# Crowdsourcing the pedestrian experience: Who's represented in the data?

Amanda Bishop[1], Victoria Fast[2], Trisalyn Nelson[3], Karen Laberee[4]

[1]Geography, University of Calgary, abbishop@ucalgary.ca
[2] Geography, University of Calgary, victoria.fast@ucalgary.ca
[3]Geography, University of California Santa Barbara, trisalyn@ucsb.edu
[4]Geography, University of Victoria, klaberee@uvic.ca

## ABSTRACT

Older adults are overrepresented in pedestrian-traffic-related collisions. These collisions represent a major public health concern because collisions with older adults lead to fatalities at triple the rate of other age cohorts. However, older adults are under-sampled in official and crowdsourced data. The persistent underrepresentation of older adults can result in their needs not being accommodated for and/or considered in pedestrian policy or planning. The goal of this study is to understand if volunteered geographic information (VGI) approaches, like WalkRollMap.org (WRM), can provide representative samples for how older adults experience incidents, missing sidewalk amenities, and hazards and concerns. The age distribution of the Canadian population is compared with WRM survey responses to visualize where contributions are over and underrepresented. Older adults represent 19% of the Canadian population but only 11.4% of all WRM respondents. Whereas adults aged 35-39 years represent 6.8% of the Canadian population but almost 20% of all WRM sample data. Consistent under- and over-sampling have the potential to create policy mismatches. Where transportation policy is directed to service those in the highest proportion of the population, but not necessarily those experiencing the greatest risks.

## 1. Introduction: High Risk and Under-Sampling

An analysis of Transport Canada's (2019) national collision database shows over 6.8 million traffic collisions have been reported since 1999. These data further reveal that in 2019, 1 in 10 of these collisions involved pedestrians and older adults (65 years and older) accounted for 16% of Canadian pedestrian-traffic-related collisions. Collisions involving older adult pedestrians often have more severe outcomes and result in fatalities at nearly triple the rate of other age cohorts. Individual and environmental characteristics can create real or perceived barriers for pedestrians that heighten collision risks and can immobilize older adults. Incidents resulting in non-fatal injuries can lead to a diminished quality of life, or increased healthcare pressures as hospital recovery times for older adults are often longer and more costly compared to younger pedestrians (Cloutier et al., 2017; Thomson, 2022). Additionally, older adults often face higher fatality risks because of pre-existing mobility and/or health conditions (Stoker et al., 2015), declining cognitive capabilities (Lord et al., 2018), and slower walking speeds that increase exposure risks (Lachapelle & Cloutier, 2017). These risks weigh heavily against the health benefits of walking, which include the reduced prevalence of cardiovascular conditions, dementia, and reduced social isolation among non-driving older adults (Berchicci et al., 2013; Kerr et al., 2012; Plaut et al., 2021).

Despite the known risks and increase fatality rates, many scholars suggest policy and planning practices are not creating pedestrian environments that accommodate for and/or consider

 Volume DOI: 10.17605/OSF.IO/N6DU5

the needs of older adults (Cloutier et al., 2017; Drilling & Neuhaus, 2019; Skinner et al., 2015). One reason is that there is limited data or research on the specific travel needs, behaviours, and risks older adults experience within these environments. Official data collection mechanisms—such as traffic counters (Lee & Sener, 2017), police and insurance reports (Branion-Calles et al., 2017), 311 reports (White & Trump, 2018), and other municipal data collection platforms— tend to underrepresent older adults (Cloutier et al., 2017). This gap represents a data blind spot regarding the (in)accessibility older adults experience in the pedestrian realm.

