A methodological framework to conduct joint zone-based analysis of traffic safety and accessibility

Mehrnaz Asadi a,b, M. Baran Ulak a, Karst T. Geurs a, Wendy Weijermars b

a University of Twente, Department of Civil Engineering, Faculty of Engineering Technology, P.O. Box 217, 7500 AE Enschede, the Netherlands
b SWOV Institute for Road Safety Research, P.O. Box 93113, 2509 AC The Hague, the Netherlands

ARTICLE INFO

Keywords:
Increasing accessibility and improving cycling safety are the primary objectives of cities and policymakers. In the literature, there are several studies examining cycling safety and few studies examining cycling accessibility. However, few studies conduct a joint examination of these two topics to understand the relationship between them. This paper aims to develop and apply a methodological framework to conduct a joint zone-based analysis of traffic safety and accessibility. The developed framework integrates route choice modelling and four-step modelling into traffic safety and accessibility analysis and the outcome is a novel zone-based crash risk estimation incorporating traffic crashes with trip characteristics (including, trip purpose, origin-destination, and alternative routes). The proposed framework was applied in a case study to examine the injury crash risk imposed on commuting cyclists and job accessibility in the city of Utrecht, the Netherlands. The results illustrate the need to embed trip characteristics into zone-based traffic safety estimation to be able to compare the spatial distribution of traffic safety with other mobility and transport-related indicators. Moreover, the findings reflect on significance of taking route choice features into account in accessibility analysis to avoid overestimating accessibility levels. They show that safety and accessibility levels are not equally distributed in the city of Utrecht. For example, low-income areas generally have lower job accessibility levels and higher cycling crash risks which imply the spatial and social inequalities in the transport system’s effects.

1. Introduction

Cycling is a healthy, environmentally friendly, and cost-effective mode of transport that has been promoted in the Netherlands through urban transport policies (Harms et al., 2016; Pucher and Buehler, 2008). Despite the acknowledged benefits of cycling, there are growing concerns about the safety of cyclists due to the increasing number of fatal and severe injury crashes involving bicycles (Adminaitė-Fodor and Jost, 2020; Schepers et al., 2017; Wegman and Schepers, 2024). New urban planning approaches such as the “15-min city” concept rely on safe and active modes of transport to maintain the mobility needs of urban residents and to improve overall accessibility (Knap et al., 2023). Therefore, providing safe infrastructure for cycling and walking is essential for the successful implementation of these plans which can improve accessibility, reduce inequalities, and promote sustainable mobility. To achieve this goal, understanding the relationship between safety and accessibility has the utmost importance. However, there is scarce research offering a methodological approach to examine these two topics jointly.

In this paper, we develop and apply a methodological framework that can be used to conduct a joint zone-based analysis of safety and accessibility to assess spatial inequality in transport for cyclists. This framework integrates the four-step modelling (FSM) and route choice modelling into traffic safety analysis. The product of this framework is a novel zone-based crash risk metric incorporating traffic crashes with trip characteristics (i.e., the purpose, origin-destination, and alternative routes). We implemented the developed framework in a case study to examine the relationship between the injury crash risk imposed on commuting cyclists and job accessibility in different zones of Utrecht, the cycling capital in the Netherlands. This framework involves 1) a four-step modelling to estimate and assign commute trips between origins and destinations (ODs); 2) a route choice modelling to identify a route choice set between ODs; 3) a traffic safety modelling to estimate crash risk on the identified routes; 4) an accessibility analysis utilizing the alternative routes and route choice behaviour; and 5) an analysis of the spatial relationship between traffic safety and accessibility levels of...
the zones in the study area along with the social equity aspects. To the authors’ knowledge, the developed framework is the first in the literature in terms of using alternative routes and route preferences in traffic safety and accessibility assessment.

1.1. Overview of literature

This section gives a brief overview of existing literature on the spatial distribution of cycling safety, cycling accessibility, and relationships between safety and accessibility; followed by an overview of limitations of background methodology for joint assessment of distribution of safety and accessibility and the equity impacts.

Transport systems and policies lead to various outcomes in terms of safety and accessibility of different mode users, including the cyclists. The distribution of these impacts on different socioeconomic groups is also not consistent. Several studies have assessed the inequalities in the spatial distribution of crash risk among cyclists (Braun et al., 2021; Dadashova et al., 2022; Tortosa et al., 2021). The underlying reason leading to those inequalities is limited access to safe cycling infrastructure and low-quality roadways (Braun et al., 2019; Gordon-Larsen et al., 2006; Noland and Laham, 2019). A combination of all these factors and outcomes is likely to discourage disadvantaged population groups from using bicycles which causes a decline in cycling (Standen et al., 2021). For example, Cunha and Silva (2023), in Lisbon, showed that the distribution of cycling infrastructure is an indicator of accessibility by bicycle, and zones with the least accessibility by bicycle were the zones with a prevalence of immigrant residents residing in rental accommodations and low-quality structures. Moreover, census data in Australia supports the notion that individuals who commute by bicycle usually have an income above the average (Standen et al., 2021).

The recent study of Odijk et al. (2023) is one of the first studies linking traffic safety and accessibility. This study gave insights into transport equity by analyzing the distribution of traffic crash risk and job accessibility by bicycle among population groups, in the Rotterdam-The Hague Metropolitan region, in the Netherlands. Their findings indicate that individuals with lower incomes encounter higher risks of bicycle crashes but also face not only challenges in job accessibility by bicycle compared to those with higher income-levels. Additionally, areas identified as most disadvantaged, requiring improvements in road safety and accessibility.

