A USDOT NATIONAL
UNIVERSITY TRANSPORTATION CENTER

# Safe by Design: Collecting Traveler Centric Data to Inform Safe Street Design

Megan S. Ryerson, Ph.D. (https://orcid.org/0000-0002-8843-3286)

FINAL RESEARCH REPORT - June 30, 2023

Contract # Contract # 69A3551747111

# Table of Contents

**Evaluating Cyclist Biometrics to Develop Urban Transportation Safety Metrics**

Megan S. Ryerson*[1,2], Carrie S. Long[1], Michael Fichman[3], Joshua H. Davidson[1], Kristen N. Scudder[1], Michelle Kim[4], Radhika Katti[5], George Poon[2], and Matthew D. Harris[1]

[1]*Department of City and Regional Planning, PennDesign, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.*

[2]*Department of Electrical and Systems Engineering, PennEngineering, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.*

[3] *PennPraxis, Weitzman School of Design, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.*

[4] *Swarthmore College, Philadelphia, Pennsylvania, U.S.*

[5] *Department of Civil and Environmental Engineering, College of Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, U.S.*

**Abstract**

The transportation safety paradigm for urban transportation – particularly safety for those walking, cycling, and scooting – relies on counting crashes to parameterize safety. This *reactive*, crash-based method is limiting: crashes are underreported and devoid of human perception, and their low frequency precludes testing countermeasure effectiveness. Building on diverse safety-critical fields, we hypothesize that urban transportation safety can be measured *proactively* with traveler biometrics, including eye and head movements, such that high readings of biometric indicators correlate with less safe areas. We collect biometric data from cyclists traversing an urban corridor with a protected, yet not continuously, cycle lane. By isolating and correlating peaks in cyclist biometric measures with infrastructure design, we develop a set of continuous variables – lateral head movements, gaze velocity, and off-mean gaze distance, both independently and as a vector – that allow for the evaluation of safe urban infrastructure proactively. The results reflect higher biometric readings correspond to less safe (i.e., unprotected) areas, indicating that safety can be measured *proactively* with biometric data.

## 1. Introduction

The U.S. Department of Transportation defines a "crash" as "an event that produces injury and/or property damage, involves a motor vehicle in transport, and occurs on a trafficway or while the vehicle is still in motion"; counting crash events and, further, counting those killed or seriously injured in crashes (KSI) is the existing, gold standard measure of safety for urban transportation (Federal Highway Administration, 2017). Aggregated crash statistics are visualized in many cities' Vision Zero plans in the form of a High Injury Network map, which highlights corridors with the highest frequency of crashes. KSI can be a sheer count, augmented based on exposure at a crossing location, or augmented with travelers' perspectives collected through surveys and focus groups (Lassarre et al., 2007). While numerous methods to modify KSI are explored in the literature, at its core KSI can only reflect revealed crashes.

The current paradigm of safety and its focus on measuring *objective* risk by counting crashes has led to a reactive, flawed method for choosing where to locate infrastructure interventions. Consider that the rules for installing traffic control devices in the U.S. mandated by the Federal Highway Administration are based on specific thresholds of pedestrian or cyclist volume or specific numbers of fatalities that must occur before safety interventions can be deployed (FHWA, 2009; Schmitt, 2019). Rather than allowing safety planning to be driven by a design variable of deaths and injuries, planning safe urban transportation systems and countermeasures requires new indicators to evaluate safety with a combination of objective and *perceived* risk, based on the experience of travelers using active transportation modes. As argued by Frank Haight (1986), it is perceived risk – the risk experienced by road users – that induces and influences macrobehaviors (route choice, seatbelt use, choice to cycle, etc.) and microbehaviors (stopping at stop signs, scanning for conflicts, etc.).

In the following manuscript, we seek to develop methods that result in empirical, continuous measures of perceived safety as defined by the perceived risk that influence microbehaviors. We explore a range of safety-critical fields and learn that by assessing the biometric variables from active travelers, in tandem with the physical characteristics of the built environment in which they occur, we can identify the infrastructure designs that are the safest and the least safe from the perspective of vulnerable road users, including active transportation modes (biking, walking, scooting, etc.) (Krauzlis, 2013; Olmos-Solis et al., 2017). Building on the research of driver safety as well as fields outside transportation that find high cognitive workload is correlated with a diminished capacity to remain safe while engaging in a safety-critical task, we hypothesize that: 1) workload of active travelers traversing the urban infrastructure can be measured with a composite of classic eye movement variables combined with head movement variables and 2) correlating workload variables with fine-grained spatial variables will establish the relationship between infrastructure design and active traveler cognitive workload. Our goal is to develop methods to measure the safety of urban transportation infrastructure for active travelers based on workload and stress. This is a fundamental departure from the reactive, crash and *objective risk*-based definition of safety that has dominated urban transportation safety research and practice since its inception.

## 2. Literature Review

### 2.1 Measuring Cognitive Workload in Safety-Critical Fields

Analyzing data captured through physiological and eye tracking sensors is an established method for identifying and measuring cognitive workload, with frequent, rapid eye movements being reflective of a high cognitive workload (Engström et al., 2005; Hogervorst et al., 2014; Marquart et al., 2015). Fields as diverse and driver safety, toddler and infant education, air traffic control, and surgical techniques assess workload through the study of eye movements (Aslin, 2012; Elsabbagh et al., 2012; First and Palfrey, 1994; Holmqvist et al., 2011; Loft et al., 2007; Tien et al., 2014; Wilson et al., 2019). The eye movement-based quantification of workload is often complemented by physical or motor measures: state-of-the-art eye tracking technology is equipped to simultaneously collect gyroscope (head orientation and movement) data; the literature reflects the integration of measures related to neck and torso flexion into the quantification of workload (Åkesson et al., 2012; Kapp, 1997; Szeto et al., 2012; Yu et al., 2017, 2016). Establishing a thorough and comprehensive understanding of workload is imperative, as workload can be indicative of one's capacity to perform tasks correctly, wherein high workload is correlated with the deterioration of safety (Yi-Fang Tsai et al., 2007).

Safety-critical fields—fields in which a failure may result in a fatality, serious injury, or other damage (Knight, 2002)—have used workload analysis research to leverage reductions in injury by identifying workload thresholds and implementing safety measures. Transportation, an essential and safety-critical field, is one of the primary arenas for this research. In the field of air traffic control, high cognitive workload is found to be associated with dangerous performance deviations and false alarms; as a result, air traffic controllers are limited to a set number of flights they can handle at one time (Callan, 1998). Motorcycle operators saw both their eye movements and their crash rate decrease in a simulated environment after completing a training course (Di Stasi et al., 2011). Studies of automobile transportation find that when drivers fixate on an area of interest away from center, their workload increases and they are more prone to crash (Hancock et al., 2003; Horberry et al., 2006; Jankowska-Karpa and Wacowska-Slezak, 2016; Lee et al., 2002; Topolšek et al., 2016). Scholars have measured driver blink frequency and duration, fixation frequency and duration, pupil diameter, and two-dimensional horizontal vergence while drivers in a simulator were asked to complete different tasks and found that decreased fixation duration (or an increased velocity of gaze) correlated with a driver committing errors in their assigned tasks (Yi-Fang Tsai et al., 2007). From the literature, we see that high cognitive workload can lead to errors and a crash, and high cognitive workload can manifest itself in these eye and head movements.

The understanding that high cognitive workload can lead to driver stress and error is well-established, and in fact, serves as a basis for current safety policies: the entire set of U.S. guidance and regulations on safe roadway development is based on driver line of sight and driver cognitive workload (Foy and Chapman, 2018). To date, researchers have mainly recruited subjects to drive in virtual and test track driving simulators while equipped with biometric sensors to collect heart rate, skin conductance, and/or respiration rate; and/or eye tracking glasses collecting data on gaze velocity (velocity of the movement of the eyes in degrees per second), fixation duration (identifying fixations and their average length in time, or areas on which a collection of gazes are located noting an area of higher processing), horizontal vergence (a depth-related measure), and pupil diameter (Klauer et al., 2014; Lee and Winston, 2016; Mannering and Washburn, 2012; Mehler et al., 2009, 2008; Stephens and Groeger, 2014). Recently, there has been a call made by scholars to employ these methods outside of closed simulators and move into field-based data collection using eye tracking devices (Asan and Yang, 2015).

While driver stress, workload, and ultimately safety is well understood from biometrics, scant related studies exist for active travelers (such as pedestrians, cyclists, scooter-riders). Pedestrian-oriented simulators study distraction yet continue to assess the binary external outcomes such as a "successful" crossings – i.e., a pedestrian is not struck in the virtual environment (Deb et al., 2017; Neider et al., 2011) – or study distraction and visual attention due to the environment and surrounding (Tapiro et al., 2020, 2018). With regard to cyclist biometric research, video analysis has shown that cyclists turn their heads more during periods of higher workload and interaction with vehicles (Räsänen, 1999), yet this has not been quantified beyond qualitative video analysis.

## 2.2 Issues in Safety Measurement for Surface Transportation

The safety-critical field of surface transportation is centered on the evaluation of drivers, and not the pedestrians and cyclists among whom numbers are increasing and for whom safety is declining. In the U.S., the number of trips made by bicycle doubled between 2001 and 2009 (Pucher et al., 2011); by 2016, the number of persons biking to work was 50 percent higher than it had been in 2005 (American Community Survey, 2016). This growth is accelerated by the expansion of active modes available, as seen in the introduction of bicycle and electric scooter-share systems. Amidst the uptick in travelers choosing active modes, the U.S. has seen a dramatic rise in crashes and fatalities for active travelers. Consider that between 2007 and 2017, driver fatalities in the U.S. experienced a 14 percent decline yet pedestrian fatalities rose 27 percent (National Center for Statistics and Analysis, 2018; Retting, 2018).

Improving the safety for these vulnerable roadway users requires well-planned safety interventions such as protected cycling lanes separated from traffic. Separated, designated bicycle facilities (both painted and protected) have been shown to reduce both the occurrence of a crash-related injury and the severity of an injury when one does occur (Chen et al., 2012; Cicchino et al., 2019; Wall et al., 2016). By correlating crash counts with the presence of different infrastructure designs in cities across the U.S., it was found that infrastructure that is physically separated from the vehicle traffic – not just by paint, but by bollards – reduced the risk of a fatal bicycle crash up to nearly one-half (Marshall and Ferenchak, 2019). Moreover, interviews regarding the specific location of a crash with emergency room patients revealed that protected lanes with driveways or alleys intersecting the lanes lead to a higher probability of a crash (Cicchino et al., 2019). A qualitative survey of self-reported comfort confirmed these findings, noting that both drivers and cyclists report greater comfort traveling in facilities that are physically separated and delineated by mode (Sanders, 2016). Yet, these assessments of safety interventions continue to be based upon one stark (and inherently biased) design variable: the count of crashes and/or fatalities before and after an intervention.

