Transport Reviews

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttrv20

Crowdsourced data for bicycling research and practice

Trisalyn Nelson, Colin Ferster, Karen Laberee, Daniel Fuller & Meghan Winters

To cite this article: Trisalyn Nelson, Colin Ferster, Karen Laberee, Daniel Fuller & Meghan Winters (2021) Crowdsourced data for bicycling research and practice, Transport Reviews, 41:1, 97-114, DOI: 10.1080/01441647.2020.1806943

To link to this article: https://doi.org/10.1080/01441647.2020.1806943

Published online: 15 Aug 2020.

TRANSPORT REVIEWS
2021, VOL. 41, NO. 1, 97–114
https://doi.org/10.1080/01441647.2020.1806943

Crowdsourced data for bicycling research and practice

Trisalyn Nelson[a], Colin Ferster[b], Karen Laberee[b], Daniel Fuller[c] and Meghan Winters[d]

[a]School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA; [b]Department of Geography, University of Victoria, Victoria, BC, Canada; [c]School of Human Kinetics and Recreation, Memorial University, St. John's, NL, Canada; [d]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

ABSTRACT

Cities are promoting bicycling for transportation as an antidote to increased traffic congestion, obesity and related health issues, and air pollution. However, both research and practice have been stalled by lack of data on bicycling volumes, safety, infrastructure, and public attitudes. New technologies such as GPS-enabled smartphones, crowdsourcing tools, and social media are changing the potential sources for bicycling data. However, many of the developments are coming from data science and it can be difficult evaluate the strengths and limitations of crowdsourced data. In this narrative review we provide an overview and critique of crowdsourced data that are being used to fill gaps and advance bicycling behaviour and safety knowledge. We assess crowdsourced data used to map ridership (fitness, bike share, and GPS/accelerometer data), assess safety (web-map tools), map infrastructure (OpenStreetMap), and track attitudes (social media). For each category of data, we discuss the challenges and opportunities they offer for researchers and practitioners. Fitness app data can be used to model spatial variation in bicycling ridership volumes, and GPS/accelerometer data offer new potential to characterise route choice and origin-destination of bicycling trips; however, working with these data requires a high level of training in data science. New sources of safety and near miss data can be used to address underreporting and increase predictive capacity but require grassroots promotion and are often best used when combined with official reports. Crowdsourced bicycling infrastructure data can be timely and facilitate comparisons across multiple cities; however, such data must be assessed for consistency in route type labels. Using social media, it is possible to track reactions to bicycle policy and infrastructure changes, yet linking attitudes expressed on social media platforms with broader populations is a challenge. New data present opportunities for improving our understanding of bicycling and supporting decision making towards transportation options that are healthy and safe for all. However, there are challenges, such as who has data access and how data crowdsourced tools are funded, protection of individual privacy, representativeness of data and impact of biased data on equity in

ARTICLE HISTORY
Received 11 December 2019
Accepted 31 July 2020

KEYWORDS
Crowdsourced; bicycling; exposure; safety; infrastructure; attitudes

CONTACT Trisalyn Nelson trisalyn.nelson@asu.edu

decision making, and stakeholder capacity to use data given the requirement for advanced data science skills. If cities are to benefit from these new data, methodological developments and tools and training for end-users will need to track with the momentum of crowdsourced data.

Introduction

Around the world, cities are increasingly prioritising bicycling. Bicycling for transportation is being hailed as a solution to transportation congestion and air pollution, and active travel provides health benefits related to physical activity (Fishman, Schepers, & Kamphuis, 2015). As a result, cities are setting goals to increase mode share of bicycling and are investing in policies and infrastructure to increase bicycling ridership. Decisions about which policies to implement and where to prioritise investment have created an unprecedented need for data and information on bicycling in cities.

The lack of data on many aspects of bicycling have limited our ability to research and implement pro-bicycling policy (DiGioia, Watkins, Xu, Rodgers, & Guensler, 2017). Although cities invest heavily in tracking vehicular volumes, safety, and infrastructure, the budgets for collection of bicycling data are limited. For example, bicycling volume data are usually only available for a handful of locations and limited time periods (Roy, Nelson, Fotheringham, & Winters, 2019). Bicycling safety incidents are underreported, on the order of >85% underreported in some cities (Winters & Branion-Calles, 2017). Given that changes to bicycling infrastructure are often incremental, they can be difficult to track.

Gaps left by traditional sources of bicycling data are being filled with crowdsourcing. Crowdsourcing is a general term that means users generate data (Xu & Nyerges, 2017); within this the term "citizen science" is often used to encompass activities that broaden participation in research and practice (Eitzel et al., 2017). In the case of bicycling, this means that all road users can contribute data on bicycling levels, safety, and conditions. Crowdsourcing and citizen science are being used, with growing popularity, to generate data on everything from bike safety incidents to infrastructure. As well, people are sharing their perceptions of bicycling, infrastructure projects, and response to policies through social media, which creates an archive of data of interest for bicycling research and practice.

Our goal is to overview and critique crowdsourced data that are being used to fill gaps and advance bicycling research around bicycling volumes, safety, infrastructure, and public attitudes. As a collaboration between data scientists and transportation researchers, we aim to explore the strengths and weaknesses of emerging crowdsourced data as they apply to bicycling research. We build on a previous review of bicycling and big data that focused on ridership data (Romanillos, Austwick, Ettema, & De Kruijf, 2016). Nearly five years on in a rapidly changing environment, we provide updates on ridership data and further introduce new domains of crowdsourced data for bicycling (Figure 1). We focus on data issues; however, we also touch on methods, given that issues of data and methods are inevitably intertwined. This is not an attempt to provide a systematic review of all literature; rather we aim to highlight dominant trends. We set the stage for discussion of each data type by reviewing the current state of more traditional data

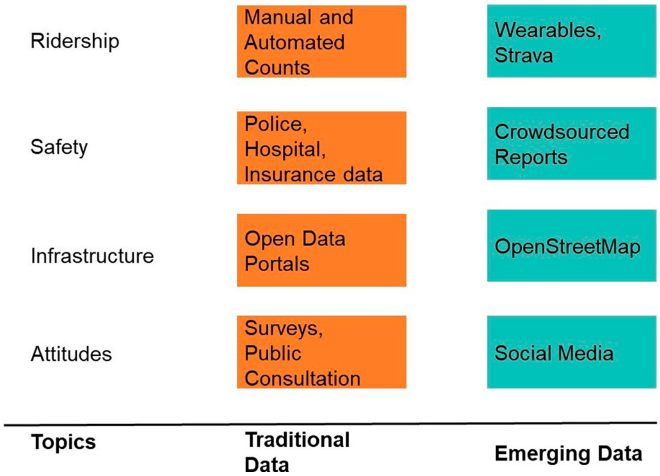

Figure 1. An overview of crowdsourced data for bicycling topics covered in this review and their juxtaposition with traditional datasets.

collection. Finally, we conclude with recommendations for advancing the use of novel data in bicycling research and planning practice.