To investigate the relationship between traffic safety and accessibility, the earlier studies focused on using proximity variables or simple accessibility indices in the safety analysis. In the literature, it was shown that increased proximity to secondary schools, grocery stores, and train stations leads to a higher frequency and probability of severe vehicle-to-bicycle crashes (Asadi et al., 2022; Schepers, 2021). Similarly, Kocatepe et al. (2019) and Mafi et al. (2019) assessed crash injury risk for children in urban units based on their distance to risk sources such as hotspots. The logic behind this approach is that the proximity to identified traffic crash hotspots would indicate the level of risk to which the residents of a census unit are exposed. However, the proximity to the source of risk may not be a risk to the residents unless they use the routes which involve the crash risk in the first place.

Therefore, approaches that are not considering the trip characteristics of the residents of a zone may lead to inaccurate estimation of zone-based crash risks. Nonetheless, it is very challenging to determine the travel-based crash risk due to the lack of data on the trip characteristics of crash victims. To address this limitation, Ding et al. (2021) used the London public rental bicycle data and developed four cycling safety models, incorporating exposure variables calculated based on trips made between origins and destinations. They demonstrated that the use of trip-based exposure variables in assessing cyclist safety outperforms other commonly used indicators, such as bicycle counts, number of bicycle trips, and bicycle kilometres travelled. In their analysis, they used bicycle-distance-travel (BDT) which was calculated using the shortest-path method based on the actual distance between origins and destinations, as well as weighted distances that accounted for road attributes influencing perceived safety and level of service in bicycle route choice. However, this study did not consider the impacts of trip characteristics, such as trip purpose, and the associations between road attributes and cyclist route preferences on safety.

In this paper, we develop a methodological framework to estimate a novel zone-based crash risk metric that is used to conduct a joint analysis of traffic safety and accessibility. For this purpose, we used a four-step modelling (FSM) approach coupled with a route choice model, and a safety model. The final product of the framework is the safety metric indicating the injury crash risk associated with bicycle commute trips. This metric allows us to assess the spatial relationships between zone-based safety and accessibility.

2. Methodology

Zone- or area-based analysis approaches have been frequently used for traffic safety assessment in the literature. Various types of analysis units such as grid cells, traffic analysis zones, census tracts, postcodes, and building blocks were employed in previous studies. These studies have incorporated aggregated values of variables, such as the total number of crashes, total exposure, and aggregated data on built environment characteristics (Abdel-Aly et al., 2013; Asadi et al., 2022; Obelheiro et al., 2020; Osama and Sayed, 2017; Quddus, 2008; Wei and Lovegrove, 2013; Wier et al., 2009).

Zone-based safety analysis is more advantageous than road or individual-level analysis when the purpose of the investigation is to identify spatial differences in safety levels. Nonetheless, the weakness of the zone-based analysis is the difficulty of associating traffic crashes with the zones since there is no direct link between a crash and the zone (i.e., residents of the zone) where it happens. To address this drawback, Kocatepe et al. (2019) and Mafi et al. (2019) assessed crash injury risk for residents of census units based on their distance to risk sources such as hotspots. The logic behind this approach is that the proximity to identified traffic crash hotspots would indicate the level of risk to which the residents of a census unit are exposed. However, the proximity to the source of risk may not pose a risk to the residents unless they use the routes which involve the crash risk in the first place.

Therefore, approaches that are not considering the trip characteristics of the residents of a zone may lead to inaccurate estimation of zone-based crash risks. Nonetheless, it is very challenging to determine the travel-based crash risk due to the lack of data on the trip characteristics of crash victims. To address this limitation, Ding et al. (2021) used the London public rental bicycle data and developed four cycling safety models, incorporating exposure variables calculated based on trips made between origins and destinations. They demonstrated that the use of trip-based exposure variables in assessing cyclist safety outperforms other commonly used indicators, such as bicycle counts, number of bicycle trips, and bicycle kilometres travelled. In their analysis, they used bicycle-distance-travel (BDT) which was calculated using the shortest-path method based on the actual distance between origins and destinations, as well as weighted distances that accounted for road attributes influencing perceived safety and level of service in bicycle route choice. However, this study did not consider the impacts of trip characteristics, such as trip purpose, and the associations between road attributes and cyclist route preferences on safety.

In this paper, we develop a methodological framework to estimate a novel zone-based crash risk metric that is used to conduct a joint analysis of traffic safety and accessibility. For this purpose, we used a four-step modelling (FSM) approach coupled with a route choice model, and a safety model. The final product of the framework is the safety metric indicating the injury crash risk associated with bicycle commute trips. This metric allows us to assess the spatial relationships between zone-based safety and accessibility.

2.1. Methodological framework

The developed methodology framework starts with origin-
destination (OD) determination and a route choice model that is used to identify the shortest path and other alternative routes between the postcode area zones. The route choice model utilizes costs other than distance to capture the preferences of cyclists given that cyclists often deviate from the shortest route (P. Chen et al., 2018; Hood et al., 2011; Khatri et al., 2016; Prato et al., 2018; Ton et al., 2018). Following the identification of alternative routes, we used the preferences of cyclists based on previous research by Ton et al. (2018) who estimated the probability that a route is chosen by a cyclist. Identification of alternative routes and accounting for route choice probabilities between ODs help to generate a weighted average travel time cost matrix for each OD pair. This matrix, along with a travel cost decay function developed for bicycle commute trips, was used to make a trip distribution matrix for commuting trips, which is used for trip assignment between ODs.