One platform being used to understand how pedestrians experience the urban environment is WalkRollMap.org (WRM), a simplified, user-oriented, map-based data collection platform that provides users a way to report the location of incidents, hazards, and missing amenities while walking and rolling. WRM reports contain open-ended questions where respondents can describe their concern(s) and identify solution(s). Additionally, socio-demographic data (age, gender, disability type, if applicable) is collected alongside each report. WRM emerged from the successful BikeMaps.org platform, which engaged thousands of cyclists to share their near-misses, collisions, and falls (Laberee et al., 2021; Nelson et al., 2015). BikeMaps.org data proved to be more complete than traditional bicycle safety data (Branion-Calles et al., 2017), providing qualitative and demographic data that help researchers and policymakers understand riders and the unique barriers they face while cycling (Ferster et al., 2017; Fischer et al., 2020).

New map-based crowdsourcing applications, such as WRM, that support the collection of volunteered geographic information (VGI) are increasingly providing powerful tools to engage citizens and fill gaps in existing data. Fast and Rinner (2014) show that active VGI systems enable data collection through time-limited, targeted, and participatory mapping interventions, which can support research and official data collection mechanisms. VGI can inform decision-making processes in situations where there are gaps in official data repositories. Official data collection mechanisms, for example, present location biases where stationary traffic counters are placed in high-priority areas and not necessarily high-risk areas (Lee & Sener, 2017). Therefore, data on near-misses with pedestrians, hazardous sidewalk conditions, and other local barriers are often missing or incomplete in *authoritative* or government-based data portals (Nelson et al., 2022). VGI can supplement authoritative data sets and minimize bias by contextualizing conflicts and providing the type of population level information often omitted and/or not collected by governments. VGI has a strong record for accuracy and completeness (Antoniou & Skopeliti, 2015; Fonte et al., 2017).

Then again, VGI has also been shown to exclude populations, perpetuate existing misrepresentations, and introduce general biases associated with sampling (Zhang & Zhu, 2018; Graham et al., 2014). As a manifestation of the digital divide, older adults are usually under-sampled because VGI technologies often contain digital tools that can be or are perceived to be difficult for older adults to use. The goal of our study is to investigate the representation of older adults in the VGI dataset generated by WRM. In this paper, we analyze and compare the age distribution of WRM survey respondents to the Canadian population, looking for potential over- and under- sampling of different age cohorts. We then discuss strategies to improve future engagement of older adult pedestrians.

## 2. Methods: Comparing WalkRollMap.org and Population Data

WalkRollMap.org (WRM) data has been collected in multiple Canadian municipalities, with over 1000 responses collected at the time of this study. Within the platform, participants can share the details of their experiences related to incidents, hazards and concerns, and missing amenities (Figure 1). To reach potential participants, the WRM team began engagement activities across the Capital Regional District (in BC). These activities included Facebook advertisements and social media posts, guided walks, and tables at festivals and markets. Additional promotion activities occurred in Saskatoon (SK), Toronto (ON), Ottawa (ON), and across the province of Quebec.

