were the entire length of road, leading to excess zeros. Third, and lastly, the excess number of zeros could also be due to the nature of bicycling in the study area; if most cyclists usually make the same trips on the same segments of road, there will be many segments that have no trips and no crashes at all.

The exponential hurdle model used, however, is able to handle with these excessive zeros. The first hurdle model predicting crash rates (per 10,000 trips) based on the number of trips alone shows that the number of trips is significant, meaning that it

explains variation in the dependent variable beyond simply chance. But when it is used

exclusively as a predictor of crash rates, it is a poor predictor (based on the very low

pseudo r-squared value). This is intuitive and fits with the framework of cyclists' safety

previously mentioned (see Figure 9); there are other factors that influence cyclists'

safety, and SiN *alone* is neither *the only* cause of nor *the only* preventer of crashes.

The marginal effects shown in Figure 13, however, do point towards the same

type of effect as seen in previous literature about SiN: as the number of trips increase

from 0 – 5,000, the number of crashes per 10,000 trips decreases from approximately 5

to approximately 1. This confirms the SiN theory as it applies to individual road

segments.

This is a significant finding, and it informs the literature in several important

ways. First, this confirms that the SiN effect does apply across all units of analysis. This is

important because previous research has assumed that the phenomenon "behaves" in

the same fashion at the national-level (i.e., crashes per year compared against trips

taken in the same year) as the segment level. These results were the first known

confirmation of these assumptions.

While the model shown in step 1 is a very poor explanation of all the variation

seen in the dependent variable, the model's ability to explain variation in crash rates

improves when other factors were considered alongside the number of trips. Results

from step 3 are summarized again in the following table in terms of positive and

significant relationships (+), negative and significant relationships (-), and non-significant results (/).

Table 8: Hurdle model (multiple factors) results summary

Variable	Sig. in Probit Model	Sig. in Conditional Model
Trips	+	-
Functional Class *(relative to Local Roads)*		
Collectors	+	/
Minor Arterials	+	-
Major Arterials	/	-
Interstate	-	/
Bike Lane	+	-
Physically Separated Bicycle Facility	+	-
Median Household Income	-	/
Speed Limit	+	+
Volume/Capacity	-	+

There were several interesting findings from these results. The number of trips is a significant predictor of rates in both models but in opposite directions. This means that trips predict both whether or not there will be a positive crash rate on a segment *and* what the crash rate will be for crash rate positive segments. More trips make the likelihood of a positive crash rate higher, but for segments that have positive crash rates, the risk per person decreases with more trips.

At first glance, it may seem counter-intuitive that the directionality for trips would be opposite between the probit model and the selection model. But it should make sense that having more trips is positively associated with having at least one crash per trip; for example, if a segment has zero trips it should also have zero crashes, but if a segment has 10,000 trips, it will very likely have one crash, and that singular crash would create a very small ratio of crashes/trip.

When considering functional class, some of the results were more surprising, especially the differences in significance and direction of the coefficient across the outcomes when compared to local roads.

- *Collectors* were significant in the probit model but not in the conditional model. This means that collectors have a higher chance of there being any crash at all compared to local roads (which follows intuition because there are likely more cars and more bicycles), but that they do not significantly predict how many crashes will occur. Predictions from step 4 of the research design also show that collectors have highest crash rates compared to other roads, even as the number of trips increases. Similarly, collectors see the one of the largest decreases in crashes from increased trips than any functional class

- *Minor arterials* are significant positive predictors of there being any crash at all (when compared to local roads) and how many crashes would occur if there are

crashes. However, the opposite directionality of the coefficients is surprising and hard to explain. When compared to local roads, minor arterials are more likely to have a crash at all, but, by this model, they will have less crashes per trip than local roads. This is also reflected in the predictions; compared to local roads, they have less crashes at 0 trips and decrease a slower rate as the number of trips increases to 5,000. It makes sense that minor arterials are more likely to have a single crash than local roads due to their design (typically higher speed limits and more lanes than local roads), but this is counter-intuitive in in the second regard; it seems that collectors would have more crashes per trip than local roads that also have crashes. Perhaps this could be ascribed to the *type* of cyclist that uses each type of road. A cyclist with less experience and confidence may choose to only ride on local roads as opposed to minor arterials, but they still have more bicycling crashes than those who would choose to rider on arterials (Mekuria, Furth, & Nixon, 2012). My research also considered that other variables that are being controlled in this context, especially speed limit, may cause these unexpected results. However, even when speed limit is left out of the analysis, the conditional model still results for minor arterials to be significant. This deserves future research that is out of the scope of this report.

