Transportation Research Interdisciplinary Perspectives 14 (2022) 100600

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

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

Communicating perceptions of pedestrian comfort and safety: Structural topic modeling of open response survey comments

Emily Bardutz [a,1], Alexander Bigazzi [b,2,*]

[a] School of Community and Regional Planning, University of British Columbia, Canada
[b] Department of Civil Engineering and School of Community and Regional Planning, University of British Columbia, Canada

ARTICLE INFO

Keywords:
Comfort
Safety
Pedestrian
Topic modeling
Perceptions
Open comments

ABSTRACT

Purely quantitative or qualitative studies of perceived comfort and safety have different strengths and limitations related to scope, reliability, and generalizability. To pursue new insights about perceptions of pedestrian comfort and safety, this study uses the novel approach of Structural Topic Modeling to organize open-response comments into prominent topics, or themes, and explore the factors affecting the prevalence and content of topics. The study data come from a survey in which participants rated short video clips of pedestrians interacting with motor vehicles or bicycles at unsignalized crosswalks, and then provided optional open-response comments about each video. Modeling results identified six topics in the text data: (1) driver behaviour and pedestrian risk, (2) pedestrian risk assessment and mitigation, (3) pedestrians not waiting before crossing, (4) cyclists leaving the bicycle lane, (5) cyclists' yielding necessity, and (6) and cyclists' speed modifications. The prevalence of the topics was significantly affected by attributes of the interaction (vehicle type—bicycle or car—and its proximity), and attributes of the perceiver (travel habits and self-reported risk aversion). Topic prevalence was unrelated to participants' comfort and safety ratings for the same interactions, indicating that topic modeling provides a fundamentally different source of information than quantitative severity ratings. The findings help to illuminate why pedestrian-cyclist interactions may attract disproportionate attention from the travelling public, despite their relatively lower risk than interactions with motor vehicle drivers.

Introduction

Increasingly, governments are encouraging their residents to choose active modes of transportation for the expected public health, environmental, and economic benefits. Improving perceptions of how comfortable and safe it is to walk and cycle is imperative to increasing uptake of active transportation modes. Factors affecting perceptions of comfort and safety that have previously been investigated for pedestrians and cyclists include environmental factors such as proximity of motor vehicles, street lighting, and facility design, and personal factors such as travel habits, socio-demographics, and experience (Bigazzi et al., 2021; Fitch & Handy, 2018; Gkekas et al., 2020; Habib et al., 2014; Kaparias et al., 2015; Landis et al., 1997; Li et al., 2012; Monsere et al., 2020; Peña-García et al., 2015). Analyses of walking determinants have historically been biased towards objective measures of the built environment, and further investigation of subjective measures is needed (Bozovic et al., 2020).

The existing research on travel perceptions and behaviour, to date, has primarily been conducted using quantitative methods (Handy et al., 2014; Næss, 2020; Poulenez-Donovan & Ulberg, 1994). Perceptions of comfort and safety are most often measured using Likert-scale ratings after study participants have either experienced a situation first-hand (Fitch & Handy, 2018; Landis et al., 1997; Li et al., 2012; Peña-García et al., 2015) or watched videos of other road users in some situation (Bigazzi et al., 2021; Fitch & Handy, 2018; Monsere et al., 2020). These closed-response, Likert-style questions restrict the dimensions of inquiry, which facilitates rigorous statistical analysis but can over-simplify respondents' concerns and relies on the researchers' *a priori* conceptualization of the issue (Poulenez-Donovan & Ulberg, 1994).

Other studies have used qualitative techniques such as in-depth interviews or focus groups to evaluate pedestrian and cyclist perceptions (Anciaes & Jones, 2018; Daley & Rissel, 2011; Fishman et al., 2012;

* Corresponding author.
E-mail addresses: e.bardutz@ubc.ca (E. Bardutz), abigazzi@civil.ubc.ca (A. Bigazzi).
[1] 0000-0003-1844-4051.
[2] 0000-0003-2253-2991.

https://doi.org/10.1016/j.trip.2022.100600
Received 22 November 2021; Received in revised form 1 April 2022; Accepted 11 April 2022
Available online 18 April 2022

Hine, 1996; Mitman & Ragland, 2007; Osborne et al., 2020; Pollack et al., 2014). Qualitative methods can provide less constrained measurements of traveller perspectives, which may be more appropriate for the investigation of complex phenomena like travel behaviour, but with a different set of challenges and limitations related to costs, sample size and representativeness, demonstrating rigour and reproducibility, and potential bias introduced by the interviewer or group (Flick, 2014; Mars et al., 2016; Næss, 2020; Poulenez-Donovan & Ulberg, 1994).

Comparisons of the results derived from closed-response questions versus qualitative techniques have found that qualitative data can complicate otherwise seemingly straightforward quantitative results (Hine, 1996; Pooley et al., 2011). For example, Pooley et al. (2011) found that most survey participants strongly rejected the notion that walking and cycling are high-risk activities, but the same participants expressed more mixed opinions during in-depth interviews. Due to their contrasting strengths, combining qualitative and quantitative methods can potentially exploit the advantages of each approach (Anciaes & Jones, 2018; Mars et al., 2016; Shay et al., 2016).

Open comment data from surveys inhabit a 'grey area' overlapping quantitative and qualitative methods (O'Cathain & Thomas, 2004). Open-ended questions in surveys provide a means to efficiently collect unconstrained text data on participants' thoughts, reactions, or opinions, avoiding some of the limitations of closed-response questions (Schuman & Presser, 1979). The open text data can then be analyzed using quantitative or qualitative techniques. Although most surveys have at least one open-response question, the text data are rarely analyzed by researchers, as resources are not often allocated to their analysis (O'Cathain & Thomas, 2004).

The objectives of this paper are to investigate the unconstrained expressions of pedestrian comfort and safety reported in open-response survey comments, and then compare those to the closed-response ratings obtained in the same survey. In addition to seeking new insights about perceptions of pedestrian comfort and safety beyond the constraints of closed-form questions, we aim to demonstrate the value of analysing open comment data using structural topic modeling (STM) in transportation research. STM is an efficient method for organizing large text datasets into interpretable topics or themes, and allows for analysis of variables that affect how often topics are discussed and the words that participants use to discuss those topics (Roberts et al., 2014). STM has previously been used in other social science disciplines, but within the domain of transportation has only been applied by Lee & Kolodge (2020) and Bennett et al. (2019) to study perceptions of autonomous vehicles.

Structural topic modelling

STM is a machine-assisted textual analysis method that simultaneously predicts topics, or themes, within a text dataset and assigns words from the dataset's vocabulary to those predicted topics (Roberts et al., 2014). Machine-assisted methods, such as STM, have two primary benefits over manual text analyses: 1) they allow for the analysis of text datasets that would be otherwise too large to manually analyze, and 2) they eliminate the need for the researcher to decide, *a priori*, on a coding scheme, allowing for the discovery of new concepts or issues (DiMaggio et al., 2013; Grimmer & Stewart, 2013; Lesnikowski et al., 2019).

In topic modeling, words constitute documents, and documents are weighted combinations of discrete topics, each of which is simply a probability distribution over all words (Landauer et al., 2013). For example, the word "speed" has a certain probability of being used to express each topic, and so the prevalence of each topic in a document determines the likelihood of "speed" appearing in that document. Estimating a topic model provides statistical information about the prevalence of themes, and about what words are used to express each theme.

Unlike other topic models, STM does not assume that topic prevalence (how often a topic is discussed) or topic content (the prevalence of words for each topic) is constant across documents. Thus, the advantage of STM is the ability to include topic prevalence and topic content covariates, facilitating investigation of the factors that influence both topics and the words used to express those topics (Roberts et al., 2014).