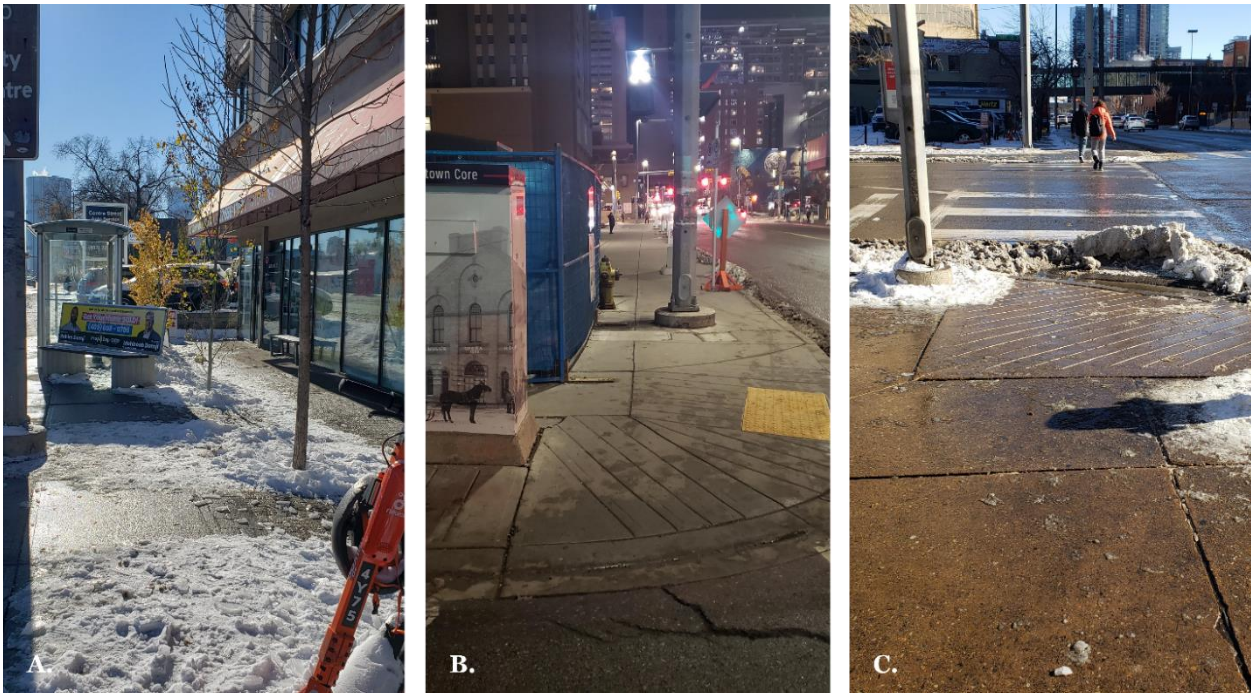


Figure 1: These images provide an example of what barriers can look like on sidewalks in Calgary (AB). Barriers can present as street furniture that is difficult to maneuver around with mobility devices (A), construction fences that narrow sidewalks (B), or uncleared snow that blocks access to intersections (C). In each of these situations, incidents are more likely to occur because barriers present hazards and concerns which may force a pedestrian to step off a sidewalk, thereby interacting with traffic to continue their journey.

The *incidents* survey collects data on conflicts between pedestrians and other road users, such as vehicles, cyclists, or animals. Data captured through this survey type reflects different severities of incidents and the associated contextual details. For example, incident surveys include information on the incident type (hit, near-miss, fall), if an injury occurred because of a pedestrian-traffic conflict, and a description of the event.

In the *hazards and concerns* survey, respondents can identify specific environmental characteristics that can contribute to pedestrian-traffic-related collisions. Hazards are categorized as crossing issues, safety/comfort concerns, sidewalk infrastructure issues, and weather-related/seasonal concerns. Respondents can also provide a detailed description of the event to further contextualize the environmental concern.

The *missing amenity* survey allows respondents to report conditions that can inhibit an individual's ability to access the built environment. Such as when sidewalks are not wide-enough for an older adult and their caregiver to walk side-by-side. With this survey, respondents can select amenity types and describe how the environment could be improved with the implementation of such an amenity. An amenity can include access to transit, auditory features (e.g., audio-enabled crossing buttons), connections between streets (e.g., cut-throughs), curb cuts, lighting, marked crosswalks, sidewalks, traffic signals, wayfinding, or other common amenities.

Data collection also includes socio-demographic information, such as gender, age, and disability of the individual involved in an event. To better understand the demographic representation of WRM data, we compare the age distribution of WRM data to the 2021 Canadian Population Census (Statistics Canada, 2022).

