

News Media Mining to Explore Speed-Crash-Traffic Association During COVID-19

Subasish Das¹  and Sobhan Sarkar² 

Transportation Research Record

1–18

© National Academy of Sciences:

Transportation Research Board 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03611981221121261

journals.sagepub.com/home/trr



Abstract

The COVID-19 pandemic has affected the lives of everyone from almost every perspective. Research communities from many domains have contributed to quantifying and mitigating the influence of the ongoing pandemic. The public depend on news media for reliable information. At the start of the pandemic in March 2020, many news media articles started to report on quieter roads resulting from travel restriction and work from home mandates, but also more excessive speeding and more traffic fatalities. Although the news media's representation of the crash-speed association is often not based on data-driven safety analysis, the extent to which the media place emphasis on such news content has never been adequately quantified. The current study performed content analysis and text mining (both text network analysis and topic modeling) to explore the representations of speed-crash association in news media (local, national, and international) to provide insights into the generated news content. Findings reveal that topics such as a surge in fatal crashes, careless driving, driving under the influence, law enforcement, and equity issues, as well as impacts on pedestrians and bicyclists, were frequently highlighted by local, national, and international news media during the pandemic era.

Keywords

policy and organization, executive management issues, communications and engagement, sustainability and resilience, transportation and society, accessible transportation and mobility, communication

To address the COVID-19 pandemic and minimize the spread of the coronavirus, many social distancing measures, such as “shelter-in-place” orders, halting non-essential business operations, limitations of gatherings, and so forth, were taken in 2020 (1). For example, there were six COVID-19-related orders in Texas, and there were four in Dallas County alone. These mitigation measures affected lives in nearly every aspect, including travel behavior. According to the Federal Highway Administration’s Office of Highway Policy Information, the vehicle miles traveled (VMT) on Interstate highways in the United States dropped drastically from mid-March 2020, to –50% at its lowest point at the beginning of April 2020 (2, 3). Recently released fatality data from the Fatality Analysis Reporting System (FARS) indicates that speeding-related crashes increased by 17.8% in 2020 compared with 2019 and 9.7% in 2020 compared with 2016 (4). Although the FARS data justifies the rationality of media reporting on speeding-

related crashes, it is important to examine the perspectives and causations behind this trend.

This extensive and prolonged disruption in travel patterns provided a unique opportunity for transportation researchers to investigate transportation-related subjects from a fresh perspective. For example, with traffic demand lower than “normal,” or pre-COVID, conditions, roadway crash frequency decreased in comparison with previous years, which reinforces the model that motor vehicle crash occurrence is a function of risk and exposure (i.e., VMT) (5, 6). In other words, fewer crashes are expected if the traffic exposure is lower.

¹Ingram School of Engineering, Texas State University, San Marcos, TX

²Information Systems & Business Analytics, Indian Institute of Management Ranchi, Ranchi, Jharkhand, India

Corresponding Author:

Subasish Das, subasish@txstate.edu

Table 1. Key Studies on Speed-Related Crashes During COVID-19

Reference	Location	Issue	Method	Key findings
Gao et al. (7)	New York and Seattle	Impact of COVID-19 on transportation system	Descriptive statistics	Crashes were low because of low traffic; fatalities increased, possibly because of higher speeds.
Aarts and Van Schagen (8)	NA	Speed-crash association during COVID-19	Literature review	High speed was associated with more severe crashes.
Qureshi et al. (9)	Missouri	Lockdown and traffic crashes	Descriptive statistics	Fatal and injury crashes are unchanged compared with earlier years. Minor injury and no injury crashes decreased.
Doucette et al. (10)	Connecticut	Impact of COVID-19 stay-at-home order on traffic crash patterns	Descriptive statistics	Single-vehicle crash rates significantly increased by 2.29 times.
Pishue (11)	U.S.A.	Effect of COVID-19 on Interstate and highway crashes	Descriptive statistics	Overall vehicle miles traveled (VMT) dropped at onset of COVID-19 with major metropolitan areas yet to recover. Lower VMT was associated with less traffic congestion and higher speeds.
Shilling and Waetjen (12)	California	Impact of COVID-19 on crashes	Descriptive statistics	Fatal and injury crashes were reduced by 50%. For certain roadways, maximum and average operating speeds increased slightly, although the difference is just a few miles per hour.
Shirani-bidabadi et al. (13)	Alabama	Within-day travel speed patterns	Unsupervised data mining	A new travel speed pattern at end of travel restriction order was identified.

Note: NA = not available.

However, other than basic information about the new trends in the transportation network, most altered patterns are still unknown. The ordinances and policy actions of different states and cities during the pandemic varied drastically according to the local conditions or perspectives toward COVID-19. Moreover, even residents' responses and demands in these conditions were widely different considering individuals' values, beliefs, or socio-economic characteristics. Because of this, diverse traveling and driving behaviors are expected from different communities and groups across the country. This study aims to identify the key topics and trends from the unstructured news media text data so that speeding-related behaviors during the COVID-19 can be examined well for future policy related changes. Reviewing the news coverage, which often reports local or global trends and issues, would be the first step to draw the attention of researchers to the public's concerns and highlight the variety of transportation patterns.