STM has previously been used in various applications, including evaluating attitudes towards de-policing (Mourtgos & Adams, 2019), climate change (Tvinnereim et al., 2017; Tvinnereim & Fløttum, 2015), and urban development projects (Anzoise et al., 2020). Within the transportation field, Lee & Kolodge (2020) and Bennett et al. (2019) both used STM to investigate attitudes towards self-driving vehicles. The method remains underutilized in transportation research, possibly due to a lack of awareness of textual analysis methods (Lesnikowski et al., 2019).

Data

Data for this study are from an online survey used to investigate varying perceptions of pedestrian comfort and safety, described in Bigazzi et al. (2021). The survey used 84 short (7–18 s) video clips of pedestrian interactions with motor vehicles and bicycles at unsignalized crosswalks in Vancouver, British Columbia, Canada. A maximum Post-Encroachment Time (PET – the time gap between when the first road user exits a conflict point and the second road user enters it) of 4 s was used to define interactions.

Three groups of participants are included in the study data: the local travelling public, members of a city advisory committee for a street redesign project, and North American transportation experts from outside British Columbia. These three groups were identified and recruited for the purposes of the original study (contrasting their quantitative ratings, as reported in Bigazzi et al. (2021)) rather than for the purposes of the analysis presented in this paper. The public group was recruited using online advertisements. Members of the advisory group and the transportation experts were recruited by direct email contact. Study methods were reviewed and approved by the University of British Columbia Behavioural Research Ethics Board.

Members of the public and the advisory groups were shown 15 randomly selected video clips, stratified on PET, manually coded from the videos in 1-second bins (<1, 1–2, 2–3, or 3–4 s), and interacting vehicle type (motor vehicle or bicycle). The transportation experts were shown all 84 video clips. For each video, participants were asked to provide their agreement rating to the following statements on a 4-point scale from Strongly Disagree to Strongly Agree, with the alternative to select 'I don't know' (coded as NA):

1. The [driver/cyclist] yielded to the pedestrian.
2. The [driver/cyclist] *should* have yielded to the pedestrian.
3. The pedestrian felt comfortable in this crossing.
4. The risk of injury for this pedestrian in this crossing was low.

A 4-point scale for the rating questions, with no neutral option, was selected after extensive pilot testing with academics and the public. The resolution of the scale was low to reduce response burden and time, and a neutral option was omitted to allow binary classification of each rating as positive or negative.

Participants were also given the option to provide an open comment for each video, prompted with "Please provide comments if you wish to clarify your rating or describe any confusion/difficulty you had with rating this video." Socio-demographic information (age, gender, household income, educational attainment), travel habits, and self-reported risk aversion were also collected during the survey. Travel habits were measured by asking how frequently the participant walked, cycled, used a motor vehicle (as a driver or passenger), or used public transportation with response options of "Never," "Monthly or less," "Several times a month," "Several times a week," and "Almost daily." Risk aversion was measured by asking the participant "Overall, how would you place yourself on the following scale?" with seven response options ranging from "Extremely comfortably taking risks" to "Extremely uncomfortable taking risks" based on Glanz et al. (2016).

The cleaned dataset includes responses from 366 participants (343 public, 17 advisory group, and 6 transportation experts). Compared to the city-wide population, the public group is younger (70% versus 50% under 40 years old), has higher educational attainment (70% versus 40% with a bachelor's degree or higher), and includes more women (60% versus 50%) (Bigazzi et al., 2021). Of the 366 total participants, 215 left at least one comment: 197 public (57%), 12 advisory group (71%), and all 6 transportation experts. On average, members of the public left comments on 36% of the videos they were shown, members of the advisory group left comments on 29%, and the transportation experts left comments on 20%. In total, 1219 comments were received, ranging from 4 to 34 comments per video.

Table 1 gives the distribution of self-reported travel model use by commenting participants. The sample is comprised of few participants (<5%) who 'Never' use transit, an automobile (as a passenger of driver), or walk – but 27% of the sample 'Never' cycle. Most (80%) of the sample regularly walks (several times a week or more), whereas 39%, 43%, and 46% regularly cycle, use an automobile, and use transit, respectively.

Methodology

Perceptions of the comfort and safety for each observed pedestrian interaction are hypothesized to be directly impacted by interaction attributes such as PET, as well as participant/perceiver attributes such as socio-demographics (Fig. 1). Perceptions of comfort and safety are then expressed in the survey through the closed-response ratings and open comments. Open comments are a combination of topics (topic prevalence), which are expressed using words (topic content). We hypothesize that topic content is determined by participant attributes, while topic prevalence is determined by perceptions of comfort and safety (and so also influenced by participant attributes).

Fig. 2 illustrates the STM method applied to this study's conceptual framework. In applying STM to the open comment data, each open comment (response in the open text box for one video clip by one participant) is a document. Each document d has a vector of p topic prevalence covariates $X_{d,p}$ that are attributes of the interaction and perceiver. The prevalence of topic k within document d, $\theta_{d,k}$, is conditional on the product of the document covariates $(X_{d,p})$ and the estimated matrix of coefficients for each covariate and topic $\gamma_{p,k}$. Each document also has its own distribution of words used to represent each topic, $\beta_{d,k,w}$, which is conditional on a topic content covariate Y_d and estimated matrix of coefficients for each topic and word, $\delta_{k,w}$, in addition to the baseline distribution of words for each topic, $\mu_{k,w}$. The words w_1 through w_V make up the corpus vocabulary of length V, which is every word that appears at least once in any comment. To generate each of the N words in a document, a topic is drawn from the multinomial distribution of $\theta_{d,k}$, and then a word is drawn to express that topic from the multinomial distribution of $\beta_{d,k,w}$ (Landauer et al., 2013; Roberts et al., 2014; Roberts et al., 2019).

Text processing

All analysis was conducted using the statistical software R, with the *stm* package for topic modelling (R Core Team, 2016; Roberts et al., 2019). The text data require processing before topics can be modeled. By manually reading the comments, spelling errors were corrected, and irrelevant comments were removed (comments not related to events in the video, such as comments about video loading errors). All punctuation, numbers, and stop words (words that primarily serve a grammatical function, such as "the," "which", or "and") were removed. Letters were reduced to lower case, and words were stemmed. Stemming is a process of reducing words to their base form (for example, 'stopping' and 'stopped' are both reduced to 'stop'). The final step was the removal of words in the vocabulary that appeared very frequently ($\geq$ 99% of documents) or very infrequently ($\leq$ 1% of documents), as these words provide little information for topic inference (Grimmer & Stewart, 2013; Roberts et al., 2014).

Table 1
Self-reported Frequency of Travel Mode Use by Commenting Participants.

Travel mode	Never	Monthly or less	Several times a month	Several times a week	Almost daily
Walk	2%	6%	13%	18%	62%
Bike	27%	25%	9%	16%	23%
Transit	1%	26%	27%	23%	23%
Auto (passenger or driver)	5%	21%	31%	27%	16%

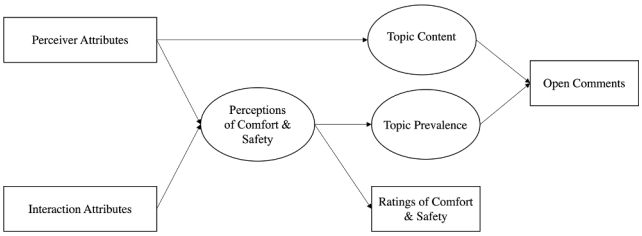

Fig. 1. Conceptual framework.