Data from all three WRM datasets contains records ranging from May 2021 – October 2022, these were downloaded as of October 27, 2022, and combined for pre-processing. Counts of respondents' ages were then further grouped into age ranges starting at 12 years, the age of eligibility for WRM respondents. Population and WRM data were normalized by calculating the relative frequency of both datasets. This included converting categorical age counts into percentages of the total Canadian population and the total number of WRM responses. Data from the Canadian Population Census is arranged in 5-year increments. Since WRM data begins at age 12, and the census data combines age 10-14 years, we assumed 60% of the 10-14 years age range to represent the WRM 12-14 category. Data were visualized using multiple line graphs to investigate the differences in the Canadian population compared to the WRM data.

## 3. Results

The Canadian age distribution (Table 1) shows young adults (12-29 years) represent 21.4% of the total Canadian population, adults (30-64 years) represent 46.8%, and older adults (65 years and older) represent 19%. Whereas older adults only represent 11.4% of all WRM responses. Representation of older adults in the WRM data varies across different survey types but data capture is lowest in the missing amenities (9.6%) survey. Adults are significantly overrepresented across all WRM responses. This age cohort surpasses the Canadian benchmark by over 20% and represents almost 70% of the entire WRM dataset. Total WRM representation of young adults is on par with the Canadian benchmark.

Table 1: Below is a representation of the Canadian population in comparison to the WalkRollMap.org (WRM) survey responses. Respondents aged 30-64 are overrepresented in comparison to the most vulnerable age cohorts, those 65 years and older. Totals may not add due to rounding. (Statistics Canada Catalogue no. 98-316-X2021001)

| Age Range | Canadian Population | Incidents | Missing Amenities | Hazards & Concerns | All WRM Surveys |
|---|---|---|---|---|---|
| Young Adults (12-29 years) | 21.4% | 13.9% | 27.1% | 19.4% | 21.3% |
| Adults (30-64 years) | 46.8% | 75.0% | 63.6% | 67.9% | 67.3% |
| Older Adults (65 years and older) | 19.0% | 11.1% | 9.6% | 12.7% | 11.4% |
| **Percentage of Total** | **87%** | **100%** | **100%** | **100%** | **100%** |

Visualizing the population distributions (Figure 2) of the data provides a clearer overview of what age ranges drive the overrepresentation of adults. Distribution amounts of the Canadian population (Figure 2, denoted by the blue bars) are similar among the middle age-ranges (adults)but are highest overall in the 55-59 years range, these begin to decline at the 60-64 years range and are lowest amongst the 75 years and older ranges.

In comparison, the 35-39 years range is consistently overrepresented in the WRM surveys. This age range represents 6.8% of the Canadian population but 20% of the missing amenities data (Figure 2A), 14% of the incident data (Figure 2B), and 18% of the hazard and concerns data (Figure 2C). The 35-39 years age range is overrepresented by an average of 11% in all WRM data, whereas older adults in all age cohorts are underrepresented by an average of 8% in WRM surveys. Data availability significantly diminishes among the 70-74 years ranges in the WRM surveys. WRM responses are virtually non-existent for individuals 80 years and older, despite this population representing 4.6% of the Canadian population.