This study can be considered as a starting point to explore the contents in news media articles reporting on the association of speed and crashes, and the overall

intention of reporting this information. The content analysis and natural language processing (NLP) applied in this study provide some evidence on the reporting contents and perceived relevant information and knowledge by mining 57 relevant online news media articles (links collected from Google News based on keyword search) published in 2020 and 2021.

Literature Review

The literature review focuses on academic reports and papers which explored the associations of crash and speed or speed and traffic volume during COVID-19. The literature review is limited to the studies related to the speed-crash association during COVID-19. Table 1 provides a brief overview of the studies recently conducted to explore the speed-crash association during COVID-19. A general overview of the studies shows that the majority of them provided a descriptive analysis to explore crash, speed, and driver behavior during COVID-19. As the traffic volume dropped, less traffic on the roads encouraged higher travel speed (7, 11).

Table 2. Numbers of Vehicles Involved in Speeding-Related Fatal Crashes

Speeding-related variables	2016	2017	2018	2019	2020
Yes, exceeded speed limit	3888	3675	3486	3296	4183
Yes, racing	84	82	79	78	143
Yes, specifics unknown	1231	1368	1280	1357	1875
Yes, too fast for conditions	4202	3966	3951	4023	4111
Total	9405	9091	8796	8754	10312

High speed is considered a factor for increased risk of being involved in a crash, particularly with more severe crashes (4, 7, 8). However, crashes are affected by many other factors and by complicated psychosocial reasons other than speed. Qureshi et al. (13) determined that property damage only (PDO) and minor injury crashes were less frequent during the pandemic. However, fatal and severe injury crashes did not change much, according to another study, although the study period could have affected the study results. Some published studies that used only data from March and April reached different conclusions. For example, a study conducted by Katrakazas et al. (6), only using the March 2020 crash data, concluded that the total number of crashes and fatal crashes decreased. This study also found evidence of frequent harsh acceleration and braking during COVID-19. Another study produced descriptive statistics on fatal and injury crash reduction (12). A study conducted in Connecticut showed an increase in the single-vehicle crash rate during COVID-19. An Alabama study (13) identified an emerging pattern of “new normal” within-day travel behavior. The findings from COVID-19 related studies (see Table 1) indicate the need for a robust exploration of the driving and travel patterns and their associations with traffic crashes.

Two recent papers documented the results of crash-speed association during the COVID-19 pandemic. Dong et al. (14) reported that the share of severe crashes (including serious injuries and fatal injuries) has increased from 6.09% to 7.63% in comparison to the Covid-19 outbreak. Moreover, Zou et al. (3) reported that the fatal crash rate increased to 1.9 from 1.4 per thousand accidents in the first three weeks of April 2020 compared with the same time of February 2020.

The literature review shows that the studies on the associations of speed and crashes and of speed and traffic volume during the COVID-19 pandemic are mostly based on descriptive statistics of incomplete data. As state-maintained crash data require quality checks before publication, many of the papers and reports are based on incomplete and less reliable safety data. Additionally, many studies fail to acquire information about operating speed or speeding-related crash

information. Additionally, a crash is an outcome of multifarious spatiotemporal and environmental factors with the inclusion of demographics and driver behavior. There is a need for an extensive scientific study which can examine this crash-speed-volume relationship during COVID-19 with the required data and proper study design. The current study aims to explore the perceived knowledge in news media content about speed-crash-volume relationships.

Methodology

Data Collection

Table 2 lists the number of vehicles involved in fatal crashes during the period from 2016 to 2020. The year 2020 experienced a higher number of fatal crashes compared with the earlier years (an 18% increase in 2020 compared with 2019 and a 10% increase in 2020 compared with 2016).

In the absence of any updated and functional sources to understand the speed-crash association during COVID-19, this study collected relevant information from online news media reports. Extracting information from online media reports has been considered a popular method for exploring insights into news trends or societal concerns. This study collected news from the Google News search engine. Search terms such as “COVID speed,” “COVID speeding,” “COVID crash,” “COVID crashes,” “COVID accident,” “COVID accidents,” “COVID over speeding,” “COVID restrictions,” “COVID traffic,” and “COVID roads” were used to collect the news URLs (i.e., URL or Uniform Resource Locator indicates the address of a webpage). Later “COVID” was replaced with “COVID-19” and “coronavirus” to collect additional URLs. The timeline of the news media collected was from March 2020 to July 2021. A list of 73 URLs was collected at the beginning. After manual checking, some of the articles were discarded because of duplication and non-relevance. After web scraping using the popular Python library BeautifulSoup (<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>), 57 articles were selected for analysis. The selected news articles are shown in Table 3.