Following the trip assignment, a traffic safety model was developed using the police-reported crash data. We used a resampling approach to remedy imbalanced data problem (Arianezhad et al., 2021; Morris and Yang, 2021), and the XG-Boost machine learning approach to estimate the crash risk on the cycling network (T. Chen et al., 2015; Morris and Yang, 2021). The traffic safety model and the trip assignment matrix were integrated to calculate the safety risk imposed on a cyclist using a specific route between an OD pair. Finally, the safety risks imposed on all the potential cyclists living in a post-code zone were calculated based on the routes from this zone to all other commuting zones to determine the zone-based crash risk metric of the postcode zone.

The estimated weighted average travel time matrix and travel cost decay function were also used to calculate the potential job accessibility of each postcode zone. The zone-based crash risk metrics and the accessibility levels were used in the final step to identify spatial associations between traffic safety and job accessibility in the analysis zones.

The overall methodological framework is illustrated in Fig. 1 and each analysis step is further explained under subsections 2.2 to 2.8.

2.2. Determination of origin and destination zones

The first step of the analysis is to classify the zones into two groups, residential (origin) and work (destination) zones to limit the size of the origin-destination matrix, which can reduce the computational complexity. For this purpose, we considered several factors including the total population aged 15–65 years old representing the active population, job density, and the share of the floor-space area dedicated to residential and work land use.

2.3. Route choice modelling

1) Route generation

Cyclists often choose routes that do not provide the shortest travel time or distance (Ehrgott et al., 2012; McArthur and Hong, 2019; Ton et al., 2017). There are several factors influencing cyclists’ route choice, including environmental features such as greenness and land use types (Chen et al., 2018; Prato et al., 2018), traffic characteristics such as speed, volume, and composition of vehicles (Khatri et al., 2016; Menghini et al., 2010; Sener et al., 2009), and road infrastructure.

Fig. 1. Methodological framework for estimation of zone-based cycling safety metric and job accessibility; the framework starts with OD determination step and ends with Spatial analysis step.
characteristics such as availability of cycling path or lane, directness, and the number of crossings (P. Chen et al., 2018; Hood et al., 2011; Ton et al., 2018). For instance, in Amsterdam, the presence of separated cycling paths/lanes, fewer crossings per kilometre, and shorter travel distances significantly impact the route choices of commuter cyclists (Ton et al., 2018). Therefore, relying solely on the shortest distance or travel time, which is mainly used in the accessibility models (Vale et al., 2016) and to some extent in road safety (Mafi et al., 2019; Ulak et al., 2019) assessments, is not sufficient to capture cyclists’ route preferences.

Various approaches are proposed for route choice set identification (Prato, 2009). Among these approaches, Ton et al. (2017) proposed two approaches suitable for cycling route choice including data-driven path identification (DDPI) and utilizing choice set generation algorithms (CSGA). The DDPI approach requires actual data on cycling routes. Whereas CSGA utilize stochastic and deterministic algorithms to generate a set of routes based on characteristics of the road link(s) in the network. The majority of choice set generation methods fall into the deterministic category and involve multiple searches for the shortest path in the network (Ton et al., 2017).

In this study, we utilized CSGA method to generate alternative routes between OD pairs based on link characteristics. We incorporated six impedance factors of actual travel distance, travel time, number of crossings, and perceived travel distances. The perceived travel distances were estimated by considering the type of cycling roads and employing three sets of weights as shown in Table 1. The weights assigned lower values to cycling lanes or separated paths, leading to shorter perceived distances. Different weight settings were tested. Setting 1 assigns equal weights to suggested and separated segments, Setting 2 gives slightly higher utility to separated segments, and Setting 3 introduces a significant perceived distance difference among the segment types. By considering these alternative characteristics in the route choice model, we aimed to capture the diverse routes preferred by cyclists. This approach provides a more comprehensive understanding of cyclists’ preferences beyond the traditional emphasis on the shortest distance or travel time.

2) Route choice probability

After determining the alternative routes for each OD pair, we proceeded to calculate the probability of selecting each route by commuter cyclists using a cycling route choice model. In this study, we employed the coefficients of a Path Size Logit model, as presented in eq. 1, which was originally developed for the route choice of commuter cyclists in Amsterdam (Ton et al., 2017). This model captures the route choice preferences and decision-making process of commuter cyclists (Eq. 1).

\[
P \left( r | CS \right) = \frac{\exp \left( -0.057 * \text{distance}_i - 0.525 * \frac{\text{crossings}}{\text{km}} - 0.029 * \text{distance}_j - 0.248 * \ln(PS_j) \right)}{\sum_{r \in CS} \exp \left( -0.057 * \text{distance}_i - 0.525 * \frac{\text{crossings}}{\text{km}} - 0.248 * \ln(PS_j) \right)}
\] (1)

Table 1

<table>
<thead>
<tr>
<th>Setting</th>
<th>Perceived length (Weights for cycling infrastructure types)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mixed traffic</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

where distance, is the travel distance (km), crossings/km is the number of crossings per kilometre, and PS, is the path size factor. By applying the Path Size Logit model, we were able to estimate the probability of selection for each alternative route between OD pairs, providing insights into the route preferences and behaviour of commuter cyclists. The outputs of the route choice model, specifically the probabilities of selecting alternative routes between origin-destination (OD) pairs, were utilized to calculate a weighted average travel time matrix. This matrix incorporated the probabilities of selecting each route and the corresponding travel times.