Ridership

Bicycling volume data are needed for planning bicycling infrastructure, monitoring changes in ridership, characterising network connectivity, and as exposure data for safety and health studies (Vanparijs, Panis, Meeusen, & de Geus, 2015). Yet collecting bicycling volumes at the street level has been limited by the availability of tools for data collection. Typically, data are collected using temporary counts at numerous locations (i.e. pneumatic tube counters or video), permanent counts at a few locations (i.e. automated in-ground counters), or periodically through count programmes facilitated by volunteers. Count programmes provide critical and complete data on bicycling volumes at a particular location and time; however, they lack the spatial and temporal detail that is often desirable to meet bicycling information needs (Roy et al., 2019).

GPS-enabled smartphones and wearables are transforming how data on movement, including bicycling, are collected and in Table 1 we provide details of some key papers that use crowdsourcing tools to collect data and analyse or map ridership. Fitness apps like Strava, MapMyRide, and Garmin Connect are used by many bicyclists to track and monitor their bicycling activity. As a result, massive global databases on bicycling ridership now exist. Strava data, for example, provide the number of bicyclists using Strava on every road segment with 15-minute resolution; thus, Strava data are continuous across space and through time (Sun & Mobasheri, 2017). Of course, Strava bicyclists are a sample of all bicyclists, and their proportion relative to all bicyclists varies across infrastructure, road type, and neighbourhood characteristics (Griffin & Jiao, 2015). However, in dense urban cores, the patterns of Strava ridership are a good proxy for overall bicycling (Jestico, Nelson, & Winters, 2016), suggesting that despite demographic bias (Strava riders are disproportionately middle-aged and men) the spatial patterns may be

Table 1. Literature on crowdsourced data for bicycling ridership.

Type of crowdsourced data	Reference	Research paper goal	Opportunities of using crowdsourced data	Select challenges that arose with crowdsourced data
Strava	Roy et al. (2019)	Develop an approach to statistically correcting bias in Strava data.	Strava was used to predict bicycling ridership within ±100 average annual bicyclists for 86% of road segments.	Method was sufficient to predict categorical cycling levels, not exact volumes.
Strava	Hochmair, Bardin, and Ahmouda (2019)	Identify built environment and sociodemographic factors associated with bicycling ridership.	High resolution of crowdsourced data, combined with a technique for bias made it possible to characterise relationships between ridership and other factors across an entire network.	Using patterns in big data it is not possible to determine causality and further experimental research and data are needed.
Strava (+OSM)	Orellano and Guerrero (2019)	Explore the effect spatial configuration of street networks on movement patterns using Strava Metro and OpenStreetMap.	Strava data enabled evaluation of a bicycling network.	Sampling bias in Strava data was not accounted for.
Strava (+Infrastructure)	Boss et al. (2018)	Detect changes in bicycling volumes after infrastructure changes.	Installation of bicycling infrastructure changes volumes of bicyclists at multiple locations in a city.	Difficult to bias correct data at a very high temporal resolution, due to limited data for training that has the same temporal resolution.
Smartphone app	Garber et al. (2019)	Quantify race and gender bias in users of smartphone apps to record bicycling ridership.	Presents an approach to using app data in a more representative way.	Demonstrates bias in standard app generated data.
Smartphone app	Pritchard et al. (2019)	Study route and mode choice of study participants before and after new cycling infrastructure using a passive smartphone app (Moves).	Using an app it is possible to collect complete routes of individual bicyclists with high accuracy.	Small samples associated with recruitment limited representativeness of sample
Smartphone app	Blanc and Figliozzi (2017)	Crowdsourced data collected by an app (developed by state DOT) to model cyclists' comfort.	Using an app it is possible to collect observed route information and quantify impacts on comfort levels of trip purpose and sources of stress along a route.	Possible overrepresentation by year-round bicycle commuters. Transferability to other areas and cities is unknown.
GPS/ Accelerometry	Brondeel et al. (2015)	Construct an algorithm to reliably predict transportation mode at trip level with data collected from GPS/accelerometers.	Using machine learning, GPS/accelerometry made it possible to correctly predict transportation mode in 90% of trips.	Mode detection remains challenging for bicycling and method inappropriate for multi-modal trips.
Bike share	Winters, Hosford, and Javahera (2019)	Characterise the super-users of a bikeshare to understand who is capitalising to understand the equity implications of a public programme.	System data with userID enabled the identification of "super-users" – 10% of users who made 50% of bike share trips. From user survey with demographic data, determined super-users were lower income with fewer transportation options than other members.	Demographic data not included in bike share ridership data; analyses relies on linkage to user survey data.
Bike share	Scott and Ciuro (2019)	Effect of weather conditions, hub attributes, temporal variables on ridership.	Because data are collected continuously, there is a full year of data across space and time to compare to weather and temporal variables.	High resolution data was buffered which creates issues with modifiable areal units and scale.
Bike share (+GPS)	Wergin and Buehler (2017)	Examine the routes taken by bike share users to identify geographic areas and individual road segments used; correlate routes with infrastructure and highlight gaps; and describe stops.	By using the on-bike GPS and the bike share system data, researchers were able to identify differences in trips by membership type	Accuracy of the GPS pings may be limited in urban areas; spatial granularity lacking to discern sidewalk from road riding.

representative. A unique benefit of Strava data is that because it is continuously collected and archived, it can be drawn on for retrospective pre–post studies. For example, if an unexpected controversy arises over new infrastructure, Strava provides a unique ability to evaluate usage before and after (Boss, Nelson, Winters, & Ferster, 2018).

Beyond fitness apps, researchers use GPS and/or accelerometry data to track bicycling trips, often to supplement or replace more travel diary tools (Kerr, Duncan, & Schipperijn, 2011). These studies invite participants to use research- and practice-oriented apps (e.g. Ride Report, Cycle Atlanta) or provide participants with GPS devices for a study period. The resulting datasets comprise millions of second-by-second data on individual travel behaviours. One of the advantages of using researcher apps is that the individual tracks of bicyclists can be mapped. Strava, for example, provides the number of bicyclists per road segment, rather than tracks, to protect the privacy of individual bicyclists who could easily be identified by frequent origin and destination locations. Yet with accelerometry activity, it remains difficult to identify bicycling activities (Broach, Dill, & McNeil, 2019). As well, use of GPS and accelerometry data require recruitment of participants to download and use an app of study-specific technology. Even recruiting a few hundred participants can require huge effort and recruiting a representative sample of participants can be extremely difficult.

