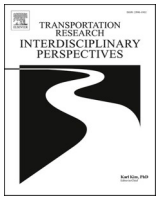
Transportation Research Interdisciplinary Perspectives 30 (2025) 101375

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives



Underreporting and selection bias of serious road traffic injuries in auto insurance claims and police reports in British Columbia, Canada

Michael Branion-Calles [a,b,*], Andrea Godfreyson [b], Kate Berniaz [c], Neil Arason [d], Herbert Chan [a], Shannon Erdelyi [a], Meghan Winters [e], Joy Sengupta [f], Mohamed Essa [f], Fahra Rajabali [g], Jeffrey R. Brubacher [a,h]

[a] Department of Emergency Medicine, The University of British Columbia, Vancouver, Canada
[b] Prevention & Health Promotion Branch, British Columbia Ministry of Health, Victoria, Canada
[c] Clean Transportation and Programs Branch, The British Columbia Ministry of Transportation and Infrastructure, Victoria, Canada
[d] Island Health Authority, Provincial Health Services Authority, Victoria, Canada
[e] Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
[f] Engineering Services Branch, Integrated Transportation and Infrastructure Services Division, The British Columbia Ministry of Transportation and Infrastructure, Victoria, Canada
[g] BC Injury Research and Prevention Unit, BC Children's Hospital, Vancouver, Canada
[h] Vancouver General Hospital, Vancouver, Canada

ARTICLE INFO

Keywords:
Data linkage
Traffic injury
Insurance claims
Selection bias
Underreporting
Inequity

ABSTRACT

Background: Administrative datasets (police reports, insurance claims, medical records), form the basis for road safety research, but suffer from under-reporting and selection bias. Data linkage can provide a fuller picture of road traffic injuries and provide insight into dataset-specific biases. We examined the overlap of serious road traffic injuries involving motor vehicles reported in hospitalization records, police reports, and insurance claims in British Columbia, Canada (2015 – 2019) and assess selection bias within each injury dataset.

Methods: We probabilistically linked police reports, insurance claims, and hospital admissions to a provincial population directory, identifying distinct persons and injuries across datasets. Injuries were linked to socio-demographic and geographic details from other government data including age, sex, low-income status, neighbourhood income and health authority. We analyzed serious injuries to drivers, cyclists and pedestrians. We assessed the proportion of injuries captured by a database (ascertainment rate) and assessed selection bias based on which sociodemographic groups were more likely to only be captured in hospital admissions.

Results: From 2015 to 2019, we estimated 57,097 motor vehicle-involved injuries (48,198 motor vehicle drivers, 2,641 cyclists, 6,258 pedestrians). Insurance claims had the highest ascertainment rate for drivers (95.7%), but lower for cyclists (83.3%) and pedestrians (76.5%). Police records and hospital admissions better captured cyclist and pedestrian injuries compared to driver injuries. Unlinked hospital admission injuries were more likely from low-income and remote populations.

Conclusions: The underreporting highlights the need for improved injury data collection especially for pedestrian and cyclists, to better capture the full injury burden, particularly among marginalized sociodemographic groups.

1. Introduction

Road safety research is often based on administrative data, including police reports, insurance claims, and medical records. Administrative data are not designed for monitoring road traffic injuries and no single dataset records all road traffic crashes and injuries that occur in a given jurisdiction. The proportion of traffic injuries that go unrecorded by a database, the underreporting, will vary for different types of road traffic injury and within different population groups (Elvik and Mysen, 1999; Soltani et al., 2022). When a road safety dataset systematically underreports injuries to a certain type of road user, sociodemographic group, or injury severity than it can be said to demonstrate selection bias. The selection bias within a road safety dataset arises from the processes for collecting the data which will differ by place, time and

* Corresponding author at: Department of Emergency Medicine, The University of British Columbia, Vancouver, Canada.
E-mail address: michael.calles@ubc.ca (M. Branion-Calles).

https://doi.org/10.1016/j.trip.2025.101375
Received 11 September 2024; Accepted 28 February 2025
Available online 6 March 2025

dataset. For instance, police of a given jurisdiction will have criteria for attending and recording the details of a traffic collisions such as thresholds for property damage, while hospital records will be based on medical necessity. Although no single dataset fully captures all road traffic injuries, linking multiple administrative datasets can improve our understanding of the injury burden and dataset-specific biases.

Police data are the focus of most data linkage studies owing to their importance in road traffic injury surveillance in many jurisdictions (Alsop and Langley, 2001; Amoros et al., 2006; Conderino et al., 2017; Elvik and Mysen, 1999; Hosseinzadeh et al., 2022; Janstrup et al., 2016; Langley et al., 2003; Short and Caulfield, 2016; Soltani et al., 2022; Watson et al., 2015). These studies link individual injuries captured in police data to injury records in hospital data, based on the notion that emergency department (ED) visits and/or hospital admissions can provide a more complete dataset for monitoring road traffic injuries, as they record injuries that require medical intervention in hospital. The main focus of these studies is to quantify the proportion of hospital records that are not recorded by police for different road user types, injury severities and in some cases, sociodemographic groups. A consistent finding in these studies are that police data fail to capture a substantial proportion of non-fatal road traffic injuries, with higher levels of underreporting for more minor injuries, as well as injuries to bicyclists in particular (Alsop and Langley, 2001; Hosseinzadeh et al., 2022; Janstrup et al., 2016; Langley et al., 2003; Lujic et al., 2008; Watson et al., 2015). The implication of these studies is that the use of police-data alone for injury surveillance will not only under-estimate the overall burden of injury but will do so to a greater extent for specific types of road injuries and for different population groups.

In the Canadian province of British Columbia, auto-insurance claims are one of the main sources of road safety data (Kamel et al., 2020; Osama and Sayed, 2019; Osama et al., 2018; Urban Systems, 2015; Urban Systems, 2012), setting it apart from most jurisdictions in countries that primarily rely on police reports. In BC, all motor vehicle owners are mandated to obtain basic auto-insurance coverage for their vehicle from the Insurance Corporation of British Columbia (ICBC), resulting in a unique dataset of auto-insurance claims that covers the entire province. Auto-insurance claims capture (self-reported) information on the crash itself and on the individuals involved. In the province, auto-insurance claims data, and police data only record injuries to cyclists or pedestrians if a motor vehicle is involved; single-bicycle crashes and pedestrian falls are not captured. Given the unique context of universal auto-insurance, investigation of potential selection bias and underreporting is warranted.

Our study links population-level insurance claims, police records, and hospital admission records for the province of BC, home to five million people (Statistics Canada, 2024). We aim to understand the degree of overlap of motor vehicle involved serious injuries captured by each of these datasets and to quantify the extent of their underreporting and selection bias across datasets stratifying by injuries to motor vehicle drivers, cyclists and pedestrians. For police and hospital admissions data, we define a serious injuries, as those resulting in a fatality and/or requiring hospital admission; within claims data serious injuries are determined by an insurance adjuster using a set list of injury types. Auto-insurance claims and police data only include injuries to pedestrians and cyclists if they resulted from a collision with a motor vehicle. As such, our analysis is restricted to injuries that involved a collision with a motor vehicle and does not include pedestrian or cyclist falls or collisions with road users other than motor vehicles.

2. Materials and methods

This analysis was developed as part of the Pedestrian and Cyclist Safety Study, commissioned by the Ministry of Health and the Ministry of Transportation and Infrastructure, using injury and sociodemographic datasets for the province of British Columbia made available by the BC Data Innovation Program (DIP) (Government of British Columbia, 2022).To protect the privacy of individuals, the DIP partners with Population Data BC (PopData), a multi-university data and education organization, to provide researchers with de-identified versions of each dataset through a secure research environment (Ark et al., 2020; Population Data BC, 2023). Each dataset contains de-identified study IDs for individual residents in the province of BC based on a probabilistic linkage conducted by PopData to a "Population Directory". The directory, created and maintained by PopData, is a comprehensive database of all residents of BC that uniquely identifies individuals (Ark et al., 2020). As a result, the assignment of a study ID to a given record within a dataset depends on the availability and quality of identifying information, such as name, address, date of birth, sex, driver's license, personal health number etc. For more details on the linkage process see Ark et al (2020).