As shown in Fig. 1, the weighted average travel time matrix was used in two ways. First, it was utilized in the configuration of the trip distribution matrix, which represents the number of trips between different zones in the study area. The probabilities of selecting alternative routes influenced the distribution of trips, taking the travel times associated with each route into account. Second, the weighted average travel time matrix was used in the accessibility model aiming to assess the level of accessibility to job locations from the residential zones.

2.4. Four-step modelling (FSM)

FSM is a universally applied model in transportation modelling and planning. This model considers travel as derived from the demand of individuals for participation in activities. The four steps of this model are as follows.

1) Trip generation

To estimate the number of produced commuting trips we used information from the travel survey data in the origin zones. For that, a fraction of the total population, who potentially make work trips by bicycle, was considered. The potential number of locations that attract work commute trips was estimated based on the number of jobs in the destination zones.

2) Trip distribution

A doubly constrained gravity model (TMIP, 1997) was estimated to create a trip distribution matrix (Ti,j) between origin zone i and destination zone j. The model is represented by Eq. 2.

\[
T_{ij} = a_i b_j P_i A_j f(c_{ij})
\] (2)

where, a and b are balancing factors for trip production and attraction in the zones, Pi indicates the total trip production in zone i, Aj indicates the total trip attraction in zone j, and f(cij) indicates the travel impedance function.

3) Modal share

According to data from the Travel Survey (OViN, 2017) in the Netherlands, approximately 21.85% of trips “to/from work” were made by bicycle in the Province of Utrecht. We use this average percentage to estimate the proportion of cycling trips to/from work for each OD pair. The estimation of cycling trips for each OD affects the estimation of the total crash risk.
4) Trip assignment

We utilized the estimated route choice probabilities from Eq. 3 to determine the proportion of work trips made by bicycle between zones i and j via route r \((W_{ij}^r)\).

\[
W_{ij}^r = T_{ij} \times P_{ij}^r \times 21.85\% \tag{3}
\]

where, \(T_{ij}\) is the number of work trips between zones i and j; and \(P_{ij}^r\) is the probability of choosing route r among others in the generated route choice set. Choosing an appropriate form for the impedance function and estimating its parameters using recent empirical travel behaviour data from the study area is crucial for reliable results (Geurs and Van Wee, 2004). In this study, we used travel behaviour survey data (OViN, 2017) to develop an impedance decay function for bicycle trips made to/from work. We considered reported travel times by respondents living in the Utrecht Province for the analysis.

2.5. Traffic safety modelling

In this study, we focused on two types of crashes, severe and slight injury crashes, to estimate the crash injury risk on the segments of the network. The estimated crash injury risks of the segments along with the calculated number of trips on identified routes are used to determine the zone-based crash risk metric. For this purpose, we employed two approaches of resampling method to address imbalanced data problems and a machine learning model to predict the number of crashes on the segments of the cycling network.

1) Resolving crash data limitations by resampling

The traffic safety analysis is commonly burdened by the imbalanced data problem, especially for less frequent types of crashes such as fatal crashes. However, understanding and exploring the factors influencing rare crash types is essential to prevent them. This issue also applies to cases where the time interval (e.g., peak hour) is limited. Recent literature showed that resampling techniques can be employed to alleviate the imbalanced data problem and can improve the performance of modelling approaches (Abou Elassad et al., 2020; Tiantian Chen et al., 2021; Morris and Yang, 2021; Wang et al., 2022). Morris and Yang (2021) demonstrated the effectiveness of different resampling techniques and found that resampling the crash data improves the modelling performance. In their study, they used a cluster-based under-sampling method along with an adaptive synthetic (ADASYN) over-sampling method, which improved the model’s performance in predicting minor crash types without hindering the prediction of the majority of crashes.

We employed the cluster-based under-sampling method to address the excess number of segments with no crashes during rush hour. Additionally, the ADASYN method was used to increase the number of segments with at least one crash (i.e., oversampling). The resampling process took segments’ characteristics such as road class, traffic, and land use variables into account. The aggregated under-sampled and over-sampled data were then used to develop safety models.

2) Machine Learning Approach for Safety Modelling

Conventional regression models, such as Poisson and Negative Binomial, as well as their extensions, are commonly used in crash data modelling. However, recent studies have shown that machine learning models can achieve comparable performance to traditional statistical approaches (Abdel-Aty et al., 2004; Tang et al., 2020).

In this study, we employed the XG-Boost modelling approach to develop a safety model to find the non-linear relationships between environmental and exposure variables and injury crash counts. This model employs a scalable machine learning technique, utilizing a boosting process to create a precise predictive model for the number of severe and slight injury crashes on road segments. A learning objective of Poisson was set for the number of crashes (with max_depth = 5 and nthread = 4). To ensure the reliability of the synthesized data and model parameters, the data resampling and modelling process was repeated 500 times. This approach allowed us to obtain more robust predictions of the number of crashes on road links. The average of the predicted number of crashes was used to estimate the crash risk for each segment.