Bike share is another source of crowdsourced data on ridership. An excellent source for bike share data is Bike Share Research (https://bikeshare-research.org/), a collaborative, open data platform linked to system data from hundreds of bike share programmes globally and allowing API accessibility. There are many papers on bike share, in part because it was one of the first detailed datasets on ridership. Recent reviews provide insights from bike share data (e.g. Hirsch et al., 2019), and there is potential for GPS movement data from bike share bicycles (Romanillos, Moya-Gómez, Zaltz-Austwick, & Lamíquiz-Daudén, 2018). However, most systems provide only origin-destination data (not routes), and the geographic coverage of systems is often limited to downtown cores. Select systems provide gender and age categories for members, demographic information that is rarely available in ridership data.

Fitness app and GPS and/or accelerometry datasets can be massive. For example, Strava data from 2016 for the Phoenix region has ~44 million records. While Strava has outstanding spatial and temporal coverage and resolution, it is difficult to wrangle the raw data into information products, as data are not stored in GIS-ready formats, and they include multiple measures of bicycling by direction travelled and differently for street segments and intersections. Cleaning data directionality and connecting aspatial data to a road network can be time consuming and require skills with handling large volumes of data because there are many time periods of data included. Similarly, with GPS and/or accelerometry data there are not yet reliable and transferable tools to detect trips/dwells and identify the mode of travel. Bicycle trips may be the most challenging to identify because bicyclists can go as fast as slow cars or as slow as fast pedestrians (Broach et al., 2019; Brondeel, Pannier, & Chaix, 2015). If these data sources are to be tapped widely for research and practice applications, substantial effort is needed to develop and standardise tools for data processing. At present, strong data science skills are needed to put the data to their best use. For example, there are legitimate concerns about bias of the Strava sample, which is disproportionally used by middle-aged, white men (Garber, Watkins, & Kramer, 2019), that have led to development of many studies of how best to use Strava

(Mcnair & Arnold, 2016). As with many big data analyses, it is critical that when using Strava data for research and practice applications that expert opinion, local knowledge, and appropriate goals and metrics are also considered (Griffin, Mulhall, Simek, & Riggs, 2020). Additionally, methods to correct the bias in Strava data have been developed (Roy et al., 2019) to map all ridership. Using official counts to train and test a model, as well as geographic covariates, average annual daily bicycling can be predicted within 100 bicyclists for 86% of street segments. Limitations of the modelling include lower accuracies in low ridership areas and the precision of data makes the results best presented as categorical ridership (low/medium/high). However, implementation of bias correction requires advanced programming and statistical expertise. The uses of these data are just beginning to be understood and patterns in data have potential to help stratify count programmes (Brum-Bastos, Ferster, Nelson, & Winters, 2019) and monitor change (Boss et al., 2018; Hong, McArthur, & Livingston, 2019). At present, cell phone GPS is not accurate enough to pick up sidewalk riding (Wergin & Buehler, 2017) and using GPS/accelerometry data to differentiate multimodal trips is problematic (Brondeel et al., 2015). For the full potential of fitness app and GPS and/or accelerometer data to be used for mainstream planning new methods and tools are needed to easily ingest and work with data.

Safety

Bicycling incidents are underreported (Medury, Grembek, Loukaitou-Sideris, & Shafizadeh, 2019) and detailed reports of safety incidents are difficult for researchers to obtain. Based on hospital reports, it has been shown that only 15% (Winters & Branion-Calles, 2017) to 30% (Teschke et al., 2014) of bicycling incidents, even those that require hospital attention, are reported to official sources such as insurance or police. Of the bicycling incidents that do get reported, most include vehicles, whereas single bicycle crashes or those that involve infrastructure or other bikes are rarely reported (Branion-Calles, Nelson, & Winters, 2017). Fatality data are collected more completely, but often lack georeferencing in detail making it difficult to link fatality data with other spatial datasets. Finally, near miss data are rarely collected, although they have been shown to cause some bicyclists to have serious psychological barriers to continued bicycling (Sanders, 2015). Near miss data can be used to address underreporting and increase predictive power of modelling (Poulos et al., 2017). Even when data has been collected, accessing data can be difficult for researchers due to limited access and licensing. A typical scenario is that the data are stripped of incident descriptions to protect the privacy of the people involved. Alternatively, the data may be spatially aggregated to intersections, to mid-block, or summarised by larger areas. As a result, safety data used for research and planning are often incomplete and lack details needed to advance planning and knowledge.

Crowdsourcing has become a tool for filling gaps in safety data and key papers are outlined in Table 2. Using web-maps, bicyclists can self-report crashes and near misses, and the data are stored as GIS or other geolocated sources. Examples are BikeMaps.org, a global tool for mapping crashes and near misses (Nelson, Denouden, Jestico, Laberee, & Winters, 2015) and Collideoscope (collideoscope.org.uk), a tool used in the UK. Related tools include www.badintersections.com and social media campaigns such as #NearDeathTO to report near misses in Toronto, Canada. Using a combination of structured and open-ended questions, these websites allow anonymous reports of bicycling incidents and reports include details of crash and near miss outcomes, such as the presence and severity of injuries (Nelson et al., 2015). From our team's experience with BikeMaps.org data, we have observed that crowdsourced data on bicycling crashes highlight locations where there are a lot of bicycles, whereas data that comes from police or insurance tends to highlight areas with high amounts of vehicle traffic. Risks associated with features like multi-use paths, regional trails, and micro barriers (e.g. railroad tracks and loose surfaces) are more likely to emerge in analyses of crowdsourced data (Jestico, Nelson, Potter, & Winters, 2017). Other apps, similar to BikeMaps.org in functionality, have been developed but tend to be more regionally specific, like ORcycle (Blanc & Figliozzi, 2017). New technology that could be of interest for crowdsourcing safety data include apps that use GPS data to track unsafe riding behaviour (Gu et al., 2017).

Crowdsourcing favours data collection from people with access to technology (Ferster, Nelson, Winters, & Laberee, 2017). Therefore, these tools do not capture the experience from some of the most vulnerable road users, such as older adults, children, and people with lower incomes. The amount of data submitted is also tied to promotion of the tool and requires consistent and ongoing promotion. As such, the data are not ideal for tracking change through time, but rather can represent spatial variation in bicycling safety and can be used to understand safety barriers to bicycling. The longevity of crowdsourced tools is also problematic, and many tools come and go due to lack of funding or energy for ongoing promotion and maintenance. While many jurisdictions want access to better safety data, investment in, and promotion of the data collection tool can be difficult to fund. BikeMaps.org is the longest running project that we are aware of and has been operating for more than five years. While the initial tools were developed with industry support, operations have been federally supported and are funded by research projects that use the data. Tools that are centrally supported and long-standing can fill critical gaps in safety data and ideally be used in combination with other existing datasets.