- *Major arterials* are not significant in the probit model, but they are negative and significant in the conditional model. These results indicate that, for local roads

and major arterials that have positive crash rates, there will be less crashes on

major arterials as compared with local roads. Similar to minor arterials, this

seems counter-intuitive, but likely due to other variables in the analysis.

- *Interstates* significantly predict less likelihood for a crash than local roads. This is

  probably because riding a bicycle on the interstate is not legal in most parts of

  the country, including Colorado, so there should be less likelihood for a crash.

Bike lanes and physically separated bicycle facilities were significant in both models

but with opposite directions in their coefficients. These results indicate that there is a

higher likelihood of a single crash per 10,000 trips on roads with bike lanes/physically

separated facilities than those without, but for crash rate positive segments, there will

be less crashes on facilities with bicycle lanes than those without them. These results

are expected. Bike lanes provide cyclists with their own right of way in a road, which can

help the cyclists feel more comfortable and potentially make them safer. However, bike

lanes offer no sort of physical protection from an oncoming vehicle and therefore may

not be powerful enough to decrease the overall likelihood of a single crash. But, if there

are, for example, 10,000 trips on a segment with a bike lane, there is a higher likelihood

of there being a single crash, but the bike lane may decrease the overall rate of crashes

on crash rate positive segments by providing, at the very least, increased visibility. The

same result is seen with physically protected bike facilities, however, which is harder

may be harder to explain. Intuition says that physically protected bike facilities would be

less susceptible to the same effect because they provide more protection for cyclists

than a regular bike lane. But even separated bike lanes must interact with motorized

traffic at some point and therefore may still not be powerful enough to decrease the

likelihood of a single crash with more trips.

Bike lane and PSBL predictions in step 4 contribute to the argument among scholars

about the relationship between bicycle safety in numbers and bicycle facilities. There

was no significant difference in the predicted number of crashes for segments with or

without bike lanes as the number of trips increased, so it does not seem that bicycle

lanes significantly reduce the added safety benefit of numbers. PSBLs, however, do see a

significant difference. My predictions showed segments with PSBLs start off with less

crashes and "end" (at 5,000 trips) with less crashes, but they benefit less from more

cyclist exposure as more cyclists use them. Said another way, the rate of decrease in

crash rates is less for segments with PSBLs, but those segments have less total predicted

crashes.

Results from the median household income variable were somewhat surprising as

well. Typically, most research has found that, all else being constant, people with lower

incomes are disproportionately affected by accidents involving vulnerable road users

potentially because of poor quality of road design and/or neglected maintenance (both

of which are not controlled for in this model). Results from this analysis show the

opposite, albeit by a very marginal amount; areas with higher median household income were (very) slightly more likely to have a bicycle crash than those in lower areas.

Speed limits as a predictor of bicycle crashes each segment have expected results. Those segments with higher speed limits are more likely to have a crash than those with lower speed limits, and they are also more likely to have more crashes if they do have crashes. The predictions from step 4 of the research design also showed that higher speed roads benefit from increased exposure more than lower speed roads, and that the effect of numbers decreases with decreasing speed limit.

Each segment's volume-to-capacity ratio also has opposite significant coefficient signs between the probit and conditional models. Roads that are more congested are less likely to have any crashes at all, but for roads that have crashes, there will be more crashes on more congested roads (which is as expected). This was supported in the predictions, and the congested roads receive a greater safety benefit from increased number of trips than roads that are not congested. It should be noted that the conditional model is likely the more reliable source for understanding bicycle safety as a result of the number of trips per segment because its results are more quantifiable (as opposed to a simpler "crash v. no crashes" segment).

In previous research that used hurdle models to investigate crash rates, similar results for conflicting coefficients have occurred. For example, when using a hurdle model for quantifying the effect of various road conditions on the number of vehicular crashes, Ma, Yan, & Weng (2015) found that the number of through lanes significantly

predicted that roads with more than two lanes were more likely to have crashes than roads with less than two lanes, but two lane (or more) roads with positive crash rates would have lower overall rates than roads with less than two lanes. As they explained, these confusing results likely reflect the variables' complex relationships with each other and with the risk of a crash.