Counting crashes (and/or KSI) is the foundation of the current safety paradigm. A census of each crash severe enough to register an injury or death is compiled by the National Highway Traffic Safety Administration (e.g., Crash Outcome Data Evaluation System [CODES] and Fatality Analysis Reporting System [FARS], U.S. Department of Transportation, 2018, 2010). Each record incorporates over 140 contributory factors, encompassing characteristics about the vehicle(s), collision type and sequence of events, person(s) involved, and surrounding environment (U.S. Department of Transportation, 2018), from which researchers identify trends or shared factors to be addressed through design or policy.

While straightforward to measure and communicate, counting KSI is highly biased and imperfect. These crash and fatality data rely on the synthesis of medical and police reports, resulting in underreporting significant enough to warrant its own extensive body of research (Abay, 2015; Asgarzadeh et al., 2017; Sciortino et al., 2005; Watson et al., 2015; Wilson et al., 2012). By extension, estimating crashes is equally fraught given the incomplete and sparse underlying data: Haight (1986, 1973) notes risk should be an expectation, a continuous random variable, yet this is statistically infeasible based on the rare event data.

Moreover, the use of crash events is an inadequate safety metric as it does not report the numerous unsafe interactions between vehicles or between vehicles and those using active transportation modes: the near misses which occur at a rate of 20,000 to every one crash (Aldred, 2016; Aldred and Goodman, 2018; Gettman et al., 2008; Sanders, 2015; Simons-Morton et al., 2015). Near misses, like crashes, are significantly limited as they are rarely documented and rely primarily on self-reports on small-scale, crowd-sourced platforms (Aldred and Goodman, 2018; Branion-Calles et al., 2017; Mattingly et al., 2017). Alternative methods require capturing observational data and discerning potential and realized conflicts, including field observations of roadways and manual coding of conflicts (Glauz and Migletz, 1980), recording video of a roadway segment and screening for near misses and conflicts with variables such as change in vehicle trajectory or closeness of passing (Levine, 2017), and recording continuous driver behavior and experience *in situ*, most notably through the Naturalistic Driving Study (NDS).

Delving into near miss data highlights a fundamental issue with counting crashes: as crashes are events attributed to a confluence of factors (U.S. Department of Transportation, 2010), so too are near misses, which simply lack at least one of the multiple necessary factors to cause a crash in that instant (Sanders, 2015; Simons-Morton et al., 2015). At its most basic, the one precipitating factor delineating a near-miss from crash territory may be directly related to cognitive workload–looking away from the road, feeling emotional or stressed, or a slowed processing of information due to impairment, age, and experience (Dingus et al., 2016). Furthermore, Rendon-Velez et al., (2016) find little to no physiological difference in stress indicators for drivers in a simulator in near-crash as opposed to crash scenarios. This lack of clear difference in crash versus near crash events suggests the need for a more holistic and proactive analytic protocol that accounts for physiological safety responses as they associate with infrastructure, not only the presence, absence, or near-presence of events.

Even if all crashes and near misses were documented, safety would still not be well represented by their sums, as perceived risk keeps pedestrians and cyclists from traveling certain routes; one cannot be struck if one does not travel (Haight, 1986). Surveyed pedestrians and cyclists have been found to modify their travel route due to their perception of personal safety rather than the actual count or probability estimate of a crash (Agrawal et al., 2008; Loukaitou-Sideris et al., 2007; Saelens and Handy, 2008); elder pedestrians choose not to travel due to a lack of good walking infrastructure rather than actual crash statistics (Wasfi et al., 2012a).

## 3. Methodology and Analytic Framework

### 3.1 Experimental Design

In our experiment, participants were presented with a single task: cycle on a specific 0.3-mile urban arterial corridor in Philadelphia, PA, the city in the U.S. with the highest mode share of cycling (American Community Survey, 2016). This corridor features a bicycle lane delineated from traffic with bollards and parked cars with intermittent zones where vehicles and cyclists cross paths. Participants were outfitted with Tobii Pro Glasses 2 100 Hz eye tracking glasses, the current state-of-the-art in eye tracking. The glasses include outward-facing and inward-facing (i.e. facing the eyes of the participant) cameras to track the pupils, which continuously capture data about users' eye position and movement, pupil dilation, in addition to a gyroscope. The glasses collect, 100 times per second for each eye, the variables that have been found to be correlated with high stress and workload in the literature: gaze velocity and position; and gyroscope rotations recorded as angular velocity about the axis. Angular velocity provides a measure of a cyclist checking over their shoulder (rotation about the y-axis, akin to checking the blind spot) and looking up and down (rotation about the x-axis, akin to checking for potholes or high signs).

At any point in space, the road segments are either characterized as **protected zones**, protected by cars or bollards, which are flexible and collapsible delineators that line the bicycle lane spaced approximately 1.5 meters apart (Schepers and den Brinker, 2011); or **conflict zones**, where bicycles are not protected from vehicles and the modes must negotiate movement through a shared space. Figure 1 notes different conflict zones along the study route, with sample images of the cyclists' perspective in these areas. Note that intersections, driveway entrances, and left-turn vehicle lanes (as the cycle lane is on the left side of the street) are all conflict zones; in these locations, there are no bollards nor a painted demarcation to identify the space for bicycle throughput; cyclists must negotiate their movement with other vehicles traversing the intersection. To relate infrastructure data to our rider measurements, we examined City of Philadelphia engineering documents and coded the location of bollards, curb bumpouts, and potential mixing areas, the "conflict zones" (See Figure 2).


**Figure 1. Study route segments with noted conflict zones.**

For example, the 1600 block of JFK Boulevard features a section of bollards that starts 22.6m from the beginning of the block (which is 44.6m from the beginning of our course) and ends 71.3m from the beginning of the block (93.3m from the beginning of our course). All rider observations when the rider is between 22.6 and 71.3m on the 1600 block of JFK Boulevard are tagged as "Bollards." We similarly tag observations as belonging to a street segment or an intersection.

Forty-three subjects were recruited to ride this mission. The research team advertised the trials through various outlets, including social media and communication with the city of Philadelphia's urban-focused groups and listservs. Given the nature of the data collection–urban cycling on one of the city's highest auto-bicycle crash corridors–the research team used a self-selected opt-in strategy. Certainly, the results then represent a sample of those already comfortable with urban cycling; we will later normalize the results to make them more generalizable.  The data were collected under IRB protocol 831623.

### 3.2 Data and Methods

The research team tagged video recordings using the event function and the area of interest function of Tobii Pro Lab, Tobii's user interface software. While reviewing each participant's video in Tobii Pro Lab, we used these functions to timestamp the riders' entrance and exit to each street segment and intersection and points at which the rider is stopped at red lights. Subsequently, data were moved into a custom software environment built using the statistical computing language R. From each rider traversing our course, raw data derived from the eye tracking glasses consists of time-stamped sensor readings for eye position and movement, gyroscope, and accelerometer.

Cleaning and processing the raw data involved reducing noise in the data and, if necessary, removing subjects with high amounts of missing observations. For each subject, observations from when the subjects were stopped were removed (this is not data loss but rather isolating instances of movement as that is the focus of this study). Sporadic data loss (i.e., when the pupils are "lost" for a period from eyes being closed) present as N/A values; these values were replaced with interpolated values consistent with other eye tracking studies (Holmqvist et al., 2011; Mack et al., 2017). For 39 of the 43 subjects, this data loss was low, sporadic, and non-systematic. Recall, the eye tracking glasses collect 100 instances/second, which can be thought of as timestamps; for four subjects, missing data points represented more than 66.67% of such time stamps meaning that upwards of 2/3rds of data cells were coded as N/A.. While data loss consistent with these four subjects – loss of more than 66.67% of data – is not uncommon in eye tracking research (Holmqvist et al., 2011), we exclude these four subjects completely from our analysis. It is possible that subjects with this high a level of data loss interfered with the eye tracking glasses after the glasses were calibrated to their pupils, or that there was some other physiological obstructions interfering with the data collection (i.e. eyelashes) (Burmester and Mast, 2010; Komogortsev et al., 2010; Nielsen and Pernice, 2009). After cleaning, these 39 riders' data consisted of 3,769,142 observations. The mean number of observations per individual was 96,644.

Raw Tobii data is timestamped and wholly independent of location. To relate observations to infrastructure, we must connect the data with the physical conditions along the course. We do this through a two-step process. First, we interpolated the locations of our timestamped observations using a process we call "distance standardization." Second, we coded the locations of the "protected zones" of infrastructure by creating and analyzing street design schematics based on the City of Philadelphia's plans and engineering documents, as well as observations of the project *in situ*. To perform distance standardization, we stretched the data of each cyclist to a uniform number of observations per meter (in this case, 24), using linear interpolation to fill in the gaps. Knowing the timestamp for each subject's entry and exit into each street segment, we interpolated each rider's sensor measurements in space and then compared our interpolated data at specific timestamps to actual data at said timestamps in order to determine the accuracy of the process. The result is a data set in which all participants' data consist of the same number of observations, and each observation is related to a distance. At each distance, we have interpolated measurements for every single rider. This amounts to 29,165 observations per rider over the length of a course just under 1,216 meters in length. These observations are then spatially coded based on location along the course and as falling within either protected or conflict zones.