Infrastructure

Cities small and large have increasing capacity to maintain spatial data files of their bicycling infrastructure. With the Open Data movement, many communities are making their bicycling infrastructure data publicly available. However, there are no standard naming conventions for bicycle infrastructure. As an example, an investigation of open datasets in 45 Canadian municipalities revealed ∼100 different terms used for bicycle infrastructure (Winters, Zanotto, & Butler, 2020). There was also great variability in the timeliness of the data, with some open datasets already up to five years old when accessed for the study. These realities mean that compiling bicycle infrastructure data directly from cities is laborious, requiring the assembly of data from disparate sources and reconciling the resolution of differences in naming conventions, timeliness, and accuracy. As such, multi-city studies that relate bicycle infrastructure to mode share and health outcomes are limited and have rarely been repeated even with substantial investments across North America in recent years (Pucher & Buehler, 2012).

OpenStreetMap (OSM – openstreetmap.org) is an emerging data source with the potential to serve as a single source of infrastructure data with global coverage. OSM is a crowdsourced map of the world that provides free spatial data for the natural and built environment, including active transportation infrastructure. With data quality

Table 2. Literature on crowdsourced data for bicycling safety.

Type of crowdsourced data	Reference	Research paper goal	Opportunities of using crowdsourced data	Select challenges that arose with crowdsourced data
Safety – BikeMaps.org	Nelson et al. (2015)	Development and deployment of a tool for crowdsourced bicycling safety data.	New technology enables collection of self-reported bicycling crash and near miss data.	Data collection requires continued promotion of crowdsourced tools.
Safety – Exposure & Risk	Strauss, Miranda-Mireno, and Morency (2015)	Estimate bicycle volumes (AADB) by combining GPS (smartphone app) with manual counts to map injury risk through a network.	Continuous spatial coverage for ridership enabled exposure and risk to be mapped across the entire network to be mapped.	Exposure maps were not bias corrected.
Safety (intersections) – Exposure	Saad, Abdel-Aty, Lee, and Cai (2019)	A demonstration of the issues of how bicycle crash modelling suffers from a lack of exposure data and that exposure data limitations can be overcome using crowdsourced data.	Continuous spatial coverage for ridership enabled exposure and risk to be mapped across the entire network. Study data statistically adjusted to address demographic bias.	Bias corrections benefit from additional SES and behavioural variables.
Safety	Aldred and Goodman (2018)	UK Near Miss Project, a crowdsource project of self-reported near miss bicycling incidents, used to characterise individual predictors that influence scariness of an incident.	Using crowdsourcing it is possible to collect diverse individual narratives.	Crowdsourcing impacts representativeness of data. More experienced bicycling and men aged 30–49.

enforced through community standards, OSM data are contributed by a wide range of people and interests, including but not limited to hobbyists, professional mappers, and hired mappers at companies that use OSM data for delivery services and navigation apps (Anderson, Sarkar, & Palen, 2019). OSM has a rapidly growing user community and completeness is quickly improving: OSM launched in 2004 and by 2019 ~5.7 million contributors have created a database with ~600 million features. Recent investigations have found that globally more than 80% of roads are complete on OSM; upwards of 95% in North American cities (Barrington-Leigh & Millard-Ball, 2017).

Applications of OSM data in research and practice are extensive and diverse. As OSM data are provided for free use, they underlie many other applications; for example, Strava Metro uses OSM data to aggregate activities across the street and trail network. In research applications, OSM provides a single global data source spanning across city, state, and national boundaries. At national scales, OSM may be used to measure bicycling infrastructure in cities across Canada (Ferster, Fisher, Manaugh, Nelson, & Winters, 2019) (Table 3). At global scales, OSM can provide insight into urbanisation around the world (Barrington-Leigh & Millard-Ball, 2017). Researchers are already tapping into OSM as a core input for developing walkability metrics at the national scale in Canada (Hermann et al., 2019). In terms of bicycle infrastructure, a study of six Canadian cities found the length of bicycle infrastructure mapped in OSM was similar to cities' open data sources in some cases (<±2%), in others differed more (± 30%) (Ferster et al., 2019). An investigation into the differences indicated that cities' data often were not up-to-date, and that more informal paths may also be mapped in OSM. Further, the

Table 3. Literature on crowdsourced data for characterising bicycling infrastructure.

Type of crowdsourced data	Reference	Research paper goal	Opportunities of using crowdsourced data	Select challenges that arose with crowdsourced data
Infrastructure – OSM	Ferster et al. (2019)	Compare the representation of bike facilities on OSM with open data provided by six cities in Canada.	OSM provides a single data source for all cities. Agreement with city data was high.	Inconsistent labelling and definitions in both OSM and open data.
Infrastructure – OSM	Hochmair et al. (2019)	Completeness of bicycle trails and lanes in two American cities on OSM and Google Maps.	OSM was more complete for bicycle trails.	To obtain the most complete representation of bicycle facilities OSM was ideally supplemented with other data sources
Infrastructure – open data	Collins and Graham (2019)	Develop a predictive model for cycling collisions using road infrastructure variables.	By using a data network it is possible to show bus lanes, 30 mph speed limits, junction density and multi-lane roads affected the collision counts.	OSM lacked information on parking and loading facilities. Open data also a snapshot at one time point whereas collisions occurred over many years.

OSM tags (open-ended and flexible labels assigned to features by citizen-mappers) were queried to differentiate types of bicycle infrastructure (cycle tracks, on-street bicycle lanes, paths (bicycle-only or multi-use), and local street bikeways). Across cities, paths (bicycle-only or multi-use) and painted bike lanes were the most common types of facility, and importantly, these were also the most well-mapped and easiest to query in OSM. Frequently the OSM data were more up-to-date than cities' data, as new infrastructure is often mapped on OSM before it is released on open data from cities. Further, OSM provides an edit history for every feature, which allows measurement of change in the built environment (Zhang & Pfoser, 2019), though few researchers have used this feature.

In crowdsourced data such as OSM, consistency can be a challenge. Similar to open data provided by cities, citizen-mappers may employ diverse labelling practices, with variation within and across cities. While it is possible to develop queries on the data, this requires cross-checking in Google Street View, and repeated training of queries for them to be valid across cities. It will be important to standardise tagging methods for different infrastructure types, especially for types of bicycling infrastructure that may be less common. Multiple tools are available to extract and process OSM data within standard GIS software. Methods of accessing OSM data include downloading OSM world files, accessing the data using application programming interfaces (APIs), or using third party data products. Downloading world files (e.g. https://planet.osm.org/) provides the most timely and direct route of obtaining data, but requires a relatively large global download and importing the data into traditional GIS applications may be a barrier to some researchers. APIs (e.g. http://overpass-api.de/) provide access to spatial and thematic subsets of the data, but these APIs rely on servers that can be overwhelmed by large or frequent requests, there may be technical barriers for using the APIs with traditional GIS software is plugins. Third parties offer data products that deliver spatial subsets of the data in formats that are easily imported into traditional GIS applications (e.g. http://download.geofabrik.de/); however, third party processing steps influence the data format (e.g. the data structure of tags). Further, this is a dynamic space, with companies and products always shifting. With respect to data extraction and processing, publishing detailed extraction and processing steps (including queries) would increase clarity and repeatability. On a promising note, it is expected that OSM data will become increasingly complete (Barrington-Leigh & Millard-Ball, 2017), given its wide acceptance and use in a wide range of applications (including commercial, research, and more) and new tools and programmes to increase the ease of contribution.