Volume DOI: 10.17605/OSF.IO/N6DU5

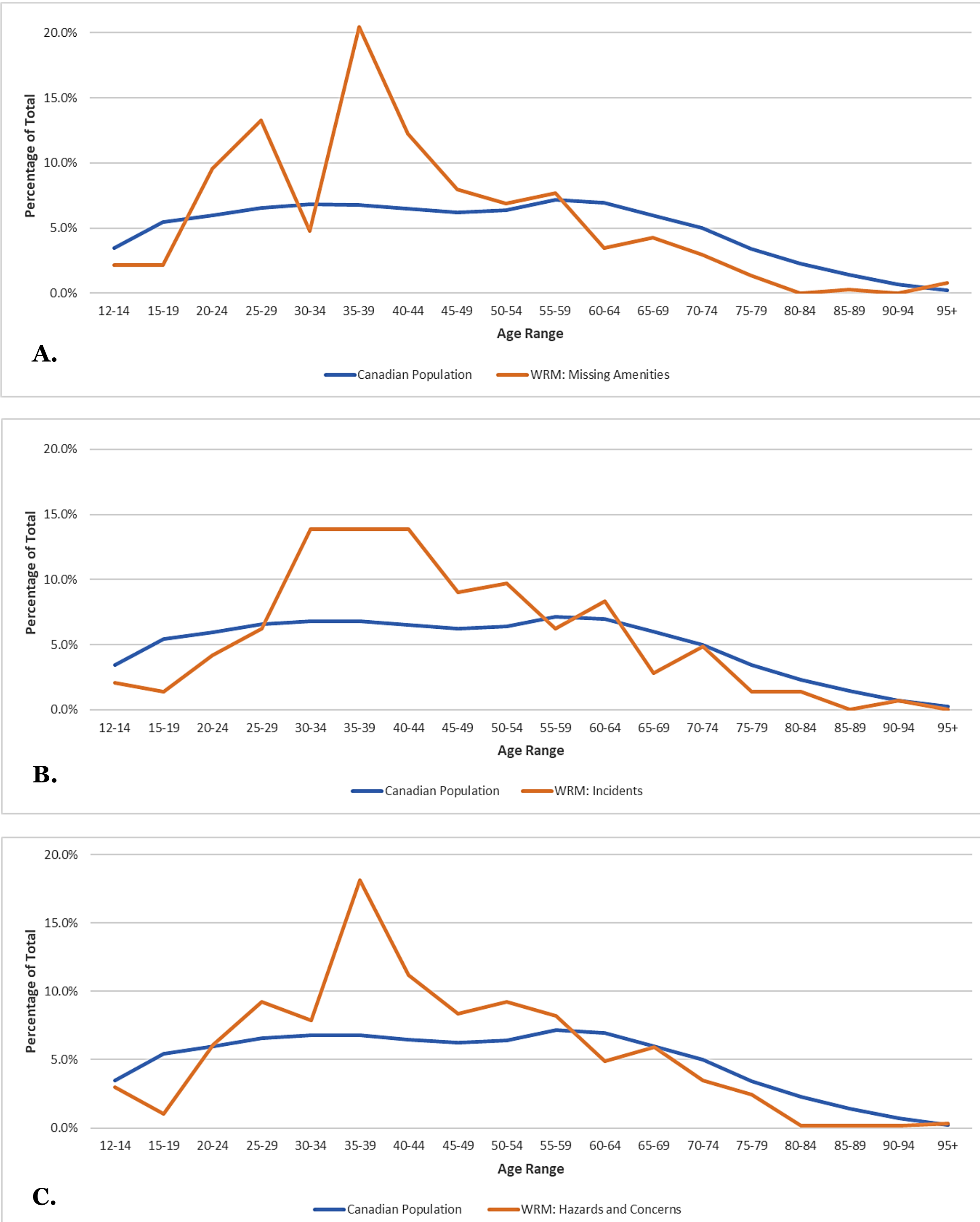


Figure 2: The above line graph visualizes the age distribution of the Canadian population in comparison to the responses collected from the missing amenities (A), incidents (B), and hazard and concerns (C) surveys on WalkRollMap.org (WRM). WRM contributions are unevenly distributed across all age cohorts. Respondents aged 35-39 are consistently overrepresented in comparison to the most vulnerable age cohorts. Contributions from older adults (65 plus) are consistently lower than the Canadian benchmark. This disparity is concerning because the challenges each cohort experiences are not adequately represented in the data.