Model specification

Number of topics

STM requires specification of how many topics are in the dataset, based on judgement and analysis (Roberts et al., 2014). The goal is to maximize topic interpretability by making a trade-off between semantic coherence within topics and word exclusivity between topics. Topics are coherent when they are comprised of similar words, and they are exclusive when a topic's most frequent words rarely appear in other topics. To make this trade-off, models are generated with varying numbers of topics and their coherence and exclusivity values are plotted. The residual variance and held-out likelihood can also be plotted as additional criteria. Once several candidate models have been selected from the generated models, the candidate models are manually inspected to ensure interpretability (Grimmer & Stewart, 2013; Roberts et al., 2014).

Selection of the number of topics was undertaken by first specifying models with topic numbers ranging from 2 to 10, with PET (as an integer) as the only topic prevalence covariate. Candidate models were then identified as those that exhibited high exclusivity, semantic coherence, and held-out likelihood, and low residual variance. Based on these metrics, the 6 and 7-topic models were considered candidate models.

Covariates

For typical quantitative modeling, all relevant variables that have a causal relationship with the dependent variable should be included in the statistical model. However, quantitative text analysis is not always improved by including every relevant variable (DiMaggio et al., 2013). The goal of topic modeling is to maximize interpretability, not only the statistical fit. This is because topic granularity may become so fine that topics are no longer interpretable when only statistical fit is maximized (Chang et al., 2009).

With this in mind, the interaction attributes that were tested as topic prevalence covariates were PET and interacting vehicle type (bicycle or motor vehicle). These were the interaction stratification variables used in the survey design, and their importance for perceptions was supported by the analysis of closed-form ratings in the dataset (Bigazzi et al., 2021). PET was used in the model as an integer variable because

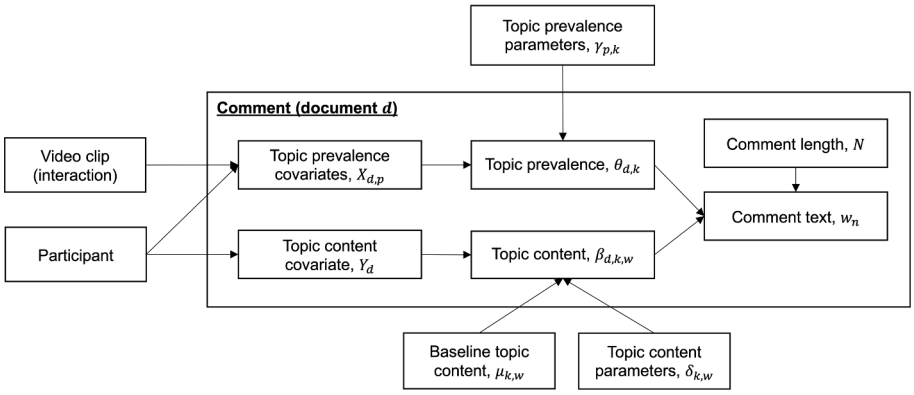

Fig. 2. Structural topic modelling in the study framework.

the *stm* package does not allow for ordered factors.

The participant attributes that were tested as topic prevalence covariates were age, gender, risk aversion, and travel habits. Age in years was included as an integer and gender was a binary variable for 'female' (as only one participant did not select either 'male' or 'female', exclusive). Risk aversion was an integer ranging from 1 = "Extremely comfortable taking risks" to 7 = "Extremely uncomfortable taking risks." Highest educational attainment and household income were not tested as prevalence or content covariates because of many missing values (9% and 19% of participants, respectively).

Travel behaviour variables were tested in several formats. Frequency of walking, cycling, and motor vehicle and transit use was recorded in the survey as 1 = "Never," 2 = "Monthly or less," 3 = "Several times a month," 4 = "Several time a week," and 5 = "Almost daily." Mode frequency for walking, cycling, transit use, and driving was first tested as a binary variable indicating whether or not participants used the mode at least 'several times a month.' Second, a multi-modal variable was tested as an integer variable indicating the number of modes the participant used at least 'monthly or less.' Third, an active transport binary variable was tested, indicating whether or not the participant walked or cycled at least 'monthly or less.' Finally, mode frequency was tested as an integer variable converted from the ordered factor.

The variables were tested as topic prevalence covariates using the 6- and 7-topic models and retained if they were significant for any of the topics at $p < 0.05$. After finalizing the variables to use as prevalence covariates, age and gender were each tested as topic content covariates. These two variables were tested one at a time because the *stm* package only allows for one content covariate to be included per model (Roberts et al., 2019).

Finally, the 6- and 7-topic models were manually inspected to determine which model was most interpretable. Reading and interpreting the results of each model involves reading the word profiles and the top 3 to 5 comments for each topic (those with the highest predicted prevalence based on estimated document-topic proportions). Word profiles are created using four measures: Highest Prob, FREX, Lift, and Score (Roberts et al., 2019). Highest prob ranks words by their probability of being used to represent each topic. FREX weights words according to their exclusivity in a given topic and their overall frequency in the corpus. Lift divides each word's frequency for a given topic by the word's frequency in other topics. Score divides the log frequency of a word in a particular topic by the log frequency of the word in the other topics.

Correlation between topics and ratings of comfort & safety

After finalizing and interpreting the topic model, the relationships between topics and closed-response survey ratings were investigated by examining pairwise correlations between document-level topic proportions and the commenter's associated rating of comfort and safety. Ratings of comfort and safety were tested as both continuous variables (strongly disagree to strongly agree) and binary variables (=1 if agree or strongly agree). A correlation coefficient greater than or equal to |0.3| would indicate that participants were more or less likely to discuss certain topics when they perceived the interaction as (un)safe or (un) comfortable (Cohen, 1988).

Results

After text processing, 1062 comments remained (87%) from 192 participants (176 public, 10 advisory group, and 6 transportation experts). Due to the imbalanced sample sizes and small numbers of commenting participants in the expert and advisory groups, we were unable to segment the statistical model by respondent group, and report the following results for the pooled sample. The mean comment response length was 28 words, and the median comment length was 23 words. The longest comment was 174 words. The vocabulary retained 227 unique words from the original 1453 words. The top 3 most frequently used words were "pedestrian" (795 occurrences), "cyclist" (484), and "cross" (329) (Fig. 3). Despite participants providing a nearly equal number of ratings for videos showing interactions with bicycles as compared to motor vehicles, bicycle interaction videos generated 57% of the 1062 comments (606), while motor vehicle interaction videos generated 43% (456). Median and mean comment lengths were the same for both types of videos. Comment frequency varied more by interaction proximity, with 58% and 42% of comments coming from videos with PET of < 2 sec vs. 2–4 sec, respectively, despite a nearly equivalent number of ratings. In other words, participants were about 40% more likely to comment on an interaction with PET < 2 sec than on an interaction with PET of 2–4 sec.

The 6-topic model was ultimately selected for its parsimony and improved topic interpretability over the 7-topic model. The resulting odds ratios are reported in Table 2, and Fig. 4 shows each topic's expected proportion in the entire corpus (including labels which are described below). The final model includes seven topic prevalence covariates (PET, interacting vehicle type, risk aversion, and frequency of walking, cycling, and motor vehicle and transit use) and no topic content covariates. Neither age nor gender significantly impacted the prevalence or content of any of the topics. That is, participants across all ages and genders discussed the 6 topics at similar rates using similar words. The expected topic proportions (Fig. 4) were compared between participant groups (public, advisory, and experts), and no meaningful differences were found (all topic proportions were within 5%).