Attitudes

Understanding public attitudes about bicycling is critical for understanding how and when to implement bicycling policies and investments. Even when there is evidence that infrastructure is a good decision for mobility and health reasons, projects cannot be successfully implemented without support from public and elected officials. Often survey data are the predominant source of data on attitudes toward bicycling (Handy, van Wee, & Kroesen, 2014). Population-wide samples may provide the most representative data including those who bicycle and those who do not; however, these are expensive and increasingly challenging as landlines become increasingly obsolete. Convenience samples, recruited through word-of-mouth or a listserv, are ultimately biased towards those most connected to the issues. Field intercept surveys will capture the users (and the most regular of users) of the infrastructure but brings few insights on those who are not current users. Generally, cities and researchers must invest on a project-by-project basis or city-specific efforts, as there are few questions related to attitudes towards national or regional travel surveys in the North American context. Further, such survey efforts happen only every few years. Including survey questions in household travel surveys or other national surveys on perceptions and attitudes is one way to ensure that representative samples can be collected with sufficient frequency to understand changes in attitudes, although endeavours such as the National Highway Traffic Safety Administration Attitudes and Behavior Survey are few and far between (Federal Highway Administration, 2017). Over the past decade, there have been increasing challenges to reach representative samples via traditional telephone phone surveys, with researchers and local governments turning to online panels, either those compiled by market research firms, or panels maintained by the organisations themselves (Dillman, Smyth, & Christian, 2014).

Social media research methods can provide timely data about attitudes towards active transportation programmes (Table 4). Social media functions as an extension of public, urban spaces (Brighenti, 2012), and social media research tools can measure communication in these digital-urban spaces. Common social media research methods include natural language processing (NLP; which uses word frequencies and co-occurrence of words in phrases to quantify meaning), sentiment analysis (where dictionaries are used to assign positive or negative sentiments to words and phrases), topic modelling (where topics are classified based on the occurrence of words), and social network graphs (where connections on social networks are analysed). Social media has been used to evaluate transportation attitudes. For example, Schweitzer (2014) found that the sentiments expressed towards public transportation agencies and patrons on Twitter were more negative than sentiments for other types of public services. However, public transit agencies that interacted with Twitter users (by answering questions and responding to statements rather than disseminating information from the top-down) had more

Table 4. Literature on crowdsourced data for characterising attitudes towards bicycling.

Type of crowdsourced data	Reference	Research paper goal	Opportunities of using the crowdsourced data.	Select challenges that arose with crowdsourced data.
Media coverage in crash reporting	Ralph, Iacobucci, Thigpen, and Goddard (2019)	Examine how local news outlets report car crashes involving pedestrians and bicyclists.	Using social media it was possible to show that local news coverage subtly but consistently blames VRUs for crashes.	Media typically gathers information about crashes from police – not planners, engineers or other experts.
Social media data & urban attitudes to bicycling	Hollander and Shen (2017)	Using a case study evaluates how bicycling facilities and city plans influence people's attitudes.	By comparing a random sample of tweets to geolocated bike-focused tweets, it was possible to relate attitudes to infrastructure.	Twitter data only collected for 5 weeks, and difficult to obtain retrospectively.
Twitter attitudes to new bike infrastructure	Ferster et al. (2020)	Evaluate attitudes on Twitter toward new bicycling facilities in two Canadian cities.	Attitudes on Twitter changed from bike advocacy before the lanes opened, to controversy after the lanes opened, towards broader acceptance over time.	Twitter is the most accessible platform to researchers, but it is only part of the discussions and messaging on social media.

positive sentiments that might impact how voters and stakeholders think about future investments. To understand user experiences for a bike share programme in Washington D.C., Das, Sun, and Dutta (2015) used Twitter data to measure sentiments, which extend beyond usual measures of service (e.g. number of trips) and can be obtained from a large audience more rapidly than traditional surveys. Ferster et al. (2020) evaluated discussion about new protected bike lanes in Edmonton and Victoria, Canada, and found that opposition to the projects was framed around weather in Edmonton and impacts to small business in Victoria. The discussion on Twitter changed over time, where Twitter was first used by bicycling advocates to promote new bike infrastructure, engaged larger audiences in sharing news articles and general discussion once the lanes were opened, and finally indicated greater general acceptance of the new infrastructure.

Social media research methods tools can rapidly measure communication from large audiences on social networks, but face challenges related to data access and sample representativeness. Key considerations for social media research are data access and data represention. In part due to access limitations, most social media research papers in active transportation have used Twitter because the Twitter Search application programming interface (API) and Twitter Streaming API provide data in a format that is accessible to researchers for free. The data provided by these tools are a subset and include biases due to internal filters, such as overrepresenting central accounts (accounts that are widely retweeted or mentioned) compared to less central accounts that represent a much larger mass of people (González-Bailóna, Wang, Riveroc, Borge-Holthoefer, & Morenoc, 2014). Continued access to the data depends on platform policies, changes to filters, and monetisation of data access. Active transportation research using other social media platforms have potential for increasing understanding of these socio-technical systems, yet research access to other platforms can be challenging. Another critical issue is representativeness. Compared to the general population of the United States, Twitter users are younger, more educated, and more often located in urban centres (Duggan, 2015). However, this younger demographic is often missing from traditional feedback mechanisms, such as townhall meetings (Einstein, Palmer, & Glick, 2019), so social media research may help provide voices in planning processes for some underrepresented groups. As well, some neighbourhoods, topics, and bicycling projects are underrepresented, as social media does not reflect any of these aspects of urban life equally. Research is needed to understand the spatial and thematic completeness of social media discussions at a neighbourhood scale. Failure to reflect representative voices, through use of social media or other crowdsourced data, can lead to decision making that is biased towards those who use technology, which are typically already privileged groups.

