Effects of COVID and spatial demography on the reporting of cyclists struck by a motor vehicle

by

Mickey Edwards, MPA, PhD
University of Illinois at Springfield
Center for State Policy and Leadership
Motor Vehicle Data Linkage Project
Hedwa4@uis.edu

May of 2023

for

The Illinois Department of Public Health and Illinois Department of Transportation

Introduction

The equitable distribution of resources intended to mitigate the damage caused to communities when a cyclist is struck by a motor vehicle depends upon an accurate accounting of incidents. The need for accurate accounting applies not only to the total number of incidents, but upon whom the burden disproportionately befalls. By relying on crash data alone, policymakers may make allocation decisions using incomplete information. As a result, those communities most in need of resources to protect vulnerable road users may be missing out because of higher rates of unreported struck cyclists. By relying on police crash reports alone by officials, unreported struck cyclists may remain officially unseen.

The body of literature investigating unreported, or discordant, cyclist crash incidents is not robust. This is not due to a lack of interest in the subject but rather a dearth of reliable data – these are, after all, by definition cases for which a police report was never created. Even reliable hospital data that include injury classification codes and descriptions of the mechanism of injury provide no information regarding reporting status. Having access to crash data provided by the Illinois Department of Transportation *and* hospital discharge data provided by the Illinois Department of Public Health, the Motor Vehicle Data Linkage (MVDL) team at the University of Illinois Springfield (UIS) is uniquely positioned to investigate the burden of unreported cyclist crashes. This paper seeks to overcome prior data obstacles to investigate the incidence of unreported struck cyclists by first linking crash and hospital data files, and then by examining those incidents in the hospital file for which no corresponding report can be found.

Research Statement

This research is a follow up to a previously published manuscript which focused strictly on pedestrians (Edwards and Gutierrez 2023), here we focus strictly on cyclists. Additionally, this manuscript utilizes two more years of crash and hospital data, 2019 and 2020, totaling five consecutive years of linked files, 2016 through 2020. The addition of data from 2020 provides the opportunity to enhance this follow up study by examining the relationship between the onset of COVID and the propensity to report cyclists stuck by a motor vehicle. With cause and means established, this manuscript asks and answers: (1) did the onset of COVID-19 and subsequent stay-at-home orders impact the reporting of cyclists struck by a motor vehicle? (2) If so, whom among the populace was disproportionately affected?

As further detailed in the Methods section of this manuscript, analysis takes the form of an interrupted time series with 2020 receiving treatment group designation. The intent of applying an interrupted time series analysis, in addition to a segmented binary logistic regression, is to investigate unreported struck cyclists during normal times and compare them to similar incidents following the onset of stay-at-home orders.

Related literature

Doggette et al. (2018) assert that this stream of research on pedestrian and bicycle crashes relies upon police and hospital data, while treating unlinked cases between the two data sets as those unreported to police. At issue is the underreporting of crashes can lead to ineffective policies and

interventions. The ten articles reviewed by Doggett et al. (2018) agree that pedestrian and cyclist crashes are commonly underreported across a wide sampling of geographies. This may be especially concerning during the COVID-19 pandemic, "… with less cars on the roads, people drive faster, making roadways less safe for drivers, passengers, pedestrians and cyclists" (McFarland 2021). The share of incidents reported is potentially quite low, Langley et al. (2003) estimate that only 22% of bicyclist crashes can be linked to a corresponding crash report – indicative of potentially systemic issues in the administration of record keeping.

Stutts and Hunter (1999) study pedestrian and cyclist crashes across a mix of urban, suburban, and rural sites in California, New York, and North Carolina. Results show that a high number of injured pedestrians and cyclists do not involve a motor vehicle, and some of them were injured in non-roadway settings, like sidewalks, parking lots, off-road trails, or private property (Ibid). Similarly, Langley et al. (2003) find that crashes on public roads involving non-motorized vehicles are rarely reported to the police. Tarko and Azam (2011) find that pedestrians are less likely to be included in the database if the crash took place on a state road, a Y intersection, or a divided highway. They also find crashes are more likely to be included if they happened while crossing a road, rather than walking or standing along a road.

Langley et al. (2003) find that pedestrian and bicycle crashes are less likely to involve insurance compensation, implying a lower propensity to report the incident to police. Similarly, Lujic et al. (2008) find that those entitled to insurance compensation were more likely to alert the police, since such a report is often required for remission by insurance companies. In those crash reports, injury classification judgements are made by police officers. Since police officers are not normally trained medical professionals, they may classify minor injuries as severe (like those that involve bleeding), or overlook more severe injuries like internal bleeding (Doggett et al. 2018).

Age, race/ethnicity, injury severity, and cumulative length of hospital stay have all been linked to whether crashes were likely to be reported (Langley et al. 2003). Tarko and Azam (2011) find that males and older pedestrians were particularly prone to the most severe injuries. They also find that vehicle type plays a role in injury severity, they assert, "size and weight of the vehicle involved in a pedestrian crash were also found to have an effect on the pedestrian injury level." Further, Edwards and Leonard (2022) find that children and seniors are significantly more likely to die when struck by an SUV and a pickup truck, respectively, than when struck by a passenger car. Along with age and sex, an association with race has been noted. Sciortino (2005) argues that Black males are less likely to be included in crash records because of a reluctance to call police, leading to an underreporting of injuries by an estimated 21%.

Sources and methods

This manuscript elevates the term *crash* over *accident* in accordance with the prevailing theory that using the latter term implies motor vehicle collisions, and resulting injury and death, are unavoidable and faultless. This choice of terms is not intended to be a normative one. But rather a more accurate description of events that allows appropriate and effective analysis and policy

intervention. Just as a medical professional must accurately diagnosis a disease for effective treatment.