There are also some interesting results for the county control variable. Boulder County is significant in both models; a segment is more likely to have a positive crash rate if it is in Boulder County, but more trips in Boulder County segments lead to lower crash rates (for crash rate positive segments). Denver and Jefferson Counties are also significant and negative in the conditional model. This is somewhat surprising. Boulder, Jefferson, and Denver Counties are the more urban of the counties, so in a sense, the county variable be capturing some of the urban/non-urban effects. It should also be noted that each county is being compared to Adams County, which is also a less urban county. The predictions show that the most rural counties benefit the most from SiN effects, but that counties that are more suburban benefit the least from SiN.

When comparing results from linear regression and exponential regression, it is clear that the exponential regression is a better fit to the model (r^2 = 0.0788 compared to r^2 = 0.03210). This is an expected result. The exponential relationship seen in many risk models predicting bicycle safety outcomes in analyses with larger units is reflected in smaller units of analysis and more control variables.

The only other known research at this unit of analysis comes from Thompson et al.'s theoretical agent-based model (2017). Their research assumed the nature of SiN (by only counting bicycles as "exposed" to vehicles if they were not using bicycle infrastructure) at this level of analysis and was aimed at characterizing how cyclists' relative risk varied with the use of physically separated infrastructure. While this the study was seeking to investigate a different aspect of SiN, the results are somewhat comparable. Their research found that cyclists' relative risk increased with less interaction with vehicular traffic. Despite the questionable assumptions built into their model, this relationship has been somewhat been supported here. Crash-positive segments with more trips saw a decrease in the number of crashes per 10,000 trips.

4.2   Implications for practice

In summary, SiN effect was confirmed to hold true at the segment level as well as a city- or country-wide unit of analysis. It also holds true when controlling for other observable factors that may affect bicycle safety. This has several major implications for urban planners as they consider promoting bicycle infrastructure and more bicycle ridership in their municipalities or in their consulting work. First, it is important for practicing planners to know that safety and numbers do go hand-in-hand. While the direction of causation may not be entirely clear from existing research, this study does

effectively control for other variables and shows that "numbers" actually do "cause safety" as opposed to "safety causing numbers."

It is more challenging to assert what these results mean for overall system safety because it depends on perspective. These results show that encouraging more people to bike can be a good practice for safety if we assume that the best definition of safety is reducing crash rates. But if we assume instead that safety means reducing overall crashes, perhaps only encouraging more people to bike is not the best solution.

However, like any planning decision, encouraging a community to cycle more should be done with careful consideration, with the ultimate safety and welfare of the community having the highest priority. These results show that "numbers" have powerful influence on bicycle crash rates, but that the most complete picture of bicycle safety includes other factors. This means that programs and policies that encourage increased ridership should be done in tandem with the appropriate bicycle infrastructure, traffic calming mechanisms, and careful planning of routes to avoid congestion/high speed roads to truly ensure cyclist safety.

These results show that planners should also consider the types of changes they can make to their communities that would encourage more people to choose bicycling for recreational and utilitarian trips. Are policies alone enough to truly encourage more people to bicycle? Unless they are very powerful and far-reaching, it seems unlikely that, based on these results, any policy *alone* would encourage enough additional people to bicycle such that a SiN effect would accrue to a particular road. It seems more

likely, instead, that policies that promote/reward cycling combined with supportive infrastructure that allows cyclists to feel safer and more comfortable during their trips is likely the better approach to truly increasing numbers (Mekuria, Furth, & Nixon, 2012).

The exponential shape of the first prediction curve (tested in step 3) also points to another interesting finding that has been posited in previous literature; it is more likely that Jacobsen's construct of the SiN effect is correct than other constructs suggested. He stated that SiN is likely caused by drivers changing their behavior in response to seeing cyclists on the road. Now verified here at the segment level, the shape of the marginal predictions curve seems to suggest that a single exposure unit (here measured in trips) is much more powerful than 2 units or 10 units. Intuitively that finding makes sense for Jacobsen's construct. For example, it would seem that a motorist seeing one cyclists is not going to drive less safely than if (s)he saw two cyclists. Simply seeing the first person (or first few people) on a bicycle would be enough to influence driving behavior, and seeing many more would not add a great deal to driver safety.

Finally, the findings presented here also speak somewhat to debate about bicycle infrastructure decreasing the effect of cyclist exposure to motor vehicles on bicycle safety. The analyses above show that even when the number of cyclists is held constant, there will be lower crash rates on segments on segments with bicycle infrastructure than on those without it (assuming that the crash rate is positive). The predictions in step four show that bike lanes do not make a notable difference in the

precicted number of crashes, but that PSBLs do. This speaks to studies that posit that bicycle infrastructure can actually reduce how safe cyclists are because they are not actually "exposed" to drivers. Instead, this study shows that cyclists can still benefit from the added safety from numbers while also receiving safety benefits from bicycle infrastructure.