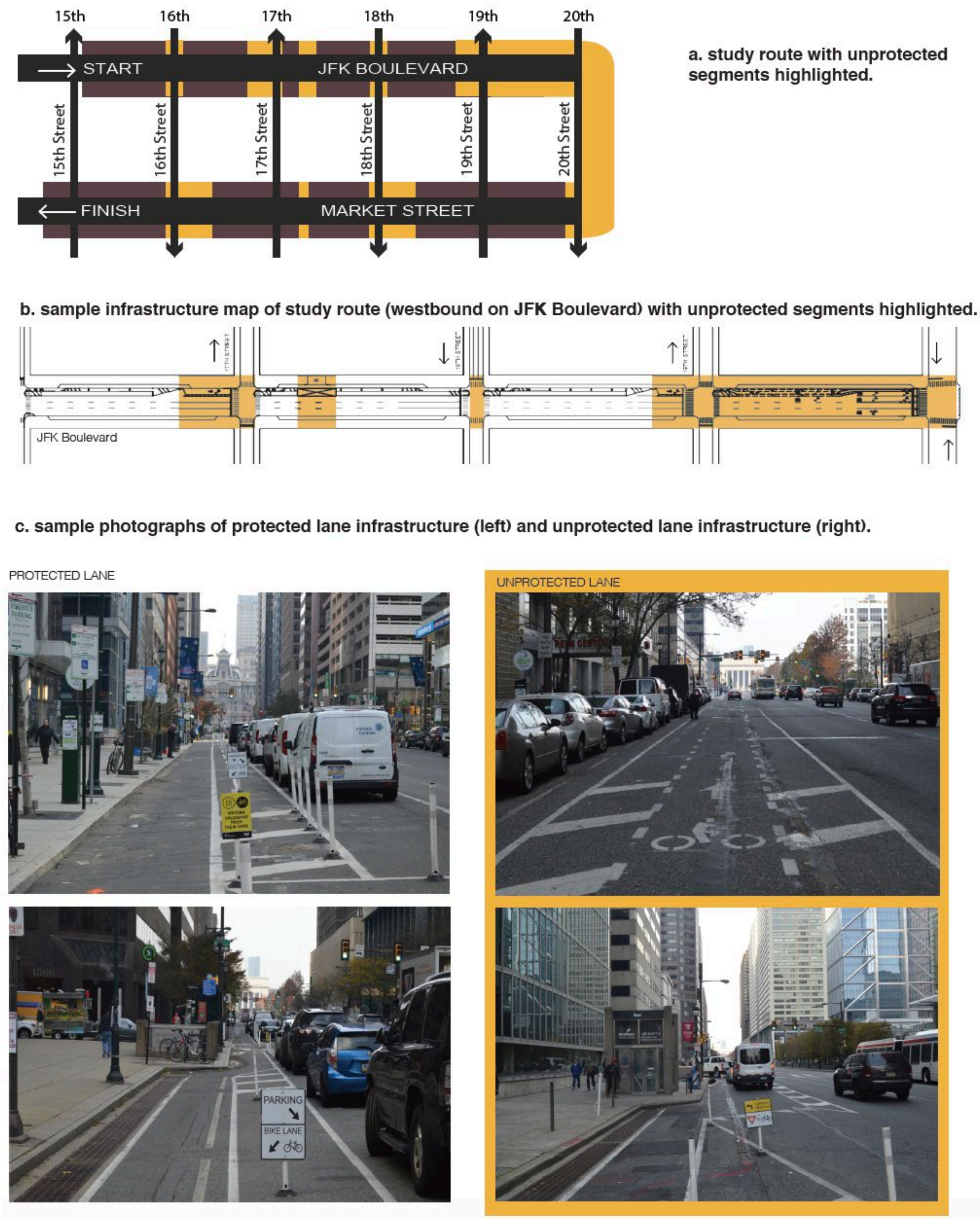



**Figure 2. Study Segment. a,** The study corridor. Road segments are either characterized as protected by bollards, which are flexible and collapsible delineators that line the bicycle lane spaced approximately 1.5 meters apart, or being conflict zones (highlighted in yellow), where bicycles and vehicles mix, and bicycles are not protected. **b,** sample infrastructure map with conflict zones highlighted**. c,** Sample photographs of protected and unprotected segment cross-sections.

After standardizing observations for distance and coding for spatial location, we then measure differences in key workload values across three dimensions: first, we describe the differences in means across all riders for the four workload indicators – gaze velocity, off mean gaze point, and gyroscope angles – when measured in protected compared to conflict zones. Second, we assess whether the range of values for workload indicators varies systematically for individual riders, and then measure *in-rider normalized* values for each of the workload variables and the number of outliers across these variables based on spatial colocation with protected versus conflict zones. Third, and finally, we employ *k*-means cluster analysis (Hastie et al., 2009) on the rider normalized workload variables to measure how these four workload indicators vary together, and categorize these groupings of movements into typologies to determine different characteristics of stress based on locations along the course.

## 4. Correlation between Workload, Infrastructure Design, and Individual Micro-Behavior

### 4.1 Aggregate Differences in Workload in Protected Versus Conflict Zones

First consider the data for each participant to occur in two groups: biometric responses when cyclists are in a lane protected by bollards, and the responses when they are not protected by bollards (Figure 3). The position points of participants' gaze in 2D space and the speed registered by the gyroscope in 2D space are more dispersed when experienced within unprotected segments compared to protected segments. For gaze position measured along the x-axis (Figure 3a), which is the distance (in coordinate units) along the horizontal plane, the data collected when bollards are not present show a greater number of points at the extreme right of the plot ($\mu$(unprotected)=948.93, $\sigma$(unprotected)=228.80). This indicates a high frequency of gazes to the right side of the cyclists; as vehicle traffic is on the right of the bicycle lane, cyclists are looking at traffic more in unprotected areas. When the data is collected in protected areas, the data appear to have a significantly more intense central tendency and less dispersion on the sides, particularly the right side ($\mu$(protected) = 939.05, $\sigma$(protected) = 183.49, $p<.001$ for both difference of means and standard deviation). Measures of off-mean gaze distance, the Euclidian measure of deviation from the riders' central tendency, and the gaze velocity, also indicate significantly smaller and more bounded gaze activity while cyclists are inside protected zones rather than unprotected.

Figure 3b presents gyroscope movements, with lateral head movements plotted along the x-axis and vertical head movements plotted along the y-axis. For vertical movements, a comparatively thicker and longer band of data is observed for the protected category ($\mu$(protected = -0.42 rad/sec, $\sigma$(protected) = 26.81 rad/sec) compared to the unprotected category ($\mu$(unprotected) = 0.65 rad/sec, $\sigma$(unprotected) = 42.22 rad/sec). Similarly, the data collected in conflict zones appear to have a larger range of lateral movements ($\mu$(protected = 0.11, $\sigma$(protected) = 14.11; $\mu$(unprotected) = 0.21, $\sigma$(unprotected) = 14.40). ANOVA tests revealing the difference in standard deviations for both lateral and vertical gyroscope readings ($p<0.001$, $p<0.001$, respectively) indicate that protected bicycle lanes are correlated with cyclists making fewer and more contained head movements. However, it is notable that the range of lateral readings is much larger than the range of vertical readings; this indicates that the need to check above (signage, traffic lights) and below (potholes) is fairly consistent across infrastructure designs, while the presence of bollards reduces the need for shoulder checks.

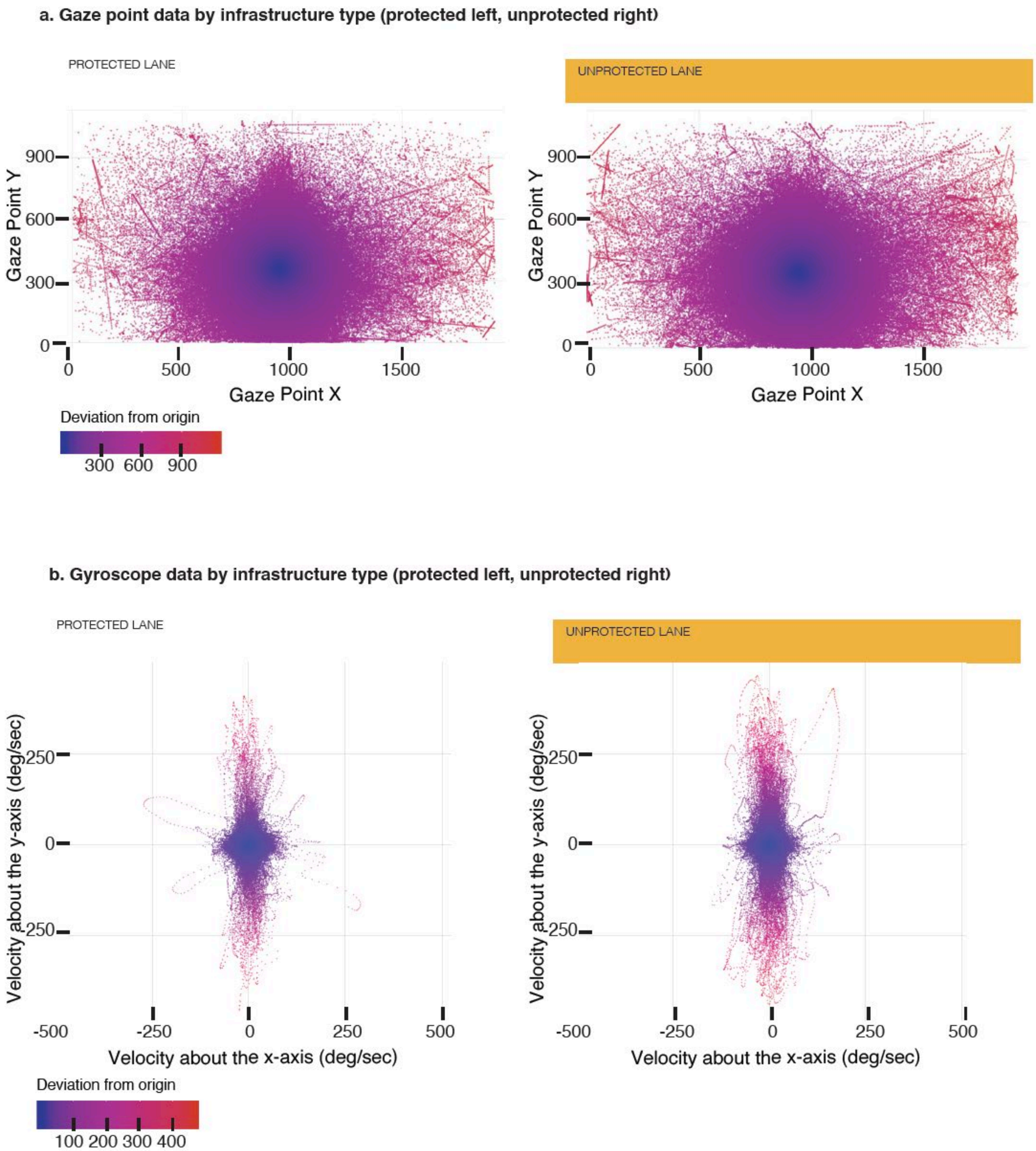



**Figure 3. Range of Biometric Responses Experienced (along Segments Only, Not Intersections). a,** Gaze position in 2D space divided into the categories of protected zones (wherein cyclists are physically separated from other modes) and conflict zones along the segment. The range of values along the x-axis is significantly larger when collected during periods of a cyclist traversing a conflict zone compared to a protected area, indicating that travelers must gaze in the direction of traffic more frequently when they are in a conflict zone. While it is hard to discern the discrepancy in the range of gaze data along the y-axis by inspecting the graph, a statistical test confirms that there is a higher workload, both in the mean and the standard deviation, while travelers are in conflict zones. **b,** Gyroscope readings with lateral head movements plotted along the x-axis and vertical head movements plotted along the y-axis divided into the categories of protected and unprotected along street segments only. The range of values along both axes are significantly larger when the data is collected in conflict zones compared to protected zones. The difference in angle range is larger for the

lateral readings. This indicates that regardless of whether a traveler is in a conflict zone or a protected zone, movements representing a ground check are similar; this is not the case for movements constituting checking over one's shoulder, which happen more frequently and to a larger magnitude when travelers are in conflict zones

### 4.2 In-rider Variation in Workload Variables

Given these signals of overall workload differences, the empirical question is then how the measures of cognitive workload can be used to identify the infrastructure designs of greatest and least safety. Once "high" workload has been established, each road segment can be characterized based on the extent and frequency of high workload movements that occurred. However, the exact character of "high workload" of movement is hard to identify. Previous research considers "high" cognitive workload as either a binary–for example, a pedestrian is looking at their phone or not (Schwebel et al., 2012)–or a based on a threshold or an action, such as classifying a distracted movement as any value above a certain threshold (Neider et al., 2011). We instead assess if, or to what extent, the angle constituting a distracted movement varies across participants. For each variable of interest, we conduct an N-wise test to study the equality of means and the equality of standard deviations across participants (Table 1). Difference of means and standard deviation tests were performed, testing n=39 participants against all others individually. In an overwhelming percentage of cases for a two-tailed test regarding the gaze variables, we reject the null hypothesis that the means and standard deviations are equal; this indicates travelers' eye movements are significantly different in terms of their absolute readings and the range of their readings. Travelers' head movements are also different in a majority of comparisons. Cyclists in our sample have biometric responses that are too different to characterize workload using a simple binary or threshold for workload. Table 1 illustrates the percent of paired participants with tests that yield a result in a difference of means or ANOVA test that are statistically significant for each variable are presented. This suggests that riders vary in their measurement ranges and central tendencies and that normalization of data by rider will help detect extreme values. By examining the 99$^{th}$ percentile of each rider's measurements in the aggregate, a conservative measure of high workload can be created.