2.6. Zone-based crash risk estimate

The zone-based crash risk metric provides a comprehensive measure of the crash risk associated with commuting trips from each zone. It considers the number of trips and the cost of crashes per cyclist in the crash risk analysis. In the final step, we developed a metric to assess crash risk in the analysis zones. To estimate the crash cost per cyclist on each route, we divided the crash cost of the route by the bicycle counts on that route obtained from the Fietstewelg (FTW) survey conducted in 2016. Due to the use of four-year crash data, we focused on 4 rush hours, in working days, in 4 years (Eq. 4). The crash risk cost metric integrates the number of commuting trips between the OD pairs on each route \((W_{ij}^r)\) and the cost of crashes per cyclist on that route (Eq. 5). The cost of severe injury crashes was considered as 1 million euros, which corresponds to the value of the Statistical Severe Injury (SSVI) in the Netherlands (Schoeters et al., 2021). Additionally, the cost of slight injury crashes was assumed as 5000 euros (Wijnen, 2022). By assigning these costs to the respective crash types, we calculated the total crash risk costs for potential work trips originating from each analysis zone (Eq. 6).

\[
\text{Total exposure}_{ij} = \sum_{l=1}^{T} \text{hourly BKT}^l_{ij} \times 4\text{hours} \times 255\text{days} \times 5\text{years} \tag{4}
\]

\[
\text{Crash risk cost}_{ij} = W_{ij}^r \times \frac{\sum_{l=1}^{T} \text{No.crash}^l_{ij} \times \text{Crash cost}}{\text{Total exposure}_{ij}} \tag{5}
\]

\[
\text{Zone – based crash risk cost,} = \sum_{f=1}^{n} \text{Crash risk cost}_{ij}^f \tag{6}
\]

2.7. Potential job accessibility in zones

To assess job accessibility by bicycle in the analysis units, we employed the potential accessibility model proposed by Hansen (1959). This widely used method has been applied in previous studies to evaluate accessibility to different destinations, including job locations. The potential accessibility measure calculates the potential accessibility from each zone (i) to all other zones (j), considering the effects of travel impedances or travel costs. The measure is calculated using Eq. 7, where smaller opportunities, such as jobs, and/or more distant opportunities have a lesser impact on accessibility.

\[
A_i = \sum_{j=1}^{n} D_{ij} f(c_{ij}) \text{ s.t.,} \quad f(c_{ij}) = \frac{1}{1 + e^{-c_{ij}}} \tag{7}
\]

where \(A_i\) is an accessibility measure in zone i to all available opportunities (jobs) \(D\) in other zones \(j\); \(c_{ij}\) represents travel impedance (e.g., travel time cost) between zones \(i\) and \(j\); \(f(c_{ij})\) is the travel cost (decay) function. We fitted a log-logistic distribution (with parameters of \(\alpha = -4.028\) and \(\beta = 1.939\)) on travel time cost of commuting trips (RSS = 0.002 and AIC = -105.23). The generalized travel time cost matrix \(c_{ij}\) was calculated by multiplying the travel time values by the value of travel time (VoTT) by bicycle for commute trips, estimated at 13.43
euros/h (van Ginkel, 2014), in the Netherlands. Previous research estimating decay functions based on NVP data revealed that s-shaped curve (log-logistic) provides a better fit for commuting in the Netherlands, also for cycling (Knap et al., 2023).

2.8. Spatial relationships between safety and accessibility of cyclists

In this study, we examined the spatial correlation between a zone-based crash risk (y) and potential job accessibility (x) using Bivariate Local Moran’s I (BLMI) method. BLMI is commonly used in research to analyse spatial relationships between two variables (Anselin et al., 2014), such as transport accessibility and regional performance (Freiria et al., 2022), healthcare accessibility and population numbers (K. Kim and Kwon, 2022), or accessibility to different public facilities (e.g., parks, healthcare centres, and public transport) and population density (Y. Chen et al., 2020; Hu et al., 2020; Sharma and Patil, 2021).

The BLMI analysis results revealed different types of relationships between job accessibility and crash risk, including High-High (HH), Low-Low (LL), High-Low (HL), Low-High (LH), and non-significant. HH represents zones where both job accessibility and crash risk are high, while LL indicates zones with low values in both variables. HL and LH highlight clusters of zones where one variable is high and the other is low. The non-significant type indicates no significant spatial relationship between the variables within the specified confidence interval.

3. Case study

3.1. Study area

The city of Utrecht, the fourth largest city in the Netherlands, was chosen as the case study area for this research. Utrecht is well-known as the best bicycle city in the world in 2022, with approximately 51% of the population using bicycles (Road-Safety-Comparator, 2021). Notably, the city has contributed to its reputation as a safe place for cyclists with a low number of fatalities and a high proportion of residents who use bicycles to commute. In 2020, the number of bicycle-related fatalities in Utrecht was 0.72 in 2021; this number for trips less than 15 km was 0.69 (CBS, 2023, June). The high-quality cycling infrastructure and facilities in Utrecht have contributed to its reputation as a safe place for cyclists (Uijtdewilligen et al., 2022). However, despite this infrastructure, the Netherlands has experienced an increased number of serious injuries and fatal (cyclist) crashes from 2010 to 2019 (Aarts et al., 2020; SWOV, 2023). The number of fatal (cyclist) crashes in Utrecht fluctuates from year to year and the number of serious road injuries cannot be determined on a city level, yet this trend is likely similar in Utrecht. This trend as well as the plans for further promoting cycling to make Utrecht a 10-minute city (Knap et al., 2023), motivated us to choose Utrecht as the case study area. The study was conducted using the level-five postcode zones (PC5) of the municipality of Utrecht, as level 5 is the smallest spatial unit for which sufficiently detailed census data is available. Utrecht was divided into 540 PCS areas, with an average of 660 inhabitants and 0.18 km² area, in 2020 (CBS, 2020). To account for the boundary effect of neighbouring zones on crash risk and job accessibility estimations, the analysis also included PC5 zones located within a 5 km buffer outside of the municipality boundary.