Table 3. Selected News Articles From Web Scraping

News media	Year	Article title
National	2020	New York City Traffic Deaths Rise During COVID-19 Pandemic
National	2020	The Roads Are Quieter due to Coronavirus, but There Are More Fatal Car Crashes
National	2020	The Coronavirus Pandemic Emptied America's Roadways. Now Speeders Have Taken Over.
National	2020	COVID-19 Cuts Car Crashes—But What About Crash Rates?
National	2020	Riskier Behavior During Pandemic Is Causing More Accidents
National	2020	This Spring, We All Drove Much Less. Yet Traffic Deaths Went Up. Why?
National	2020	New Study Finds Coronavirus Lockdowns Dramatically Cut Traffic Accidents and Fatalities
National	2020	Drunk and Impaired Driving Awareness During COVID-19
National	2020	On Nearly Emptied Roads, Motor Vehicle Fatality Rate Spikes by 14% in March
National	2020	Work Zone Crashes Climb During Pandemic, Even as Traffic Ebbs
National	2020	California Crashes and Traffic Down by Half During Shelter-in-Place for COVID-19, Saving State \$40M/Day
National	2020	2019 U.S. Traffic Deaths Lowest Since 2014, but 2020 Numbers Aren't Looking Good
National	2021 ^a	Black People Are More Likely To Die in Traffic Accidents. COVID Made It Worse
National	2021	Rate of Traffic Deaths Rises During COVID as Open Roads Tempt Drivers to Speed up
National	2021	Car-Crash Death Rate Surged in 2020
National	2021	Another Way that COVID Can Kill: Car Crashes
National	2021	Traffic Deaths Rose 8% in 2020, Even as Americans Drove Fewer Miles During Pandemic
National	2021	Fast & Furious: Pandemic Speeding Has Deadly Consequences
National	2021	The Pandemic Was the Bloodiest Year for Driving in over a Decade
National	2021	Alcohol-Involved Deaths on Roads Doubled During COVID Year
National	2021	Despite Reduced Traffic, Accidents Have Increased
National	2021	Traffic Deaths Rose 8% in 2020, Even As Americans Drove Fewer Miles During Pandemic
National	2021	A Dark Conundrum Americans Are Driving Less, but More Are Dying in Accidents
National	2021	How COVID's Toll Compares With Other Things That Kill Us
National	2021	Road Deaths Down as COVID-19 Pandemic Cuts Traffic
National	2021	2020 Accident Rate Raises Questions About COVID's Safety Effect
National	2021	Chicago Traffic Deaths Were up 45% Last Year as Speeding Increased During COVID
National	2021	Pedestrian Fatalities Spike During Pandemic
National	2021	'Mind-Boggling': Pedestrian Deaths Surged in 2020, Despite Fewer Cars on the Road
National	2021	Despite Fewer Cars on the Road, Pedestrian Fatalities Are up During COVID
National	2021	Why Emptier Streets Meant an Especially Deadly Year for Traffic Deaths
National	2021	Car Culture Disproportionately Kills Black Americans. The Pandemic Made Things Worse
National	2021	Where Traffic Deaths Surged in Wisconsin During the Pandemic
Local	2020	Virginia Sees Uptick in Unbelted and Speed-Related Crashes During COVID-19 Pandemic
Local	2020	Wisconsin Has Seen More Car Crash Deaths During the Pandemic—Even With Less Driving
Local	2020	Are Higher Traffic Fatalities Another Symptom of COVID in Rhode Island?
Local	2020	Study: Risk of Death or Injury Is Greater When Roads Are More Clear
Local	2020	Colorado's Roads Are emptier, but Deadlier so far This Year
Local	2020	Teenagers Nearly 4 Times More Likely To Die in a Car Crash than by COVID-19 This Summer
Local	2020	'Alarming Increase': L.A. Traffic Deaths Surge Back to Pre-Pandemic Levels
Local	2020	Coronavirus Stay-At-Home Order Saves State Taxpayers \$1 Billion After Car Crashes Cut by 60%
Local	2020	How Bike Crashes Shifted out of Downtown DC During the Pandemic
Local	2021	New AAA Study Reveals how Driving Changed During COVID-19 Pandemic
Local	2021	More Fatal Crashes on NJ's Roads in COVID-laced 2020—why?
Local	2021	Despite a Massive Drop in Crashes, Texas Sees Rise in Speed-Related Road Deaths During the Pandemic
Local	2021	Some Fear Rise in Deadly Crashes as Connecticut Teens Hit the Road This Summer With COVID-19 in Rearview Mirror
Local	2021	Why Are There So Many Seattle Traffic Fatalities When So Few People Are Commuting?
Local	2021	Texas Roadway Deaths "Going The Wrong Way" As 2021 Looks To Be Risky Year for Driving
Local	2021	Reckless Driving Increases Since COVID-19
Local	2021	Cyclist Deaths Remain High During COVID-19
International	2020	U.S. Traffic Deaths Fell After Coronavirus Lockdown, But Drivers Got Riskier
International	2020	Traffic, Crashes and Congestion Rebounding After COVID-19 Spring Slump
International	2020	COVID-19: Crashes More Likely To Be Fatal When Roads Are Clearer, Finds New Study
International	2021	Using COVID-19 Lockdown Road-Crash Data to Inform Transport Safety Policy, Cali, Colombia
International	2021	Road Accidents, Deaths Drop by Over 20% In 2020 Partly due to Covid-19: Gadkari
International	2021	Report: Traffic Fatalities Increased in Sarasota and Manatee Counties During COVID-19
International	2021	Road Deaths in Japan Fell to Record Low in 2020 Amid COVID-19 Stay Home Requests

^a 2021 indicates January to July of 2021.