**Table 1. Difference of Means and Outlier Values for Biometric Variables.**

| Variable n(riders)=39 | T-test comparisons $p < 0.05$ (%) | ANOVA comparisons $p < 0.05$ (%) | Outlier Value Mean | Outlier Percentile Mean (Minimum) |
|---|---|---|---|---|
| Angular Velocity | 85.7 | 92.4 | 34.84 | 96.86 (86) |
| Off Mean Gaze Distance | 93.5 | 95.9 | 398.52 | 98.99 (96) |
| Lateral Gyroscope Angle | 46.3 | 63.0 | 38.33 | 99.80 (97) |
| Vertical Gyroscope Angle | 59.4 | 76.8 | 25.62 | 96.97 (86) |

### 4.3 Spatial Locations of Workload Changes

With the understanding that individual riders are quite different across each workload indicator, we can revisit linking micro behavior to infrastructure designs while accounting for rider differences. Understanding that riders vary in physical stature, cycling experience, and agility, we conducted "in-rider normalization" such that each subject's most extreme stress measurements could be detected. We map each data point for each traveler to a percentile measurement from the pooled data set. In Figure 4a we describe the interquartile range of in-rider normalized values for each of the four workload variables, against the location of conflict zones along the course. We find a general pattern of higher values across all indicators at locations of higher conflict such as intersections, and areas without bollards.

Next, we analyze the location of values above 1.5 times the interquartile range, which we determine to be substantial outliers and signals of high stress (1.5IQR). Values above 1.5IQR represent percentile values of 96-99 percent; outliers are in the top 1-4% of values experience for each participant for each variable. A conservative estimate of an extreme value for each variable is the 99th percentile. At each point along the travel corridor, we sum the count of top 1% of values experienced across travelers and plot them for each variable (Figure 4b). We see that a relatively high sum of top percentile values is experienced in conflict zones. The count of top percentile values experienced when travelers are in protected zones are significantly smaller. Isolating extreme values for gyroscope angle (vertical and horizontal) and off-mean gaze points allows for the identification of significant peaks for each variable. Elevated occurrence rates of gaze velocity, gyroscope angle (vertical and horizontal), and off-mean gaze points are seen when approaching intersections, leaving protected zones, or turning corners. These results confirm findings from qualitative survey data (Sanders, 2016, 2015) that had yet to be empirically validated. If such a correlation exists, specifically if it exists across all participants, then infrastructure design directly correlates with cyclists' cognitive workload.

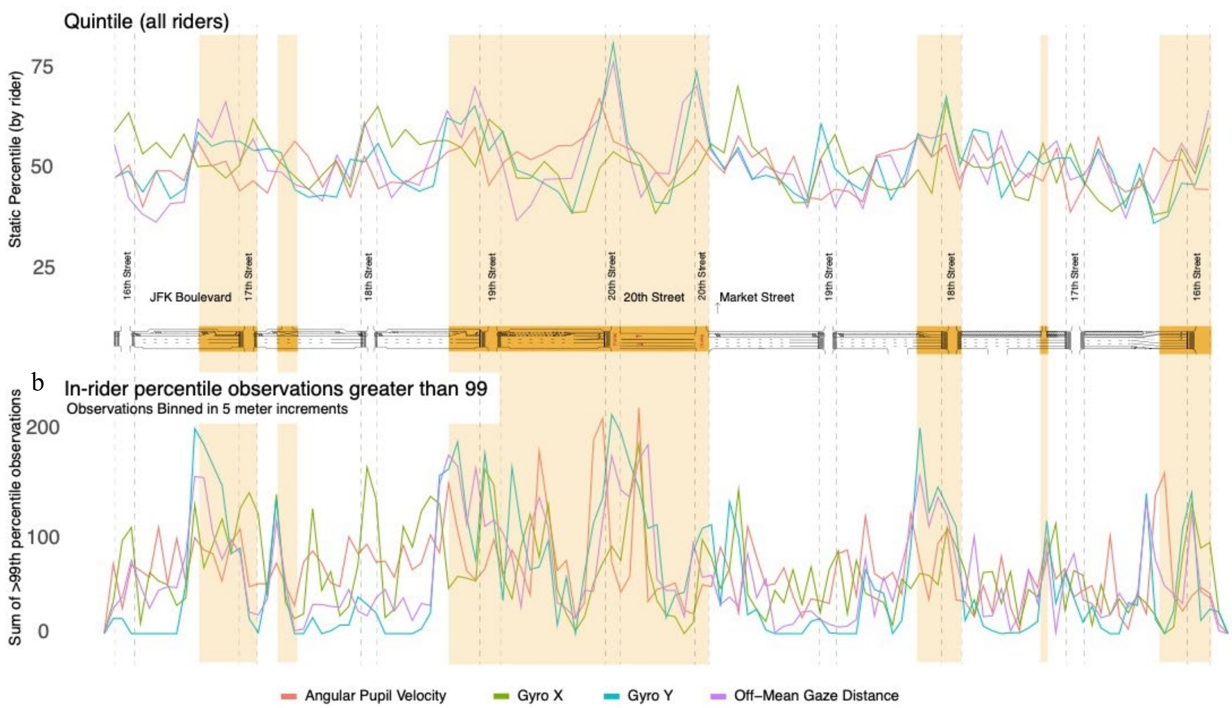



**Figure 4. Comparison of riders' biometric responses while traversing the study corridor; yellow areas signify "conflict zones".** a. The inter-quartile range of values for each rider, for each of the variables of interest, are calculated and plotted over the length of the corridor. The interquartile range describes where the middle range of values. We find that even when excluding outliers the mid-50% tend to exhibit higher values across indicators at areas of higher conflict. b. The top percentile values for each rider, for each of the variables of interest are calculated and plotted over the length of the corridor. Isolating extreme values for gyroscope acceleration (vertical and horizontal) and off-mean gaze points allows for the identification of significant peaks for each variable. These peaks are observed most in the conflict zones; in contrast, there are very few extreme values observed over the space the bicycle lane is protected by bollards.

### 4.4 Workload Movement Typologies

Typologies of movement allow us to identify the ways in which workload metrics – gaze velocity, off mean gaze point, and gyroscope angles – are experienced jointly, and how these covarying movements correlate with changes in the built environment. We conduct a *k*-means unsupervised clustering analysis to classify "types" of observations using the *mclust* package in R (Scrucca et al., 2016). We start by building a downsampled data set of 10,000 observations, and then input our stress indicators as cluster variables. We then apply the *k*-means "Lloyd" algorithm to this group of variables and generate 1,2…20 clusters, in 50 iterations. One of the key features of a *k*-means analysis is that the researcher must determine the number of *k* clusters *a-priori.* In order to determine the optimal number of clusters to describe our data, we utilize an "elbow plot" to locate at what number of clusters *k* do we begin to lose the ability to lower the within cluster sum of squared errors (WSS). This is, roughly, the "elbow" of the curve when plotting WSS against the number of clusters *k*. This is an established, highly utilized method for identifying the optimal number of clusters in the statistical literature (Hastie et al., 2009; Tibshirani et al., 2001).

Figure 5. describes the elbow plot used for determining the optimal number of clusters utilizing the stress indicators as cluster variables (left plot), as well as a plot of the lag of WSS (right plot). Both plots indicate that the reduction in WSS significantly tapers between 5 and 10 clusters; the WSS lag shows two clear inflection points at *k*=6 and *k*=8.

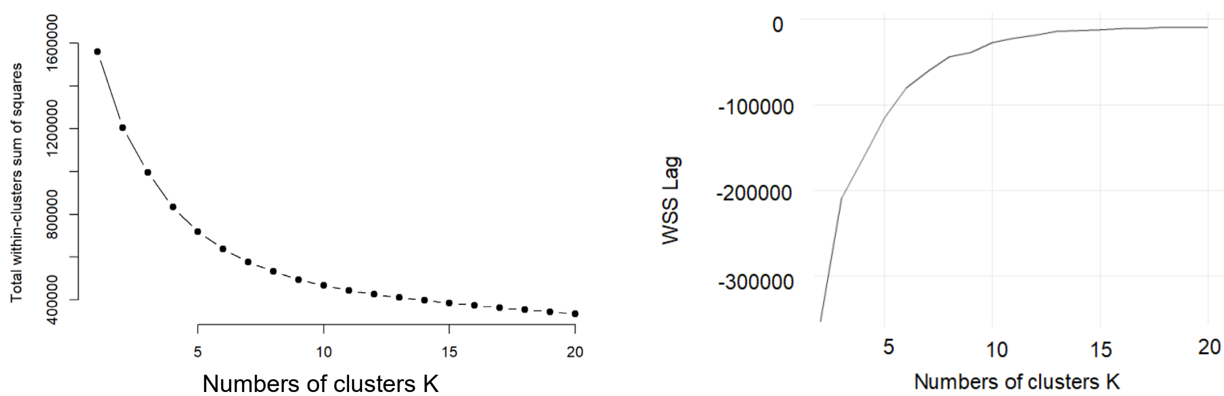



**Figure 5. "Elbow plot" of the within sum of squares (left) and the lag of the within sum of squares vs. the number of clusters "k".**

The cluster analysis identified that an optimal number of clusters hovers between six and eight. It is common practice when conducting cluster analysis to, when faced with range of possible levels of *k,* utilize the smaller, rather than larger, number of clusters. A smaller number of clusters allows the researcher to better glean differences between typologies, while higher values of *k* can be overly-specific at the expense of distinction between groups in the data (Hastie et al., 2009). As the goal is to isolate high and low workload clusters, rather than determine a gradient of workload values; we utilize the characteristics of clusters when *k*=6, the smaller value of the range of values for k.