4.3   Assumptions, Validity, and Limitations of Research; Need for Future Research

It is important to recognize potential threats to validity—internal, content, and external— and to identify the limits of research arising from each step of the research design and from the data sources. Generally, content validity and external validity have been addressed through using conceptions and methods that are prevalent in literature and that statistically measure the probability of risk. However, there are other threats to validity and limitations that should be addressed.

First, the limitations of the data sources used in this study should be addressed. Strava data, while robust in the information it contains about some cyclists, is not a comprehensive review of all cyclists and their trips along each segment. It only contains geospatial data from those who use the app to record their trips, therefore creating a notable sample bias. This means that a portion of the population of all cyclists on each segment is not accounted for in the data set and Strava riders are likely more fit and experienced than the average population, which is likely reflected in their route choice.

There may be ways to compensate for (or at least discern the extent of) this

shortcoming, including using data from the American Community Survey to determine

the proportion of trips captured in the Strava data, that are out of the scope of this

research. In this case, the Strava data is really being used as a measurement of ridership

within a given segment. It is assumed that the proportion of Strava users compared to

all trips is similar across all segments and so Strava can be used as a tool for determining

segments' relative amount of ridership.

This report also makes assumptions about the habits of ridership over the last 10

years. The Strava data used in this research contains data about trips within the

calendar year of 2016, and the crash reporting is only available for the years 2005 to

2015. By using the Strava data to define areas of high ridership and lower ridership, this

research assumes that the general routes most used by cyclists have not changed

drastically since 2005. If it is assumed that the number of trips has increased somewhat

proportionally between 2005 and 2016, then the Strava data will generally portray

accurate representation of the amount of cycling on each segment. I believe that this

does not compromise my results in any significant way.

Similar to the Strava data, the crash data obtained from the DRCOG is robust in

its granularity in that each recorded crashed is geospatially referenced, but the data sets

also will not contain all crashes in all places. As noted in the literature, bicycle crashes

are systemically underreported; in most cases, crashes are not reported unless there is a

major injury or fatality. It is important to recognize that this may skew results in a way

that does not completely capture what happens in reality. It is not clear from the literature, however, to what extent underreporting would affect my results. For example, it is not clear whether or not crashes are underreported in certain types of places more than others; urban areas, for example, may suffer equally from underreporting as rural areas, or they might not. Future research should focus on understanding the extent to which underreporting truly affects overall crash rates. The important consideration in light of this reality is whether crashes tend to be more (or less) underreported in areas with higher (or lower) ridership.

There is also another potentially confounding variable that has been mentioned in some of the existing literature: attitude towards cyclists. There is no known research that categorizes the study areas' residents' perspectives on biking, so this factor cannot be included in the multiple regression analysis. It should be noted, though, that my research compares abutting counties in the same state; while perspectives towards cyclists are likely different in each county, it is less of a concern if this data was compared against data in different states or countries. Nonetheless, this presents a need for future research that characterizes attitudes of both riders and non-riders as it affects bicycle safety.

This research also does not consider the potentially important spatial component of SiN. Segments do not exist in space alone, but are rather connected to other segments, and, by extension, to an entire network of varying road conditions and travel patterns, which would suggest that each segment is likely not spatially

independent from other segments. This may be somewhat controlled from using counties as an independent variable, but it is likely that these units are too large to fully capture all of those effects.

This hypothesis for spatial autocorrelation is based on the reality that the cyclists and drivers who travel on one segment are likely the same people who travel on adjacent segments. Because of this, the exposure measure could be serving as a proxy of other variables that are not explicitly measured or understood here and that autocorrelates over space. Results from researching bicycle crash data spatially would inform existing literature in several ways. First, understanding the spatial component of SiN (if one exists) would inform many of the analyses that use least squares analyses. One of the underlying assumptions of these "line of best fit" tests is that the observed data are independent of one another. This assumption would be invalidated if spatial dependence (which is reflected by spatial autocorrelation) is verified through statistical tests. This could mean that some of the underlying assumptions purposed to underpin the SiN effect and potentially some of the assumptions made from that data might not be valid. Second, it would help characterize the nature of SiN and how the probability of a crash varies over space (if at all). If there is any spatial structure found through statistical testing, understanding and empirically modeling that structure could inform the literature to further characterize the nature of SiN.