The specific datasets used from the DIP include *ICBC – Crash* (insurance claims for motor-vehicle involved crashes), *ICBC – Traffic Accident System* (police-reports from motor vehicle involved crashes), *Ministry of Health – Registry and Demographics* (sociodemographic information), *Ministry of Health – Discharge Abstract Database* (hospital admissions), *Ministry of Social Development and Poverty Reduction – BC Employment and Assistance* (income assistance). These datasets are explained in more detail in the following sections, however, visit the BC Data Catalogue for further information: https://catalogue.data.gov.bc.ca/organization/data-innovation-program-dip.

All inferences, opinions, and conclusions in these materials are those of the authors. They do not reflect the opinions or policies of the provider(s) of the data upon which they are based.

2.1. Insurance claims

ICBC provided insurance claims data on motor vehicle collisions from 2015 to 2019, the latest available data from ICBC at the time of this analysis. These data include information for each person involved in a crash including the mode of travel of the individual (driver, passenger, cyclist, pedestrian) and the injuries sustained by each. For each person involved in an insurance claim, ICBC categorized the severity of injuries into four distinct categories: none, any injury, serious injury, and fatal. ICBC claims adjustors categorize the injuries based on their initial assessment of whether the described injuries are "minor". If an injury is initially assessed as minor but involves an injury to the spine it is reassigned as "serious". Where the adjustor does not make an assessment, a serious injury is defined by a proprietary predefined list of injuries and a description of the severity associated with it.

2.2. Police reports

The traffic accident system (TAS) contains data collected from all police-reported crashes in BC. Police attendance of nonfatal collisions is discretionary. TAS records were available from 2010 to 2019. All injuries reported in TAS have at least one motor vehicle involved and took place on a public road; TAS excludes crashes on industrial roads, private driveways, or forest service roads. Police assign one of the following injury severity categories to involved parties: no apparent injury, minor injury or major injury (defined as an overnight stay at a hospital) and a fatal injury.

2.3. Hospital admissions

The Discharge Abstract Database (DAD) has recorded all hospitalizations from acute care facilities in all Canadian provinces and territories, except for Quebec, since April 1, 1985. Road traffic injuries were identified using the Canadian version of the International Classification of Diseases, 10th Revision (ICD-10-CA) external cause codes. The initial cases we identified used external cause codes ranging from V01.1 to V79.9. We included a broad definition of road traffic injuries for potential linkage but subsequently removed injuries that either occurred

off-road or that did not involve a motor vehicle, unless they were linked to an insurance claim or police report.

2.4. Sociodemographic characteristics

The DIP maintains a database of sociodemographic information for individuals who receive health services through the BC's public health insurance program (Medical Services Plan) (Data Innovation Program (DIP), 2023b; Data Innovation Program (DIP), 2023c). The *Central Demographics Collection* includes age, sex and postal code of residence. Because the data specifically codes for sex, categorizing individuals as either male or female, we employ this terminology but we recognize gender encompasses a broader spectrum of identities that are not represented in the binary data collected. Within the DIP, postal code of residence is not directly available to external researchers. Instead, DIP analysts use postal code to link individuals to the Statistics Canada Postal Code Conversion File and assign each individual with a range of geographic variables including regions and census tract income quintiles, which are made available to researchers. The DIP makes BC Employment and Assistance program data from the Ministry of Social Development and Poverty Reduction available, enabling the identification of individuals who are part of a family unit that received income assistance in each month since January 1989 (Data Innovation Program (DIP), 2023a). The population eligible for income assistance include those temporarily in need of assistance (out of work, awaiting other income, unable to work, in need of immediate assistance due to hardship, medical illness) as well as those with permanent disabilities that render them unable to work (BC Ministry of Social Development and Poverty Reduction, 2014, 2022).

2.5. Data processing

We established specific criteria to exclude certain injury records from our analysis before linkage across injury datasets. We noted that motor vehicle passengers in police reports could not be reliably linked to the population directory due to frequently missing identifiers. Therefore, we excluded injuries to passengers and only included injuries to pedestrians, cyclists, and motor vehicle drivers. We limited our analysis to crashes and injuries that occurred between 2015 and 2019, which are the years common to each injury dataset. For injuries involving motor vehicle drivers, we only included persons eligible for a driver's license (aged 16 and over).

We reclassified injury severity across each dataset. For police data injury severity definitions were modified by renaming "major injuries" as "serious injuries" and combining them with fatal injuries, while insurance claims naming conventions remained unchanged but were combined with fatal injuries, and all hospital admissions were considered serious injuries.

Before linkage, duplicate records within each database were removed. For police and insurance claims, duplicate records were defined as injuries involving the same person on the same date. When we found discrepancies between duplicate records, we retained the record with the fewest missing variables. If the number of missing variable rates were identical, we retained the record with higher injury severity. In cases where the number of missing variables and injury severity were the same, a record was randomly selected. For hospital admission data, we removed records belonging to the same individual if they were labeled as a transfer or occurred within 30 days of the previous discharge date.

2.6. Linking injuries across databases

Though each injury event within a given database was assigned a unique study ID by an initial linkage to a population directory, to understand the overlap of injuries across databases, a second linkage was conducted based on matching an injury record with the same study ID (e.g. a unique individual in the province) and date of the crash or hospital admission. We considered two records to match if they involved the same individual (i.e. matching study ID) and occurred within +- three days of one another. Where there were multiple potential matches, we assumed the correct match was the one with the closest date. Injured individuals were linked to sociodemographic characteristics through the study ID as described above.

When there were discrepancies in injury severity assigned in different databases, we used the following hierarchy to assign injury severity: (i) all events captured in hospitalization records were considered "serious" (ii) for events captured in both claims and police records but not in hospital records, the police-assigned severity was given priority. If there were discrepancies across databases for the mode of transport we used the same hierarchy as injury severity to assign mode of transport.

After the linkage process, we removed hospital admissions where the injury occurred off-road, unless they were matched to an on-road crash within the claims or police datasets. Furthermore, to exclude minor injuries, we removed individuals who were not admitted to hospital and were classified in both TAS and claims data as having no injury or minor injury. Our analysis focused on serious injuries due to limitations in BC healthcare data, which lacks details about injury mechanisms for ED visits. As a result, we don't have a reliable database to assess the frequency of minor injuries (those treated outside of hospital or in the ED without an overnight hospital stay) in the province, preventing comparisons with police and insurance minor injury data.