The *k*-means cluster analysis on the four workload indicators as a vector identified six (*k*=6) distinct clusters, or typologies, of movement along the 1,137,396 four-dimension vectors. We analyze each of the clusters to assess their associations with stress metrics and stress frequency, as well as spatial location relative to infrastructure changes. These six typologies of biometric responses to infrastructure at any one point in space can be represented by the centroid mean values each variable across the cluster, as described in Figure 6a. Figure 6 also shows the distribution of percentile values for each variable represented in each cluster (b) and the sum of observations within each cluster plotted over the study corridor (c).

Overall, we see that the clusters indicating high cognitive workload – clusters with high mean values across 3-4 variables – feature data most frequently in areas classified as conflict zones, while the clusters that indicate lower cognitive workload values occur most frequently in protected zones. The sum of observations in each cluster plotted over space indicates that in conflict zones (Figure 6c, areas highlighted in yellow), the typologies capturing high movement, cluster 1 and cluster 6, in general have the highest membership, while in protected zones, the typology capturing few and slow movements, cluster 2, has the highest membership.

In Figure 6c, the three highest membership clusters are highlighted, while the others appear in the background; below we discuss the highest membership clusters in detail and the others more briefly.

### 4.4.1 Correlation with Conflict Zones

***High Overall Movements* (*Cluster 1, Membership 18.59% of total*):** The mean values for the data points in this cluster are among the highest of all clusters across the four variables of interest. The distribution of the values present in this cluster are heavily right-skewed, indicating high values for all four variables. Of the points in this cluster, the plurality are experienced in conflict zones, which require a high workload: navigation, interaction of cars and bicyclists, and maneuvering.

***Ground Checking and Ocular Scanning* (*Cluster 6, Membership 14.24% of total*):** The mean values tend to have high gaze velocity and off-mean gaze position as well vertical movements with relatively low values for activity regarding lateral movements. This cluster captures those values which indicate travelers are checking the ground and moving their eyes quickly, but not looking over their shoulders. The occurrence of this cluster peaks at conflict zones, though in a less pronounced way compared with the "High Overall Movement" Cluster.

***Shoulder Checking and Ocular Scanning (Cluster 3, Membership 15.06% of total)*:** Variables in this cluster generally have high lateral movements and high gaze velocity with relatively lower quantities of the other variables.

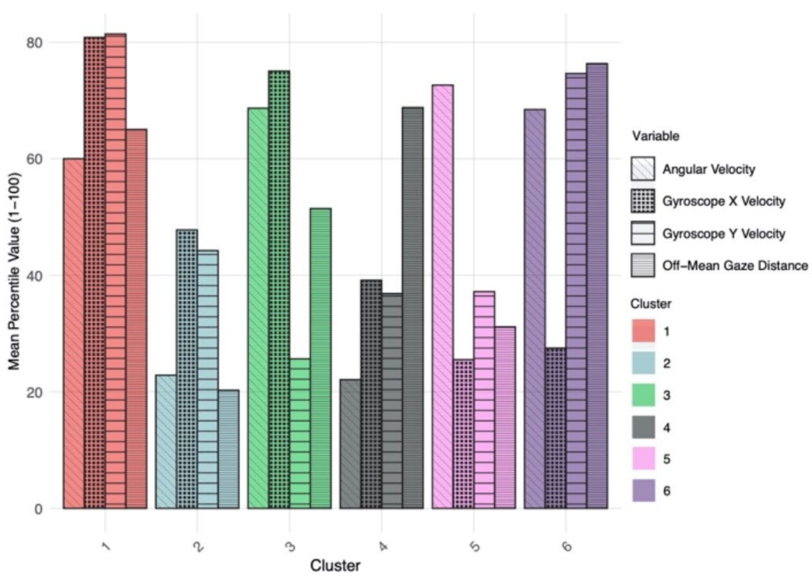
#### 4.4.2 Dispersed Correlation/Correlation with Protected Zones

***Low Overall Movements* (*Cluster 2, Membership 21.37% of total*):** The mean values for the data points that reside in this cluster are among the lowest of all clusters across the four variables of interest. The distributions for all four variables in this cluster are heavily left-skewed. The plurality of protected zone data points fall in cluster 2. The complementary nature of Cluster 1 and 2 indicate that, overall, conflict zones elicit extreme and frequent head and eye movements – the cyclist in high workload – while the protected zones are associated with a comparatively lower or more moderate cyclist workload.

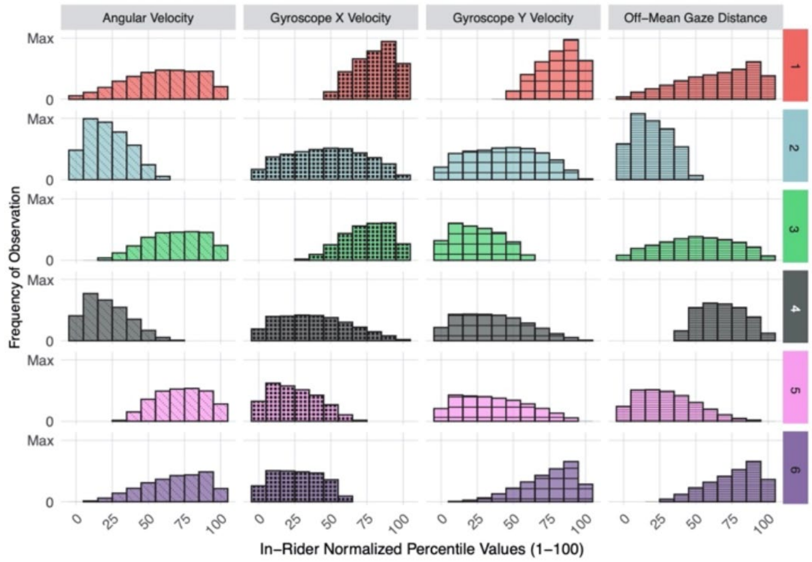
***High Off-Mean Gaze Distance (Cluster 4, Membership 15.63% of total)*:** Variables in this cluster generally have a high off-mean gaze distance and low quantities of the other variables. As can be expected, there are small spikes as a cyclist begins to traverse a segment and must take in the surrounding environment; otherwise, this cluster is experienced fairly constantly.

***High Angular Velocity of Gaze (Cluster 5, Membership 15.06% of total)*:** Variables in this cluster have a high velocity of gaze distance and low quantities of the other variables. There are sharp declines during intersections and then relatively stable, but higher, values during segments. In short, cyclists scan the area with their gaze relatively consistently while they cycle, and rest during intersections.

**a. Mean Percentile Values for In-rider Normalized Statistics by k-means Cluster**



**b. Distribution of Stress-Metric Values for Six k-means Clusters**




**c. Sum of Clusters Observations Plotted over Space (conflict/unprotected areas highlighted)**

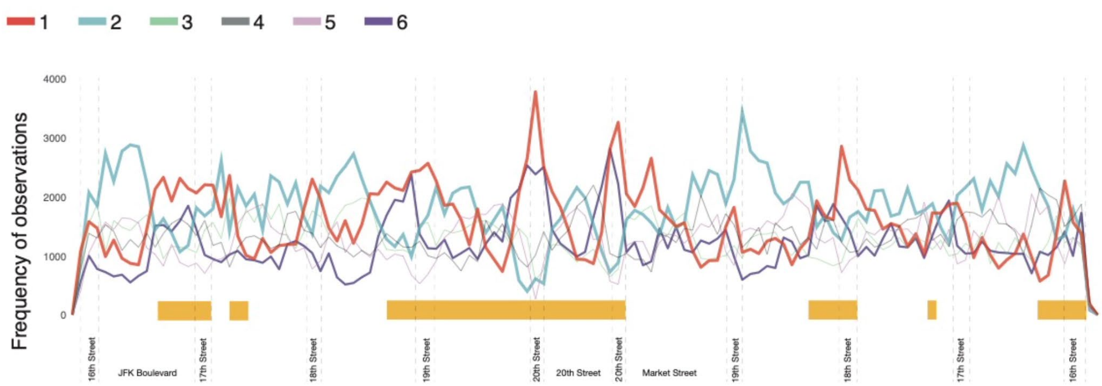



**Figure 6. Cluster Analysis of Standardized Variable Values. a,** A *k*-means cluster analysis resulted in *k=6* clusters representing like combinations of the four variables. These 6 typologies of biometric responses to infrastructure at any one point in space can be represented by the mean values each variable across the cluster. **b,** The distribution of each variable within each cluster shows the skewness or centrality of clusters. **c**, The sum of observations in each cluster plotted over space; yellow areas signify "conflict zones."

## 5. Discussion

Our manuscript pioneers a new method through which cyclist cognitive workload and, by extension, perceived safety levels can be measured and ultimately incorporated into urban transportation infrastructure design. We capture and study cyclists' microscopic behaviors in the field, on dynamic and changing urban roadways. We then employ statistical methodologies to formalize cyclists' responses at diverse articulations of urban infrastructure designs and varied traffic levels. We ultimately group these data to define new composite metrics for safety based on cyclist cognitive workload, using biometric indicators such as classic head and eye movement variables. We pair the resulting workload variables with physical attributes of the infrastructure on which they occurred (e.g., connecting workload variables with protected or unprotected lane typologies) towards quantifying the safety of different infrastructure designs with respect to how travelers behave while traversing the urban infrastructure.

The seminal literature of Frank Haight (1986) on objective and subjective risk provides a motivation to define and focus on the perception of safety. Objective risk can be properly measured by KSI, while surrogate safety measures seek to deepen our understanding of this risk by probing the characteristics of near-crashes as opposed to strictly counting crash events. Our scholarship moves this conversation into the second dimension of Haight's seminal discussion –subjective risk. Haight specifically highlights how subjective risk—or perceived risk—is the risk that is experienced by road users, and that this is the risk that ultimately informs key decisions such route choice, mode choice, crossing location, and divergence. Haight's nuanced definition of risk and perceived risk sheds light on, for example, why cyclists might avoid certain roadways Sanders' (2015) or why walkers modify their behavior to improve their perceived personal safety (Loukaitou-Sideris et al., 2007; Wasfi et al., 2012b). However, up to this point scholars have tended to only *qualitatively* confirm this realm of Haight's theory. In short, our research quantitatively defines what has been qualitatively considered in the literature for decades: that there is a latent demand for non-motorized transportation options that is being dampened by a latent risk.