Data linkage

Crash data from the Illinois Department of Transportation (IDOT) and hospital data from the Illinois Department of Public Health (IDPH) were obtained for the years 2016 through 2020 by the University of Illinois Springfield (UIS) by way of an interagency data use agreement. Upon receipt of the data files, UIS established a probabilistic linkage methodology appropriate for the type of variables common among the disparate files. Data file linkage was accomplished using the software LinkSolv – which applies methods developed in the early 2000's by the National Highway Traffic Safety Administration's Crash Outcome Data Evaluation System program (McGlincy, 2021). The LinkSolv software is especially useful for the type of data produced by states with a primate city – as Chicago is to Illinois. For example, Cook County, home to Chicago, is also home to some 40% of Illinois residents – rendering county a relatively indiscriminate field for data linking purposes.

Five data fields common to both files were determined to be those with the greatest linkage success rate: date of birth, county, crash date, age, and sex. Spatiotemporal tolerances were permitted and specified within the software between the crash and hospital files to allow for some lag between the incident (crash file) and subsequent treatment (hospital file). For example, crash date tolerances one day into the future were specified to allow for the passage of time before the crash victim could reach the hospital. Hospitals in counties bordering the county where the crash occurred were also tolerated for linking purposes, as those may have been the nearest appropriate facility.

The hospital files include rich (yet not personally identifying) individual patient data who were treated under urgent, emergency, and trauma admission types. Individual patient race, ethnicity, sex, and age are included as fields in the hospital files, among many others. A diagnosis of the presence of intoxicating substances conducted at the hospital is also included as a data field and investigated as it relates to cyclists being struck. Patient home zip code is also a field included in the files which permits the study of socioeconomic factors inferred by 2019 5-Year ACS estimates (U.S. Census Bureau 2019). All references to zip codes throughout the manuscript are made to the patient home zip code. References to average poverty and carless household rates were calculated at the zip code level to be 11.7% and 5.8%, respectively.

Data independence

Prior to data analysis, a check for independence between the linked and unlinked data files was performed. Several Chi-squared (χ^2) tests were performed on variables within, and common across, data sets that may affect discordance rates. The tested variables included two of the fields used in the data linkage process, age and sex, and were each found to have significant alpha values. A series of Cramer's V (ϕ_c) tests were also performed to estimate the strength of association between the crash and hospital files using the same variables (**Table 1**).

Table 1: Chi-squared and Cramer's V tests of unlinked struck cyclists

Characteristic	χ^2	Φ_c	p
Age	498	.477	<.01
Race	88.4	.201	<.001
Sex	14.5	.081	.013

Discordance rate

Discordant, or unlinked, records present in the hospital file that were unable to be successfully matched with a corresponding record in the crash file were not necessarily unreported to police. In the interest and pursuit of privacy, personally identifiable information, which would enable a greater data linkage success rate, were omitted from the data files in accordance with our interagency data use agreement. Necessarily, our probabilistic linking methodology relies upon matching a combination of factors unique to each incident yet sufficiently general as to permit false positives and negatives – rendering such factors relatively indiscriminate. Still, false positives or false negatives are unlikely to exist in such great quantities as to corrupt the magnitude or direction of presented findings. The reader should be heartened by the presentation of statistical significance by several measures throughout the manuscript.

The share of discordant, or unlinked (sometimes referred to here as unreported), files is calculated using Equation 1 as applied by Watson et al. (2015) and among others elsewhere:

Equation 1

$$Discordance\ \% = \left(1 - \frac{c}{b}\right) \cdot 100$$

Where c is the number of cases successfully linked, and b is the number of candidate cases in the hospital file identified as a cyclist struck by a motor vehicle.

Interrupted time series

For public health researchers studying interventions imposed at the population level over a clearly defined time period, the interrupted time series (ITS) method of analysis is growing in its frequency of application (Bernal et al., 2017; Edwards, 2022). This manuscript builds upon and adds to those increasing uses of ITS by applying it to uncommon access to a half-decade of linked crash and hospital files to examine the burden of unreported incidents involving vulnerable road users.

Illinois based gubernatorial stay-at-home orders were issued effective March 21, 2020 to blunt the spread and "flatten the curve" of a novel coronavirus named COVID-19 just a month earlier. Those orders expired on May 29, 2020 and though no longer compelled, unless necessary most maintained the restricted movement and behavior learned during the period commonly known as the "lockdown." These and myriad other actions helped to construct the circumstances, and potentially the outcomes, of a naturally occurring social experiment – or simply a natural experiment. The year 2020 has received treatment group designation as a natural experiment by

others studying suicide trends (Pirkis et al., 2021), HIV care (Dorward et al., 2021), elementary schoolers' physical activity (Burkart et al., 2022), and motor vehicle crashes (Doucette et al., 2021) among others. With the onset of stay-at-home orders in Illinois being treated here as an intervention at the population level in 2020, the prior study years of 2016 through 2019 receive placebo group treatment.

Weather (snow and rain events), holidays (consumption of alcohol and increased travel), and other social events (annual sporting events) can impact crash frequency and outcomes. To help control for seasonal and other time-confounding unobservable factors, data are stratified to coincide with the onset of stay-at-home orders. As many of the behaviors learned under the stay orders continued post-expiration, only crashes and injuries occurring prior to March 21 of each study year (2016-2020) are omitted. The remaining data for each of the five study years spans from March 21 through December 31.

Predictive modeling of cyclist characteristics

Roadway crashes are often statistically normalized by scientists by relating crash counts to some measure of exposure to risk, or opportunity for crashes – like fatalities per 100 million vehicle miles traveled. While in public health discourse disease, injury, and death are commonly communicated and related in terms of prevalence within the population, or per capita. The latter frames the issue of injury and death caused by a motor vehicle as a health concern in a manner easy to assess personal risk. The former frames the issue as a matter of a traffic problem and presents numbers too large for risk to be easily assessed. This manuscript employs the language of the latter, so a statistical predictive model is fitted to estimate the likelihood of a struck cyclist to reside in a zip code of above average incidents normalized per capita.