## 5. Discussion: Persistent Underrepresentation

Analysis of the results shows evidence of under-sampling among older adults (65+) and young adults (12-29), and over-sampling of adults aged 35-39 years in the WRM data. The results further show a pressing need to increase the representation of older adults in sample data through equitable data collection strategies. Collecting representative samples from older adults, however, can be difficult even with the assistance of digital platforms (Liljas et al., 2017; van Deursen et al., 2017). Access to these services is based, in part, on the assumption that individuals know about the existence of such digital platforms and have the digital literacy skills required to comfortably use them. The challenge in closing the digital divide can be magnified when data collection strategies are biased towards younger cohorts. For example, overrepresentation of respondents in the 35-39 years cohort could reflect researcher bias. In this case, the WRM platform may have been promoted among networks of similar-aged individuals or on digital platforms that attract younger cohorts (i.e., Facebook, Twitter, Instagram).

The consistent underrepresentation of older adults, paired with the overrepresentation of the 35-39 years cohort, has the potential to create a policy mismatch. Where transportation policy is directed to service those with the highest proportion of the population, but not necessarily those experiencing the greatest risks. Planning and policy practices are typically implemented with population-neutral decision-making practices (Golub et al., 2013), or with data that may not have a representative sample of vulnerable populations. Graham et al. (2014) argue representative data is needed to disrupt the "uneven geographies of codified information…because [these] shape what is known and what can be known" (p. 748), and how knowledge is produced and reproduced. Put another way, those who create, and control data also influence how our built environment operates, what features are developed in it, and whom it is meant to serve. Therefore, without representative data from vulnerable populations, we run the risk of further excluding these needs when improvements are made to an area.

By identifying and calling out this data injustice, we can begin to address the divisions in the production of knowledge and policy. For example, when evidence emerged about the underrepresentation of women and girls in mapping (YouthMappers, 2022), initiatives like "Everywhere She Maps" purposefully targeted female led data collection to combat sampling bias in open-map data. Sampling bias with VGI approaches has previously been seen with gender-based knowledge gatekeeping (Stephens, 2013), and in cycling studies where "data is [often] bias[ed] towards recreational" (Roy et al., 2019, p. 2), or male riders (Ferster et al., 2017). Representation biases can reduce the relevance of data sources and public realm improvements for older adults.

Taking a lesson from these examples, deploying new and creative approaches can engage older adults and help them report barriers they encounter. This may involve recognizing the different ways older adults consume and share information. Ferster et al. (2017), suggest different promotion strategies are "critical for the uptake and use of crowdsourced" tools (p. 13). They discovered promotion through traditional print media, for example, was associated with more responses from older males. Targeted approaches to collecting sample data on older adults can help fill knowledge gaps and influence policy direction to provide age friendly-streets (World Health Organization, 2007).

Representing older adult populations may require a multi-prong approach that involves educating older adults on the platform's existence and helping bridge the data literacy gap through targeted public talks and/or training opportunities. In Liljas et al. (2017) study of engagement strategies with hard-to-reach populations, they recommend overcoming barriers by building trust through in-person engagements between communities of interest and researchers. The next steps

 Volume DOI: 10.17605/OSF.IO/N6DU5

in this research will be applying best practices in participatory mapping and community engagement, such as targeted and in-person-based education approaches to build trust within senior communities and broaden the representation of older adults.

## 6. Conclusion

In this paper, we analyze the demographic information of WalkRollMap.org responses. We note that older adults are underrepresented in the data. Targeted interventions and participatory mapping strategies need to be deployed to include these voices. Wider data collection is required to close data gaps and combat selective policy implementation that may omit risks experienced by vulnerable populations in favour of utilitarian arguments—choosing to serve the greatest demand and not necessarily those experiencing the greatest risks.

In the meantime, WRM remains operational (at the time of publication) and is open to new responses—from all age cohorts, demographics, and dis/abilities. The data on barriers (not demographics) is also open and available for download and analysis. We encourage the open use and active promotion of this platform to broaden the range of data capture across all demographics.

Acknowledgments
This research is supported, in part, by the Public Health Agency of Canada.