Our approach begins to fill an existing and significant gap in the literature: we integrate the authentic user experience into a comprehensive and proactive paradigm for safety. We build on the call for expanding the safety research agenda to field-based research (Paschalidis et al., 2021); this approach comes with certain limitations but also allows data to be collected from travelers *in situ* rather than in a controlled environment with no real risk. In doing so, we set the foundation for a methodology that utilizes an epidemiological perspective to treat infrastructure changes as a safety intervention. Haight (1986, 1973) long called for such methods quantify risk, both perceived and objective, and writes that risk should be measured as a continuous variable rather than a stark count as this is more reflective of the way travelers consider safety. A continuous measure of risk could help transportation planners identify unsafe locations in the transportation system as they are perceived by non-automotive users; moreover, behavior-based methods can help define the most effective countermeasures as they will be used once implemented. For example, the findings from this study can inform the safety-utility of bollards, as compared to unprotected cycle lanes in an urban environment. Similar experiments could be conducted to measure any number of differences, such as bike lane color, sign placement, road segregation, and more. These studies are statistically challenging with counts of KSI (Lassarre et al., 2007).

As an additional dimension to the safety literature, we believe that new techniques and approaches must be developed to understand the characteristics of different populations of roadway users. As we have shown in this paper, our metrics are robust in *proactively* determining the stress levels of various street infrastructures on cyclists. The nature of these metrics as "before-the-fact," field-based, and in conversation with well-practiced, but at the same time conversational, statistical methodologies, allows our findings to make dramatic and immediate contributions to existing evaluations of roadway safety for cyclists, as well as the methodological literature that expands on surrogate safety metrics for non-motorized users. Our approach builds on simulator studies including (Paschalidis et al., 2019), which centers on classifying widely varying data from diverse users into recognizable groups that can inform policy changes to promote more safe environments for travel.

Confidence in the new method established herein comes from validation from numerous studies that protected cycle lanes reduce crash risk. The literature to date has correlated characteristics of bicycle lanes with data on crashes; our study has correlated cognitive workload measures with the characteristics of bicycle lanes. However, previous research is limited by the binary and retroactive nature of crash statistics. The method and approach we describe here yields continuous variables that allow for a proactive approach to measuring and defining safe infrastructure for active transportation. We find that lateral head movements and off-mean gaze distance, in concert and independently, indicate high workload, as does the combination of lateral head movements, off-mean gaze distance, and gaze velocity. These findings provide new, continuous variables on which infrastructure designs can be evaluated – either as is, or in a before and after study of new designs and countermeasures.

The biometric data collected and analyzed herein represents a fundamental shift in the way safety is measured by the entire transportation and public health community. Our study moves away from reliance on crashes and fatalities as the design variable. Instead, new variables for workload stemming from eye movements and head movements are used to understand the workload and safety from the perspective of a traveler. Unlike counting crashes, which do not tell us *why* a particular stretch of road or intersection is unsafe, these new safety metrics encompassing cyclist stress and workload indicate safe and unsafe locations with spatial precision. Studying and applying these new variables through design evaluations of existing infrastructure – and ultimately, through integration into policy guidance that informs roadway design– will equip planners and engineers with a more comprehensive understanding of where and how to build safer infrastructure.

## 6. References


Abay, K.A., 2015. Investigating the nature and impact of reporting bias in road crash data. Transp. Res. Part Policy Pract. 71, 31–45. https://doi.org/10.1016/j.tra.2014.11.002

Agrawal, A.W., Schlossberg, M., Irvin, K., 2008. How Far, by Which Route and Why? A Spatial Analysis of Pedestrian Preference. J. Urban Des. 13, 81–98. https://doi.org/10.1080/13574800701804074

Åkesson, I., Balogh, I., Hansson, G.-Å., 2012. Physical workload in neck, shoulders and wrists/hands in dental hygienists during a work-day. Appl. Ergon. 43, 803–811. https://doi.org/10.1016/j.apergo.2011.12.001

Aldred, R., 2016. Cycling near misses: Their frequency, impact, and prevention. Transp. Res. Part Policy Pract. 90, 69–83. https://doi.org/10.1016/j.tra.2016.04.016

Aldred, R., Goodman, R., 2018. Predictors of the frequency and subjective experience of cycling near misses: Findings from the first two years of the UK Near Miss Project. Accid. Anal. Prev. 110, 161–170. https://doi.org/10.1016/j.aap.2017.09.015

American Community Survey, 2016. B08006: Sex of Workers by Means of Transportation to Work.

Asan, O., Yang, Y., 2015. Using Eye Trackers for Usability Evaluation of Health Information Technology: A Systematic Literature Review. JMIR Hum. Factors 2. https://doi.org/10.2196/humanfactors.4062

Asgarzadeh, M., Verma, S., Mekary, R.A., Courtney, T.K., Christiani, D.C., 2017. The role of intersection and street design on severity of bicycle-motor vehicle crashes. Inj. Prev. 23, 179–185. https://doi.org/10.1136/injuryprev-2016-042045

Aslin, R.N., 2012. Infant Eyes: A Window on Cognitive Development. Infancy 17, 126–140. https://doi.org/10.1111/j.1532-7078.2011.00097.x

Branion-Calles, M., Nelson, T., Winters, M., 2017. Comparing Crowdsourced Near-Miss and Collision Cycling Data and Official Bike Safety Reporting. Transp. Res. Rec. J. Transp. Res. Board 2662, 1–11. https://doi.org/10.3141/2662-01

Burmester, M., Mast, M., 2010. Repeated Web Page Visits and the Scanpath Theory: A Recurrent Pattern Detection Approach. J. Eye Mov. Res. 3. https://doi.org/10.16910/jemr.3.4.5

Callan, D.J., 1998. Eye Movement Relationships to Excessive Performance Error in Aviation. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 42, 1132–1136. https://doi.org/10.1177/154193129804201516

Chen, L., Chen, C., Srinivasan, R., McKnight, C.E., Ewing, R., Roe, M., 2012. Evaluating the Safety Effects of Bicycle Lanes in New York City. Am. J. Public Health 102, 1120–1127. https://doi.org/10.2105/AJPH.2011.300319

Cicchino, J.B., McCarthy, M.L., Newgard, C.D., Wall, S.P., DiMaggio, C.J., Kulie, P.E., Arnold, B.E., Zuby, D.S., 2019. Not all protected bike lanes are the same: infrastructure and risk of cyclist collisions and falls leading to emergency department visits in three U.S. cities. Insurance Institute for Highway Safety, Arlington, VA.

Deb, S., Carruth, D.W., Sween, R., Strawderman, L., Garrison, T.M., 2017. Efficacy of virtual reality in pedestrian safety research. Appl. Ergon. 65, 449–460. https://doi.org/10.1016/j.apergo.2017.03.007

Di Stasi, L.L., Contreras, D., Cándido, A., Cañas, J.J., Catena, A., 2011. Behavioral and eye-movement measures to track improvements in driving skills of vulnerable road users: First-time motorcycle riders. Transp. Res. Part F Traffic Psychol. Behav. 14, 26–35. https://doi.org/10.1016/j.trf.2010.09.003

Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M., Hankey, J., 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proc. Natl. Acad. Sci. 113, 2636–2641. https://doi.org/10.1073/pnas.1513271113

Elsabbagh, M., Mercure, E., Hudry, K., Chandler, S., Pasco, G., Charman, T., Pickles, A., Baron-Cohen, S., Bolton, P., Johnson, M.H., 2012. Infant Neural Sensitivity to Dynamic Eye Gaze Is Associated with Later Emerging Autism. Curr. Biol. 22, 338–342. https://doi.org/10.1016/j.cub.2011.12.056

Engström, J., Johansson, E., Östlund, J., 2005. Effects of visual and cognitive load in real and simulated motorway driving. Transp. Res. Part F Traffic Psychol. Behav., The relationship between distraction and driving performance: towards a test regime for in-vehicle information systems 8, 97–120. https://doi.org/10.1016/j.trf.2005.04.012

Federal Highway Administration, 2017. Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities [WWW Document]. URL https://safety.fhwa.dot.gov/ped_bike/tools_solve/fhwasa17041/ch2.cfm (accessed 5.29.19).

FHWA, 2009. MUTCD, 2009 Edition. ed.

First, L.R., Palfrey, J.S., 1994. The Infant or Young Child with Developmental Delay. N. Engl. J. Med. 330, 478–483. https://doi.org/10.1056/NEJM199402173300708

Foy, H.J., Chapman, P., 2018. Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. Appl. Ergon. 73, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

Gettman, D., Pu, L., Sayed, T., Shelby, S., 2008. Surrogate Safety Assessment Model and Validation: Final Report (No. FHWA-HRT-08-051). Federal Highway Administration, McLean, VA.

Glauz, W.D., Migletz, D.J., 1980. Application of traffic conflict analysis at intersections, Report - National Cooperative Highway Research Program ; 219. Transportation Research Board, National Research Council, Washington, D.C.

Haight, F.A., 1986. Risk, especially risk of traffic accident. Accid. Anal. Prev. 18, 359–366. https://doi.org/10.1016/0001-4575(86)90009-6

Haight, F.A., 1973. Induced exposure. Accid. Anal. Prev., Special Issue on Accident Exposure 5, 111–126. https://doi.org/10.1016/0001-4575(73)90019-5

Hancock, P.A., Lesch, M., Simmons, L., 2003. The distraction effects of phone use during a crucial driving maneuver. Accid. Anal. Prev. 35, 501–514. https://doi.org/10.1016/S0001-4575(02)00028-3

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd edition. ed. Springer.

Hogervorst, M.A., Brouwer, A.-M., van Erp, J.B.F., 2014. Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. Front. Neurosci. 8. https://doi.org/10.3389/fnins.2014.00322

Holmqvist, K., Nyström, N., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye Tracking: A Comprehensive Guide To Methods And Measures. Oxford University Press, Oxford, UK.

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J., Brown, J., 2006. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. Accid. Anal. Prev. 38, 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Jankowska-Karpa, D., Wacowska-Slezak, J., 2016. Safe school transportation of children supported by intelligent transport systems. Inj. Prev. 22, A252–A252. https://doi.org/10.1136/injuryprev-2016-042156.702

Kapp, E.Z., 1997. Evaluation of the Risk for Work-Related Upper Extremity Musculoskeletal Symptoms in USAF Air Traffic Controllers: A Pilot Feasibility Study. THE UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER AT HOUSTON SCHOOL OF PUBLIC HEALTH.

Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A., 2014. Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers. N. Engl. J. Med. 370, 54–59. https://doi.org/10.1056/NEJMsa1204142

Knight, J.C., 2002. Safety critical systems: challenges and directions, in: Proceedings of the 24th International Conference on Software Engineering. ICSE 2002. Presented at the Proceedings of the 24th International Conference on Software Engineering. ICSE 2002, pp. 547–550.