The following sections describe each topic in detail, including the prevalence covariate parameters, exemplar words and comments, and interpretation. The type of interacting vehicle (bicycle vs. motor vehicle) is the topic prevalence covariate with the largest magnitude effect for all six topics (Table 2). Therefore, the discussion below splits the topics into two groups: three topics (1, 2, and 3) significantly more

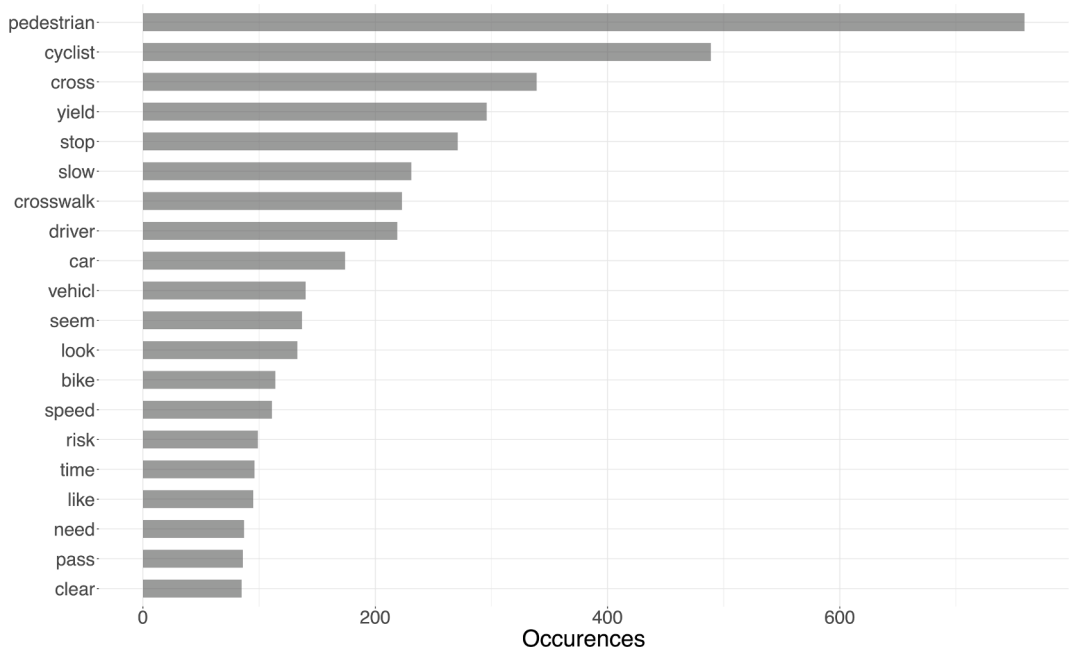

Fig. 3. Word frequencies for the twenty most frequently used words in the vocabulary.

Table 2
Estimated Odds Ratios (* indicates $p < 0.05$).

Variable	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
Intercept	1.310*	1.320*	1.285*	1.266*	1.024	0.943*
PET	0.987*	1.017*	0.988*	0.977*	1.020*	1.012*
Type of vehicle (=bicycle)	0.733*	0.958*	0.906*	1.093*	1.024*	1.404*
Risk aversion	1.009*	0.993*	0.998	0.986*	1.004	1.012*
Walk frequency	1.004	0.997	0.982*	1.008	1.011*	0.999
Bicycle use frequency	1.008*	0.994	1.005	0.994*	0.997	1.001
Motor vehicle use frequency	1.005	0.989*	1.014*	0.997	0.995	1.000
Transit use frequency	1.006*	0.992	1.008	0.982*	1.005	1.008*

likely to be discussed when crossing pedestrians interacted with motor vehicles (i.e., odds ratios < 1.0 for 'Type of vehicle = bicycle' in Table 2), and three topics (4, 5, and 6) significantly more likely to be discussed when crossing pedestrians interacted with bicycles (i.e., odds ratios > 1.0 for 'Type of vehicle (=bicycle)' in Table 2).

Topics associated with motor vehicle interactions

Topic 1: Driver behaviour and pedestrian risk

Topic 1 is about driver behaviour and its effect on pedestrian risk (Table 3). The most consistent exemplar words for this topic (appearing in at least 3 of the 4 word profile measures) are driver, vehicl, and cross. The exemplar comments all link the driver or vehicle actions with the pedestrian's risk of injury (in positive or negative ways). The comments position the driver as the risk-controlling agent in the interactions, primarily through speed choices (e.g., "no effort of slowing down was displayed", "driver can easily slow down and yield, but chose not to",

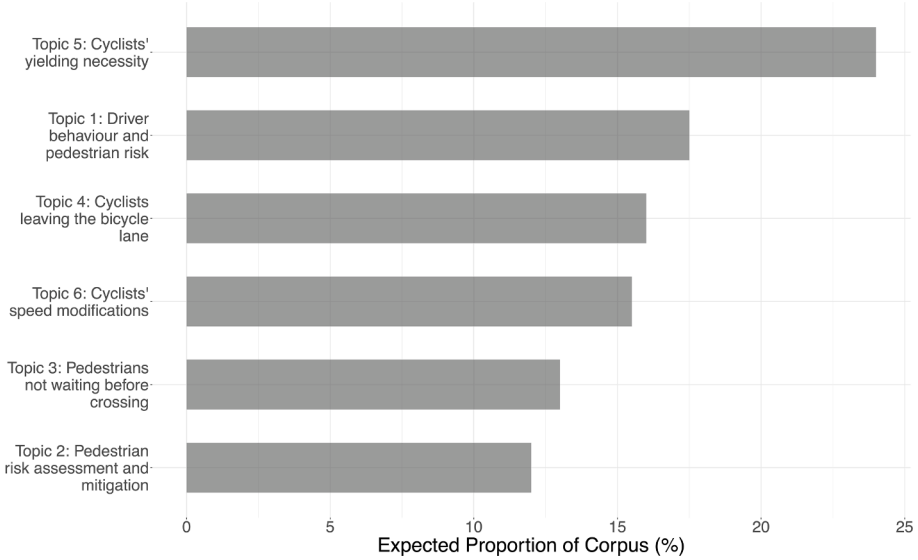

Fig. 4. Topic prevalence across all comments.

Table 3
Topic 1 Exemplar Words and Comments.

Exemplar Words	Exemplar Comments (as written)
Highest prob: pedestrian, driver, cross, vehicl, intersect **FREX:** driver, vehicl, intersect, cross, approach **Lift:** light, driver, vehicl, roll, phone **Score**: driver, light, vehicl, cross, pedestrian	"The first approaching red vehicle did not yield, as no effort of slowing down was displayed. Given the speed of the vehicle and the amount of reaction time from when the pedestrian showed clear intention to cross, the driver can easily slow down and yield, but chose not to. The pedestrian had to slow down and had to be more aware of the traffic. The risk of injury may be higher in this crossing, as there is more traffic in this area." "The driver sped up to attempt to cross before the pedestrian, showing no signs of yielding even when given enough reaction time and adequate braking distance. Vehicles should always yield to pedestrians at crossings when reasonable. The pedestrian was not comfortable at this crossing. Having the urge to pull out his phone to take a picture of the license plate indicates that he has seen this behaviour before and is unhappy with the drivers, perhaps at this crossing. The risk of injury seems to be a lot higher here as vehicles tend not to yield." "The driver crossed through the intersection before the pedestrian arrived at it, so in this case I would agree that the risk of injury was low"

and "driver sped up to attempt to cross before the pedestrian"). The prevalence of this topic significantly decreased with PET (see odds ratios in Table 2), indicating it was more likely to be discussed in closer interactions. Prevalence significantly increased with participant risk aversion, cycling frequency, and transit use frequency, indicating greater likelihood of discussion by risk-averse individuals and non-drivers.