## References

Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: An overview. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. 2, pp. 345–351). Hellenic Military Geographical Service, Athens, Greece. https://doi.org/10.5194/isprsannals-II-3-W5-345-2015

Berchicci, M., Lucci, G., & di Russo, F. (2013). Benefits of physical exercise on the aging brain: The role of the prefrontal cortex. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, *68*(11), 1337–1341. https://doi.org/10.1093/gerona/glt094

Branion-Calles, M., Nelson, T., & Winters, M. (2017). Comparing crowdsourced near-miss and collision cycling data and official bike safety reporting. *Transportation Research Record*, *2662*(1), 1–11. https://doi.org/10.3141/2662-01

Cloutier, M. S., Lachapelle, U., d'Amours-Ouellet, A. A., Bergeron, J., Lord, S., & Torres, J. (2017). "Outta my way!" Individual and environmental correlates of interactions between pedestrians and vehicles during street crossings. *Accident Analysis and Prevention*, *104*(April), 36–45. https://doi.org/10.1016/j.aap.2017.04.015

Drilling, M., & Neuhaus, F. (2019). The city, aging and urban planning. *Urban Planning*, *4*(2). www.cogitatiopress.com/urbanplanning

Fast, V., & Rinner, C. (2014). A systems perspective on volunteered geographic information. *International Journal of Geo-Information*, *3*, 1278–1292. https://doi.org/10.3390/ijgi3041278

Ferster, C., Nelson, T., Laberee, K., Vanlaar, W., & Winters, M. (2017). Promoting crowdsourcing for urban research: Cycling safety citizen science in four cities. *Urban Science*, *1*(4), 21. https://doi.org/10.3390/urbansci1020021

Fischer, J., Nelson, T., Laberee, K., & Winters, M. (2020). What does crowdsourced data tell us about bicycling injury? A case study in a mid-sized Canadian city. *Accident Analysis and Prevention*, *145*(July), 105695. https://doi.org/10.1016/j.aap.2020.105695

Fonte, C. C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J. J., Bayas, J. C. L., ... & Vatseva, R. (2017). Assessing VGI data quality. *Mapping and the citizen sensor*, 137-163.

Golub, A., Marcantonio, R. A., & Sanchez, T. W. (2013). Race, space, and struggles for mobility: Transportation impacts on African Americans in Oakland and the East Bay. *Urban Geography*, *34*(5), 699–728. https://doi.org/10.1080/02723638.2013.778598

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers*, *104*(4), 746–764. https://doi.org/10.1080/00045608.2014.910087

Kerr, J., Rosenberg, D., & Frank, L. (2012). The role of the built environment in healthy aging: Community design, physical activity, and health among older adults. In *Journal of Planning Literature* (Vol. 27, Issue 1, pp. 43–60). https://doi.org/10.1177/0885412211415283

Laberee, K., Nelson, T., Branion-Calles, M., Ferster, C., & Winters, M. (2021). Crowdsourced bicycling crashes and near misses: trends in Canadian cities. *Urban, Planning and Transport Research,* 9(1), 449–463. https://doi.org/10.1080/21650020.2021.1964376

Lachapelle, U., & Cloutier, M. S. (2017). On the complexity of finishing a crossing on time: Elderly pedestrians, timing and cycling infrastructure. *Transportation Research Part A: Policy and Practice*, 96, 54–63. https://doi.org/10.1016/j.tra.2016.12.005

Lee, K., & Sener, I. N. (2017). Emerging data mining for pedestrian and bicyclist monitoring: a literature review report. Texas A&M Transportation Institute. https://safed.vtti.vt.edu/wp-content/uploads/2020/07/UTC-Safe-D_Emerging-Data-Mining-for-PedBike_TTI-Report_26Sep17_final.pdf

Liljas, A. E. M., Walters, K., Jovicic, A., Iliffe, S., Manthorpe, J., Goodman, C., & Kharicha, K. (2017). Strategies to improve engagement of 'hard to reach' older people in research on health promotion: a systematic review. *BioMed Central Public Health*, *17*(349), 1–12. https://doi.org/10.1186/s12889-017-4241-8