Komogortsev, O.V., Gobert, D.V., Jayarathna, S., Koh, D.-H., Gowda, S., 2010. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. IEEE Trans. Biomed. Eng. 57. https://doi.org/10.1109/TBME.2010.2057429

Krauzlis, R.J., 2013. Chapter 32 - Eye Movements, in: Squire, L.R., Berg, D., Bloom, F.E., du Lac, S., Ghosh, A., Spitzer, N.C. (Eds.), Fundamental Neuroscience (Fourth Edition). Academic Press, San Diego, pp. 697–714. https://doi.org/10.1016/B978-0-12-385870-2.00032-9

Lassarre, S., Papadimitriou, E., Yannis, G., Golias, J., 2007. Measuring accident risk exposure for pedestrians in different micro-environments. Accid. Anal. Prev. 39, 1226–1238. https://doi.org/10.1016/j.aap.2007.03.009

Lee, J.D., McGehee, D.V., Brown, T.L., Reyes, M.L., 2002. Collision Warning Timing, Driver Distraction, and Driver Response to Imminent Rear-End Collisions in a High-Fidelity Driving Simulator. Hum. Factors 44, 314–334.

Lee, Y.-C., Winston, F.K., 2016. Stress induction techniques in a driving simulator and reactions from newly licensed drivers. Transp. Res. Part F Traffic Psychol. Behav. 42, 44–55. https://doi.org/10.1016/j.trf.2016.06.019

Levine, R., 2017. Bike lanes on Bloor Street: U of T partners with Miovision, City of Toronto to help evaluate pilot and track traffic safety [WWW Document]. Univ. Tor. Transp. Res. Inst. URL https://uttri.utoronto.ca/news/bike-lanes-bloor-street-u-t-partners-miovision-city-toronto-help-evaluate-pilot-track-traffic-safety/ (accessed 4.29.19).

Loft, S., Sanderson, P., Neal, A., Mooij, M., 2007. Modeling and Predicting Mental Workload in En Route Air Traffic Control: Critical Review and Broader Implications. Hum. Factors 49, 376–399. https://doi.org/10.1518/001872007X197017

Loukaitou-Sideris, A., Liggett, R., Sung, H.-G., 2007. Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. J. Plan. Educ. Res. 26, 338–351.

Mack, D.J., Belfanti, S., Schwarz, U., 2017. The effect of sampling rate and lowpass filters on saccades – A modeling approach. Behav. Res. Methods 49, 2146–2162. https://doi.org/10.3758/s13428-016-0848-4

Mannering, F.L., Washburn, S.S., 2012. Principles of Highway Engineering and Traffic Analysis, 5th Edition. Wiley Global Education.

Marquart, G., Cabrall, C., de Winter, J., 2015. Review of Eye-related Measures of Drivers' Mental Workload. Procedia Manuf. 3, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783

Marshall, W.E., Ferenchak, N.N., 2019. Why cities with high bicycling rates are safer for all road users. J. Transp. Health 13, 100539. https://doi.org/10.1016/j.jth.2019.03.004

Mattingly, S., Casey, C., Johnson, T., 2017. App-based Crowd Sourcing of Bicycle and Pedestrian Conflict Data (No. TRCLC15-7). Transportation Research Center.

Mehler, B., Pohlmeyer, A., Reimer, B., Coughlin, J.F., 2008. The association between heart rate reactivity and driving performance under dual task demands in late middle age drivers. Adv. Transp. Stud. Int. J, 53−70.

Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A., 2009. Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. Transp. Res. Rec. J. Transp. Res. Board 2138, 6–12. https://doi.org/10.3141/2138-02

National Center for Statistics and Analysis, 2018. 2017 Fatal motor vehicle crashes: Overview. (Traffic Safety Facts Research Note. No. Report No. DOT HS 812 603). National Highway Traffic Safety Administration, Washington, DC.

Neider, M.B., Gaspar, J.G., McCarley, J.S., Crowell, J.A., Kaczmarski, H., Kramer, A.F., 2011. Walking & Talking: Dual-Task Effects on Street Crossing Behavior in Older Adults. Psychol. Aging 26, 260–268. https://doi.org/10.1037/a0021566

Nielsen, J., Pernice, K., 2009. Eyetracking Web Usability, 1st edition. ed. Publisher : New Riders Pub.

Olmos-Solis, K., van Loon, A.M., Los, S.A., Olivers, C.N.L., 2017. Chapter 1 - Oculomotor measures reveal the temporal dynamics of preparing for search, in: Howard, C.J. (Ed.), Progress in Brain Research, Temporal Sampling and Representation Updating. Elsevier, pp. 1–23. https://doi.org/10.1016/bs.pbr.2017.07.003

Paschalidis, E., Choudhury, C.F., Hess, S., 2021. From Driving Simulator Experiments to Field-Traffic Application: Improving the Transferability of Car-Following Models. J. Transp. Eng. Part Syst. 147, 04020145. https://doi.org/10.1061/JTEPBS.0000468

Paschalidis, E., Choudhury, C.F., Hess, S., 2019. Combining driving simulator and physiological sensor data in a latent variable model to incorporate the effect of stress in car-following behaviour. Anal. Methods Accid. Res. 22, 100089. https://doi.org/10.1016/j.amar.2019.02.001

Pucher, J., Buehler, R., Seinen, M., 2011. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. Transp. Res. Part Policy Pract. 45, 451–475. https://doi.org/10.1016/j.tra.2011.03.001

Räsänen, 1999. Car Driver and Bicyclist Behavior at Bicycle Crossings Under Different Priority Regulations. J. Safety Res. 30, 67–77. https://doi.org/10.1016/S0022-4375(98)00062-0

Rendon-Velez, E., van Leeuwen, P.M.., Happee, R., Horváth, I., van der Vegte, W.F., de Winter, J.C.F., 2016. The effects of time pressure on driver performance and physiological activity: A driving simulator study. Transp. Res. Part F Traffic Psychol. Behav. 41, 150–169. https://doi.org/10.1016/j.trf.2016.06.013

Retting, R., 2018. Pedestrian Traffic Fatalities by State. Governors Highway Safety Association, Washington, D.C.

Saelens, B.E., Handy, S.L., 2008. Built Environment Correlates of Walking: A Review. Med. Sci. Sports Exerc. 40, S550–S566. https://doi.org/10.1249/MSS.0b013e31817c67a4

Sanders, R.L., 2016. We can all get along: The alignment of driver and bicyclist roadway design preferences in the San Francisco Bay Area. Transp. Res. Part Policy Pract. 91, 120–133. https://doi.org/10.1016/j.tra.2016.06.002

Sanders, R.L., 2015. Perceived traffic risk for cyclists: The impact of near miss and collision experiences. Accid. Anal. Prev. 75, 26–34. https://doi.org/10.1016/j.aap.2014.11.004

Schepers, P., den Brinker, B., 2011. What do cyclists need to see to avoid single-bicycle crashes? Ergonomics 54, 315–327. https://doi.org/10.1080/00140139.2011.558633

Schmitt, A., 2019. Seattle Tosses Out Rulebook to Protect Pedestrians. Streetsblog USA. URL https://usa.streetsblog.org/2019/02/05/seattle-tosses-out-the-rulebook-to-protect-pedestrians/ (accessed 5.29.19).

Schwebel, D.C., Stavrinos, D., Byington, K.W., Davis, T., O'Neal, E.E., de Jong, D., 2012. Distraction and Pedestrian Safety: How Talking on the Phone, Texting, and Listening to Music Impact Crossing the Street. Accid. Anal. Prev. 45, 266–271. https://doi.org/10.1016/j.aap.2011.07.011

Sciortino, S., Vassar, M., Radetsky, M., Knudson, M.M., 2005. San Francisco pedestrian injury surveillance: Mapping, under-reporting, and injury severity in police and hospital records. Accid. Anal. Prev. 37, 1102–1113. https://doi.org/10.1016/j.aap.2005.06.010

Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E., 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8, 205–233.

Simons-Morton, B.G., Klauer, S.G., Ouimet, M.C., Guo, F., Albert, P.S., Lee, S.E., Ehsani, J.P., Pradhan, A.K., Dingus, T.A., 2015. Naturalistic Teenage Driving Study: Findings and Lessons Learned. J. Safety Res. 54, 41.e29-41.e44. https://doi.org/10.1016/j.jsr.2015.06.010

Stephens, A.N., Groeger, J.A., 2014. Following slower drivers: Lead driver status moderates driver's anger and behavioural responses and exonerates culpability. Transp. Res. Part F Traffic Psychol. Behav. 22, 140–149. https://doi.org/10.1016/j.trf.2013.11.005

Szeto, G.P.Y., Cheng, S.W.K., Poon, J.T.C., Ting, A.C.W., Tsang, R.C.C., Ho, P., 2012. Surgeons' static posture and movement repetitions in open and laparoscopic surgery. J. Surg. Res. 172, e19-31. https://doi.org/10.1016/j.jss.2011.08.004

Tapiro, H., Oron-Gilad, T., Parmet, Y., 2020. Pedestrian distraction: The effects of road environment complexity and age on pedestrian's visual attention and crossing behavior. J. Safety Res. 72, 101–109. https://doi.org/10.1016/j.jsr.2019.12.003

Tapiro, H., Oron-Gilad, T., Parmet, Y., 2018. The effect of environmental distractions on child pedestrian's crossing behavior. Saf. Sci. 106, 219–229. https://doi.org/10.1016/j.ssci.2018.03.024

Tibshirani, R., Walther, G., Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. J. R. Stat. Soc. Ser. B Stat. Methodol. 63, 411–423. https://doi.org/10.1111/1467-9868.00293

Tien, T., Pucher, P.H., Sodergren, M.H., Sriskandarajah, K., Yang, G.-Z., Darzi, A., 2014. Eye tracking for skills assessment and training: a systematic review. J. Surg. Res. 191, 169–178. https://doi.org/10.1016/j.jss.2014.04.032

Topolšek, D., Areh, I., Cvahte, T., 2016. Examination of driver detection of roadside traffic signs and advertisements using eye tracking. Transp. Res. Part F Traffic Psychol. Behav. 43, 212–224. https://doi.org/10.1016/j.trf.2016.10.002

U.S. Department of Transportation, 2018. Fatality Analysis Reporting System (FARS) Analytical User's Manual (No. DOT HS 812 602). U.S. Department of Transportation, Washington, D.C.

U.S. Department of Transportation, 2010. The Crash Outcome Data Evaluation System (CODES) and Applications to Improve Traffic Safety Decision-Making (No. DOT HS 811181). National Highway Traffic Safety Administration, Washington, D.C.