Table 4. Numbers of Selected News Articles From Web Scraping

Media type	2020	2021	Total
National	12	21	33
Local	9	8	17
International	3	4	7
Total	24	33	57

Natural Language Processing (NLP)

Text mining is a popular data mining methods. In recent years, many transportation engineering studies adopted different NLP methods in their analytical framework (15–31). It is a branch of NLP that provides exploratory analysis of unstructured datasets. Text mining can be extended to complicated statistical analysis based on the research question or research needs. Text network analysis, a tool of text mining, can provide important insights into hidden trends of unstructured text data. This method consists of three major processes:

- Step 1: Text cleaning and token development: The first steps require some text mining steps that are necessary to prepare the data for analysis. These steps include text cleaning by using: (i) stop word removal and removal of study-specific needs such as removal of redundant punctuation, number, special text, or URLs, (ii) token or phrase generation, and (iii) perform stemming (i.e., word root generation; for example, the stem of ‘injury’ and ‘injuries’ is ‘injur’) or lemmatization (i.e., word generation by morphological analysis; for example, the lemma of ‘injury’ and ‘injuries’ is ‘injury’) based on the study needs.
- Step 2: Node and edge generation: This step involves a word search algorithm. The algorithm will at first find two words or n words sentence by sentence in a corpus (i.e., a document with words and sentences). For the first combination, a node with a word and edge between two words or n words will be created. The search will continue to acquire such a pair or pairs for the following sentences. If such a pair or pairs exist, the weight on the node and edge will increase by a unit weight. With the presence of a new word or new pair, the network will grow and follow the steps of weighting throughout the corpus.
- Step 3: Network analysis: Network analysis involves an explanation of the nodes and edges based on a network plot. Usually, a large node indicates the presence of a keyword that is associated with many other words with some specific high edge association with particular words in the

form of higher co-occurrence in the corpus. The presence of such nodes can be considered as clusters or key topics determined by the word with a large node and higher edges for different combinations of word pairs.

- Step 4: Topic modeling: This study trained Latent Dirichlet Allocation (LDA) topic models with the R mallet package (32). Statistical topic modeling, such as LDA (33), extracts a coherent theme, which is a probability distribution across a vocabulary assuming documents are formed of many themes. Each theme (or topic) is typically represented by the words (which we refer to as topic words) that appear most frequently in the relevant documents, as well as by the papers that best illustrate the theme. In recent years, many studies have applied topic modeling to extract crucial trends and patterns.

The news articles were divided into several groups to perform text network analysis and topic modeling. First, the articles were divided into three groups based on the media types: local, national, and international. These groups were analyzed for both text network analysis and topic modeling. Additionally, the news articles were divided into two groups based on publication year: 2020 and 2021. Only text network analysis was performed on this group. Using the R package (15), all the texts were transformed to lower case, normalized white spaces, and erased punctuation, non-alphanumeric characters, and short terms of just one or two characters after selecting petitions written in English. This study uses an English stop words dictionary contained in the “mallet” package to delete less informative words such as “a,” “the,” and “of,” which exist in virtually all English publications; “amp” is added to the stop words dictionary to eliminate “amp,” which is a processed form of ampersand (&).

Results

Content Analysis

Table 4 lists the selected news articles gathered from web scraping. The number of articles by type (national, local, and international) and by year (2020 and 2021) are presented. For the national type, there are 12 from 2020 and 21 from 2021 (a total of 33), for the local type there are nine from 2020 and eight from 2021 (a total of 17), and for the international type there are three from 2020 and four from 2021 (a total of 7). Of the total number of articles (which is 57), around 42.1% are from 2020 and about 57.9% are from 2021. Of the national total (which is 33), around 36.4% are from 2020 and about 63.6% are from 2021. Of the local total (which is 17), around 52.9% are from 2020 and about 47.1% are from 2021.

Table 5. Measures of Article Length in Words

Media type	Number of articles	Number of words			
		Maximum	Minimum	Mean	Standard deviation
National	33	1979	153	717.1	455.0
Local	17	1213	144	656.0	305.1
International	7	786	339	568.7	198.0

Of the international total (which is 7), about 42.9% are from 2020 and about 57.1% are from 2021.

Table 5 lists the lengths of the articles through the number of words within them. National, local, and international groups are considered. National typically had the most words (with a mean of 717.1 words), whereas international typically had the fewest (with a mean of 568.7 words). Local, which was in between the two, had a mean of 656.0 words. International had both the lowest maximum number of words (786, compared with 1,979 words for national and 1,213 words for local) and the highest minimum number of words (339 words, compared with 153 words for national and 144 words for local). National had the highest standard deviation at 455.0, compared with 305.1 for local and 198.0 for international.

Content analysis was performed by using the data storage and analysis software Tagouette (34). During the COVID-19 pandemic era, cities experienced another crisis: a rise in extreme speeding, reckless driving behavior, and consequently a surge in car crashes, injuries, and fatalities. Reviewing the local, national, and international reports indicates that there are some specific topics that were mainly discussed about transportation safety during the pandemic. Below, each topic is explored and the related main points are discussed.