3.2. Databases

This study used various types of data on vehicle-to-bicycle crashes, cycling and road infrastructure, cycling, and motor-vehicle volume data. Moreover, data on socio-demographic characteristics and job density in the zones, as well as survey data on travel behaviour are utilized in the analysis.

3.2.1. Zonal demographic and job database

We utilized national databases of Statistics Netherlands, updated yearly and publicly available (CBS, 2020), containing sociodemographic and economic characteristics of the postcode-level (PC) zones. The PC zones are the geographical zones from level 6 (PC6) aggregated to bigger size zones (PC5 and PC4). We extracted information on the density of the population in five different age groups within the PC5 zones. Furthermore, we used the National Job Information System (LISA), which collects and presents job distribution data across detailed geographical zones (PC6) in the Netherlands. For this study, the total number of jobs in the study area was available PC4 zones (LISA, 2014). To align the data with the PC5 level, which is a smaller division than PC4, we allocated the job positions to the corresponding PC5 zones by taking the effects of land use characteristics on jobs into account (Harbers et al., 2019). For that, we created a Negative-binomial model for the number of jobs in PC4 zones. Based on this model the total area of working spaces, including offices, accommodations, industrial, and healthcare areas had a positive relationship with the number of jobs in the zones. Therefore, we used the total area of these classes in the PC5 zones and number of jobs in PC4 zones to estimate the number of jobs in PC5 zones. An apportion technique (ArcGIS-Pro) is used for this purpose.

Fig. 2 shows the distribution of the number of jobs and the density of the working-age population number (aged between 15 and 65 years), in the PC5 zones. As the figure shows, the majority of the zones with a high density of active population and high job density are located in the city centre and the middle-west part of the city. Furthermore, the most eastern part of the city involves a lot of number of jobs as the campus of (Medical) University of Utrecht are located there. Thus, these areas are candidates to generate and also attract more commute trips. This also applies to the city centre area located in the middle-east part of the city with relatively high population and job density.

3.2.2. Dutch national travel behaviour database

The Dutch national travel behaviour database is a comprehensive survey that collects data on the daily travel behaviours of a fixed group of households and individuals aged 6 years or older. With over 45,000 respondents, this survey has been conducted annually since 1978, providing a rich source of information. For this study, we focused on the province of Utrecht and utilized data on the commute trips of respondents in this region. Specifically, we analyzed data from a total of 369 individuals who participated in the survey (OVIN, 2017). This dataset enables us to gain insights into the commuting patterns and bicycle use of residents in the province of Utrecht and develop a travel cost decay function for commute trips by bicycle.

3.2.3. Network and exposure database

In this study, we utilized the Bicycle Count Week database, Fietstellingenweek, 2016), providing the bicycle counts on roads and cycling speeds throughout the Netherlands collected from a sample of 51,000 participants during one week (19 to 25 September 2016). The FTW data are provided based on the Open Street Map (OSM) network. Considering that the FTW data only consists of a sample of cyclists, previous studies recommended modifying the FTW cyclist counts. For this purpose, we used a fixed of 41.12, as suggested by Uijtdewilligen et al. (2022). That is, cyclist counts provided in FTW data was multiplied by 41.12 to approximate the actual number of cyclists on the network. Moreover, as this study focuses on commuting trips, only the rush hour (i.e., 07:00–09:00 and 16:00–18:00) counts were used. For this purpose, hourly cyclist counts available from loop detectors were used to calculate a scaling factor that can estimate the number of cyclists during this time. Analysis of the daily and rush hour bicycle counts from the loop detectors indicated that the scaling factor is Lognormally distributed (μ = −1.109, σ² = 0.345). We used the identified distribution to stochastically generate scaling factors that can approximate cycling counts.

In addition to the FTW data, we retrieved other important characteristics of the cycling network from the Cycling Union database (Fietserbond, 2016). This database offers comprehensive information on various attributes of the cycling network, including geometrical features attributed to each link, road types (such as separated cycling paths, suggested cycling lanes, and mixed traffic roads), and one/two-
way streets, covering the entire Netherlands. This information was applied in route choice modelling. To complete our analysis, we also incorporated motor-vehicle (MV) traffic characteristics. MV volume data estimated by a traffic model were obtained from the municipality of Utrecht, allowing us to gain insights into the traffic flow patterns and volume of motor vehicles in the city. Additionally, we included speed limit data from Rijkswaterstraat (2020) to consider the varying speed limits on different road segments.

3.2.4. Cycling crash database

This research utilized 4 years of police-reported road crash data (BRON, 2015-2018) which is the primary source of traffic crash information in the Netherlands. The dataset contains various attributes of crashes, including their geographical location, severity level, time and date, and the modes involved. As the focus of the study was on commute trips, crashes that occurred during rush hours (7:00–9:00 and 16:00–18:00) were selected for safety modelling. Specifically, we concentrated on crashes involving at least one motor vehicle, such as a
In the BRON data, crash locations are registered at four accuracy levels: 1) exact coordinates, 2) intersection level, 3) street level, and 4) municipality level \cite{Rijkswaterstaat-CIV2021}. To ensure the accuracy of the analysis, we excluded crashes reported at the street and municipality levels. The final dataset used for the analysis included actual injury vehicle-to-bicycle crashes, consisting of 87 severe and 69 slight injuries.