2.7. Overlapping patterns of police, claims, and hospitalization records

We assessed the effectiveness of each database in capturing injuries by counting the number of common reports in the following sources: (1) police, claims and hospitalization; (2) police and claims; (3) police and hospitalization; (4) hospitalization and claims; (5) police only; (6)

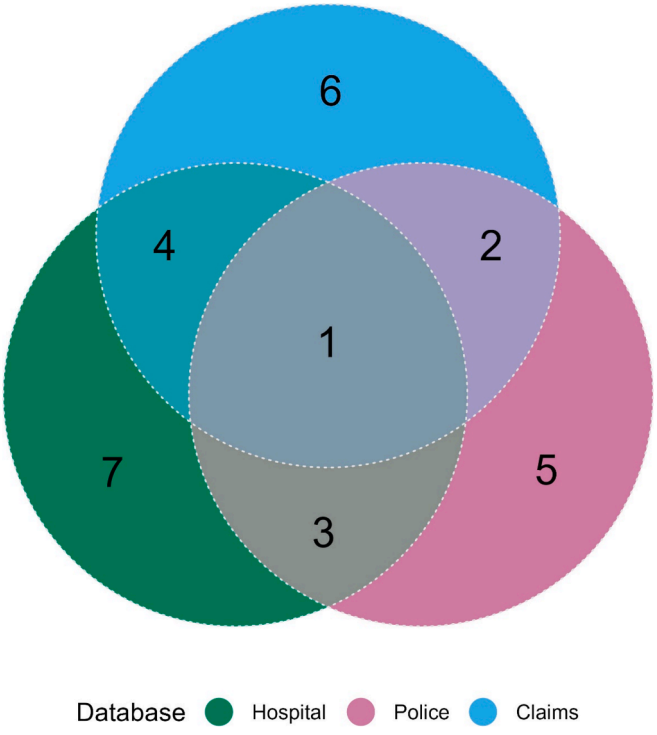


Fig. 1. Illustration of the patterns of overlap we counted in our linked injury database. These categories are mutually exclusive and include the injuries that were captured within 1) police and claims and hospitalization; 2) police and claims; 3) police and hospitalization; 4) hospitalization and claims; 5) police only; 6) claims only; 7) hospitalization only.

claims only; and (7) hospitalization only (Fig. 1). We calculated ascertainment rates for each database (stratified by transport mode) by dividing the total number of linked cases within a database by the number of unique cases across all databases. For both insurance claims and police records we separately calculated the proportion that could be linked to a hospital admission.

2.8. Selection bias

To assess selection bias in each database, we calculated ascertainment rates, stratified by sociodemographic variables. For each transport mode, we stratified by age group, sex, participation in provincial income assistance programs during the five years prior to the injury (a proxy for low-income status), and characteristics of place of residence (based on 6-digit postal codes) including neighbourhood income quintile, rural or urban region, and health authority.

We used logistic regression to quantify the association between events that resulted in hospital admission but could not be linked to either claims or police data and selected sociodemographic characteristics. This analysis aimed to identify sociodemographic factors associated with underreporting of injuries requiring hospitalization, adjusting for other factors.

3. Results

3.1. Linkage to personal identifier

All injured motor vehicle drivers were successfully linked to the Population Directory by PopData, receiving a unique study ID (Table 1), with lower linkage rates observed for pedestrians and cyclists across all databases. Within auto-insurance claims, linkage rates were 88.4 % and 85.7 % for pedestrians with minor and serious injuries respectively; cyclists had rates of 91.7 % and 90.7 %. Hospital admissions had linkage rates of 95.3 % for both pedestrians and cyclists. Just over half of the people with hospital admission records that could not be linked to the population directory were not residents of B.C.; insurance claims and police data do not record this information.

3.2. Overlapping patterns of police, claims, and hospitalization records

From 2015 to 2019, we estimated 48,198 serious injuries due to collisions among motor vehicle drivers, 2,641 among cyclists and 6,258 among pedestrians across the three databases (see Fig. 2). Auto-insurance claims captured 95.7 % of injured motor vehicle drivers, 76.5 % of cyclists, and 83.3 % of pedestrians (Table 2). Police records captured 12.4 % of injured drivers, 25.2 % of cyclists and 36.5 % of pedestrians (Table 2). Hospital admissions follow a similar pattern to police records with 10.1 % of drivers, 28.4 % of cyclists and 36.8 % of seriously injured pedestrians captured in hospital admissions data (Table 2).

Many motor vehicle collisions that resulted in serious injuries which required hospitalization were not recorded by police or insurance claims. Of the 4,850 drivers who were hospitalized after a motor vehicle collision, 2,979 (61.4 %) were captured in police data and 3,492 (72.0 %) in insurance claims (Fig. 2). Out of the 751 cyclists who required hospital admission after a motor vehicle collision, 344 (45.8 %) were captured in police records and 515 (68.6 %) in insurance claims (Fig. 2). Similarly, for the 2,305 pedestrians injured in collisions with motor vehicles and admitted to hospital, only 1,109 (48.1 %) were captured in police records and 1,656 (71.8 %) in auto-insurance claims (Fig. 2).

Conversely, insurance claims recorded 46,136 serious injuries to drivers in total, of which 3,492 (7.6 %) were linked to a hospital admission (Fig. 2). Of the 2,201 serious cyclist injuries captured in insurance claims 515 (23.4 %) were matched to a hospital admission. For pedestrians, insurance claims captured 4,787 serious pedestrian injuries, and 1,656 (34.6 %) of them were linked to a hospital admission.

Within police reports we identified 5,976 serious injuries to motor vehicle drivers with 2,979 (49.8 %) resulting in hospital admission (Fig. 2). Of the 666 serious cyclist injuries within police records, 344 (51.7 %) were linked to a hospitalization. For pedestrians, police reports captured 2,283 serious injuries, of which 1,109 (48.6 %) were linked to a hospital admission.

Ascertainment rates and selection bias by sociodemographic groups

3.2.1. Driver injury

There were variations across sociodemographic groups in the ascertainment rates of serious injuries involving motor vehicle drivers (Table 3). Within insurance claims, groups with lower ascertainment rates included individuals recently receiving income assistance, adults aged 65 and over, residents of rural areas, and residents of the Northern Health authority. For police records, the ascertainment rate was lowest for the 35–44 age group, females, residents of urban areas and people within the Vancouver Coastal Health authority. The ascertainment rate in police records for those requiring income assistance was higher than for the general population. In hospital data, older adults (65 +), the income assistance population, residents of rural areas and those within the Northern Health Authority region had higher ascertainment rates.

Of the 4,850 motor vehicle driver injuries requiring hospital admission, 1,166 (24.0 %) were not linked to any insurance claim or police record (Fig. 2). Amongst injuries requiring hospital admission, adults aged 65 years and older had almost13 times higher odds of being unreported in claims and police data (Adjusted Odds Ratio [aOR] = 12.59, 95 % Confidence Interval [CI] = 10.21 – 15.68), relative to persons aged 45–54 years (Table 4). Individuals who had received income assistance were more than twice as likely to go unreported in insurance claims and police data than the general population (aOR = 2.48. 95 % CI = 2.02 – 3.03). Residents of the lowest income neighbourhoods were nearly 50 % more likely to go unreported in insurance claims and police data compared to those in the highest income neighbourhoods (aOR = 1.46, 95 % CI = 1.20 – 1.79). Likewise, residents of rural neighbourhoods were 36 % more likely to go unreported in insurance claims and police data compared to their urban counterparts (aOR = 1.36, 95 % CI = 1.13 – 1.64). Furthermore, regions beyond the Fraser and Vancouver Coastal Health authorities (Northern, Island and Interior) also had greater odds of hospital admissions going unreported in

Table 1
Linkage rates to the population directory representing subset of records eligible for linkage across available datasets.