Two segmented binary logistic regression models were fitted, one for the years 2016-2019 and another for 2020. The binary response variable was specified as a struck cyclist's home zip code cohort designation: 1 indicates they live in a zip code of above average unreported struck cyclists, while a 0 indicates they do not. The independent, or predictive, variables take the following form (Equation 2): proportion of carless households in zip code (*crlshh*), poverty rate of zip code (*povrte*), childhood poverty rate of zip code (*chpov*), Hispanic binary of individual (*hisp*), White binary of individual (*white*), Black binary of individual (*black*), Asian binary of individual (*Asian*), "other" race of individual (*other*), median household income of zip code (*incme*), and binary interaction term for treatment group designation (*inter*).

Equation 2

$$N_k = \beta_0 + \beta_1 crlshh + \beta_2 povrte + \beta_3 chpov + \beta_4 hisp + \beta_5 white + \beta_6 black + \beta_7 asian + \beta_8 other + \beta_9 incme + \beta_{10} inter + \varepsilon$$

The model for study years 2016-2019 was estimated to be statistically significant, and able to explain 22% of the variation (Nagelkerke R^2) in cohort designation. For 2020, the model was also estimated to be statistically significant, and to explain a little more than 20% of the variation (Nagelkerke R^2) in whether or not the struck cyclists resided in a zip code of above average incident rates. Full results are presented below in the Results section.

Results

Discordance rates across time

For the years 2016 through 2019 there were approximately 11,906 records of cyclists receiving hospital treatment following being struck by a motor vehicle, as identified through ICD-10 coding. For those same years, 3,138 of those incidents were able to be successfully linked to a corresponding crash record. The result is a discordance rate of 73.6% for cyclists struck by a motor vehicle across Illinois, similar to findings elsewhere (Langley et al., 2003). In 2020 alone, 2,887 records of struck cyclists were identified in the hospital file with 702 successfully linked to a corresponding crash file. The resulting discordance rate for cyclists in 2020 is 75.7%, or about two percentage points higher than the prior four-year mean. Pedestrians over the same time period of 2016 through 2019 had a discordance rate of 55.4%, with 15,820 struck pedestrians identified in the hospital data and 7,053 linked to a corresponding crash file. Some 3,398 struck pedestrians in 2020 received treatment at a hospital for their injuries, yet only 1,422 of those cases were able to be linked to a corresponding crash file. The result is a discordance rate for struck pedestrians of 58.2% in 2020, or 2.8 percentage points higher than the prior four-year mean.

These findings suggest two insights: first, cyclists struck by a motor vehicle are much less likely to report the incident to police compared to pedestrians – by a margin of up to 20 percentage points. Second, both pedestrians and cyclists were less likely to report being struck to police in 2020 compared to previous years.

Relative to pedestrian injuries

A reasonably intuitive explanation for the disparity in reporting between struck cyclists and pedestrians might simply be that being struck on foot results in elevated injury severity relative to those struck on a cycle (Monfort and Mueller 2023). The logic being that the greater the injury endured the greater the likelihood of the need of emergency services that results in the generation of a police report. The data somewhat support this logic, though the slim margins in injury severity differences likely do not rise to the scale of the disparity in reporting between the two cohorts. From 2016 to 2019 the average Maximum Abbreviated Injury Scale (MAIS; 0 implies no injury, 6 implies maximum injury) of discordant struck cyclists was 1.17, and 1.28 for the same cohort in 2020. The average MAIS of discordant struck pedestrians from 2016 to 2019 was 1.12, and 1.20 for the same cohort in 2020. So, unreported struck cyclists were actually reported by hospitals as having suffered a slightly more severe injury relative to unreported struck pedestrians. However, among reported incidents, struck cyclists from 2016 to 2019 averaged an MAIS of 1.19 and 1.28 in 2020. Reported struck pedestrians had an average MAIS of 1.27 from 2016 to 2019 and 1.38 in 2020, once more enforcing the notion of elevated injury severity among vulnerable road users following the onset of COVID-19. Though interesting to note, the differences in reported injury severity are slight enough that drawing definitive conclusions from this line of inquiry requires caution.

The built environment

Rural-urban commuting area (RUCA) codes are applied here at the zip code level to control for the differences in risk exposure a cyclist might face between varying built environments and traffic patterns. The U.S. Department of Agriculture (USDA) stratifies RUCA codes across four community types: metropolitan, micropolitan, small town, and rural commuting areas. USDA classifies communities based upon population density, urbanization, and the size and direction of prevailing daily commuting movement.

Table 2 disaggregates discordance rates by RUCA classifications (built environment type), cyclist demographics, and injury severity. Cell counts of 10 and fewer have been redacted to secure patient anonymity and remain in compliance with our interagency data use agreement. Writ-large, **Table 2** shows that discordance rates among struck cyclists balloons by more than 20 percentage points as geography transitions from denser metropolitan areas to more rural places. In other words, metropolitan cyclists struck by a motor vehicle are much more likely to report the incident to police than their rural counterparts. Still, there is variation and nuance among demographics and crash circumstances as they relate to a cyclists' propensity to report.

Injury severity

In the aggregate, discordance rates increased modestly in 2020 relative to the prior four-year mean regardless of built environment type. Though the picture is less clear in RUCA categories 4 through 10 because the numbers become small, in Metropolitan Areas (RUCA 1 through 3) more severe injuries consistently have greater rates of reporting. This is true for both head and thorax injuries, with head injuries of 3 on the MAIS scale climbing to a nearly 53% reporting rate – some 20 percentage points higher than the overall mean. Those increasing reporting rates for more severe injuries, especially the head, seem logical as the sense of urgency among those involved in the crash elevates.