Topic 2: Pedestrian risk assessment and mitigation

In contrast to Topic 1, Topic 2 is about pedestrian risk assessment and mitigation (Table 4). The most consistent exemplar words for this topic are look and law. The exemplar comments ascribe agency to the pedestrian in assessing their personal risk in a crossing situation and behaving accordingly to mitigate their risk to a level they are comfortable with. The topic includes references to pedestrians "looking both ways" as a manifestation of this agency, and an argument that pedestrian capacity for risk assessment is more important for safety than legal prescriptions ("laws"). The comments provide evidence of the importance of pedestrian gaze and head movement in participants' evaluations of comfort and safety (e.g., "they were checking both directions much more, so I rated their risk as lower").

The prevalence of Topic 2 was significantly higher for interactions with motor vehicles than bicycles, although to a lesser extent than for Topic 1. Unlike Topic 1, the prevalence of this topic significantly increased with PET, indicating greater focus on driver agency for closer interactions (Topic 1 with lower PET) and greater focus on pedestrian agency in less severe interactions (Topic 2 with higher PET). Also in contrast to Topic 1, the prevalence of this topic was significantly less for risk averse participants, indicating that with greater risk aversion the agency focus tends to shift from the pedestrian to the driver. Finally, prevalence for Topic 2 decreased significantly with more frequent motor vehicle use. Overall, discussions of driver behaviour and pedestrian risk (Topic 1) were more prevalent than discussions of pedestrian risk assessment and mitigation (Topic 2), as shown in Fig. 4.

Topic 3: Pedestrians not waiting before crossing

Topic 3 is about pedestrians not waiting before crossing (Table 5). The most consistent exemplar words for this topic are stop, crosswalk, and wait. The exemplar comments describe or imply an obligation or need for the pedestrian to wait or stop before entering the crosswalk. The motivations for this topic seem to vary; some commenters state or imply that the pedestrian should yield to the vehicle because the driver could not safely stop, while others suggest that waiting is the prudent action for the pedestrian to take because the driver chose not to stop or had not sufficiently communicated their intention to stop. The prevalence of this topic was significantly higher for closer interactions (lower PET), in addition to being primarily associated with motor vehicles (not bicycle) interactions. This topic was more likely to be discussed by frequent users of motor vehicles and less likely to be discussed by frequent walkers, indicating a logical bias associated with individuals' travel habits.

Table 4
Topic 2 Exemplar Words and Comments.

Exemplar Words	Exemplar Comments (as written)
Highest prob: look, like, right, comfort, make **FREX:** make, law, turn, look, made **Lift:** last, concern, contact, eye, law **Score**: last, look, law, eye, contact	"I can't get a read on this or last video in terms of whether pedestrian looks comfortable. In first video, they were checking both directions much more, so I rated their risk as lower, as there was more evidence of them mitigating it. Less clear in this one whether they're looking both ways." "again, one car, why make him stop, the laws need to be changed pedestrians should only have the right of way when there is no other solution, only when traffic is high ONE car should not be stopped ever! I know you don't believe it, but people are capable of risk assessment so long as you don't dumb them down with stupid laws that gives them the rightaway infront of a two ton car there in lies your problem charge the idiots who dreamed up the law giving pedestrians the right away without question don't give pedestrians right of ways but make it law that a driver has to stop if the pedestrian can't cross safely without doing so in high traffic situations." "As with the last video, it is difficult for me to interpret how safe the pedestrians FELT. In both videos they appeared to feel safe with how the situation played out, but I would not feel entirely safe in their situation (always assume the worst and that drivers or cyclists won't see me trying to cross)."

Table 5
Topic 3 Exemplar Words and Comments.

Exemplar Words	Exemplar Comments (as written)
Highest prob: stop, crosswalk, car, wait, time **FREX:** stop, wait, enter, crosswalk, complet **Lift:** engag, full, enter, stop, complet **Score:** engag, stop, car, crosswalk, wait	"The pedestrian should have waited for the car to stop before entering the roadway. He seemed to assume that the car would stop since it was slowing." "Since there was a median in the crosswalk, the blue car could have rightly assumed that the pedestrian would wait there until he passed. The pedestrian (as usual) kept on walking regardless of the blue car and just expected the blue car to stop for her." "The lady was not at the curb. There was no time for the car to stop in time." "Same thing – the pedestrian just took for granted that the car would stop. Myself, I would have waited until the car passed." "It seemed the car didn't want to stop, assuming the pedestrian would have stopped if the car kept going forward."

Topics associated with bicycles

Topic 4: Cyclists leaving the bicycle lane

Topic 4 is about cyclists leaving the bicycle lane (Table 6). The most consistent exemplar words for this topic are bike and lane. All the exemplar comments refer to bicycle lanes, and most focus on the cyclist leaving the bicycle lane to maneuver around the crossing pedestrian. Some comment on this as problematic, while others describe it as an appropriate way to safely negotiate the interaction. The comments connect this maneuver to both yielding and risk. In addition to path deviations, the comments provide evidence of the importance of bicycle traveling speed (e.g., "cyclist was going way too fast"), and proximity (e.g., "cut unnecessarily close to the pedestrians") in participants' evaluations of comfort and safety.

The prevalence of this topic was significantly higher for closer interactions (lower PET), which is likely inherent to the maneuver. The prevalence of this topic was also significantly higher among risk-tolerant participants and participants who infrequently bicycle, use a motor vehicle, or use transit. The association of this topic's prevalence with travel habits of participants may reflect how familiar they are with this cycling maneuver; unfamiliarity may have led to a higher likelihood of commenting on this behaviour, or any given behaviour.

Topic 5: Cyclists' yielding necessity

Topic 5 is about cyclists' yielding necessity (Table 7). The most consistent exemplar words for this topic are yield, need, enough, and far. The exemplar comments explain why a yield was not or would not be needed. The participants focus primarily on cyclist distance from the pedestrian, and secondarily on cyclist speed, as determining yielding necessity. They also focus on slowing as the primary indicator of cyclist yielding (vs. stopping, path deviation, or crossing order, for example). The comments provide evidence of the importance of cyclist proximity, speed, and speed deviation in participants' evaluations of yielding, but do not connect that explicitly to pedestrian comfort and safety. The prevalence of this topic is significantly higher for less proximate interactions (higher PET), in addition to being more prevalent among more frequent walkers. This topic completes a cycle pattern related to perceiver travel habits: participants who use motor vehicles more frequently focus more on pedestrian yielding (Topic 3), participants who more frequently cycle focus more on driver yielding (Topic 1), and participants who more frequently walk focus more on cyclist yielding (Topic 5).

Topic 6: Cyclists' speed modifications

Topic 6 is about cyclists' speed modifications (Table 8). The most consistent exemplar words for this topic are cyclist and pedal. The exemplar comments refer to speed changes by the interacting cyclist as a means of yielding and mitigating risk. The comments provide evidence of the importance of pedalling in evaluations of cyclist behaviour and pedestrian safety. The prevalence of this topic was significantly higher for less proximate interactions (higher PET). This topic is also more prevalent among risk averse participants and frequent transit users.

Table 6
Topic 4 Exemplar Words and Comments.

Exemplar Words	Exemplar Comments (as written)
Highest prob: bike, lane, close, walk, person **FREX:** bike, lane, fast, close, path **Lift:** swerv, bike, cycl, fast, biker **Score**: everyon, bike, lane, person, biker	"cyclist was going way too fast for that area and wasn't in the bike lane and cut unnecessarily close to the pedestrians." "Bike should be in the bike lane and required to follow the rules of the road." "The person cycled across before the person walking was even close to the cycle lane therefore there yielding wasn't necessary." "same as the last, the biker kept the person safe but left the bike lane onto an empty road" "The biker yielded in their own way, not textbook as they veered onto the car lane but they got the job done and its a non issue because there were no cars around, - not a best practice for the biker though"

Table 7
Topic 5 Exemplar Words and Comments.