		Linked to Population Directory					
		Insurance Claims		Police Reports		Hospital Admissions	
Injury Severity	Transportation Mode	Total	% linked	Total	% linked	Total	% linked
Minor	Driver	212,100	100	41,352	100		
	Cyclist	5,340	91.7	2,338	70.5		
	Pedestrian	7,135	88.4	3,889	62.8		
Serious	Driver	46,136	100	5,976	100	4,850	100
	Cyclist	1,999	90.8	511	76.7	716	95.3
	Pedestrian	4,117	86	1,658	72.6	2,197	95.3

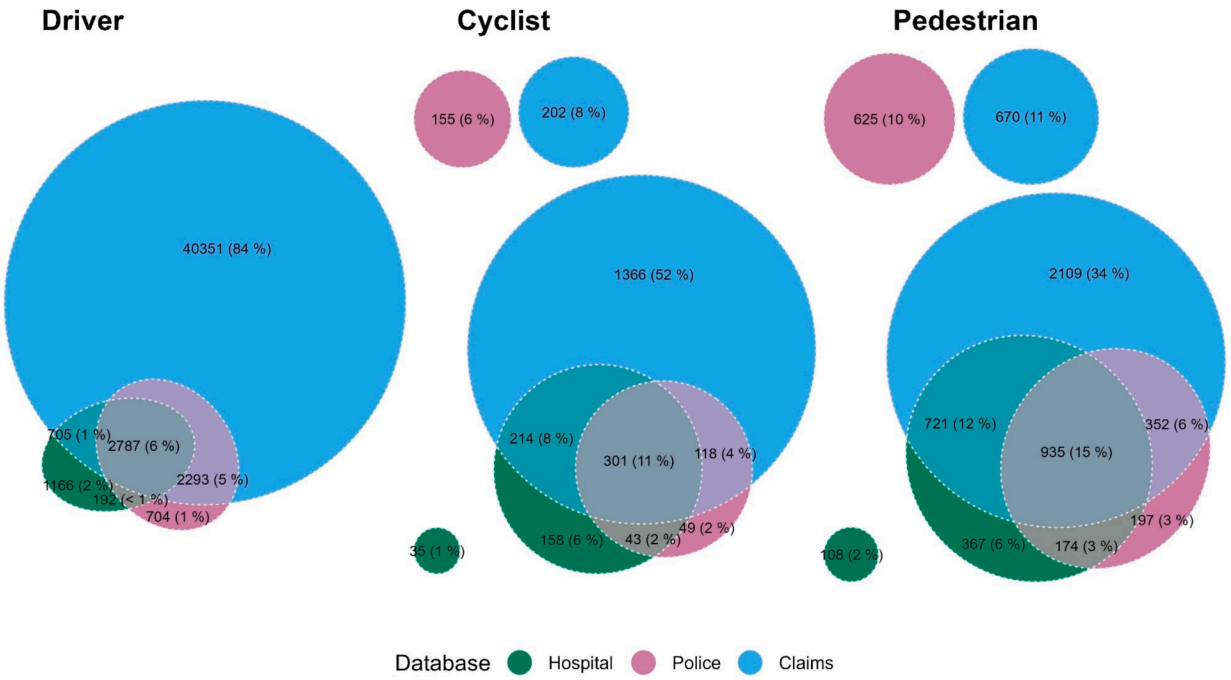


Fig. 2. Euler diagram of overlap of serious injury capture stratified by transport mode. In a Euler diagram the area of an ellipse is proportional to the percentage of injuries represented by the set relationship. Injuries that could not be linked across databases due to insufficient identifying information are represented by the separate ellipses.

Table 2
Distribution of serious injuries by mode and database of capture. Ascertainment rate is the total injuries from all three databases divided by the total captured in that database.

Mode	Serious Injury Total	Total Captured by			Ascertainment Rate		
		Insurance Claims	**Police**	**Hospital**	**Insurance Claims**	**Police**	**Hospital**
Driver	48,198	46,136	5,976	4,850	95.7 %	12.4 %	10.1 %
Cyclist	2,641	2,201	666	751	83.3 %	25.2 %	28.4 %
Pedestrian	6,258	4,787	2,283	2,305	76.5 %	36.5 %	36.8 %

insurance claims and police data, especially hospital admissions from the Northern Health authority which were over 4.5 times more likely to go unreported in both insurance claims and police reports relative to Fraser Health (aOR = 4.69, 95 % CI = 3.69 – 5.93).

3.2.2. *Cyclist injury*

There were variations across sociodemographic groups in the ascertainment rates of serious cyclist injuries involving motor vehicles in each database (Table 5). For insurance claims, underreported groups included children, people who received income assistance, residents of rural areas, and people living within the Northern Health authority region. In police reports, ascertainment rates were lower for females and children. Within hospital admissions, groups with lower ascertainment rates included the 25–34 age group, females, residents of urban areas and those living within the Vancouver Coastal and Vancouver Island Health authority regions. Ascertainment rates were higher in older adults and residents within the Northern Health authority.

Of the 716 cyclist injuries that required hospital admission and could be linked to the population registry, 158 (22.1 %) could not be linked to either an auto-insurance claim or police report (Fig. 2). The logistic models identified sociodemographic factors associated with cyclist injuries that led to hospitalization but weren't recorded in police or insurance databases (Table 6). The likelihood of a cyclist injury that required hospitalization going unreported in police or insurance claims databases was over 2 times higher for those who received income assistance, compared to those who did not (aOR = 2.18, 95 % CI = 1.45 – 3.26). Compared to residents of Fraser Health authority, hospitalized cyclists were more likely to go unreported in police and insurance claims in Northern Health authority (aOR = 2.97, 95 % CI = 1.34 – 6.23) and in Interior Health authority (aOR = 1.94, 95 % CI = 1. 13 – 3.31).

1.1.1. *Pedestrian injury*

In each database, there are discrepancies across sociodemographic groups in the ascertainment of serious pedestrian injuries involving motor vehicles (Table 7). In auto-insurance claims, groups with lower than average ascertainment rates included children under 15 years, males, people who recently received income assistance, and residents of the Northern Health authority region of BC. The ascertainment rate for police report data was more consistent across demographic groups, with the exception of injuries to older adults for which the ascertainment rate was higher than for other age groups, and the population who had received income assistance, for which the ascertainment rate in police data was also higher. Hospital admissions data captured a higher proportion of injuries to children, injuries to adults aged 55–64 compared to age groups in between, as well as injuries in people receiving income assistance, while people living within the Fraser and Vancouver Coastal Health authority regions had lower ascertainment rates than other regions.

Of the 2,197 pedestrian injuries that required hospital admission and could be linked to the population registry, 367 (16.7 %) could not be

Table 3
Driver injury ascertainment rate by sociodemographic variables. Ascertainment rate is the total injuries captured by a given database divided by the total number of distinct injuries captured across all three databases.

Variable	Level	Injury Total (%)	Claims Total	Police Total	Hospital Total	Claims Asc.	Police Asc.	Hospital Asc.
Age Group	16–24	5,836 (12.1)	5,654	822	571	96.9	14.1	9.8
	25–34	10,913 (22.6)	10,652	995	609	97.6	9.1	5.6
	35–44	10,261 (21.3)	10,067	849	556	98.1	8.3	5.4
	45–54	9,598 (19.9)	9,375	925	624	97.7	9.6	6.5
	55–64	6,744 (14)	6,410	973	766	95.0	14.4	11.4
	65+	4,846 (10.1)	3,978	1,412	1,724	82.1	29.1	35.6
Sex	Females	24,280 (50.4)	23,486	2,068	1,994	96.7	8.5	8.2
	Males	23,918 (49.6)	22,650	3,908	2,856	94.7	16.3	11.9
Income Assistance in Previous 5 years	No	44,984 (93.3)	43,156	5,286	4,244	95.9	11.8	9.4
	Yes	3,214 (6.7)	2,980	690	606	92.7	21.5	18.9
Neighbourhood Income Quintile (Census Tract)	1-Lowest	9,405 (19.5)	8,939	1,225	1,114	95.0	13.0	11.8
	2-Medium low	10,725 (22.3)	10,338	1,256	1,060	96.4	11.7	9.9
	3-Medium	10,023 (20.8)	9,677	1,097	962	96.5	10.9	9.6
	4-Medium high	9,222 (19.1)	8,896	1,077	834	96.5	11.7	9.0
	5-Highest	7,870 (16.3)	7,590	979	823	96.4	12.4	10.5
	Missing	953 (2)	696	342	57	73.0	35.9	6.0
Regional Classification	Urban	44,885 (93.1)	43,406	4,492	3,804	96.7	10.0	8.5
	Rural	2,794 (5.8)	2,413	1,245	1,012	86.4	44.6	36.2
	Missing	519 (1.1)	317	239	34	61.1	46.1	6.6
Health Authority	Fraser	26,009 (54)	25,450	1,961	1,564	97.9	7.5	6.0
	Interior	4,904 (10.2)	4,450	1,522	1,316	90.7	31.0	26.8
	Northern	1455 (3)	1,229	572	485	84.5	39.3	33.3
	Vancouver Coastal	10,672 (22.1)	10,397	712	619	97.4	6.7	5.8
	Vancouver Island	4,675 (9.7)	4,327	973	834	92.6	20.8	17.8
	Missing	483 (1)	283	236	32	58.6	48.9	6.6

Asc. = Ascertainment Rate.