Sex

However, the picture becomes less clear when the data are disaggregated, with some characteristics associated with higher reporting rates and others lower. For example, across geographical classifications females had a consistently higher discordance rate when compared to males. Further, the unreported rate among females generally increased following stay-at-home orders while simultaneously decreasing among males. It is not clear why a greater share of females would be less likely to report either before or after the imposition of stay-at-home orders. Though by raw counts, males typically outnumber females as an unreported struck cyclist by a factor of three to four. So, imparting meaningful insights from this line of inquiry should be accompanied with caution.

Race and ethnicity

Native Hawaiians or other Pacific Islanders had the highest discordance rate among all races in 2020, and American Indian or Alaska Natives had the highest in the prior four-years. Though shown to be statistically significant, these groups represent a small portion of all incidents so results should be interpreted carefully. The discordance rate among Black or African Americans

jumped 10 percentage points from Metropolitan to Micropolitan designated areas from 2016 through 2019, and nearly 18 points in 2020. Among Whites, discordance rates also jumped nearly 11 percentage points from Metros to Micros from 2016 through 2019, but only by 10 points in 2020. While it is clear that reluctance to report increases outside of urban areas, the issue appears most acute among Black struck cyclists.

Given the population share of Hispanics across Illinois, one would expect a somewhat proportionate representation among unreported cyclist crashes. Yet the Hispanic population is not proportionately represented. Some Hispanics may not identify with one of the presented options for race, prompting the data-collector to select "other." The U.S. Census Bureau recently reported that almost 44% of Hispanics selected the other category in the 2020 Census or did not answer the race question at all because they did not identify with any of the categories (U.S. Census, 2023). This data collection failure would help explain the large share of "other" in the race field. Finally, the "other" race category accounted for 15.5% of unreported struck cyclists in the data file.

Intoxicating substances

Discordance rates among struck cyclists who later tested positive for one of six intoxicating substances (alcohol, cannabis, opioid, cocaine, hallucinogen, stimulant, and "other drug") in 2020 were generally unremarkable relative to the prior four-year mean. Polysubstance use (a combination of any two or more substance categories) was up slightly, and opioid use among unreported struck cyclists was up almost eight percentage points in 2020. Though once more, the sample size is limited in numbers so results should be interpreted with care.

Table 2: Discordance rates by geography and demography

Cyclist characteristic		Metropolitan Area (RUCA 1-3)						Micropolitan Area (RUCA 4-6)						Small Town (RUCA 7-9)						Rural Area (RUCA 10)					
		Cases in hospital file		Unlinked hospital cases		Discordance rate		Cases in hospital file		Unlinked hospital cases		Discordance rate		Cases in hospital file		Unlinked hospital cases		Discordance rate		Cases in hospital file		Unlinked hospital cases		Discordance rate	
		2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020	2016-2019	2020
All		10,475	2718	7527	2001	71.9	73.6	733	157	620	135	84.6	86.0	563	143	507	131	90.1	91.6	104	16	95	15	91.3	93.8
Age	<18	3471	929	2683	729	77.3	78.5	302	68	260	62	86.1	91.2	307	84	290	76	94.5	90.5	66	13	61	12	92.4	92.3
	18-64	6398	1589	4397	1046	68.7	65.8	401	78	337	58	84.0	74.4	230	55	195	43	84.8	78.2	32	-	29	-	90.6	-
	65+	606	200	447	133	73.8	66.5	30	11	23	-	76.7	-	26	-	22	-	84.6	-	-	-	-	-	-	-
Sex	Male	8070	2014	5748	1380	71.2	68.5	561	112	472	90	84.1	80.4	398	107	356	86	89.4	80.4	75	-	69	-	92.0	-
	Female	2403	704	1779	529	74.0	75.1	172	45	148	38	86.0	84.4	165	36	151	35	91.5	97.2	29	-	26	-	89.7	-
Ethnicity	Hispanic/Latino	1830	476	1263	329	69.0	69.1	26	-	23	-	88.5	-	20	-	14	-	70.0	-	-	-	-	-	-	-
	Non-Hispanic	8645	2188	6216	1580	71.9	72.2	707	151	620	125	87.7	82.8	543	140	493	119	90.8	85.0	103	16	94	15	91.3	93.8
Race	American Indian or Alaska Native	67	-	50	-	74.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Asian	280	73	191	52	68.2	71.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Black or African American	1870	439	1293	284	69.1	64.7	82	17	65	14	79.3	82.4	15	-	13	-	86.7	-	-	-	-	-	-	-
	Native Hawaiian or other Pacific Islander	54	14	39	11	72.2	78.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	White	6158	1713	4555	1238	74.0	72.3	591	129	502	106	84.9	82.2	518	129	470	109	90.7	84.5	91	16	84	15	92.3	93.8
	Other	1931	415	1326	287	68.7	69.2	50	-	43	-	86.0	-	22	-	18	-	81.8	-	-	-	-	-	-	-
	Two or more	72	29	48	19	66.7	65.5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Declined or Unkown	43	25	25	14	58.1	56.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Substances	Alcohol	93	37	73	29	78.5	78.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Cannabis	89	26	58	17	65.2	65.4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Opioid	81	27	56	21	69.1	77.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Cocaine	53	-	37	-	69.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Hallucinogen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Stimulant	11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Other Drug	38	-	31	-	81.6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	Polysubstance	55	15	40	11	72.7	73.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Head Injury Severity	0	8080	2011	5861	1445	72.5	71.9	584	128	500	105	85.6	82.0	447	117	403	100	90.2	85.5	78	11	71	-	91.0	-
	1	1775	512	1292	349	72.8	68.2	121	20	98	16	81.0	80.0	91	20	83	17	91.2	85.0	22	-	20	-	90.9	-
	2	412	125	257	83	62.4	66.4	15	-	13	-	86.7	-	16	-	13	-	81.3	-	-	-	-	-	-	-
	3	201	68	113	32	56.2	47.1	12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thorax Injury Severity	0	9543	2444	6887	1731	72.2	70.8	654	139	561	111	85.8	79.9	508	129	461	110	90.7	85.3	94	15	88	14	93.6	93.3
	1	623	151	440	105	70.6	69.5	58	14	47	13	81.0	92.9	39	11	36	-	92.3	-	-	-	-	-	-	-
	2	257	94	174	56	67.7	59.6	16	-	-	-	-	-	12	-	-	-	-	-	-	-	-	-	-	-
	3	50	30	26	17	52.0	56.7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Socioeconomic indicators where discordant rates are high