Exemplar Words	Exemplar Comments (as written)
Highest prob: yield, slow, didnt, need, think **FREX:** far, need, enough, didnt, yield **Lift:** far, attent, care, enough, need **Score:** attent, yield, far, enough, slow	"The cyclist was far enough away it doesn't seem necessary to have yielded." "The cyclist seemed too far away and going to slowly to need to slow down to yield" "This is a challenging one since we don't see the cyclist until the pedestrian is fully through the crosswalk. We don't know if they slowed down or changed their behaviour when they saw the pedestrian/when the pedestrian completed the crossing, nor do we know if they even saw the pedestrian." "Hard to tell if they needed to slow down based on the camera angle, therefore, I can't agree or disagree if they yielded. However, if they needed to slow down then they definitely should have yielded." "the cyclist was a far enough distance that he didn't need to slow down"

Topic correlation with ratings

Table 9 gives correlation coefficients between topic prevalence in each comment and the associated ratings of comfort and risk (see Section 3 for comfort and risk statement wordings). All correlations between expected topic proportions and ratings of comfort and risk were less than $|0.2|$, and most less than $|0.1|$, indicating that specific topics were not substantially more likely to be discussed when the commenter rated comfort or safety as high or low. Or put another way, topic prevalence was independent of rated perceptions of interaction severity. This result is reflected in the tendency of topics to include both positive and negative evaluations of the associated behaviour. For example, Topic 4 (Cyclists' Leaving the Bike Lane) included comments that were both disapproving (i.e., "cyclist was going way too fast for that area and wasn't in the bike lane") and approving of the behaviour (i.e. "The biker yielded in their own way, not textbook as they veered onto the car lane but they got the job done and its a non issue").

The independence of topic prevalence from comfort and safety ratings suggests that the topics represent a consistent set of primary issues

Table 8
Topic 6 Exemplar Words and Comments.

Exemplar Words	Exemplar Comments (as written)
Highest prob: cyclist, pedestrian, cross, speed, seem **FREX:** cyclist, pedal, speed, distanc, pace **Lift:** wrong, pedal, pace, cyclist, maintain **Score:** cyclist, wrong, cross, pedal, pedestrian	"Neither pedestrian or cyclist modified their speeds or trajectory. Cyclist did not pedal actively through intersection." "The interaction between cyclist and pedestrian was at considerable distance but it appeared the cyclist had slowed and then began to effort more after the pedestrian crossed." "Slow speed of cyclist appears to lower risk of injury and allowed for fair pedestrian crossing" "The cyclist stopped pedalling from a distance to allow the pedestrian ample time to cross, resumed pedalling as she cleared the crossing"

Table 9
Topic Prevalence Correlation with Ratings of Risk and Comfort.

Topic	Correlation with: Low risk rating As integer*	As binary**	Comfort rating As integer*	As binary**
1. Driver behaviour and pedestrian risk	−0.01	−0.05	−0.04	−0.07
2. Pedestrian risk assessment and mitigation	0.10	0.07	0.03	−0.01
3. Pedestrians not waiting before crossing	0.02	0.00	0.01	0.05
4. Cyclists leaving the bicycle lane	0.01	0.07	0.05	0.06
5. Cyclists' yielding necessity	−0.10	−0.14	−0.05	−0.06
6. Cyclists' speed modifications	−0.02	0.06	−0.02	0.05

* 1 = Strongly Disagree, 2 = Disagree, 3 = Agree, and 4 = Strongly Agree.
** TRUE = Agree or Strongly Agree.

that participants focussed on when evaluating the pedestrian interactions, and do not reflect the sentiment that participants had toward those issues. Hence, analysis of the open text comments provides a fundamentally different source of information than quantitative severity ratings. Open text comments can be analysed to discover new issues in emerging fields that can later be tested for their effect size and direction using traditional quantitative survey methods.

Overall, the analysis results support the hypothesis that topic prevalence is determined by personal attributes (travel habits and risk aversion), but fail to support the hypothesis that either topic prevalence or topic content is determined by participant/perceiver socio-demographic attributes. And although the topics relate to perceptions of comfort and safety, topic prevalence is not determined by quantitative severity ratings.

Discussion

The topics reveal a consistent set of primary issues in third-party evaluation of pedestrian interactions, such as head movement by pedestrians and pedaling by cyclists. Speed and speed deviations were key factors for evaluating yielding and risk in both motor vehicle and bicycle interactions. The appropriateness of bicycle path deviations out of the bicycle lane is disputed, as is the obligation for pedestrians to wait before entering the crosswalk. More generally, observers varied in their perceptions of the reasonableness of the observed behaviours—particularly those that were inconsistent with traffic rules or laws.

Some of these factors have been examined in past research on pedestrian comfort, particularly vehicle proximity, volume, and speed (Bigazzi et al., 2021; Kaparias et al., 2015; Landis et al., 1997; Li et al., 2012). Several studies have also explored the effect of non-verbal communication between pedestrians and drivers on driver behaviour (Guéguen et al., 2015; Sucha et al., 2017), but not the effect on pedestrian perceptions of comfort and safety. Differences between formal and informal "rules of the road", and how compliance with each affect pedestrian comfort, is also under-studied – particularly for interactions with cyclists. For example, the effects on perceptions of comfort and safety when cyclists leave the bicycle lane is not well understood.

Interactions involving bicycles elicited discussion of a different set of topics related to pedestrian comfort and safety than interactions involving motor vehicles. For motor vehicle interactions, the topics mainly concerned risk, whereas for bicycle interactions they mainly concerned cyclist yielding. Pedestrian agency or responsibility for risk mitigation was also more of a focus for interactions with motor vehicles than interactions with bicycles. Interactions with bicycles generated more and longer comments than interactions with motor vehicles, hence the cyclist-focused topics had greater overall prevalence in the corpus (Fig. 4). Pedestrian-focused topics had the least overall prevalence (Fig. 4).

The focus on risk and pedestrian agency for evaluations of motor vehicle interactions may be because drivers were dehumanized by participants more than cyclists. This is evidenced by comments that attribute actions of the driver to the motor vehicle (e.g., "there was no time for the car to stop"). Also note that 'car' and 'vehicle' combined were used more frequently than 'driver' (Fig. 3), whereas 'cyclist' was used almost five times more frequently than 'bike'. These patterns in word choice were consistent across participants, as there were no systematic differences in the words participants used to discuss similar topics (i.e., no significant topic content covariates).

Whether or not cyclists need to yield in a given situation is both a disputed and primary issue (Topic 5 was the most prevalent topic), and so can be expected to spark controversy for the public. The focus on cyclist behavior may be a consequence of less cycling activity by the study sample (Table 1) and by the population. Perceiver travel habits, but not socio-demographics, significantly affected the topics raised in evaluating pedestrian comfort and safety, which is consistent with the findings of a quantitative analysis of comfort and safety ratings from the same dataset (Bigazzi et al., 2021). In open comments, participants tended to focus more on the behaviour of those using travel modes they themselves use less often, and vice-versa. This finding is similar to past research in which cyclists were more likely to blame drivers than other cyclists during traffic conflicts, and cyclists with access to a motor vehicle were more likely to blame pedestrians than were cyclists without access to a motor vehicle (Paschalidis et al., 2016).