Table 4
Logistic regression results for modelling the odds of not being captured by claims or police data for serious motor vehicle driver injury.

Variable	Levels	Unadjusted OR (95 % CI)	Adjusted OR (95 % CI)
Age	16–24	0.90 (0.65–1.23)	0.90 (0.65–1.24)
	25–34	0.82 (0.62–1.08)	0.81 (0.61–1.07)
	35–44	**0.68 (0.51–0.91)**	**0.68 (0.51–0.92)**
	45–54	Reference	
	55–64	**2.53 (2.00–3.22)**	**2.40 (1.89–3.06)**
	65+	**13.13 (10.72–16.24)**	**12.59 (10.21–15.68)**
Sex	Females	Reference	
	Males	1.11 (0.99–1.25)	0.99 (0.88–1.12)
Income Assistance in Previous 5 years	No	Reference	
	Yes	**1.77 (1.46–2.13)**	**2.48 (2.02–3.03)**
Neighbourhood Income Quintile (Census Tract)	1-Lowest	**1.48 (1.22–1.79)**	**1.46 (1.20–1.79)**
	2-Medium low	1.03 (0.85–1.26)	1.13 (0.92–1.38)
	3-Medium	1.05 (0.86–1.29)	1.15 (0.94–1.42)
	4-Medium high	0.96 (0.78–1.18)	1.11 (0.89–1.37)
	5-Highest	Reference	
Regional Classification	Urban	Reference	
	Rural	**3.43 (2.92–4.02)**	**1.36 (1.13–1.64)**
Health Authority	Fraser	Reference	
	Interior	**4.25 (3.61–4.99)**	**2.31 (1.92–2.77)**
	Northern	**6.37 (5.12–7.89)**	**4.69 (3.69–5.93)**
	Vancouver Coastal	**1.33 (1.11–1.58)**	1.14 (0.95–1.36)
	Vancouver Island	**3.60 (3.03–4.27)**	**2.24 (1.87–2.68)**

linked to either an insurance claim or police report (Fig. 2). Our logistic regression models identified sociodemographic factors associated with injuries that required hospital admission but were not recorded in police reports or insurance claims, including age, sex, income assistance participation and region (Table 8). Compared to injuries involving individuals aged 45–54, injuries to individuals aged 15 years or younger were twice as likely to go unreported (aOR = 2.13, 95 % CI = 1.29 – 3.46), as well as injuries to individuals aged 65 and over (aOR = 1.88, 95 % CI = 1.29 – 2.78). Injuries to males were 67 % more likely to be unreported (aOR = 1.64, 95 % CI = 1.31 – 2.07). Likewise, relative to the wider population, injuries to participants of income assistance programs were three times as likely to be unreported (aOR = 3.12, 95 % CI = 2.43 – 4.00).

Discussion

Our study used a linked, provincial database of road traffic injuries with linkages to government datasets that enabled measurement of sociodemographic characteristics not typically available in traditional road safety data. This unique, population-based dataset enabled us to understand road traffic injury ascertainment rates for insurance claims, police records and hospital admissions, and to quantify reporting bias by identifying population groups that were more likely to go unreported in each dataset. Based on this data, we estimated that there were 57,097 distinct serious injuries to pedestrians, cyclists and motor vehicle drivers resulting from a motor vehicle collision across insurance claims, police records and hospital admissions data. Insurance claims were the most complete dataset capturing 96 % of serious driver injuries, 83 % of serious cyclist injuries, and 77 % of serious pedestrian injuries. Police records captured 12 %, 25 % and 37 % of serious injuries to drivers, cyclists and pedestrian injuries, while hospital admissions captured 10 %, 28 %, and 37 % within the same modes. Though insurance claims are the most complete dataset, they still leave out a substantial number of serious injuries that are captured in the other datasets: 28 % of pedestrians, 31 % of cyclists and 28 % of drivers who were admitted to hospital could not be linked to an insurance claim. While hospital

Table 5
Cyclist injury ascertainment rate by sociodemographic variables. Ascertainment rate is the total injuries captured by a given database divided by the total number of distinct injuries captured across all three databases.

Variable	Level	Injury Total (%)	Claims Total	Police Total	Hospital Total	Claims Asc.	Police Asc.	Hospital Asc.
Age Group	0–15	96 (3.6)	73	13	32	76.0	13.5	33.3
	16–24	261 (9.9)	225	51	71	86.2	19.5	27.2
	25–34	547 (20.7)	515	91	109	94.1	16.6	19.9
	35–44	371 (14)	331	80	106	89.2	21.6	28.6
	45–54	464 (17.6)	409	120	176	88.1	25.9	37.9
	55–64	308 (11.7)	270	86	118	87.7	27.9	38.3
	65+	202 (7.6)	176	70	104	87.1	34.7	51.5
	Missing	392 (14.8)	202	155	35	51.5	39.5	8.9
Sex	Female	635 (24)	585	108	159	92.1	17.0	25.0
	Male	1614 (61.1)	1,414	403	557	87.6	25.0	34.5
	Missing	392 (14.8)	202	155	35	51.5	39.5	8.9
Income Assistance	No	1843 (69.8)	1,648	395	543	89.4	21.4	29.5
	Yes	441 (16.7)	351	116	208	79.6	26.3	47.2
	Missing	357 (13.5)	202	155	0	56.6	43.4	0.0
Neighbourhood Income Quintile	1-Lowest	526 (19.9)	453	136	203	86.1	25.9	38.6
	2-Medium low	444 (16.8)	399	90	148	89.9	20.3	33.3
	3-Medium	440 (16.7)	400	90	136	90.9	20.5	30.9
	4-Medium high	375 (14.2)	347	67	116	92.5	17.9	30.9
	5-Highest	332 (12.6)	297	85	104	89.5	25.6	31.3
	Missing	524 (19.8)	305	198	44	58.2	37.8	8.4
Urban/Rural	Rural	84 (3.2)	64	27	42	76.2	32.1	50.0
	Urban	2075 (78.6)	1,870	453	669	90.1	21.8	32.2
	Missing	482 (18.3)	267	186	40	55.4	38.6	8.3
Health Authority	Fraser	498 (18.9)	439	112	176	88.2	22.5	35.3
	Interior	219 (8.3)	185	57	106	84.5	26.0	48.4
	Northern	57 (2.2)	40	18	32	70.2	31.6	56.1
	Vancouver Coastal	979 (37.1)	897	188	264	91.6	19.2	27.0
	Island	409 (15.5)	375	105	134	91.7	25.7	32.8
	Missing	479 (18.1)	265	186	39	55.3	38.8	8.1

Asc. = Ascertainment Rate.

Table 6
Logistic regression results for modelling the odds of not being captured by claims or police data for serious cyclist injury.