The streets and neighborhoods comprising Logan and Palmer Squares and surrounding neighborhoods that constitute the 60647 zip code in Chicago, realized the highest number of unreported struck cyclists in Illinois for the years 2016-2019 at 126 such incidents. For 2020, 60622, a bordering zip code to the south comprising Wicker Park and Humboldt Park had the most unreported struck cyclists with 30 incidents. While those are both densely populated urban environments with elevated incidence counts, when evaluated on a per capita basis it is the low density, rural communities that stand out with high incidence rates. The 62626 zip code encircling Carlinville in rural southwestern Illinois had the most unreported struck cyclists per capita in the state for the years 2016-2019, most of the cases being among children. On a per capita basis for 2020 among zip codes with at least 10 incidents (to comply with our DUA), 62454 which encircles Robinson in rural southeastern Illinois had the highest rate.

A visible uptick in discordant struck cyclists within better-off zip codes following travel restrictions in 2020 suggests an increase in cycling within these communities following the onset of COVID relative to the prior four-year mean. **Table 3** shows remarkably similar standard deviations between year cohorts and by economic indicator, implying a shift in user demographic and an absence of meaningful outliers – which is also supported by the distribution of observations in **Figure 1**. The shift in unreported struck cyclists during the onset of stay-at-home orders appears to have been in the direction of those who live in more economically secure zip codes – or at least those zip codes with lower rates of poverty and carlessness. The typical poverty rate of the home zip code of an unreported struck cyclist in 2020 was nearly three percentage points lower than the prior four-year mean. The typical carless household rate was 2.2 percentage points lower in 2020 among unreported struck cyclists relative to the previous four-years.

Still, the typical unreported struck cyclist in 2016-2019 and 2020 lived in a zip code with 3 and 2.6 times, respectively, the rate of carless households compared to the Illinois state average. Compared to the average Illinois poverty rate, incidents occurred among residents of zip codes with an average rate nearly seven percentage points higher in 2016-2019, and just under four percentage points higher in 2020. So, while there was an adjustment in the profile of the typical unreported struck cyclist toward the economically better-off, incidents still disproportionately occurred among those in higher poverty areas with unreliable access to a car.

*Table 3: Descriptive statistics of discordant struck cyclists measured at the zip code level**

Cohort	*Mean*	*Median*	*Standard Deviation*
2016-19 Carless Household Rate	17.3%	14.5%	10.6%
2016-19 Poverty Rate	18.5%	17.6%	9.10%
2020 Carless Household Rate	15.1%	11.6%	10.7%
2020 Poverty Rate	15.6%	13.5%	9.01%

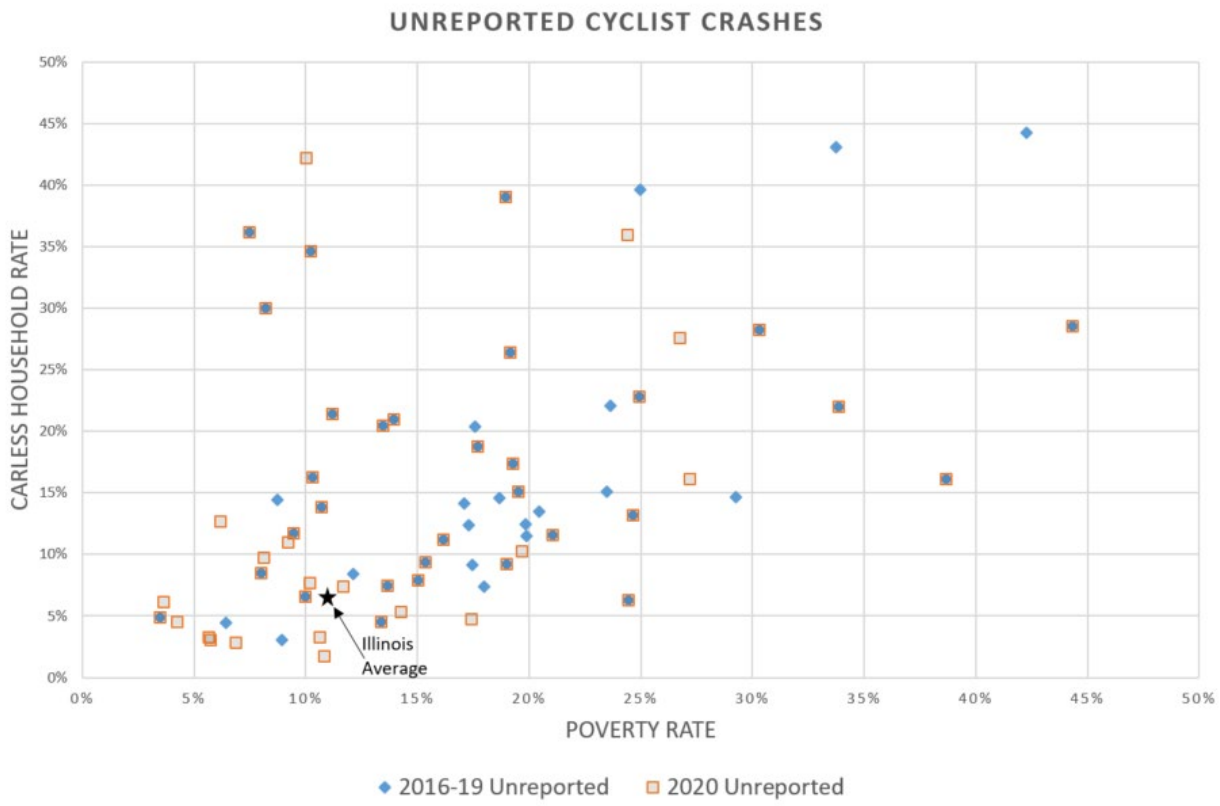
**2019 5-Year American Community Survey estimates*