Discussion

Transportation and planning research and practice need to keep pace with the backdrop of smart cities, big data, and burgeoning technology. Trends in bicycling data are an example of what is happening with city data in general; sources, variety, and volume of data are exploding, offering huge potential to help us better understand how city characteristics vary over space and through time. While emerging sources of bicycling data create opportunities, they also bring challenges. Many of the challenges of using crowdsourced data are common regardless of application (e.g. see Mooney & Pejaver, 2018), but here we frame five broad challenge areas with specific regard to bicycling.

Challenge 1 – access and funding

Emerging bicycling data, like many new datasets on cities, bring challenges of access, ownership, and funding. Companies collecting data as part of smartphone apps have business models that include packaging and selling data. In the case of Strava, data costs are based on the size of the Strava data sample (number of users and recorded rides). The costs of data purchase can seem high but may be efficient as compared to manual count programmes if the information needs benefit from data that are continuous in space and time.

Alternately, projects that use crowdsourcing to improve safety data are open and anyone can access and benefit from data. However, these projects lack a business model, and as a result many have come and gone. While cities clearly want more complete and diverse safety data, it is unclear who will fund the long-term maintenance of crowdsourced data collection tools and cities must invest in promotion for ongoing data collection. In contrast, projects like OSM have taken off globally. The sheer breadth of applications of OSM may guarantee the longevity and ongoing popularity of such projects. One potential solution to funding crowdsourced projects may be for cities to join together and form consortiums around projects that are working well. Rather than build one-off tools for crowdsourced data collection, tools that have the buy in of many cities will be beneficial and can have a shared model.

Challenge 2 – privacy

Most of the technology associated with crowdsourced bicycling data includes smartphones or mapping. When location information is captured, privacy of individuals can be compromised. Businesses selling data, like Strava, protect privacy by providing only aggregated data (Roy et al., 2019). While aggregated data are useful for volume estimates, the need to protect privacy makes it difficult to use data for analyses of individual paths or trajectories. The only way to access individual level data is to use apps that have a research purpose, although such studies may struggle with the number of participants due to recruitment issues. The ethics requirements facing university-based research projects serve to protect privacy (i.e. BikeMaps.org), by ensuring anonymity in data collection (Nelson et al., 2015). In the current era, we have seen a huge shift toward sharing of information on the internet. Twitter and other social media are good examples of the willingness of people to share personal details and even their location in digital public space. While Twitter users can turn off their geolocations, hashtags and place names can still be used to link people to places. We urge researchers and practitioners to use good ethical judgement in how they use and protect data (Breslin, Shareck, & Fuller, 2019). Issues of privacy will grow more complex as smart phone technology adds face and voice recognition. Privacy is a critical area of research that requires legal and ethical research. Research into and guidelines on the characteristics of technology that have appropriate protection for privacy are urgently needed and will be a useful guide to the industry as we determine which technologies to adopt.

Challenge 3 – representativeness and equity

Data collection efforts that opportunistically use technology result in biased depictions of the population. When transportation decisions are made using data biased towards those who use technology the most, vulnerable road users – children, older adults, and people with low incomes, may be excluded (Garber et al., 2019). Using biased data may exacerbate urban equity issues; it is insufficient to invest in infrastructure only where those with access report issues. Research that uses crowdsourced ridership data continue to identify bias as problematic (e.g. Orellano & Guerrero, 2019) and identify tradeoffs between sampling bias and sample size (Pritchard, Bucher, & Frøyen, 2019). However, with effort it can be possible to reduce bias in data using statistical approaches. There is promising research into bias correction, however bias correction is only possible when there are good data for training correction models and results should be used as categorical (low, medium, high ridership) rather than considered precise to the number of bicycles (Garber et al., 2019; Roy et al., 2019). To improve representativeness during data collection, in our experience, also requires someone to go into the community and work one-on-one with people that do not have access to technology or who do not know how to contribute data. Depending on the nature of the phenomena represented by data, it can also be possible to correct the bias from sampling by analytic methods. Bias correction is only possible when "truth" can be sampled – this may be the case of ridership (e.g. counts; Roy et al. (2019)), but not for other data needs such as safety. Finally, when possible, more sources of data are often better and using crowdsourced data in complement with traditional data sets will allow a more comprehensive understanding of the challenges and can be less biased than a single source.

Challenge 4 – analytics and data uncertainty

The focus of our paper is crowdsourced data, but it is a natural next step to consider analytics, especially given that crowdsourced datasets are usually large and can be complex. There are several emerging issues with analysis of crowdsourced bicycling data. First, most of the tools used do not sufficiently manage data uncertainty (Woodcock, Givoni, & Morgan, 2013). Lack of robust sampling design increase the uncertainties in data and the outcome can be analytical results that are more uncertain than they initially appear (Rojas-Rueda, de Nazelle, Tainio, & Nieuwenhuijsen, 2011). Bayesian machine learning methods are one group of approaches that can be used to better address uncertainty. Across fields that use crowdsourced data there is a need to develop tools that handle data uncertainty in more robust ways.

Challenge 5 – open methods

In this rapidly emerging space, researchers should be packaging tools to be usable as widely as possible. Built-in data visualisation tools can often help and more automated tools such as Tableau and Esri dashboards are helping fill this gap. Sharing annotated code (e.g. markdown documents) on Github and other platforms and developing standardised tools (i.e. packages or libraries for open source software) will help facilitate repeatable research and advance the capacity of planners and researchers to use data in future. These open tools could include machine learning and data visualisation methods specific to the various emerging data sources we discuss. While open data and methods are helping advance bicycling research and practice, collaboration between data scientists are needed to ensure that the tools developed are generating meaningful solutions and can be deployed in support of decision making.

Challenge 6 – stakeholder capacity

Crowdsourced data in bicycling is beneficial when it supports better decision making or enables monitoring of policy outcomes, but skilled practitioners are needed to translate data into information. Technology for data collection has outpaced analytics and methodological developments are needed to convert data into information products that can be utilised by practitioners. Most cities that need better bicycling data for planning do not have the in-house data science capacity to use emerging data to their full potential. Capacity is needed, in the form of both tools and training. Even with packaged tools, urban planning programmes need to be training future practitioners with data science skills. As well, partnerships between universities and regions can help fill the gap in knowledge needed to make good decisions from data.

Conclusion

In sum, crowdsourced data are a huge opportunity for advancing bicycle research and practice. There are emerging datasets relevant to bicycling ridership, safety, infrastructure, and attitudes. Tapping into these requires an understanding of their strengths and weaknesses. As well, we need to be proactive in championing solutions to challenges of access and funding, privacy, representativeness and equity, analytics, open methods, and stakeholder capacity. More specifically we recommend the following be considered to advance the effective use of crowdsourced data for supporting research on and for pro-bicycling policy development:

(1) Crowdsourced data collection tools should be developed with many stakeholders to ensure adequate funding and longevity.
(2) Research into and guidelines on the characteristics of technology that have appropriate protection for privacy are urgently needed.
(3) Representativeness of data can be improved through engagement of underrepresented populations and/or modelling. However, we must avoid the temptation to treat crowdsourced data in the same way we would a statistically robust sample. In other words, the data are not precise, but are more suitably represented as categorical data.
(4) Across fields that use crowdsourced data, including bicycling research, there is a need to develop tools that handle data uncertainty in more robust ways.
(5) Open data and methods are required to advance our understanding and use of bicycling data.
(6) Use of crowdsourced data often relies on data scientists, but we must translate the tools to enable use by practitioners.