Variable	Level	Unadjusted OR (95 % CI)	Adjusted OR (95 % CI)
Age Group	0–15	1.45 (0.68–2.86)	1.70 (0.78–3.47)
	16–24	0.88 (0.49–1.54)	1.07 (0.58–1.93)
	25–34	**0.43 (0.24–0.73)**	0.60 (0.33–1.06)
	35–44	0.74 (0.43–1.26)	0.73 (0.41–1.26)
	45–54	Reference	
	55–64	1.03 (0.61–1.73)	1.11 (0.64–1.90)
	65+	1.30 (0.73–2.26)	1.74 (0.95–3.14)
Sex	Female	Reference	
	Male	**1.47 (1.01–2.21)**	1.15 (0.78–1.76)
Income Assistance Participation	No	Reference	
	Yes	**2.67 (1.89–3.74)**	**2.18 (1.45–3.26)**
Neighbourhood Income Quintile	1-Lowest	1.51 (0.92–2.55)	1.12 (0.65–1.96)
	2-Medium low	0.94 (0.53–1.67)	0.86 (0.48–1.56)
	3-Medium	0.88 (0.49–1.57)	0.86 (0.47–1.56)
	4-Medium high	0.88 (0.48–1.60)	0.96 (0.52–1.78)
	5-Highest	Reference	
Urban/Rural	Urban	Reference	
	Rural	**2.25 (1.14–4.10)**	1.43 (0.69–2.78)
Health Authority	Fraser	Reference	
	Interior	**2.10 (1.25–3.52)**	**1.94 (1.13–3.31)**
	Northern	**3.53 (1.66–7.12)**	**2.97 (1.34–6.23)**
	Vancouver Coastal	0.89 (0.59–1.39)	1.20 (0.77–1.90)
	Vancouver Island	0.57 (0.31–1.02)	0.64 (0.34–1.16)

admissions and police data together account for a similar number of severe injuries, police records miss approximately half of all injuries requiring hospitalization for each mode of transport. We also found that amongst the injuries requiring hospital admission, certain demographic groups were more likely to be missing a corresponding insurance claims or police record. Specifically, injuries to child cyclists or child pedestrians (under age 16) and to low-income populations. That serious injuries requiring hospitalization are disproportionately missing amongst these demographic groups, emphasizes the need for a more inclusive and comprehensive approach to traffic injury data collection in the province.

Hospital admissions data are an important source of information for road safety surveillance in the province as they capture all cases that require hospital admission (Canadian Institute for Health Information, 2022) and provide the most detailed information on injury type and severity. Unfortunately, hospital data does not provide any information on the geographic location where the collision occurred, as insurance claims and police data do. The spatial location of a collision is critical for traffic safety analyses, particularly when evaluating the effectiveness of built environment interventions (e.g., specific infrastructure designs) or local policy changes (e.g., automated speed enforcement). Integrating ambulance data into B.C.'s linked datasets may enable the 'spatialization' of hospital data, as ambulance data records dispatch location, serving as a collision location proxy (Soltani et al., 2022). Ambulance data may also capture minor injuries that do not require hospital admission.

Insurance claims data capture far more serious injuries than hospital admissions data. Given the high quality of the hospital admissions data in terms of completeness, the serious injuries not captured by hospital admissions data but found in insurance claims are likely the result of insurance claims using a different definition of *serious* injury which includes injuries that did not require treatment in hospital. Many injuries deemed serious by the insurance provider but not linked to a hospital admission record likely would have required treatment in an emergency department. Unfortunately, B.C. emergency department data does not reliably capture mechanism of injury and this data gap significantly limits road traffic injury surveillance in the province. An Irish study that linked road traffic injury claims to a national "Injuries Board" found that auto-insurance claims that went unreported to police and did not require

Table 7
Pedestrian injury ascertainment rate by sociodemographic variables. Ascertainment rate is the total injuries captured by a given database divided by the total number of distinct injuries captured across all three databases.

Variable	Level	Injury Total (%)	Claims Total	Police Total	Hospital Total	Claims Asc.	Police Asc.	Hospital Asc.
Age Group	0–15	287 (4.6)	180	81	144	62.7	28.2	50.2
	16–24	685 (10.9)	574	224	270	83.8	32.7	39.4
	25–34	777 (12.4)	674	204	244	86.7	26.3	31.4
	35–44	548 (8.8)	480	165	187	87.6	30.1	34.1
	45–54	707 (11.3)	617	230	287	87.3	32.5	40.6
	55–64	684 (10.9)	608	239	323	88.9	34.9	47.2
	65+	1167 (18.6)	984	515	742	84.3	44.1	63.6
	Missing	1403 (22.4)	670	625	108	47.8	44.5	7.7
Sex	Female	2344 (37.5)	2,076	748	1,019	88.6	31.9	43.5
	Male	2511 (40.1)	2,041	910	1,178	81.3	36.2	46.9
	Missing	1403 (22.4)	670	625	108	47.8	44.5	7.7
Income Assistance	No	3872 (61.9)	3,255	1,322	1,728	84.1	34.1	44.6
	Yes	1091 (17.4)	862	336	577	79.0	30.8	52.9
	Missing	1295 (20.7)	670	625	0	51.7	48.3	0.0
Neighbourhood Income Quintile	1-Lowest	1450 (23.2)	1,225	516	741	84.5	35.6	51.1
	2-Medium low	1082 (17.3)	940	340	511	86.9	31.4	47.2
	3-Medium	824 (13.2)	706	280	389	85.7	34.0	47.2
	4-Medium high	658 (10.5)	579	230	295	88.0	35.0	44.8
	5-Highest	533 (8.5)	464	151	238	87.1	28.3	44.7
	Missing	1711 (27.3)	873	766	131	51.0	44.8	7.7
Urban/Rural	Rural	269 (4.3)	220	86	132	81.8	32.0	49.1
	Urban	4373 (69.9)	3,773	1,453	2,053	86.3	33.2	46.9
	Missing	1616 (25.8)	794	744	120	49.1	46.0	7.4
Health Authority	Fraser	1890 (30.2)	1,671	600	841	88.4	31.7	44.5
	Interior	526 (8.4)	455	199	284	86.5	37.8	54.0
	Northern	169 (2.7)	127	67	94	75.1	39.6	55.6
	Vancouver Coastal	1436 (22.9)	1,227	449	632	85.4	31.3	44.0
	Vancouver Island	626 (10)	517	226	337	82.6	36.1	53.8
	Missing	1611 (25.7)	790	742	117	49.0	46.1	7.3

Asc. = Ascertainment Rate.

Table 8
Logistic regression results for modelling the odds for serious pedestrian injury of not being captured by claims or police data.

Variable	Level	Unadjusted OR (95 % CI)	Adjusted OR (95 % CI)
Age Group	0–15	1.59 (0.99–2.53)	**2.13 (1.29–3.46)**
	16–24	1.12 (0.75–1.68)	1.47 (0.97–2.24)
	25–34	1.08 (0.73–1.60)	1.26 (0.84–1.90)
	35–44	0.87 (0.55–1.36)	0.92 (0.57–1.45)
	45–54	Reference	
	55–64	0.86 (0.56–1.31)	0.81 (0.52–1.25)
	65+	1.19 (0.84–1.71)	**1.88 (1.29–2.78)**
Sex	Females	Reference	
	Males	**1.75 (1.41–2.19)**	**1.64 (1.31–2.07)**
Income Assistance	No	Reference	
	Yes	**2.84 (2.28–3.54)**	**3.12 (2.43–4.00)**
Neighbourhood Income Quintile	1-Lowest	1.01 (0.72–1.43)	0.85 (0.60–1.23)
	2-Medium low	0.74 (0.51–1.08)	0.74 (0.51–1.10)
	3-Medium	0.83 (0.56–1.23)	0.85 (0.57–1.28)
	4-Medium high	**0.62 (0.40–0.96)**	**0.63 (0.40–0.99)**
	5-Highest	Reference	
Urban/Rural	Urban	Reference	
	Rural	1.46 (0.96–2.15)	1.49 (0.94–2.28)
Health Authority	Fraser	Reference	
	Interior	0.96 (0.63–1.41)	0.77 (0.50–1.17)
	Northern	**2.44 (1.50–3.84)**	**1.89 (1.13–3.06)**
	Vancouver Coastal	**1.41 (1.08–1.83)**	**1.47 (1.12–1.92)**
	Island	**1.62 (1.17–2.23)**	**1.48 (1.06–2.05)**

hospital admission made up approximately 55 % of injuries in their linked database (Short and Caulfield, 2016). They also found that amongst "bad injuries" (defined as an injury resulting in a settlement > 20,000 euros, or injuries with MAIS3 +), the ascertainment rate in insurance claims data was 65.6 %. While this injury board dataset is not directly comparable to insurance claims from a universal provider, it does support our finding that many serious traffic injuries are not captured by police or hospital admissions data.