Driving Behavior. With the COVID-19 restrictions and lack of entertainment, individuals' need to get out of stay-at-home situations apparently led drivers to release their pent-up emotions out on the streets and highways. In addition to this mental circumstance, the road network also was unusually empty because of the shelter-in-place orders throughout the cities. Therefore, as could be expected, driving behavior would have been changed because of this altered situation. According to transportation officials, speeding and other reckless behavior on the cities' emptier roads, especially at night, caused the increase. Numerous reports have addressed the extreme increase in vehicles' average speed, specifically at night. Drivers with a passion for racing even used the advantage of empty roads to hold illegal racing competitions. Concerning incidence of reckless driving behaviors has been widely reported.

Number of Crashes. The total number of crashes fell during the major lockdowns in the U.S.A. The accident rate ratio of single-vehicle crashes increased during the pandemic, but multi-vehicle crashes reduced notably. However, the drop was not as sharp when it was about the crashes being caused by risky driving behaviors, such as driving over the speed limit or impaired driving.

Fatalities. Although the number of VMT by drivers has decreased notably, the rate of traffic crash fatalities per mile driven has raised. There are more fatal car crashes involving teen/younger drivers, as teens are more prone to drive with unstructured time. Drivers probably get the mindset that there are fewer cars out there and therefore there is more freedom for maneuvers. Moreover, reports declare that fewer individuals involved in crashes had used their seat belts.

Equity. Specific shelter-in-place mandates forced more people to stay at home, yet lower-income earners took the most daily trips and therefore were more exposed to crashes. Moreover, according to the reports, elderly residents and homeless individuals are disproportionately affected by fatal crashes. Specifically, Black individuals were involved in the largest rise in fatal crashes in 2020/21 than other racial/ethnic communities. According to the administration's reports, the rate of Black individuals who lost their lives in traffic crashes was 23% higher than in 2019, the largest increase in traffic deaths among racial groups. A lack of proper infrastructure such as sidewalks or effective traffic signs might be the reason behind this issue.

Driving Under Influence. According to data from the Utah Department of Transportation, the number of people who died in alcohol-involved traffic crashes doubled in 2020 compared with 2019. The reason behind this surge might be drivers' assumption of the reduced level of enforcement during the shutdown. Driving under the influence of alcohol and drugs can possibly increase the severity of crashes and consequently cause more fatalities.

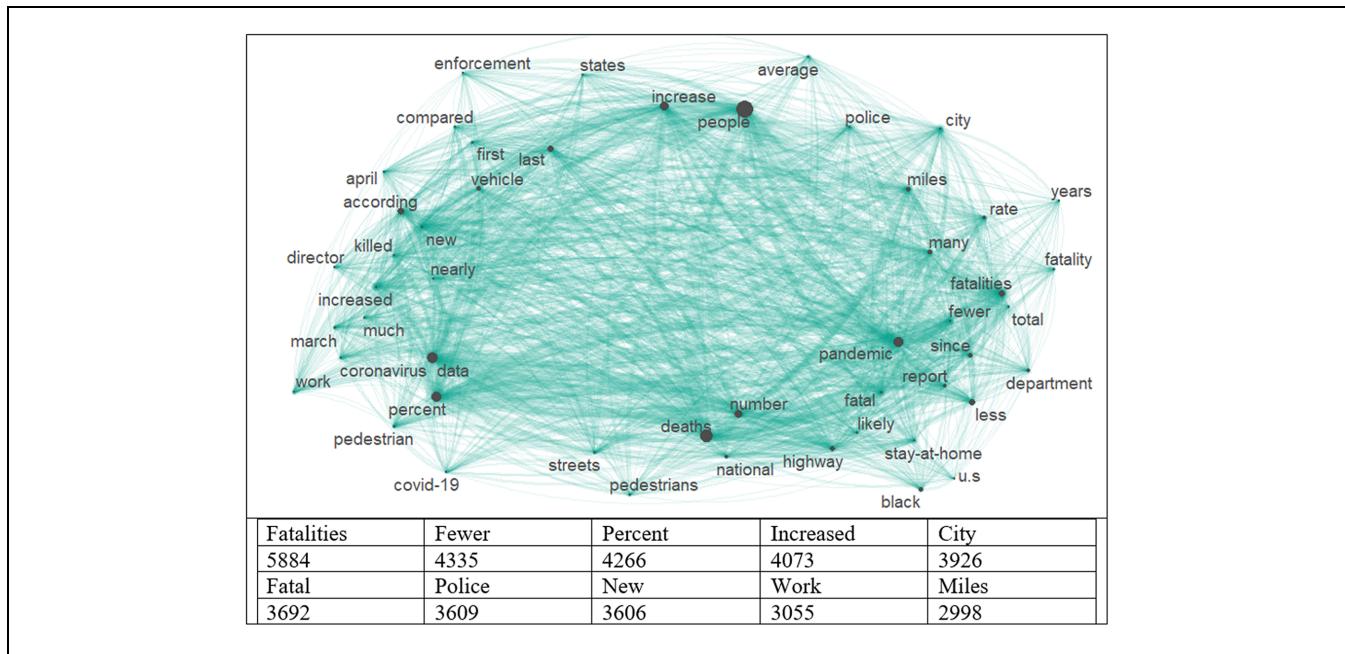


Figure 1. Text network from national news media text corpus.