Wall, S.P., Lee, D.C., Frangos, S.G., Sethi, M., Heyer, J.H., Ayoung-Chee, P., DiMaggio, C.J., 2016. The Effect of Sharrows, Painted Bicycle Lanes and Physically Protected Paths on the Severity of Bicycle Injuries Caused by Motor Vehicles. Safety 2, 26. https://doi.org/10.3390/safety2040026

Wasfi, R.L., Levinson, D., El-Geneidy, A.B., 2012a. Measuring the transportation needs of seniors. J. Transp. Lit. 6, 8–32.

Wasfi, R.L., Levinson, D., El-Geneidy, A.B., 2012b. Measuring the transportation needs of seniors. J. Transp. Lit. 6, 8–32.

Watson, A., Watson, B., Vallmuur, K., 2015. Estimating under-reporting of road crash injuries to police using multiple linked data collections. Accid. Anal. Prev. 83, 18–25. https://doi.org/10.1016/j.aap.2015.06.011

Wilson, B., Hoffman, J., Morgenstern, J., 2019. Predictive Inequity in Object Detection. ArXiv190211097 Cs Stat.

Wilson, S.J., Begg, D.J., Samaranayaka, A., 2012. Validity of using linked hospital and police traffic crash records to analyse motorcycle injury crash characteristics. Accid. Anal. Prev., PTW + Cognitive impairment and Driving Safety 49, 30–35. https://doi.org/10.1016/j.aap.2011.03.007

Yi-Fang Tsai, Viirre, E., Strychacz, C., Chase, B., Jung, T.-P., 2007. Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload 78, 10.

Yu, D., Dural, C., Morrow, M.M.B., Yang, L., Collins, J.W., Hallbeck, S., Kjellman, M., Forsman, M., 2017. Intraoperative workload in robotic surgery assessed by wearable motion tracking sensors and questionnaires. Surg. Endosc. 31, 877–886. https://doi.org/10.1007/s00464-016-5047-y

Yu, D., Lowndes, B., Morrow, M., Kaufman, K., Bingener, J., Hallbeck, S., 2016. Impact of novel shift handle laparoscopic tool on wrist ergonomics and task performance. Surg. Endosc. 30, 3480–3490. https://doi.org/10.1007/s00464-015-4634-7

# Rethinking Urban Transportation Safety Metrics: From Counting Crashes to Estimating Cognitive Workload

Megan S. Ryerson*[1,2] and Carrie S. Long[1]

[1]*Department of City and Regional Planning, PennDesign, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.*

[2]*Department of Electrical and Systems Engineering, PennEngineering, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.*

## 1. Introduction

The introduction of new shared urban transportation modes—enabled by a combination of technological, institutional, and cultural forces—have expanded and transformed mobility. In 2018, travelers across the U.S. took 84 million trips on bike share and electric scooter (e-scooter) share, more than twice the number of trips taken the previous year. Of these trips, nearly half were on shared e-scooters; a remarkable two-year growth given that docked bikeshare ridership took nearly a decade to reach that level. Yet greater numbers of e-scooterists, as well as cyclists, pedestrians, and ride-hail vehicles sharing the rights-of-way has resulted in conflicts, incidents, and crashes.

Crashes amongst travelers in the urban environment is not new; e-scooters joined a transportation system that was already struggling with the degradation of safety. In 2018, as e-scooters made their debut, the U.S. grappled with critical safety concerns: pedestrian fatalities had risen 27% since 2007, and in one year alone (2017-18) cyclist fatalities rose 10%, ranking as the deadliest year ever recorded[1]. As shared and private e-scooters proliferated, reports of significant injuries and fatalities followed. Comprehensive safety studies using crash counts from a combination of trauma registries and emergency room complaint data[2–4] and the National Electronic Injury Surveillance System (NEISS)[5] have been published, summarizing the types of injuries sustained and correlating the relationship between crashes and demographics. The Centers for Disease Control and Prevention (CDC) analysis of injured e-scooterists in Austin, TX over a two month period coupled the count and nature of crashes with data from the Mobility Data Standard (MDS), an open source tool that enables governments to collect and analyze mobility data in aggregate (number of trips, number of miles, etc.). In doing so, the CDC calculated a trip- based injury rate of 20 injured e-scooterists per 100,000 trips in Austin[3].

E-scooters present a transportation safety conundrum: shared e-scooter systems offer a new mode that is broadly accessible with a low barrier to entry as anyone with a driver's license can rent and ride; yet, many cities lack networks of separated shared lanes that mitigate the risk of crashes with cars or pedestrians, and officials are unsure if e-scooters, traveling up to 15mph, belong in those lanes. E-scooters bring the lack of safety in our urban transportation system into sharp focus, giving impetus for cities to contend with core, fundamental questions: what is the state of safety in our urban transportation system, and what countermeasures, improvements, or regulations should we enact to improve safety outcomes and reduce the risk of injury? More broadly, what are – and aren't – the safety studies telling us?

## 2. Changing the Safety Framework from Crashes to Workload

The e-scooter safety studies (as well as a long line of multimodal transportation studies on cycling and pedestrian safety[6]) are founded on counts of crashes and injuries. Engineers and planners count crashes, injuries, or fatalities (known as killed or seriously injured, or KSI) to characterize the safety of the transportation system. There are, however, fundamental limitations to the crash-based method. It is well known that travelers in all modes modify or restrict their travel routes due to their perception of personal safety, and one cannot be struck if one does not travel; the function of safety as an enabler of or impedance to mobility is not captured by KSI. Crashes are significantly underreported such that there is a field dedicated to statistically correcting the counts[7]. Finally, a reported crash only tells us the result of unsafe interactions, not the full story of the event. The low – and depressed – count of crashes and the lack of information about their underlying cause precludes testing the effectiveness of safety countermeasures.

Crash events do not represent all unsafe interactions: according to the Federal Highway Administration, there are 20,000 near-misses (unsafe interactions between roadway users) to every one crash[8]. Near misses are more common than crashes, but rarely documented and, as a result, the field of transportation misses out on critical insights on the safety of roads and intersections. Consider that near misses are a cornerstone of proactive safety in other fields: the medical community celebrates near misses as “good catches” and encourages health care providers to report near misses and data; groups of physicians debrief and analyze these data and insights at Morbidity and Mortality Conferences in order to learn how to prevent such events from happening again.

It is worth challenging the efficacy of using KSIs as a standard metric and, to do so, we compare KSIs to a corollary in another mode of transportation. Measuring air transportation safety as a function of KSI is both impossible (due to the low numbers) and politically infeasible (air transportation crashes are not an acceptable occurrence in modern society, in comparison to surface transportation crashes. Instead, the Federal Aviation Administration (FAA) uses a proactive approach that *ensures* safety, rather than *measures* safety: the FAA routinely estimates the cognitive workload of the air traffic controllers (ATCs) tasked with keeping the aircraft safely separated and managed. In doing so, the FAA measures safety proactively by examining how much their operators can mentally handle, and designing the system accordingly. The medical community also uses cognitive workload as a way to proactively manage safety: numerous studies investigate a physician’s ability to handle tasks, finding a relationship between cognitive workload and error rate[9].

## 3. Safety as Workload that is Minimized and Managed

The fields of medicine and air transportation have *established* that a person engaged in a safety critical act is more prone to error with a higher cognitive workload, and they have *implemented* operational cultures and procedures to minimize and manage this workload. While drivers in simulators have confirmed the relationship between driver workload and errors, the field has not implemented a workload-based safety culture. As a result, the federally-mandated transportation guidance on the design of countermeasures does not include a hierarchy of designs or an evaluation process through which the effectiveness of countermeasures can be measured.

The current paradigm of safety focused on measuring objective risk by counting crashes has led to a reactive, flawed method for both prioritizing the placement of infrastructure interventions and for evaluating their impact on safety. The planners, engineers, and decision makers responsible for the safety of our transportation system need new methods as the system is being infused with e-scooters: a new micromobility mode with a learning curve[3] operated at relatively high speed that can both unlock mobility and endanger the traveling public.

The public health community has set the framework for new methods to measure safety of our urban transportation infrastructure based on workload and stress; moreover, they have built a safety-critical culture of learning from rare incidents and near-misses. This proactive and workload-based understanding of safety could provide invaluable information within transportation setting, towards prediction and prevention. The community of transportation officials–federal and state officials crafting guidance and local officials implementing this guidance–should consider building a system that minimizes and manages workload proactively. With a *reactive*, crash-based model we can see tragic cost of unsafe infrastructure, as measured in lives lost and changed; thorough analyses of the factors associated with each event are published as a proxy for causation. It is time to shift to a *proactive* paradigm where human factors, workload, and perceived risk shape how we understand, design for, and prioritize transportation safety.

## 4. References


1. National Center for Statistics and Analysis. *2017 Fatal Motor Vehicle Crashes: Overview.* Washington, DC: National Highway Traffic Safety Administration; 2018. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812603.
2. Kobayashi LM, Williams E, Brown CV, et al. The e-merging e-pidemic of e-scooters. *Trauma Surg Acute Care Open*. 2019;4(1). doi:10.1136/tsaco-2019-000337

3. Morano L. *Dockless Electric Scooter-Related Injuries Study — Austin, Texas, September–November 2018*. Austin, TX: Austin Public Health; 2019:15. http://austintexas.gov/sites/default/files/files/Health/Web_Dockless_Electric_Scooter-Related_Injury_Study_final_version_EDSU_5.14.19.pdf. Accessed November 21, 2019.
4. Trivedi TK, Liu C, Antonio ALM, et al. Injuries Associated With Standing Electric Scooter Use. *JAMA Netw Open*. 2019;2(1):e187381-e187381. doi:10.1001/jamanetworkopen.2018.7381
5. Aizpuru M, Farley KX, Rojas JC, Crawford RS, Moore TJ, Wagner ER. Motorized scooter injuries in the era of scooter-shares: A review of the national electronic surveillance system. *Am J Emerg Med*. 2019;37(6):1133-1138. doi:10.1016/j.ajem.2019.03.049
6. Saelens BE, Handy SL. Built Environment Correlates of Walking: A Review. *Med Sci Sports Exerc*. 2008;40(7 Suppl):S550-S566. doi:10.1249/MSS.0b013e31817c67a4
7. Haight FA. Risk, especially risk of traffic accident. *Accid Anal Prev*. 1986;18(5):359-366. doi:10.1016/0001-4575(86)90009-6
8. Gettman D, Pu L, Sayed T, Shelby S. *Surrogate Safety Assessment Model and Validation: Final Report*. McLean, VA: Federal Highway Administration; 2008:322. https://www.fhwa.dot.gov/publications/research/safety/08051/08051.pdf.
9. Beasley JW, Wetterneck TB, Temte J, et al. Information Chaos in Primary Care: Implications for Physician Performance and Patient Safety. *J Am Board Fam Med*. 2011;24(6):745-751. doi:10.3122/jabfm.2011.06.100255