In our analysis, unlike insurance claims, police records and hospital admissions share the same definition of a serious injury: an injury requiring hospital admission. Despite the shared definition of injury severity, there is a substantial lack of overlap between injuries captured by police data and in hospital admissions. The discrepancy between these datasets results from either police not attending a traffic crash and missing the injury event, or by police misclassifying the injury severity of the injured person in their collision report. Our results find that just over a third of hospitalized drivers, and over half of hospitalized pedestrians and cyclists could not be linked to a corresponding police report (irrespective of how the police report classified the injury severity), reflecting the proportion of injuries requiring hospitalization that are missed by police. Additionally, we found that approximately half of injuries defined as serious by police could not be linked to a corresponding hospital admission record, suggesting a misclassification of injury severity in their report. Previous research of police records linked to hospital admissions in BC from 2005 to 2015 for motor vehicle driver injury found similar disagreement in classification of serious injury by police and actual requirement for hospital admission (Brubacher et al., 2019). Specifically, the authors found only 56.2 % of injuries to motor vehicle drivers classified as severe (i.e. requiring a hospital admission) by police were admitted to hospital. Our results and previous research suggest that police do not reliably distinguish between injuries that require hospitalization and those that do not. This gap could be addressed by protocols to improve information sharing between police and hospital while protecting confidential personal information.

We analysed selection bias in auto-insurance claims and police data by studying hospital admissions for road trauma that were not matched to a corresponding police report or insurance claim. This analysis showed sociodemographic variation in persons with unreported injuries across transport modes. Specifically, unreported injuries were associated with younger and older age groups for pedestrians and cyclists. For drivers, older age was strongly associated with unreported injuries. Males were more likely than females to go unreported for pedestrian injury and, to a lesser extent, for cycling injury. We estimated that the unadjusted association between residing in rural areas and unreported injuries was attenuated in adjusted models, but remained positive, though only statistically significant for driver injury. Areas beyond the more densely populated Vancouver Coastal and Fraser Health authority regions, especially the Interior and/or Northern Health authorities, showed a higher likelihood of injuries going unreported in insurance and police data. Previous studies comparing police records to hospital data also found that police data are less likely to capture child injury (Alsop and Langley, 2001; Short and Caulfield, 2016) and injuries to older adults (Amoros et al., 2006; Watson et al., 2015). A single study linked hospital admissions data to injury claims data and found lower ascertainment rates for child injuries (0–14) and older adults (65 +) compared to those aged 15 – 65 years, as well as for males compared to females (Short and Caulfield, 2016). Previous research has also found that road traffic injuries within rural areas are more likely to go unreported (Alsop and Langley, 2001; Langley et al., 2003; Watson et al., 2015) in comparison to injuries occurring in urban areas. Lower reporting in rural areas may reflect the difficulty police have attending rural collisions; this problem could be rectified by improved information sharing protocols between police and hospital.

In our research, we estimated that serious injuries to pedestrians and cyclists were underreported in both insurance claims and police reports. However, because we included minor injuries (as determined in claims or police data) that were linked to a hospital admission, and because we excluded road traffic injuries not involving a motor vehicle, the underreporting in our study was significantly less pronounced than reported in other data linkage studies (Alsop and Langley, 2001; Amoros et al., 2006; Conderino et al., 2017; Elvik and Mysen, 1999; Janstrup et al., 2016; Langley et al., 2003; Short and Caulfield, 2016; Soltani et al., 2022; Watson et al., 2015). Previous studies have identified falls (not involving a motor vehicle) as a significant contributor to the injury burden for both cyclists and pedestrians, advocating for a more inclusive definition of pedestrian injury that encompasses falls on the roadway (Methorst et al., 2017; Schepers et al., 2017). Methorst et al. (2017) emphasized the importance of pedestrian falls in road safety research, citing their frequency and severity, and estimated that such falls occur 4–9 times more frequently for pedestrians than injuries from collisions with motor vehicles (Methorst et al., 2017). Australian research estimated that falls accounted for 48 % of on-road bicycle injuries requiring hospital admission (Beck et al., 2019). Previous research in Vancouver, B.C. has shown that motor vehicles are involved in only one quarter of injuries to either cyclists or pedestrians presenting to emergency departments (Brubacher et al., 2017a; Brubacher et al., 2017b; Teschke et al., 2014). More research is needed to examine the magnitude of injuries resulting from falls on public roads and consideration of how these injuries can be captured using emergency department data or other medical databases is required.

Our study offers a unique contribution by comparing two socioeconomic status metrics: neighbourhood income and participation in income assistance, although the associations observed are mixed. After adjusting for other demographic characteristics at the individual level, we estimated that participation in an income assistance program was strongly associated with a road traffic injury requiring hospitalization going unreported in police records and insurance claims, across all travel modes. We also found under-reporting of driver injuries in residents of low-income neighbourhoods; other neighbourhood income associations exhibited mixed results and wide confidence intervals. Previous research in Copenhagen, Denmark, found that higher income victims of a bicycling injury were more likely to believe that "bicycle accident reporting is useless" compared to lower income victims, while distrust of police was higher for people with a lower income (Janstrup et al., 2016). Given the vast majority of police-reported incidents in B.C. are recorded directly by police (there is no legal obligation to self-report a traffic crash to police in BC) and insurance claims rely on injured participants to file a claim, more research is needed on the motivators and deterrents for filing insurance claims in the province of B.C. and how certain populations may be more or less likely to do so. The disproportionate number of serious road injuries in low-income populations that go unreported by police or insurance claims can lead to inequitable outcomes when prioritizing locations for improvement in infrastructure or evaluating the efficacy of road safety interventions.

Our study possesses several strengths. We incorporated auto insurance claims in our data linkage, facilitating a more comprehensive and multi-faceted examination of serious traffic injuries in the province than would otherwise be possible. We were able to leverage novel linkages to unique government datasets including income assistance and postal code information, allowing for exploration of biases, and underreporting within disadvantaged populations (defined by need for income assistance or neighbourhood income), not possible in previous research. Furthermore, our research is population-based, with each dataset covering the entire province of B.C.

Our study also has limitations. Ascertainment rates in this study are potentially underestimated because of the number of injured people who could not be linked to the population directory. Approximately 22 % of total serious pedestrian injuries (48 % from insurance claims, 45 % from police records, and 9 % from hospital admissions) and 15 % of total serious cyclist injuries (52 % from insurance claims, 40 % from police records, and 9 % from hospital admissions) could not be linked to the population directory. This is either due to an individual not being in the population directory (e.g. an out of province resident) or due to a lack of identifying information for the injured person in the database. We found just over half of admissions data unlinked to the population directory were from out-of-province residents (police and claims data do not have provincial residency information). Better practices are needed to record identifying information (name, birthdate, address, gender and possibly driver's license or personal health number), for all injured pedestrians and cyclists, and particularly for motor vehicle passengers in insurance claims and police records. Another limitation of our study is that we exclusively focused on injuries classified as "serious," excluding minor injuries. Compared to hospital and police data, insurance claims use a different definition for injury severity, limiting comparisons to other datasets. Emergency department data is a vital source of injury data in many jurisdictions, because it captures both injuries that require hospital admission as well as injuries that require treatment but not overnight admission to hospital (Amin et al., 2022). However, in B.C., emergency department data from most hospitals does not include injury mechanisms that are required to identify road traffic injuries and transport mode, thus preventing its incorporation in this study. To explore injuries not requiring hospitalization, data from emergency department visits, enabling the identification of injuries by mode of transport, would be essential. Ambulance data is not integrated into the Data Innovation Program at this time which restricted its inclusion in this analysis. The ability to integrate data that is missing from police and insurance claims to provide a more complete capture of injuries is an important consideration for future road safety research.