**Fig. 4.** Peak-hourly cycling intensity (BKT) on road segments.

**Fig. 5.** (Predicted) Injury crash risk costs (euros) on road segments.
injury crashes, involving 88 and 70 cyclists, respectively. No fatal crashes occurred within the time and spatial scope of this study. Additionally, we excluded property damage only (PDO) crashes, as they are often underreported in the BRON data (SWOV, 2016). It is important to note that despite potential inaccuracies, the BRON dataset is a high-quality source and the best available crash data in the Netherlands.

4. Results

4.1. Determination of origin and destination zones

The classification of the PCS zones resulted in a total of 1111 residential zones (origins) and 416 work zones (destinations) within the analysis area located in the boundaries of the municipality of Utrecht and a buffer area of 5 km. The residential zones consist of the zones with residential space areas greater than 5% of the zone area or with a total active population of more than 9 people. The work zones consist of the zones with workspace areas greater than 5% of the zone area or with the total number of jobs more than 4 jobs. In total 780 of the zones are fitted into both residential and work zones in the study area as they are candidates for generating and attracting commute trips.

4.2. Route generation and route choice modelling

As explained in the methodology of this paper, to generate a route choice set for cyclists, the study employed a method that minimized the travel impedances mentioned earlier. Cut-off values were established to determine routes based on travel distance, travel time, and number of crossings were 15 km, 60 min, and 50 crossings, respectively. Subsequently, any duplicate routes were eliminated from the dataset, resulting in a total of 1,430,335 unique routes between 427,220 origin-destination pairs. On average 3 to 4 (maximum 6) unique routes are generated in the route choices set for each OD pair. Fig. 3 provides examples of these unique alternative routes.

4.3. Zone-based cycling safety

As mentioned in subsection 2.5 this study used a resampled set of data applied in a machine learning model. Then we used the predicted number of crashes based on the XG-Boost model to calculate crash risk costs on the generated routes. The estimated values on routes revealed that the total cost associated with cycling crash risk majority of routes was less than 5 euros per year. Also, the average cycling crash risk cost per commute trip is low, less than 2 cents per trip. This result highlights that cycling in the cycling capital of Utrecht is generally safe.

We estimated crash cost on routes based on the (predicted) number of crashes and exposure on links estimated based on hourly bicycle intensity during peak hours, as shown in Fig. 4. As Fig. 5 shows, the cost of crash risk on cyclists during peak hours is unevenly distributed throughout the road network. Based on this figure the links located areas located in the middle-east of Utrecht (close to the city centre area) and areas in the middle-west are riskier areas for cyclists during the peak hours.

Fig. 6, illustrates the estimated zone-based crash risk associated with bicycle commute trips originating from the analysis zones. The zones are color-coded from dark green to dark red, indicating the level of total injury crash costs from the safest to the riskiest zones. Despite high exposure to road links in the city centre area, it appears that the inhabitants of these zones are relatively safe during their commutes to work by bicycle. This is because these areas have a high population and job density, resulting in shorter commute distances and potentially less exposure in terms of the length of trips to work. Additionally, traffic congestion and the slower speed of motor vehicles during peak hours contribute to improved cycling safety in these areas.

In contrast to the city centre, the east and north-west parts of the city present higher risks for cyclists’ commuters. Despite having high population density in this area, these zones have a lower concentration of job opportunities. As a result, bicycle commuters from these areas are compelled to take longer routes to reach their workplaces. This

Fig. 6. Total average yearly injury crash risk costs associated with bicycle commute trips (per Origin).
increased travel distance, which in turn cyclists in these zones exposed to higher crash risks compared to other areas. This finding also highlights the importance of integrating work trip weights into the safety analysis.

4.4. Potential job accessibility

In the analysis of this study, we utilized the weighted average travel time cost of the alternative routes between OD pairs. To calculate the weighted average travel time cost between the ODs we estimated the total travel time cost on each route multiplied by the route choice

**Fig. 7.** Potential job accessibility by bicycle (per Origin) estimated based on the shortest travel time between ODs (mean≈347 k).

**Fig. 8.** Potential job accessibility by bicycle (per Origin) estimated based on weighted average travel time between ODs (mean≈330 k).
probability of that specific route among other alternatives. We also estimated the potential accessibility by the shortest travel time to show and compare the effect of using only the shortest travel time on the accessibility.

Fig. 7 demonstrates that job accessibility values calculated based on the shortest travel times exhibit a clustered pattern. The highest accessibility is observed in the city centre, which gradually declines as the distance from the city centre increases. A similar result was found by Knap et al. (2023). Fig. 8, on the other hand, presents a different pattern when using the weighted average travel time in assessing job accessibility. The values appear to be more scattered across the zones. This finding highlights the significance of factors such as the number of crossings and the type of road infrastructure in cyclists' route choices, which contribute to higher (perceived) safety for cyclists (Ding et al., 2021; McArthur and Hong, 2019). Consequently, in the city centre areas, the probability of selecting routes other than the shortest routes, specifically the routes with a smaller number of crossings, might be greater.

Fig. 9 shows the percentage of decline in job accessibility estimated based on weighted average travel time and shortest travel time between OD pairs. The zones were found to have high job accessibility based on only the shortest travel time representing a bigger amount of overestimation in job accessibility. These zones mainly have a greater number of jobs and are located in the city centre. Moreover, the comparative analysis emphasizes the importance of considering alternative routes between origins and destinations, as well as incorporating route choice behaviour in the accessibility assessments. Neglecting these factors can lead to overestimation in the assessment of job accessibility.