Conclusion

Understanding perceptions of comfort and safety is necessary to the success of efforts to increase uptake of active transportation modes. In addition to seeking new insights about perceptions of pedestrian comfort and safety beyond the constraints of closed-form questions, this study aimed to demonstrate the value of analysing open comment data from transportation surveys using structural topic modelling. The six identified topics reflect coherent themes discussed by participants related to the comfort and safety of pedestrians interacting with a bicycle or motor vehicle at unsignalized crosswalks:

1. Driver behaviour and pedestrian risk,
2. Pedestrian risk assessment and mitigation,
3. Pedestrians not waiting before crossing,
4. Cyclists leaving the bicycle lane,
5. Cyclists' yielding necessity, and
6. Cyclists' speed modifications.

Analysis of the comments provided additional information beyond what was provided by the participants' quantitative ratings.

Several main implications for future research arise from this study's findings. First, pedestrian interactions with motor vehicles are perceived as fundamentally different than interactions with bicycles. Researchers should consider different measures and focus on different interaction features when seeking to evaluate pedestrian comfort in interactions with each. Second, future research should explore the role of perceived agency in influencing road user perceptions of comfort and safety in interactions with other road users. This clearly differentiates bicycle from motor vehicle interactions, and could be particularly relevant for research on interactions of pedestrians with autonomous vehicles. Related to this, future research should examine differences between the effects of non-compliance with formal vs. informal "rules of the road" on perceptions of comfort and safety. Lastly, transportation researchers should consider topic modeling as a tool to make use of open response survey data. This method would be particularly effective for emerging fields of study because it allows researchers to evaluate *a priori* assumptions about the important variables to measure for quantitative analysis.

The findings have implications for policy as well. Organizations conducting public consultations intended to improve the comfort of a facility or neighbourhood should consider representation by travel mode in addition to socio-demographic groups. Because people focus more on the behaviour of travelers using modes they themselves do not use, it can be expected that public discussion and complaints will be biased against travelers using modes with smaller mode shares. Policy-makers should thus be cautious about responding to public indignation against users of non-dominant and emerging travel modes. Finally, efforts to improve comfort in intermodal interactions should focus not on strict compliance to a homogenous set of traffic rules (often developed for automobiles), but instead consider the unique characteristics of different travel modes and the varying perceptions of what constitutes reasonable behaviour in negotiating shared space (i.e., informal rules of the road).

This study is subject to limitations that should be considered when interpreting its findings. First, the results of STM are reliant on the number of topics chosen to be modeled and the interpretation of the researcher; choosing to model more or fewer topics may have led to different results and interpretations (Roberts et al., 2019). Second, the results of this study are subject to the participants' willingness and ability to articulate their honest opinions and rationales in open text comments. It should not be assumed that failure to discuss a particular issue is due to indifference towards that issue; participants may have simply forgotten to comment on some aspect of the interaction that impacted their rating of comfort and safety, been unable to articulate their reasoning, or considered some aspect of their decision to be trivial or obvious (Garcia et al., 2004). Although socio-demographic variables were considered in the analysis, there may be other, unobserved factors influencing the provision comments by participants, such as English language skills. Comments from the three groups of participants were pooled in the statistical analysis, and although their comment topic proportions were similar (indicating similar content), there was variation in their frequency of commenting. Third, the topics generated for this study are limited by the contents of the 84 videos participants were shown. The findings from this study may only be relevant to unsignalized crosswalks on minor arterial streets. Lastly, the results of this study were limited by the capabilities of the *stm* package. Survey weights and random effects were not able to be included in the model, despite the statistical motivations for their inclusion. Future research should continue to investigate open comment data to establish the replicability of this study's findings.

CRediT authorship contribution statement

Emily Bardutz: Methodology, Software, Formal analysis, Investigation, Writing – original draft, Visualization, Data curation, Project administration. **Alexander Bigazzi:** Conceptualization, Resources, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

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

Acknowledgements

We would like to acknowledge the time and input from Dr. Meghan Winters and Gurdiljot Gill, as well as the study participants, for developing the dataset used for this study. This research was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and uses data from a project funded by the City of Vancouver, under Contract Number PS20181727. The views expressed in this paper are those of the authors and do not represent the views of the project funders.

References

Anciaes, P.R., Jones, P., 2018. Estimating preferences for different types of pedestrian crossing facilities. Transport. Res. Part F: Traffic Psychol. Behav. 52, 222–237. https://doi.org/10.1016/j.trf.2017.11.025.

Anzoise, V., Slanzi, D., Poli, I., 2020. Local stakeholders' narratives about large-scale urban development: The Zhejiang Hangzhou Future Sci-Tech City. Urban Studies 57 (3), 655–671. https://doi.org/10.1177/0042098019828997.

Bennett, R., Vijaygopal, R., Kottasz, R., 2019. Attitudes towards autonomous vehicles among people with physical disabilities. Transport. Res. Part A: Pol. Pract. 127, 1–17. https://doi.org/10.1016/j.tra.2019.07.002.

Bigazzi, A., Gill, G., Winters, M., 2021. Contrasting Perspectives on the Comfort and Safety of Pedestrians Interacting with Other Road Users. Transp. Res. Rec. 2675 (3), 33–43. https://doi.org/10.1177/0361198121992272.

Bozovic, T., Hinckson, E., Smith, M., 2020. Why do people walk? Role of the built environment and state of development of a social model of walkability. Travel Behav. Soc. 20, 181–191. https://doi.org/10.1016/j.tbs.2020.03.010.

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., Blei, D.M., 2009. Reading Tea Leaves: How Humans Interpret Topic Models. Adv. Neural Informat. Process. Syst. 288–296.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. 10.4324/9780203771587.

Daley, M., Rissel, C., 2011. Perspectives and images of cycling as a barrier or facilitator of cycling. Transp. Policy 18 (1), 211–216. https://doi.org/10.1016/j.tranpol.2010.08.004.

DiMaggio, P., Nag, M., Blei, D., 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. Poetics 41 (6), 570–606. https://doi.org/10.1016/j.poetic.2013.08.004.

Fishman, E., Washington, S., Haworth, N., 2012. Barriers and facilitators to public bicycle scheme use: A qualitative approach. Transport. Res. Part F: Traffic Psychol. Behav. 15 (6), 686–698. https://doi.org/10.1016/j.trf.2012.08.002.

Fitch, D.T., Handy, S.L., 2018. The Relationship between Experienced and Imagined Bicycling Comfort and Safety. Transp. Res. Rec. 2672 (36), 116–124. https://doi.org/10.1177/0361198118787635.

Flick, U., 2014. An Introduction to Qualitative Research. SAGE.

Garcia, J., Evans, J., Reshaw, M., 2004. ''Is There Anything Else You Would Like to Tell Us'' – Methodological Issues in the Use of Free-Text Comments from Postal Surveys. Qual. Quant. 38 (2), 113–125. https://doi.org/10.1023/B:QUQU.0000019394.78970.df.

Gkekas, F., Bigazzi, A., Gill, G., 2020. Perceived safety and experienced incidents between pedestrians and cyclists in a high-volume non-motorized shared space. Transport. Res. Interdiscipl. Perspect. 4, 100094 https://doi.org/10.1016/j.trip.2020.100094.

Glanz, B. I., Greeke, E., LaRussa, A., Stuart, F., Rintell, D. J., Chitnis, T., & Healy, B. C. (2016). Risk attitudes and risk perceptions in individuals with multiple sclerosis. Multip. Scler. J. – Exp., Transl. Clin., *2*, 2055217316665406. 10.1177/2055217316665406.

Grimmer, J., Stewart, B.M., 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 21 (3), 267–297. https://doi.org/10.1093/pan/mps028.