Conclusions

This is the first study to evaluate underreporting and selection bias in insurance claims data from a universal auto-insurance provider. We provide a novel contribution by quantifying the likelihood of underreporting by socioeconomic status, enabled by a unique provincial government program designed to facilitate data linkages across

governmental institutions. Our findings show patterns of underreporting and selection bias in insurance claims data in B.C. where low-income populations are less frequently captured by insurance claims and police reports. We recommend that future efforts to improve traffic injury data collection focus on better collection of personal identifiers for all injured road users, standardization of injury severity definitions across datasets, collecting mechanism of injury in emergency department visit data and integrating ambulance and emergency department data into linked data analyses.

Disclosure

The authors report no conflicts of interest.

CRediT authorship contribution statement

Michael Branion-Calles: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft. **Andrea Godfreyson:** Funding acquisition, Project administration, Writing – review & editing. **Kate Berniaz:** Funding acquisition, Writing – review & editing. **Neil Arason:** Conceptualization, Funding acquisition, Writing – review & editing. **Herbert Chan:** Conceptualization. **Shannon Erdelyi:** Methodology. **Meghan Winters:** Writing – review & editing. **Joy Sengupta:** Writing – review & editing. **Mohamed Essa:** Writing – review & editing. **Fahra Rajabali:** Writing – review & editing. **Jeffrey R. Brubacher:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Funding

This work was supported by the BC Ministry of Transportation and Infrastructure and the BC Ministry of Health. MBC was supported by a CIHR Health System Impact Fellowship (#467148).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Alsop, J., Langley, J., 2001. Under-reporting of motor vehicle traffic crash victims in New Zealand. Accid. Anal. Prev. 33, 353–359.
Amin, K., Skyving, M., Bonander, C., Krafft, M., Nilson, F., 2022. Fall- and collision-related injuries among pedestrians in road traffic environment - A Swedish national register-based study. J. Saf. Res. 81, 153–165.
Amoros, E., Martin, J.-L., Laumon, B., 2006. Under-reporting of road crash casualties in France. Accid. Anal. Prev. 38, 627–635.
Ark, T.K., Kesselring, S., Hills, B., McGrail, K., 2020. Population Data BC: Supporting population data science in British Columbia. International Journal of Population. Data Sci. 4.
BC Ministry of Social Development and Poverty Reduction, 2014. Your Guide to Employment and Assistance.
BC Ministry of Social Development and Poverty Reduction, 2022. Persons with Disabilities.
Beck, B., Stevenson, M.R., Cameron, P., Oxley, J., Newstead, S., Olivier, J., Boufous, S., Gabbe, B.J., 2019. Crash characteristics of on-road single-bicycle crashes: an under-recognised problem. Inj. Prev. 25, 448–452.
Brubacher, J.R., Chan, H., Erdelyi, S., 2019. Injury severity in police collision reports correlates poorly with requirement for hospital admission. J. Transp. Health 14.
Brubacher, J.R., Sutton, G., Lam, C., Trajkovski, A., Yip, R., Liu, T., Chan, H., 2017a. P023: Emergency department data provides a realistic count of pedestrian injuries. CJEM 19, S85–S.
Brubacher, J.R., Yip, R., Trajkovski, A., Lam, C., Sutton, G., Liu, T., Chan, H., 2017b. MP15: Profile and circumstances of cycling injuries: Data from an urban emergency department. CJEM 19, S70–S.
Canadian Institute for Health Information, 2022. Discharge Abstract Database metadata (DAD).
Conderino, S., Fung, L., Sedlar, S., Norton, J.M., 2017. Linkage of traffic crash and hospitalization records with limited identifiers for enhanced public health surveillance. Accid. Anal. Prev. 101, 117–123.
Data Innovation Program (DIP), 2023a. Metadata for BC Employment and Assistance - E06.
Data Innovation Program (DIP), 2023b. Metadata for Census Geodata (Central Demographics) - E05 Data Innovation Program (DIP).
Data Innovation Program (DIP), 2023c. Metadata for Health Registry and Demographics (Central Demographics) - E04 Data Innovation Program (DIP).
Elvik, R., Mysen, A., 1999. Incomplete accident reporting: meta-analysis of studies made in 13 countries. Transportat. Res. Record: J. Transport. Research Board 1665, 133–140.
Government of British Columbia, 2022. Data Innovation Program.
Hosseinzadeh, A., Karimpour, A., Kluger, R., Orthober, R., 2022. Data linkage for crash outcome assessment: linking police-reported crashes, emergency response data, and trauma registry records. J. Saf. Res. 81, 21–35.
Janstrup, K.H., Kaplan, S., Hels, T., Lauritsen, J., Prato, C.G., 2016. Understanding traffic crash under-reporting: linking police and medical records to individual and crash characteristics. Traffic Inj. Prev. 17, 580–584.
Kamel, M.B., Sayed, T., Bigazzi, A., 2020. A composite zonal index for biking attractiveness and safety. Accid. Anal. Prev. 137, 105439-105439.
Langley, J.D., Dow, N., Stephenson, S., Kypri, K., 2003. Missing cyclists. Inj. Prev. 9, 376–379.
Lujic, S., Finch, C., Boufous, S., Hayen, A., Dunsmuir, W., 2008. How comparable are road traffic crash cases in hospital admissions data and police records? An examination of data linkage rates. Aust. N. Z. J. Public Health 32, 28–33.
Methorst, R., Schepers, P., Christie, N., Dijst, M., Risser, R., Sauter, D., van Wee, B., 2017. 'Pedestrian falls' as necessary addition to the current definition of traffic crashes for improved public health policies. J. Transp. Health 6, 10–12.
Osama, A., Sayed, T., 2019. A novel approach for identifying, diagnosing, and treating active transportation safety issues. Transp. Res. Rec. 2673, 813–823.
Osama, A., Sayed, T., Sacchi, E., 2018. A novel technique to identify hot zones for active commuters' crashes. Transp. Res. Rec. 2672, 266–276.
Population Data BC, 2023.
Schepers, P., den Brinker, B., Methorst, R., Helbich, M., 2017. Pedestrian falls: a review of the literature and future research directions. J. Saf. Res. 62, 227–234.
Short, J., Caulfield, B., 2016. Record linkage for road traffic injuries in Ireland using police hospital and injury claims data. J. Saf. Res. 58, 1–14.
Soltani, S., Schwarcz, L., Morris, D., Plevin, R., Dicker, R., Juillard, C., Nwabuo, A., Wier, M., 2022. What is counted counts: an innovative linkage of police, hospital, and spatial data for transportation injury prevention. J. Saf. Res. 83, 35–44.
Statistics Canada, 2024. Census Profile, 2021 Census of Population. Statistics Canada Catalogue no. 98-316-X2021001.
Teschke, K., Frendo, T., Shen, H., Harris, M.A., Reynolds, C.C.O., Cripton, P.A., Brubacher, J., Cusimano, M.D., Friedman, S.M., Hunte, G., Monro, M., Vernich, L., Babul, S., Chipman, M., Winters, M., 2014. Bicycling crash circumstances vary by route type: a cross-sectional analysis. BMC Public Health 14, 1–10.
Urban Systems, 2015. City of Vancouver Cycling Safety Study: Final Report, Vancouver.
Urban Systems, Alta Planning and Design, Acuere Consulting, Sayed, T., 2012. Pedestrian Safety Study: Final Report, pp. 1-18.
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.