Despite the imperfections within this research and the opportunities it presents for future study, the results here clearly suggest that cyclist exposure to vehicular traffic

does cause roads to be to have lower risk of crashes per rider. To further these findings,

it is crucial that more research on this scale of analysis be conducted. As cities continue

to collect more and more data about bicycle and traffic safety, better datasets can be

used to investigate these open questions and improve cycling safety.

REFERENCES

Anderson, M. (2017). "Bike commuting growth has leveled off-- but not everywhere in

    the US."Retrieved from https://usa.streetsblog.org/2017/09/14/bike-commuting-

    growth-has-leveled-off-but-not-everywhere-in-the-u-s/

Berg, H. Y. (2006). Reducing crashes and injuries among young drivers: What kind of

    prevention should we be focusing on? *Injury Prevention : Journal of the*

    International Society for Child and Adolescent Injury Prevention, 12 Suppl 1, i15-8.

    doi:12/suppl_1/i15 [pii]

Bhatia, R., & Wier, M. (2011). "Safety in numbers" re-examined: Can we make valid or

    practical inferences from available evidence? *Accident Analysis & Prevention, 43*(1),

    235-240.

Botchwey, N. D., Trowbridge, M., & Fisher, T. (2014). Green health: Urban planning and

    the development of healthy and sustainable neighborhoods and schools. *Journal of*

    Planning Education and Research, 34(2), 113-122.

Campbell, R., & Wittgens, M. (2004). The business case for active transportation.

    Gloucester: Go for Green. Print.

City of Berkeley. (2010). *Bicycle and pedestrian safety in berkely*. City Report. Berkeley,

    CA.

Cragg, J. G. (1971). Some statistical models for limited dependent variables with

    application to the demand for durable goods. *Econometrica: Journal of the*

    Econometric Society, 829-844.

Denver Region Council of Governments, (2017). *APPENDIX 7 2040 MVRTP Active Transportation Component*. Denver, CO. Retrieved from: https://drcog.org/sites/drcog/files/resources/2040%20MVRTP%20-%20Appendix%207%20-%20Active%20Transportation%20Component.pdf

Dill, J. (2009). Bicycling for transportation and health: The role of infrastructure. *Journal of Public Health Policy, 30*(1), S95-S110.

Douglas, E. (2017). "What you need to know about bike to work day--hint: Plan for breakfast, prizes, and lots of company". *The Denver Post* Retrieved from https://www.denverpost.com/2017/06/27/bike-to-work-day-2017/

Elvik, R. (2009). The non-linearity of risk and the promotion of environmentally sustainable transport. *Accident Analysis & Prevention, 41*(4), 849-855.

Elvik, R., & Bjørnskau, T. (2017). Safety-in-numbers: A systematic review and meta-analysis of evidence. *Safety Science, 92*, 274-282.

Elvik, R., & Mysen, A. (1999). Incomplete accident reporting: Meta-analysis of studies made in 13 countries. *Transportation Research Record: Journal of the Transportation Research Board,* (1665), 133-140.

Ewing, R., & Hamidi, S. (2014). Measuring urban sprawl and validating sprawl measures. *Washington, DC: National Institutes of Health and Smart Growth America.*

"FAQs: What are Pseudo R-Squareds" (2011) UCLA: Statistical Consulting Group. Retrived from: https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/ (accessed August 22, 2016).

Fishman, E., Schepers, P., & Kamphuis, C. B. M. (2015). Dutch cycling: Quantifying the health and related economic benefits. *American Journal of Public Health, 105*(8), e13-e15.

Glaeser, E. L., Kahn, M. E., & Rappaport, J. (2008). Why do the poor live in cities? the role of public transportation. *Journal of Urban Economics, 63*(1), 1-24.

GeoDa Software (Producer), & Anselin, L. (Director). (2006). *"Spatial autocorrelation basics".* [Video/DVD] Urbana-Champagne: GeoDa Software.

Guide, U. B. D. (2011). National association of city transportation officials. *New York, 8*

Hauer, E. (1982). Traffic conflicts and exposure. *Accident Analysis & Prevention, 14*(5), 359-364.

Hyer, V. (2014). "Biking to work increases 60 percent over last decade, census bureau reports." Retrieved from https://www.census.gov/newsroom/press releases/2014/cb14-86.html

Jacobsen, P. L. (2003). Safety in numbers: More walkers and bicyclists, safer walking and bicycling. *Injury Prevention : Journal of the International Society for Child and Adolescent Injury Prevention, 9*(3), 205-209.