4.5. Spatial analysis

The spatial analysis was conducted using GeoDa Software (Anselin and McCann, 2009), considering neighbouring zones within a 5 km buffer outside of the municipality boundary. We considered the 100 nearest-neighbour zones to each zone for exploring the types of clusters.

Fig. 10 depicts the results of the BMLI analysis. The HL areas indicate zones with high job accessibility and low crash risk, primarily located in the city centre and the eastern part of the city. On the other hand, a smaller number of zones (29 zones) are clustered as LH zones, representing areas with poor job accessibility and high injury crash risk. In the western part of Utrecht, the majority of zones exhibit high job accessibility and high crash risk values. Conversely, clusters of zones with low job accessibility and high crash risk are mainly concentrated in the eastern part of the city.

The assessment of the demographic and socioeconomic characteristics of the population in the above-mentioned clustered zones (Table 2) reveals that the areas with the highest accessibility and safety are inhabited by the largest percentage of the population with Dutch nationality and the lowest percentage of the migrant population. Housing values, indicating the socioeconomic levels, are also the highest in these areas. In contrast, the proportion of non-western migrants is the highest in the zones located in the least advantaged zones where the economic level is also the lowest compared to the other clustered zones. The comparisons also confirm that low-income populations are at higher cycling crash risk. Moreover, zones that host a higher proportion of the native Dutch population and higher economic levels indicated by housing values provide higher accessibility to jobs for cyclists.

5. Discussion

5.1. Limitations and future research

This study raises several questions for further investigation, including the exploration of alternative cycling route choice factors and route generation approaches, such as the bicycle routing engine suggested by Hardinghaus and Nieland (2021). It is also worth noting that the findings may not be generalized to other regions. Therefore, future research with a similar approach could delve into the inequalities in the distribution of transportation system (dis)advantages across different geographical regions.
The differences between the socioeconomic groups and job sectors are important for the accessibility analysis; however, these differences are not considered in this study. Nonetheless, it is important to consider the socioeconomic and demographic characteristics of the population as well as environmental factors when implementing safety and accessibility-improving policies to achieve balanced and inclusive mobility. Therefore, more detailed investigations on the distribution of transport impacts over populations with different demographic and economic backgrounds (Pereira et al., 2017) in future studies will provide a comprehensive understanding of how different mode users experience the transport system.

Analyzing the spatial relationships between safety and accessibility provides valuable information for prioritizing transportation measures and interventions in areas with higher crash risks and/or lower accessibility levels. This knowledge should be explored in future research which can inform targeted interventions to enhance both cycling safety and job accessibility, promote sustainable commuting choices, and contribute to the overall well-being of cyclists and the transportation system as a whole.

### 5.2. Transferability of the methodological framework

The findings of the conducted case study are specific to the location of analysis and it is hard to generalize these findings to other regions. Nonetheless, the proposed and developed methodological framework in this paper can be generalized and is transferable to conduct new case studies. The components of the framework, namely route generation, traffic assignment, safety modelling, accessibility analysis, and spatial analysis can be conducted with different levels of complexity and data requirements. For instance, we utilized a machine learning model for the safety modelling, but this could also be done with a more traditional method such as a regression approach. Similarly, we generated alternative routes and used a developed route choice model for traffic assignment on routes; however, simpler techniques could be also employed such as an all-or-nothing approach. In the same way, more detailed and advanced techniques can be replaced with the used methods in this paper. However, the main structure of the proposed framework would be still valid.

The adopted methods in the developed methodological framework are generally data-demanding approaches requiring high-quality data. However, given the increasing abundance of data associated with traffic, road infrastructure, land use, travel behaviour, and socioeconomics, such data-demanding approaches become more feasible.

### 6. Conclusions

This study proposes a methodological framework to make zone-based traffic safety and accessibility assessments using a crash risk metric. The framework was applied to the city of Utrecht to examine the relationships between road safety and job accessibility levels for...
commuter cyclists. The findings of this study emphasize the importance of exploring the associations between road safety and accessibility considerations in transportation planning and policy-making. Even for the Dutch cycling capital city of Utrecht, accessibility and safety levels for cyclists are not equally distributed. Low-income areas generally have lower job accessibility levels and higher cycling crash risks, implying spatial and social inequalities.

This study shows that incorporating a route choice set with multiple routes and route choice behavior in the accessibility modeling significantly affects the accessibility levels and the spatial distribution of accessibility. This implies that the common practice of using only the shortest route in accessibility analysis will create biases in the estimates, as travel times are typically underestimated. In conclusion, joint consideration of safety and accessibility measures by accounting for the socioeconomic characteristics in transportation planning can lead to a better understanding of spatial and social inequalities. By understanding the relationships between road safety and accessibility, policymakers can continue to improve cycling not only as a sustainable but also as a safe and inclusive transport mode in cities.

CRediT authorship contribution statement

Mehrnaz Asadi: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Validation, Writing – original draft, Writing – review & editing.
Baran Ulak: Conceptualization, Data curation, Methodology, Project administration, Supervision, Validation, Writing – review & editing.
Karst T. Geurs: Conceptualization, Data curation, Methodology, Resources, Supervision, Validation, Writing – review & editing.
Wendy Weijermars: Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT 3.5 and Grammarly in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and took(s) full responsibility for the content of the publication.

Data availability

The authors do not have permission to share data.

References