Of course, data are just one of the components needed to shape the cities of our future to be sustainable, equitable, and healthy places. Supports are needed for tools and training such that data can be packaged in a timely manner into stakeholder-relevant products. Through such efforts, crowdsourcing projects that engage people to share their information, stories, and experiences, can translate into compelling evidence for decision making to support bicycling.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by a Public Health Agency of Canada [grant number #1516-HQ-000064]; Canada Research Chairs program [grant number # 950-230773]; Michael Smith Foundation for Health Research Scholar Award and the Arizona State University Foundation.

References

Aldred, R., & Goodman, A. (2018). Predictors of the frequency and subjective experience of cycling near misses: Findings from the first two years of the UK Near Miss Project. *Accident Analysis & Prevention*, *110*, 161–170.

Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *8*(5), 232.

Barrington-Leigh, C., & Millard-Ball, A. (2017). The world's user-generated road map is more than 80% complete. *PLoS One*, *12*, e0180698.

Blanc, B., & Figliozzi, M. (2017). Safety perceptions, roadway characteristics and cyclists' demographics: A study of crowdsourced smartphone bicycle safety data. *Transportation Research Board 96th Annual Meeting*, 17-03262.

Boss, D., Nelson, T., Winters, M., & Ferster, C. J. (2018). Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. *Journal of Transport & Health*, *9*, 226–233.

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.

Breslin, S., Shareck, M., & Fuller, D. (2019). Research ethics for mobile sensing device use by vulnerable populations. *Social Science & Medicine*, *232*, 50–57.

Brighenti, A. M. (2012). New media and urban motilities: A territoriologic point of view. *Urban Studies*, *49*(2), 399–414.

Broach, J., Dill, J., & McNeil, N. W. (2019). Travel mode imputation using GPS and accelerometry data from a multi-day travel survey. *Journal of Transport Geography*, *78*, 194–204.

Brondeel, R., Pannier, B., & Chaix, B. (2015). Using GPS, GIS, and accelerometer data to predict transportation modes. *Medicine & Science in Sports & Exercise*, *47*(12), 2669–2675.

Brum-Bastos, V., Ferster, C., Nelson, T., & Winters, M. (2019). Where to put bike counters? Stratifying bicycling patterns in the city using crowdsourced data. *Transport Findings*, *November*, 10828.

Collins, D. J., & Graham, D. J. (2019). Use of open data to assess cyclist safety in London. *Transportation Research Record*, *2673*(4), 27–35.

Das, S., Sun, X., & Dutta, A. (2015). Investigating user ridership sentiments for bike sharing programs. *Journal of Transportation Technologies*, *5*(2), 69–75.

DiGioia, J., Watkins, K. E., Xu, Y., Rodgers, M., & Guensler, R. (2017). Safety impacts of bicycle infrastructure: A critical review. *Journal of Safety Research*, *61*, 105–119.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys. The tailored design method*. Hoboken, NJ: John Wiley & Sons.

Duggan, M. (2015). *Mobile messaging and social media – 2015*. Pew Research Center. Retrieved from: https://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/

Einstein, K. L., Palmer, M., & Glick, D. M. (2019). Who participates in local government? Evidence from meeting minutes. *Perspectives on Politics*, *17*(1), 28–46.

Eitzel, M. V., Cappadonna, J. L., Santos-Lang, C., Duerr, R. E., Virapongse, A., West, S. E., & Jiang, Q. (2017). Citizen science terminology matters: Exploring key terms. *Citizen Science Theory and Practice*, *2*, 1–20.

Federal Highway Administration. (2017). *Exposure analysis on specific transportation facilities. Synthesis of Methods for Estimating Pedestrian and Bicyclist Exposure to Risk at Areawide Levels and on Specific Transportation Facilities*. Ch. 3. Technical Report. Retrieved from: https://safety.fhwa.dot.gov/ped_bike/tools_solve/fhwasa17041/ch3.cfm#ss2

Ferster, C. J., Fisher, J., Manaugh, K., Nelson, T., & Winters, M. (2019). Using OpenStreetMap to inventory bicycle infrastructure: A comparison with open data from cities. *International Journal of Sustainable Transportation*, *14*(1), 64–73.

Ferster, C. J., Laberee, K., Nelson, T., Thigpen, C., Simeone, M., & Winters, M. (2020). From advocacy to acceptance: social media discussions of protected bike lane installations. Manuscript submitted for publication.

Ferster, C. J., Nelson, T., Winters, M., & Laberee, K. (2017). Geographic age and gender representation in volunteered cycling safety data: A case study of BikeMaps.org. *Applied Geography*, *88*, 144–150.

Fishman, E., Schepers, P., & Kamphuis, C. B. M. (2015). Dutch cycling: Quantifying the health and related economic benefits. *American Journal of Public Health*, *105*(8), e13–e15.

Garber, M. D., Watkins, K. E., & Kramer, M. R. (2019). Comparing bicyclists who use smartphone apps to record rides with those who do not: Implications for representativeness and selection bias. *Journal of Transport & Health*, *15*, 100661.

González-Bailóna, S., Wang, N. B., Riveroc, A., Borge-Holthoefer, J., & Morenoc, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, *38*, 16–27.

Griffin, G. P., & Jiao, J. (2015). Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport and Health*, *2*(2), 238–247.

Griffin, G. P., Mulhall, M., Simek, C., & Riggs, W. W. (2020). Mitigating bias in Big Data for transportation. *Journal of Big Data Analytics in Transportation*, *2*, 49–59.

Gu, W., Liu, Y., Zhou, Y., Zhou, Z., Spanos, C. J., & Zhang, L. (2017). BikeSafe: bicycle behavior monitoring via smartphones. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers* (pp. 45–48).

Handy, S., van Wee, B., & Kroesen, M. (2014). Cycling for transport. Research needs and challenges. *Transport Reviews*, *34*(1), 4–24.

Hermann, T., Gleckner, W., Wasfi, R. A., Thierry, B., Kestens, Y., & Ross, N. A. (2019). A pan-Canadian measure of active living environments using open data. *Health Reports*, *30*(5), 16–25.