Johan de Hartog, J., Boogaard, H., Nijland, H., & Hoek, G. (2010). Do the health benefits of cycling outweigh the risks? *Environmental Health Perspectives, 118*(8), 1109-1116. doi:10.1289/ehp.0901747 [doi]

Johnson, M., Oxley, J., Newstead, S., & Charlton, J. (2014). Safety in numbers? investigating australian driver behaviour, knowledge and attitudes towards cyclists. *Accident Analysis & Prevention, 70*, 148-154.

Lavetti, E., & McComb, S. (2014). Examining bicycle safety on a college campus: Observations and rationale for unsafe cycling. Paper presented at the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, , 58.* (1) pp. 1371-1375.

Ma, L., Yan, X., & Weng, J. (2015). Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter. *Journal of Advanced Transportation, 49*(8), 928-940.

Mekuria, M. C., Furth, P. G., & Nixon, H. (2012). *Low-stress bicycling and network connectivity.* Print.

Montgomery, C. (2013). *Happy city: Transforming our lives through urban design* Macmillan.

Morris, E. A., & Guerra, E. (2015). Mood and mode: Does how we travel affect how we feel? *Transportation, 42*(1), 25-43.

Price, A. M. (2016). Bicycling and walking in the central deep south states: Why is it rare and risky?

Pucher, J., & Buehler, R. (2012). *City cycling* MIT Press.

Räsänen, M., & Summala, H. (1998). Attention and expectation problems in bicycle–car collisions: An in-depth study. *Accident Analysis & Prevention, 30*(5), 657-666.

Richards, L. (2014, Putting people first 10 steps toward pedestrian-friendly suburbs. *Land Lines, July*, 10-18.

Root, E. (2011). *Spatial autocorrelation and spatial regression*. University of Colorado-Boulder: Spatial Autocorrelation and Spatial Regression Elisabeth Root.

Sanders, R. L. (2014). Roadway design preferences among drivers and bicyclists in the bay area. Paper presented at the *TRB 93rd Annual Meeting Compendium of Papers,* (14-5454)

Sanders, R. L. (2015). Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accident Analysis & Prevention, 75*, 26-34.

Sanders, R., & Cooper, J. (2013). Do all roadway users want the same things? results from roadway design survey of san francisco bay area pedestrians, drivers, bicyclists, and transit users. *Transportation Research Record: Journal of the Transportation Research Board,* (2393), 155-163.

Schepers, J., & Heinen, E. (2013). How does a modal shift from short car trips to cycling affect road safety? *Accident Analysis & Prevention, 50*, 1118-1127.

Schepers, P., Hagenzieker, M., Methorst, R., Van Wee, B., & Wegman, F. (2014). A conceptual framework for road safety and mobility applied to cycling safety. *Accident Analysis & Prevention, 62*, 331-340.

Schepers, P., Twisk, D., Fishman, E., Fyhri, A., & Jensen, A. (2017). The dutch road to a high level of cycling safety. *Safety Science, 92*, 264-273.

Separated bike lane planning and design guide(2015). . Cambridge, MA: Massachusetts Department of Transportation.

Sorensen, P., Wachs, M., Min, E. Y., Kofner, A., & Ecola, L. (2008). *Moving los angeles: Short-term policy options for improving transportation* Rand Corporation.

Thompson, J., Wijnands, J. S., Savino, G., Lawrence, B., & Stevenson, M. (2017). Estimating the safety benefit of separated cycling infrastructure adjusted for

behavioral adaptation among drivers; an application of agent-based modelling. *Transportation Research Part F: Traffic Psychology and Behaviour, 49*, 18-28.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography, 46*(sup1), 234-240.

Twaddle, H., Schendzielorz, T., Fakler, O., & Amini, S. (2014). Use of automated video analysis for the evaluation of bicycle movement and interaction. Paper presented at the *IS&T/SPIE Electronic Imaging,* pp. 90260U-90260U-13.

U.S. Census Bureau; American Community Survey (2016). American Community Survey 1-Year Estimates; using American FactFinder. Retrieved from: http://factfinder2.census.gov

Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.

Wegman, F., Zhang, F., & Dijkstra, A. (2012). How to make more cycling good for road safety? *Accident Analysis & Prevention, 44*(1), 19-29.