Figure 1 plots poverty along the *x* axis and carlessness along the *y* of the 50 zip codes with the highest discordant rates – or roughly the highest 4% of zip codes in Illinois. The distribution of discordant struck cyclists cuts a much wider swath across socioeconomic status relative to discordant struck pedestrians. Without exception, the zip codes with the highest discordant pedestrian rates are among members of communities with poverty and carless household rates far above the Illinois average. Whereas, the zip codes with the highest discordant struck cyclists are much more evenly distributed across communities of both high and low rates of poverty and carlessness.

Figure 1 makes clear that some overlap exists between the placebo (2016-2019) and treatment (2020) groups. Several zip codes with high discordant rates from the placebo group were once more among the highest in 2020. However, upon closer examination, subtle differences in the distribution of unreported cyclist crashes begin to emerge. The highest discordant rate zip codes in 2020 are better off in a socioeconomic sense relative to those of the 2016-2019 cohort, though as discussed above on average these discordant struck cyclists are worse off than the typical Illinois resident. **Figure 1** shows how the 2020 cohort is distributed more frequently below and to the left of the Illinois average – implying less poverty and carlessness. Of note, especially among the 2020 cohort, are the high discordant rate zip codes with high (to very high) rates of carlessness but below average poverty rates.

Rather than being forced into alternative transport modes like cycling because of a financial inability to procure a reliable car – the car-*less*, some unreported strikes may be among the financially secure in pursuit of a car-*free* lifestyle (Brown 2017). In plain terms, some unreported cyclist crashes may be among people who can afford a car but choose not to. Incidents among residents of the 60610 zip code encircling the Near North Side and Gold Coast neighborhoods of Chicago with its easy access to lakeside paths and ample public transit may provide some examples of the car-*free* struck cyclist.

Figure 1: Poverty and carless household rates of 50 most discordant struck cyclist zip codes in 2020 compared to 2016-2019*



**2019 5-year American Community Survey estimates*

Segmented binary logistic regression model

Table 4 displays the results of dual segmented binary logistic regression models in which the dependent (or response) variables take the binary form of above average incidence rates aggregated at the zip code level. Models are segmented two-dimensionally by time alone. First, data are segmented to conform to Illinois' imposition of stay-at-home orders beginning March 21st of each study year. Second, year cohorts are segmented by pre-pandemic years dating to 2016 through 2019 as the placebo group, and the pandemic year of 2020 as the treatment group. **Table 4** presents the *B* coefficients for each variable along with the calculated odds ratio – which is simply the exponential value of *B*. The value added by including the odds ratio is ease of interpretation of results. It represents the estimated change in the likelihood of the event occurring (a 1 outcome in the binary dependent variable – here a zip code with above average incidents of unreported struck cyclists) for a unit change in the predictor variable. An odds ratio of 1 suggests a unit change in the predictor variable has no effect on the likelihood of the event occurring. While an odds ratio of less than 1 suggests a diminished likelihood, and an odds ratio of greater than 1 an increased likelihood, of event occurrence.

Since hospital data contain a field identifying the patient's home zip code, each observation in the models represents information germane to that patient's home neighborhood economic situation, including the prevalence of incidents involving unreported struck cyclists. Another factor controlled for in the model at the observation level is a proxy for exposure to risk based on the built environment and traffic flow intensity using RUCA classifications. Model results imply a modest increase to the likelihood of an unreported cyclist crash occurring as the built environment deurbanizes. The available data are silent on response times by emergency personnel, though it is conceivable that an anticipated prolonged arrival time due to sprawling terrain serves as motivation to seek unofficial means of transport for medical treatment.

Perhaps most surprising among results is the inverted effect of carless households on the propensity for incidents to go unreported. For the study years of 2016-2019 zip code carless household rate had a strong positive correlation with increased incidents per capita – more households without a car are related to more struck cyclists. This result seems logical since those without car access would disproportionately seek out alternative transportation modes, and in turn expose them to more risk from motor vehicles. For 2020 that effect reversed as the model estimates a strong negative relationship – more households without a car in a zip code is related to fewer struck cyclists. This inversion could be due to an increased interest in cycling throughout 2020 by high income earners pursuing a car-free lifestyle discussed throughout this manuscript.