Guéguen, N., Meineri, S., Eyssartier, C., 2015. A pedestrian's stare and drivers' stopping behavior: A field experiment at the pedestrian crossing. Saf. Sci. 75, 87–89. https://doi.org/10.1016/j.ssci.2015.01.018.

Habib, K.N., Mann, J., Mahmoud, M., Weiss, A., 2014. Synopsis of bicycle demand in the City of Toronto: Investigating the effects of perception, consciousness and comfortability on the purpose of biking and bike ownership. Transport. Res. Part A: Pol. Pract. 70, 67–80. https://doi.org/10.1016/j.tra.2014.09.012.

Handy, S., van Wee, B., Kroesen, M., 2014. Promoting Cycling for Transport: Research Needs and Challenges. Transport Reviews 34 (1), 4–24. https://doi.org/10.1080/01441647.2013.860204.

Hine, J., 1996. Pedestrian travel experiences. J. Transp. Geogr. 4 (3), 179–199. https://doi.org/10.1016/0966-6923(96)00003-8.

Kaparias, I., Bell, M.G.H., Biagioli, T., Bellezza, L., Mount, B., 2015. Behavioural analysis of interactions between pedestrians and vehicles in street designs with elements of shared space. Transport. Res. Part F: Traffic Psychol. Behav. 30, 115–127. https://doi.org/10.1016/j.trf.2015.02.009.

Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W., 2013. Handbook of Latent Semantic Analysis. Psychology Press. http://www.taylorfrancis.com/books/handbook-latent-semantic-analysis-thomas-landauer-danielle-mcnamara-simon-dennis-walter-kintsch/e/10.4324/9780203936399.

Landis, B.W., Vattikuti, V.R., Brannick, M.T., 1997. Real-Time Human Perceptions: Toward a Bicycle Level of Service. Transp. Res. Rec. 1578 (1), 119–126. https://doi.org/10.3141/1578-15.

Lee, J.D., Kolodge, K., 2020. Exploring Trust in Self-Driving Vehicles Through Text Analysis. Hum. Factors 62 (2), 260–277. https://doi.org/10.1177/0018720819872672.

Lesnikowski, A., Belfer, E., Rodman, E., Smith, J., Biesbroek, R., Wilkerson, J.D., Ford, J. D., Berrang-Ford, L., 2019. Frontiers in data analytics for adaptation research: Topic modeling. WIREs Clim. Change 10 (3), e576. https://doi.org/10.1002/wcc.576.

Li, Z., Wang, W., Liu, P., Ragland, D.R., 2012. Physical environments influencing bicyclists' perception of comfort on separated and on-street bicycle facilities. Transport. Res. Part D: Transport Environ. 17 (3), 256–261. https://doi.org/10.1016/j.trd.2011.12.001.

Mars, L., Arroyo, R., Ruiz, T., 2016. Qualitative Research in Travel Behavior Studies. Transp. Res. Procedia 18, 434–445. https://doi.org/10.1016/j.trpro.2016.12.057.

Mitman, M.F., Ragland, D.R., 2007. Crosswalk Confusion: More Evidence Why Pedestrian and Driver Knowledge of the Vehicle Code Should Not Be Assumed. Transp. Res. Rec. 2002 (1), 55–63. https://doi.org/10.3141/2002-07.

Monsere, C.M., McNeil, N.W., Sanders, R.L., 2020. User-Rated Comfort and Preference of Separated Bike Lane Intersection Designs. Transp. Res. Rec. 2674 (9), 216–229.

Mourtgos, S.M., Adams, I.T., 2019. The rhetoric of de-policing: Evaluating open-ended survey responses from police officers with machine learning-based structural topic modeling. J. Crim. Justice 64, 101627. https://doi.org/10.1016/j.jcrimjus.2019.101627.

Næss, P., 2020. Validating explanatory qualitative research: Enhancing the interpretation of interviews in urban planning and transportation research. Applied Mobilities 5 (2), 186–205. https://doi.org/10.1080/23800127.2018.1464814.

O'Cathain, A., Thomas, K.J., 2004. "Any other comments?" Open questions on questionnaires – a bane or a bonus to research? BMC Med. Res. Method. 4 (1), 25. https://doi.org/10.1186/1471-2288-4-25.

Osborne, R., Horberry, T., Young, K.L., 2020. Pedestrian distraction from Smartphones: An end-user perspective on current and future countermeasures. Transport. Res. Part F: Traffic Psychol. Behav. 73, 348–361. https://doi.org/10.1016/j.trf.2020.07.007.

Paschalidis, E., Basbas, S., Politis, I., Prodromou, M., 2016. "Put the blame on…others!": The battle of cyclists against pedestrians and car drivers at the urban environment. A cyclists' perception study. Transportat. Res. Part F: Traffic Psychol. Behav. 41, 243–260. https://doi.org/10.1016/j.trf.2015.07.021.

Peña-García, A., Hurtado, A., Aguilar-Luzón, M.C., 2015. Impact of public lighting on pedestrians' perception of safety and well-being. Saf. Sci. 78, 142–148. https://doi.org/10.1016/j.ssci.2015.04.009.

Pollack, K.M., Gielen, A.C., Mohd Ismail, M.N., Mitzner, M., Wu, M., Links, J.M., 2014. Investigating and improving pedestrian safety in an urban environment. Injury Epidemiology 1 (1), 11. https://doi.org/10.1186/2197-1714-1-11.

Pooley, C.G., Horton, D., Scheldeman, G., Tight, M., Jones, T., Chisholm, A., Harwatt, H., Jopson, A., 2011. Household decision-making for everyday travel: A case study of walking and cycling in Lancaster (UK). J. Transp. Geogr. 19 (6), 1601–1607. https://doi.org/10.1016/j.jtrangeo.2011.03.010.

Poulenez-Donovan, C.J., Ulberg, C., 1994. Seeing the trees and missing the forest: Qualitative versus quantitative research findings in a model transportation demand management program evaluation. Transp. Res. Rec. 1 (1).

R Core Team. (2016). *R: A language and environment for statistical computing* (3.3.0) [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/.

Roberts, M.E., Stewart, B.M., Tingley, D., 2019. stm: R Package for Structural Topic Models. J. Stat. Softw. 41.

Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S.K., Albertson, B., Rand, D.G., 2014. Structural Topic Models for Open-Ended Survey Responses. Am. J. Polit. Sci. 58 (4), 1064–1082. https://doi.org/10.1111/ajps.12103.

Schuman, H., Presser, S., 1979. The Open and Closed Question. Am. Sociol. Rev. 44 (5), 692–712. https://doi.org/10.2307/2094521.

Shay, E., Combs, T.S., Findley, D., Kolosna, C., Madeley, M., Salvesen, D., 2016. Identifying transportation disadvantage: Mixed-methods analysis combining GIS mapping with qualitative data. Transp. Policy 48, 129–138. https://doi.org/10.1016/j.tranpol.2016.03.002.

Sucha, M., Dostal, D., Risser, R., 2017. Pedestrian-driver communication and decision strategies at marked crossings. Accid. Anal. Prev. 102, 41–50. https://doi.org/10.1016/j.aap.2017.02.018.

Tvinnereim, E., Fløttum, K., 2015. Explaining topic prevalence in answers to open-ended survey questions about climate change. Nat. Clim. Change. https://doi.org/10.1038/nclimate2663.

Tvinnereim, E., Fløttum, K., Gjerstad, Ø., Johannesson, M.P., Nordø, Å.D., 2017. Citizens' preferences for tackling climate change. Quantitative and qualitative analyses of their freely formulated solutions. Global Environ. Change 46, 34–41. https://doi.org/10.1016/j.gloenvcha.2017.06.005.