Hirsch, J. A., Stratton-Rayner, J., Winters, M., Stehlin, J., Hosford, K., & Mooney, S. J. (2019). Roadmap for free-floating bikeshare research and practice in North America. *Transport Reviews*, *39*(6), 706–732.

Hochmair, H. H., Bardin, E., & Ahmouda, A. (2019). Estimating bicycle trip volume for Miami-Dade county from Strava tracking data. *Journal of Transport Geography*, *75*, 58–69.

Hollander, J. B., & Shen, Y. (2017). Using social media data to infer urban attitudes about bicycling: An exploratory case study of Washington DC. In A. Karakitsiou, S. Rassia, & P. Pardalos (Eds.), *City networks. Springer optimization and ITS applications* (Vol. 128). Cham: Springer.

Hong, J., McArthur, D. P., & Livingston, M. (2019). The evaluation of large cycling infrastructure investments in Glasgow using crowdsourced cycle data. *Transportation*, doi:10.1007/s11116-019-09988-4

Jestico, B., Nelson, T. A., Potter, J., & Winters, M. (2017). Multiuse trail intersection safety analysis: A crowdsourced data perspective. *Accident Analysis & Prevention*, *103*, 65–71.

Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*, *52*, 90–97.

Kerr, J., Duncan, S., & Schipperijn, J. (2011). Using global positioning systems in health research. A practical approach to data collection and processing. *American Journal of Preventive Medicine*, *41*(5), 532–540.

Mcnair, H., & Arnold, L. (2016). Crowd-sorting: reducing bias in decision making through consensus generated crowdsourced spatial information. In *International Conference on GIScience Short Paper Proceedings* (Vol. 1, No. 1).

Medury, A., Grembek, O., Loukaitou-Sideris, A., & Shafizadeh, K. (2019). Investigating the underreporting of pedestrian and bicycle crashes in and around university campuses− a crowdsourcing approach. *Accident Analysis & Prevention*, *130*, 99–107.

Mooney, S. J., & Pejaver, V. (2018). Big data in public health: Terminology, machine learning, and privacy. *Annual Review of Public Health*, *39*, 95–112.

Nelson, T. A., 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*, 53.

Orellano, D., & Guerrero, M. L. (2019). Exploring the influence of road network structure on the spatial behaviour of cyclists using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*, *46*(7), 1314–1330.

Poulos, R. G., Hatfield, J., Rissel, C., Flack, L. K., Shaw, L., Grzebieta, R., & McIntosh, A. S. (2017). Near miss experiences of transport and recreational cyclists in New South Wales, Australia. Findings from a prospective cohort study. *Accident Analysis and Prevention*, *101*, 143–153.

Pritchard, R., Bucher, D., & Frøyen, Y. (2019). Does new bicycle infrastructure result in new or rerouted bicyclists? A longitudinal GPS study in Oslo. *Journal of Transport Geography*, *77*, 113–125.

Pucher, J., & Buehler, R. (Eds.). (2012). *City cycling*. Cambridge, MA: MIT Press.

Ralph, K., Iacobucci, E., Thigpen, C. G., & Goddard, T. (2019). Editorial patterns in bicyclist and pedestrian crash reporting. *Transportation Research Record*, *2673*(2), 663–671.

Rojas-Rueda, D., de Nazelle, A., Tainio, M., & Nieuwenhuijsen, M. J. (2011). The health risks and benefits of cycling in urban environments compared with car use: Health impact assessment study. *British Medical Journal*, *343*, d4521.

Romanillos, G., Austwick, M. Z., Ettema, D., & De Kruijf, J. (2016). Big data and cycling. *Transport Reviews*, *36*(1), 114–133.

Romanillos, G., Moya-Gómez, B., Zaltz-Austwick, M., & Lamíquiz-Daudén, P. J. (2018). The pulse of the cycling city: Visualising Madrid bike share system GPS routes and cycling flow. *Journal of Maps*, *14*(1), 34–43.
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.
Saad, M., Abdel-Aty, M., Lee, J., & Cai, Q. (2019). Bicycle safety analysis at intersections from crowdsourced data. *Transportation Research Record*, *2673*(4), 1–14.
Sanders, R. L. (2015). Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accident Analysis & Prevention*, *75*, 26–34.
Schweitzer, L. (2014). Planning and social media: A case study of public transit and stigma on twitter. *Journal of the American Planning Association*, *80*(3), 218–238.
Scott, D. M., & Ciuro, C. (2019). What factors influence bike share ridership? An investigation of Hamilton, Ontario's bike share hubs. *Travel Behaviour and Society*, *16*, 50–58.
Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. *Accident Analysis & Prevention*, *83*, 132–142.
Sun, Y., & Mobasheri, A. (2017). Utilizing crowdsourced data for studies of cycling and air pollution exposure: A case study using Strava data. *International Journal of Environmental Research and Public Health*, *14*(3), 274.
Teschke, K., Frendo, T., Shen, H., Harris, M. A., Reynolds, C. C., Cripton, P. A., & Winters, M. (2014). Bicycling crash circumstances vary by route type: A cross-sectional analysis. *BMC Public Health*, *14*(1), 1205.
Vanparijs, J., Panis, L. I., Meeusen, R., & de Geus, B. (2015). Exposure measurement in bicycle safety analysis: A review of the literature. *Accident Analysis & Prevention*, *84*, 9–19.
Wergin, J., & Buehler, R. (2017). Where do Bikeshare bikes actually go? Analysis of Capital Bikeshare trips with GPS data. *Transportation Research Record*, *2662*(1), 12–21.
Winters, M., & Branion-Calles, M. (2017). Cycling safety: Quantifying the under reporting of cycling incidents in Vancouver, British Columbia. *Journal of Transport and Health*, *7*, 48–53.
Winters, M., Hosford, K., & Javahera, S. (2019). Who are the 'super-users' of public bike share? An analysis of public bike share members in Vancouver, BC. *Preventive Medicine Reports*, *15*, 100946.
Winters, M., Zanotto, M., & Butler, G. (2020). The Canadian Bikeway comfort and safety (Can-BICS) classification system: A proposal for developing common naming conventions for cycling infrastructure. *Health Promotion and Chronic Disease Prevention in Canada: Research, Policy and Practice (the HPCDP Journal)*, *40*(9).
Woodcock, J., Givoni, M., & Morgan, A. S. (2013). Health impact modelling of active travel visions for England and Wales using an Integrated Transport and Health Impact Modelling Tool (ITHIM). *PloS One*, *8*(1), e51462.
Xu, J., & Nyerges, T. L. (2017). A framework for user-generated geographic content acquisition in an age of crowdsourcing. *Cartography and Geographic Information Science*, *44*(2), 98–112.
Zhang, L., & Pfoser, D. (2019). Using openstreetmap point-of-interest data to model urban change – A feasibility study. *PLoS One*, doi:10.1371/journal.pone.0212606