Table 4: Unreported struck cyclist characteristics association with high incidence per capita: results of segmented binary logistic regression model

Variable	*Coefficient (2016-2019)*	*Odds Ratio (2016-2019)*	*Coefficient (2020)*	*Odds Ratio (2020)*
Built Environment	0.315*	1.37*	0.391*	1.48*
Carless Household Rate	5.63*	278*	-4.77*	0.008*
Poverty Rate	6.90*	988*	7.32*	1515*
Child Poverty Rate	-3.40**	0.033**	-0.430	0.651
Hispanic	-0.122	0.886	-0.657*	0.518*
White	-0.181	0.834	0.414	1.51
Black	-0.521*	0.594*	-0.167	0.846
Asian	-0.888*	0.412*	-0.311	0.732
Other (race)	-0.152	0.859	0.823***	2.28***
Median Household Income	0.00*	1.00*	0.00*	1.00*

**Significant at 1%; **Significant at 5%; ***Significant at 10%*

Proposal for weighting data

This section proposes using historical linked and unlinked data to weight, or adjust, the reported count of cyclists struck by a motor vehicle. Weighting the reported count is intended to more

accurately communicate the true number of cyclists struck by a motor vehicle by applying lessons learned through analyzing the linked and unlinked files. Generally, the crash data file contains an undercount of incidents. Two methods are developed depending on the needs of the user and data available to them. The methods described are of working status – that is, they are a work in progress and not intended to be officially implemented or relied upon to make policy or allocation decisions.

Table 5 shows unreported cyclist crashes were consistently trending downward from 2016 through 2019; that changed in 2020, ticking up to the highest level since 2017. It is not clear what caused the trend to reverse or whether it will persist. Still, a potential contributing factor for the reversal in 2020 could have been a reluctance to engage with emergency personnel outside of a controlled hospital setting for fear of contracting the then-nascent COVID-19. The ongoing civil unrest and distrust in law enforcement manifest in American culture and street protests following the killing of George Floyd undoubtedly also played a role in a reluctance to engage emergency personnel. Interesting to note that for the year 2016 the crash file contained 105 additional cases than the annual estimated total.

Table 5: annual estimated total of struck cyclist by other data sources

Year	Linked	Unlinked	Crash File Total	Annual Estimated Total
2016	852	2403	3360	3255
2017	796	2205	2813	3002
2018	777	2090	2554	2867
2019	716	2070	2550	2786
2020	703	2185	2200	2888

Equation 3 can be applied to adjust the total number of linked struck cyclists over the course of a year for the entire state of Illinois.

Equation 3

$$ac = lc(wf) + lc$$

Where *lc* is the count of linked struck cyclists, *wf* is the weighting factor (current value of 2.79), and *ac* equals the adjusted annual count of struck cyclists. The weighting factor (*wf*) is calculated as the ratio of unlinked struck cyclists to linked cases; and should be updated and averaged annually as each new tranche of linked hospital and crash data sets become available. As was shown throughout this manuscript, 2020 was an abnormal year for reporting cyclist crashes and was subsequently left out of calculating the current value of *wf* – which used data from 2016 through 2019.

Alternatively, if the linked crash and hospital files are not available to apply as specified in *Equation 3*, a similar method can be applied to crash data alone using *Equation 4* below.

Equation 4

$$ac_1 = lc_1(wf_1)$$

Where lc_1 is the count of all cyclists in the crash file, and wf_1 is the ratio of the sum of linked and unlinked cyclist crashes (annual estimated total in **Table 5**) to the count of cyclists in the crash file (lc_1). In calculation, when the annual estimated total is known, the lc_1 variables cancel each other out and what remains is simply the annual estimated total. However, the intent is that wf_1 be used when the annual estimated total is not known and the user only has access to crash data. The variable wf_1 in this construct is intended to be greater than 1. Here the weighted average factor for the years of 2016 through 2019 comes out to $wf_1 = 1.06$. In a typical year dating back to 2016, the sum of linked and unlinked cyclist crashes was about 200 incidents greater than indicated in the crash file alone.

As proposed here the adjusted annual count of struck cyclists is broadly applied to data aggregated at the state level and calendar year. The weighting factor can be further refined to calculate the adjusted count of unreported struck cyclists with increased fidelity applied both spatially and temporally. For example, *wf* can be calculated for a specific zip code during a specified timeframe, like the school year.

Discussion

Not to be overlooked as a potential contributing factor to the trends outlined here is the fast-paced expansion of bikeshare and shared mobility infrastructure and participation across the Chicago region around the same timeframe of this study. The City of Chicago reports that in 2019 nearly 4 million bike and scooter trips were completed through the shared mobility company Divvy (CDOT 2023). In 2020 Divvy along with the City of Chicago further expanded their shared mobility network into large swaths of the southern region of the city – further increasing adoption and use. Though as discussed in the findings section, metropolitan cyclists were much more likely to report being struck by a motor vehicle compared to their rural counterparts.

Car ownership and access to reliable transportation plays an important role in exposure to vehicular violence by vulnerable road users. For example, there is an inequitable U-shaped distribution of carless households across Illinois: where the very low-income earners would like to own a car but cannot, and the very high income earners can own a car but choose not to while in pursuit of their preferred lifestyle. For example, the 60602 zip code bordering Michigan Avenue to the west and covering parts of Millennium Park in dense downtown Chicago has a median household income of over $190,000, a very high carless household rate of nearly 45%, and plenty of nearby transit options including rail. While the less-dense 60624 zip code encompassing Garfield Park and a portion of Interstate 290 to the south has a median household income of just over $23,000, also a very high carless household rate of a little over 44%, and fewer, further flung, transit options.

Discordant cyclist crashes when examined in the manner described within this manuscript can prove useful to policymakers, transportation planners and engineers, and community members

when conducting road safety assessments. Having more accurate data on the true scale of cyclists struck by a motor vehicle can result in a more equitable distribution of mitigation funding and efforts within the communities most in need. And those communities most in need are commonly of a low socioeconomic status, the impoverished and the carless, and across a diverse geography – among the urban and the rural.

Additionally, while this manuscript provides more accurate information on the burden of struck cyclists compared to relying upon crash data alone, it is unable to suggest motivating factors for the act itself of not reporting. As demonstrated above, less severe crashes are more likely to go unreported to police compared to more severe crashes, and the location of the crash also plays a role. Still, there are likely larger sociodemographic factors and economic motivations at work that lead some to decide not to report. Future studies should focus on investigating the dynamics underlying outlined factors and motivations in a manner that permits community outreach to encourage the reporting of all struck cyclists. Future work should also focus on the Hispanic population, in particular the manner in which race and ethnicity are recorded by medical professionals.