Law Enforcement. Various reports have suggested different levers to reduce the consequences of negatively altered driving behaviors during the pandemic. Increasing automated technologies, such as speed cameras, and public awareness of their use was discussed as one of the potential remedies for lowering the number of speeding problems by warning drivers that they are under surveillance of the law. Moreover, police were encouraged to raise the number of citations, although the number of speeding tickets has already risen considerably. It was also reported that some states have implemented other actions, such as altering the pedestrian walk signal and switching the traffic signals into the nighttime mode, which lowers the probability of successive green lights (green wave) for individual cars.

Pedestrians. Despite the rise in crash fatalities, pedestrian deaths hit one of their lowest recorded levels during the pandemic, considering that individuals were more encouraged to walk rather than use public transit services because of health issues. On the other hand, business closures and working from home also lowered the number of times residents decided to walk.

Text Network Analysis Results

The entire analysis was performed in R-statistical software (35), using the quanteda and igraph packages, respectively (36, 37). Figures 1 to 3 show text networks

for the three different text corpora: national, local, and international.

National News Media. Figure 1 displays the text network for the national group. The top features include “fatalities,” “fewer,” “percent,” “increased,” “city,” “fatal,” “police,” “new,” “work,” and “miles.” Visually, the words “increase,” “data,” “deaths,” “people,” “pandemic,” and “percent” appear to be closest to the center of the figure. Words such as “U.S.,” “enforcement,” and “years” appear to be on the outermost parts of the figure. It is notable that certain words have larger nodes when compared with others. “People” appears to have the largest node, followed by “deaths,” “percent,” and “coronavirus.” This appears to convey that the number of deaths by COVID-19 was a topic often discussed by national news.

Collocation indicates a set of words occurring together in a text corpus (i.e., a document with texts). Collocation helps to understand unambiguous meaning or connotation for a word group from an unstructured dataset by extracting information. This study extracts collocations in the form of rigid collocations by extracting n-grams (in this study, bigram and trigram, which means groups of two or three words together) that always occur side by side by keeping the same order. The lambda (λ) computed for an n-word collocated group is the coefficient for the n-way interaction parameter in the saturated log-linear model fitted to n-word counts. High z-values of the collocated word groups indicate the statistical significance

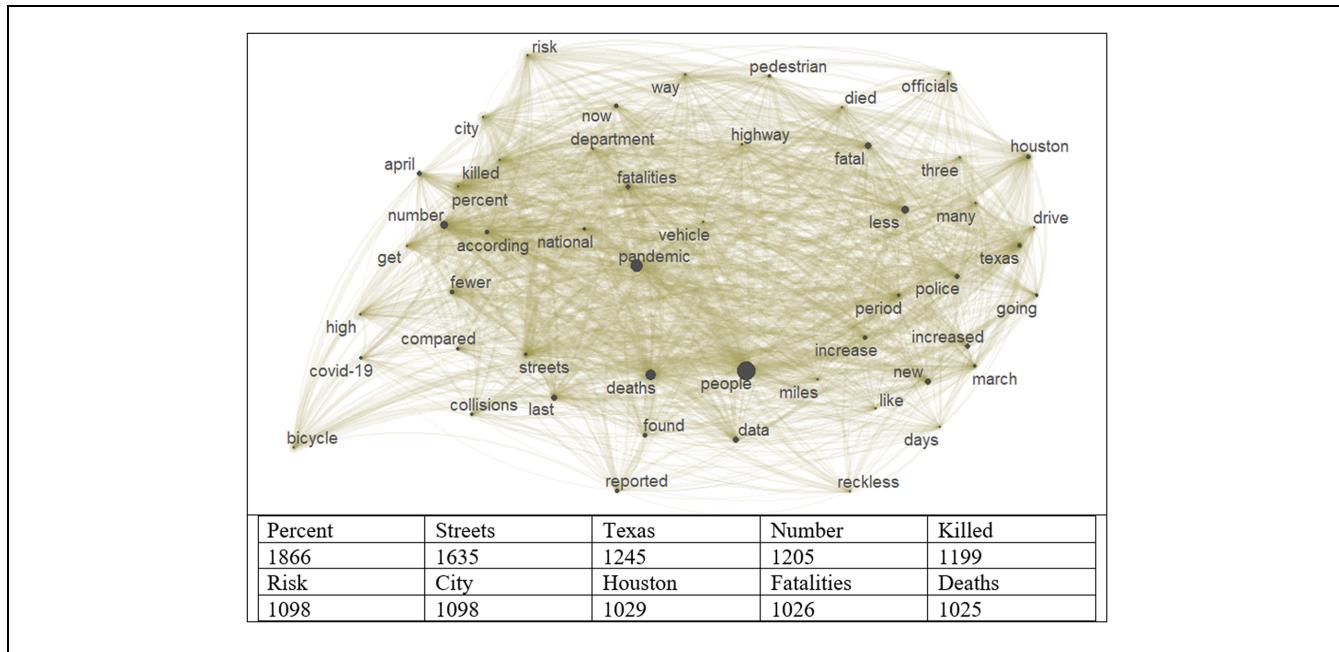


Figure 2. Text network from local news media text corpus.