Lord, S., Cloutier, M. S., Garnier, B., & Christoforou, Z. (2018). Crossing road intersections in old age—With or without risks? Perceptions of risk and crossing behaviours among the elderly. *Transportation Research Part F: Traffic Psychology and Behaviour*, 55, 282–296. https://doi.org/10.1016/j.trf.2018.03.005

Nelson, T., Denouden, T., Jestico, B., Laberee, K., & Winters, M. (2015). BikeMaps.org: A global tool for collision and near miss mapping. *Frontiers in Public Health*, 3(MAR), 1–8. https://doi.org/10.3389/fpubh.2015.00053

Nelson, T., Goodchild, M. F., & Wright, D. (2022). Accelerating ethics, empathy, and equity in geographic information science. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(19), 1–12. https://doi.org/10.1073/pnas.2119967119

Plaut, P., Shach-Pinsly, D., Schreuer, N., & Kizony, R. (2021). The reflection of the fear of falls and risk of falling in walking activity spaces of older adults in various urban environments. *Journal of Transport Geography*, *95*. https://doi.org/10.1016/j.jtrangeo.2021.103152

Roy, A., Nelson, T. A., Fotheringham, A. S., & Winters, M. (2019). Correcting bias in crowdsourced data to map bicycle ridership of all bicyclists. *Urban Science*, *3*(2), 62. https://doi.org/10.3390/urbansci3020062

Skinner, M. W., Cloutier, D., & Andrews, G. J. (2015). Geographies of ageing: Progress and possibilities after two decades of change. *Progress in Human Geography*, *39*(6), 776–799. https://doi.org/10.1177/0309132514558444

Statistics Canada. (2022). Census profile. 2021 Census of population. Catalogue no. 98-316-X2021001. [Dataset]. Statistics Canada. Ottawa. https://www150.statcan.gc.ca/n1/en/catalogue/98-316-X2021001

Stephens, M. (2013). Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. *GeoJournal*, *78*, 981–996. https://doi.org/10.1007/s10708-013-9492-z

Stoker, P., Garfinkel-Castro, A., Khayesi, M., Odero, W., Mwangi, M. N., Peden, M., & Ewing, R. (2015). Pedestrian safety and the built environment: A review of the risk factors. *Journal of Planning Literature*, *30*(4), 377–392. https://doi.org/10.1177/0885412215595438

Thomson, S. (2022). By 2042, one in every four Canadians will be a senior. Our health-care system isn't ready. The Hub. https://thehub.ca/2022-12-05/by-2042-one-in-every-four-canadians-will-be-a-senior-our-health-care-system-isnt-ready/?utm_medium=paid+social&utm_source=twitter&utm_campaign=boost&twclid=27a0pujododxdnihor3299fgvn

Transport Canada. (2019). *National collision database online*. [Dataset]. Transport Canada. https://wwwapps2.tc.gc.ca/Saf-Sec-Sur/7/NCDB-BNDC/p.aspx?l=en&l=en

van Deursen, A. J. A. M., Helsper, E. J., Eynon, R., & van Dijk, J. A. G. M. (2017). The compoundness and sequentiality of digital inequality. In *International Journal of Communication* (Vol. 11). http://ijoc.org.

White, A., & Trump, K. S. (2018). The promises and pitfalls of 311 data. *Urban Affairs Review*, *54*(4), 794–823. https://doi.org/10.1177/1078087416673202

World Health Organization. (2007). *Global age-friendly cities: A guide*. World Health Organization.

YouthMappers. (2022). *Everywhere she maps*. YouthMappers. https://www.youthmappers.org/everywhereshemaps

Zhang, G., & Zhu, A. X. (2018). The representativeness and spatial bias of volunteered geographic information: a review. *Annals of GIS*, 24(3), 151-162. https://doi.org/10.1080/19475683.2018.1501607