Conclusion

A commonly applied (in economic data) and virtually cost-free policy intervention would be to adopt a weighted, or adjusted, supplementary count of struck cyclists. A supplementary report of the estimated true count of struck cyclists that adjusts the count taken from the crash file by using historical data of unreported cases would empower communities and policymakers to address road safety head-on. An adjustment method as described in this manuscript, or similar method, would be a good first step in addressing transportation inequities by providing more accurate data regarding the true burden borne by communities. Once more, the official data can be adjusted across time and space so that communities can better understand which neighborhoods have elevated rates of cyclists being struck and whether any seasonal variations are in effect. Following a more complete understanding of the true scale of struck cyclists the important work can begin to prevent them from occurring altogether.

References

Bernal, James Lopez, Steven Cummins, and Antonio Gasparrini. 2017. “Interrupted time series regression for the evaluation of public health interventions: a tutorial.” *International Journal of Epidemiology,* Volume 46, Issue 1, Pages 348–355, https://doi.org/10.1093/ije/dyw098.

Brown, A.E. 2017. “Car-less or car-free? Socioeconomic and mobility differences among zero-car households.” *Transport Policy.* 60:152-159.

Burkart, Sarah, Hannah Parker, R. Glenn Weaver, Michael W. Beets, Alexis Jones, Elizabeth L. Adams, Jean-Philippe Chaput, and Bridget Armstrong. 2022. "Impact of the COVID-19 pandemic on elementary schoolers' physical activity, sleep, screen time and diet: A quasi-experimental interrupted time series study." *Pediatric obesity* 17, no. 1: e12846.

Chicago Department of Transportation. *Divvy for the Entire City: Divvy Service Hits All 50 Wards.* March, 2023. https://www.chicago.gov/city/en/depts/cdot/provdrs/bike/news/2023/april/divvy-for-the-entire-city--divvy-service-hits-all-50-wards.html#:~:text=Ridership%20is%20up%20nearly%2040,rides%20for%20income%2Dqualifying%20Chicagoans.

Couto, Antonio, Marco Amorim, & Sara Ferreira. 2016. “Reporting road victims: Assessing and correcting data issues through distinct injury scales.” *Journal of Safety Research*. Volume 57. p 39-45. ISSN 0022-4375. https://doi.org/10.1016/j.jsr.2016.03.008.

Dorward, Jienchi, Thokozani Khubone, Kelly Gate, Hope Ngobese, Yukteshwar Sookrajh, Siyabonga Mkhize, Aslam Jeewa et al. 2021. "The impact of the COVID-19 lockdown on HIV care in 65 South African primary care clinics: an interrupted time series analysis." *The Lancet* HIV 8, no. 3: e158-e165.

Doucette, Mitchell L., Andrew Tucker, Marisa E. Auguste, Amy Watkins, Christa Green, Flavia E. Pereria, Kevin T. Borrup, David Shapiro, and Garry Lapidus. 2020. “Initial impact of COVID-19’s stay-at-home order on motor vehicle traffic and crash patterns in Connecticut: an interrupted time series analysis.” *Injury Prevention*. 27:3-9, http://dx.doi.org/10.1136/injuryprev-2020-043945.

Edwards, Mickey. 2022. “An Interrupted Time Series Analysis of the Sociodemographics of Crash Victims during the Illinois Stay-at-home Order.” *Findings*, September. https://doi.org/10.32866/001c.38490.

Edwards, Mickey and Daniel Leonard. 2022. “Effects of Large Vehicles on Pedestrian and Pedal-Cyclist Injury Severity.” *Journal of Safety Research*. Volume 82: 275-282.

Edwards, Mickey and Manuel Gutierrez. 2023. “The incidence burden of unreported pedestrian crashes in Illinois.” *Traffic Injury Prevention.* 24:1, 82-88, DOI 10.1080/15389588.2022.2143236

McGlincy, Michael H. 2021. LinkSolv. Strategic Matching. Morrisonville, NY.

Monfort, Samuel S., and Becky C. Mueller. 2023. *Bicyclist crashes with cars and SUVs: injury severity and risk factors.* Insurance Institute for Highway Safety (IIHS): https://www.iihs.org/api/datastoredocument/bibliography/2277.

Pirkis, Jane, Ann John, Sangsoo Shin, Marcos DelPozo-Banos, Vikas Arya, Pablo Analuisa-Aguilar, Louis Appleby et al. 2021. "Suicide trends in the early months of the COVID-19 pandemic: an interrupted time-series analysis of preliminary data from 21 countries." *The Lancet* Psychiatry 8, no. 7: 579-588.

U.S. Census Bureau. 2019. 2015-2019 5-year American Community Survey.

U.S. Census Bureau. 2023. *Racial Identification for the Self-Reported Hispanic or Latino Population: 2010 and 2020 Census.* March 28, 2023. Accessed 4/27/2023. https://www.census.gov/data/tables/time-series/demo/hispanic-origin/racial-identification.html

USDA Rural-Urban Commuting Area Codes. https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/.

Watson A, Watson B, Vallmuur K. 2015. Estimating under-reporting of road crash injuries to police using multiple linked data collections. Accident Analysis and Prevention 83:18–25. doi: 10.1016/j.aap.2015.06.011. PMID: 26162640.

Disclaimer

Funding for this research was made possible (in part) by the Illinois Department of Public Health (DPH) through funds from the Illinois Department of Transportation (IDOT), 22-0343-03. The views expressed in written conference materials or publications and by speakers and moderators do not necessarily reflect the official policies of IDPH or IDOT, nor does the mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.