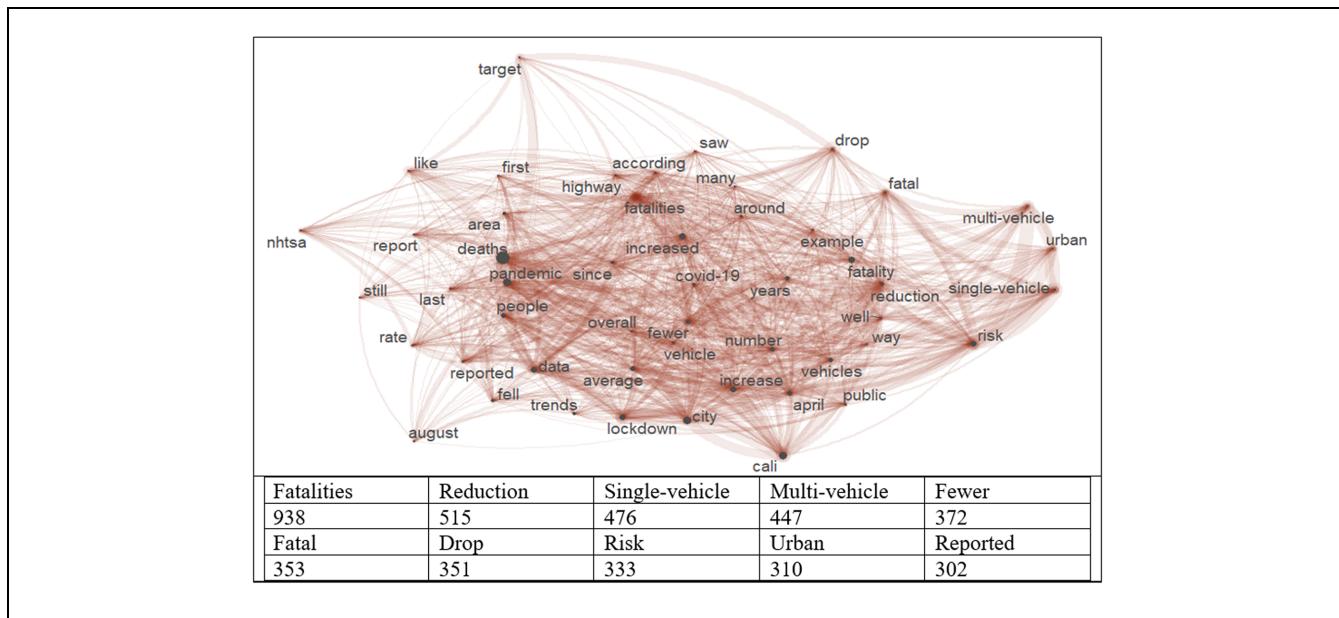


Figure 3. Text network from international news media text corpus.

of occurrences at the 95% confidence level. Table 5 shows the top 20 collocations for the national news media text corpus, organized by count, length, lambda, and z-values. “National highway” is ranked in the first position (with a z-value of 14.88), but it does not have the highest count or lambda value. “People died” has the highest count and “Americans drove” has the highest lambda value, but

both have a lower z-score and are ranked lower (8 and 15, respectively). The top 15 collocations have a word length of two, whereas the bottom five have a word length of three. The z-value of the bottom five three-word collocations also significantly drops down to below two, compared with the first 10 two-word collocations. This drop in the z-value is related to the length of the words. It

Table 6. Collocation Measures of Bigrams and Trigrams of National News Media Text Corpus

No.	Collocation	Count	Length	Lambda	z
1	National highway	16	2	5.49	14.88
2	Fatality rate	14	2	5.87	14.49
3	York City	13	2	5.99	13.86
4	National council	10	2	5.81	12.76
5	Work zone	13	2	7.46	12.66
6	Across country	9	2	6.66	12.60
7	Miles driven	12	2	5.94	12.29
8	People died	18	2	5.70	12.22
9	COVID-19 pandemic	12	2	4.50	12.02
10	Stay-at-home order	8	2	6.31	11.85
11	First half	8	2	5.52	11.66
12	Highway administration	18	2	7.77	11.18
13	Mile driven	7	2	7.32	11.09
14	Work zones	7	2	6.24	10.92
15	Americans drove	6	2	7.94	10.49
16	Drove fewer miles	3	3	3.80	1.49
17	Total deaths percent	3	3	2.36	1.37
18	Largest increase deaths	3	3	2.14	1.17
19	Pandemic consequences way	4	3	2.89	1.13
20	Number speed-related fatalities	3	3	2.05	1.04

is notable that in Table 6 the collocation “York City” appears. This does not appear in the other tables of collocations, meaning it is more relevant to the national news than to local or international. This could be because the national news discusses the areas that were the most affected by the COVID-19 pandemic.

Local News Media. The text network for the local news media text corpus is displayed in Figure 2. “Percent,” “streets,” “Texas,” “number,” “killed,” “risk,” “city,” “Houston,” “fatalities,” and “deaths” are the displayed top features. “Pandemic,” “deaths,” “number,” “fatalities,” and “people” visually appear to be closest to the center of the text network. Some notable words lying on the outside of the network include “officials,” “reported,” “COVID-19,” “April,” and “drive.” The words “people,” “pandemic,” “deaths,” and “number” have the largest nodes, relative to the rest. Similar to Figure 1, Figure 2 seems to have a focus on deaths and the number of deaths from the pandemic. “People,” “pandemic,” “deaths,” and “number” are repeated in the two figures as words close to the center. Both Figures 1 and 2 have similar larger nodes, with “people” and “deaths” repeated between the two. Figure 1, however, has “coronavirus” as a large node, whereas Figure 2 has “pandemic.” Despite most likely conveying the same thing (deaths caused by COVID-19), it is interesting to note the difference in language used by national news versus local news. As to the top features, the words “percent,” “city,” and “fatalities” are repeated between the two. However, the local top features (Figure 2) clearly have

more of a focus on specific locations, referencing “Houston” and “Texas.”

The top 20 keywords for the local news media text corpus, organized by count, length, lambda, and z-values, are displayed in Table 6. ‘People died’ is ranked first, with a z-value of 9.20, but it does not have the highest lambda. Its count of eight is tied with “many deaths,” “vision zero,” “Rhode Island,” and “New Jersey” for the highest count. “Rhode Island” has the highest lambda, at 12.13. The top 15 collocations have a length of two, whereas the bottom five have a length of three. The bottom five also have an equal count of two. Table 6 includes “bicyclists” deaths,” “vision zero,” “New Jersey,” “according preliminary data,” and “bicycle rates ward.” This is interesting as these collocations do not appear in the national or international news media text corpora. Table 7 shows a notable focus on bicycles within these collocations.

International News Media. Figure 3 shows the text network for the international news media text corpus. It has key features of “fatalities,” “reduction,” “single-vehicle,” “multi-vehicle,” “fewer,” “fatal,” “drop,” “risk,” “urban,” and “reported.” Visually, it appears that “numbers,” “increased,” “increase,” “lockdown,” and “deaths” are the closest to the center. There are a few words lying notably far from the center, including “trend,” “many,” and “NHTSA.” “Pandemic” and “deaths” both have larger nodes relative to the rest. Compared with Figure 2, which seems to have a heavier focus on words such as “deaths” and “fatalities,” Figure

Table 7. Collocation Measures of Bigrams and Trigrams of Local News Media Text Corpus

No.	Collocation	Count	Length	Lambda	z
1	People died	8	2	4.79	9.20
2	National highway	5	2	5.42	9.07
3	Many death	8	2	8.83	8.80
4	Stay-at-home order	5	2	7.51	8.76
5	Outside bike-maps	5	2	8.13	8.73
6	Bicyclist deaths	6	2	4.91	8.51
7	COVID-19 pandemic	5	2	4.64	8.09
8	Number fatalities	5	2	3.43	6.90
9	Vision zero	8	2	10.52	6.63
10	Behind wheel	7	2	10.39	6.53
11	Number deaths	5	2	3.08	6.33
12	Miles hour	7	2	9.29	6.17
13	Rhode Island	8	2	12.13	5.98
14	New Jersey	8	2	8.83	5.96
15	Los Angeles	7	2	12.00	5.90
16	Number cyclist deaths	2	3	2.21	0.99
17	People take precautions	2	3	-0.11	-0.04
18	According preliminary data	2	3	-0.76	-0.37
19	Bicycle rates ward	2	3	-1.41	-0.59
20	Last year's fatal	2	3	-1.30	-0.61

Table 8. Collocation Measures of Bigrams and Trigrams of International News Media Text Corpus

No.	Collocation	Count	Length	Lambda	Z
1	Fatality rate	3	2	5.82	6.71
2	COVID-19 pandemic	3	2	4.65	6.30
3	Public awareness	2	2	6.28	6.02
4	Multi-vehicle urban	2	2	5.69	5.97
5	Urban multi-vehicle	2	2	5.69	5.97
6	Compared pre-pandemic	2	2	7.13	5.88
7	Miles hour	2	2	7.13	5.88
8	Three months	2	2	7.13	5.88
9	Single-vehicle rural	2	2	5.82	5.88
10	Single-vehicle multi-vehicle	2	2	5.24	5.84
11	Late August	2	2	6.79	5.84
12	Metropolitan area	2	2	6.79	5.84
13	Death toll	2	2	7.64	5.80
14	April compared	2	2	5.51	5.71
15	Deaths fell	3	2	4.20	5.64
16	Severity increased pandemic	2	3	0.25	0.09
17	According preliminary data	2	3	-1.10	-0.37
18	Cent drop deaths	2	3	-1.37	-0.64
19	Stay-at-home order began	2	3	-4.42	-1.42
20	Partnership healthy cities	2	3	-5.00	-1.52

3 appears to be more focused on words such as “increase” and “numbers.” Figure 1 has a focus similar to both Figures 2 and 3, including both “increase” and “deaths” near its center. The key features of all three include “fatalities,” but the key features of Figure 3 much more closely resemble the key features of Figure 1 (both containing many of the same words) compared with Figure 2 (only having “fatalities” as the same word).

Similar to Figures 1 and 2, the node size in Figure 3 also relates back to deaths during the pandemic. Just like Figure 2, the word “pandemic” has a larger node size in Figure 3, rather than “coronavirus” (like Figure 1).

Table 8 shows the top 20 collocations for the international news media text corpus. “Fatality rate” is placed as number one, with a z-value of 6.71. “Death toll” has the highest lambda at 7.64. It is notable that all 20 have

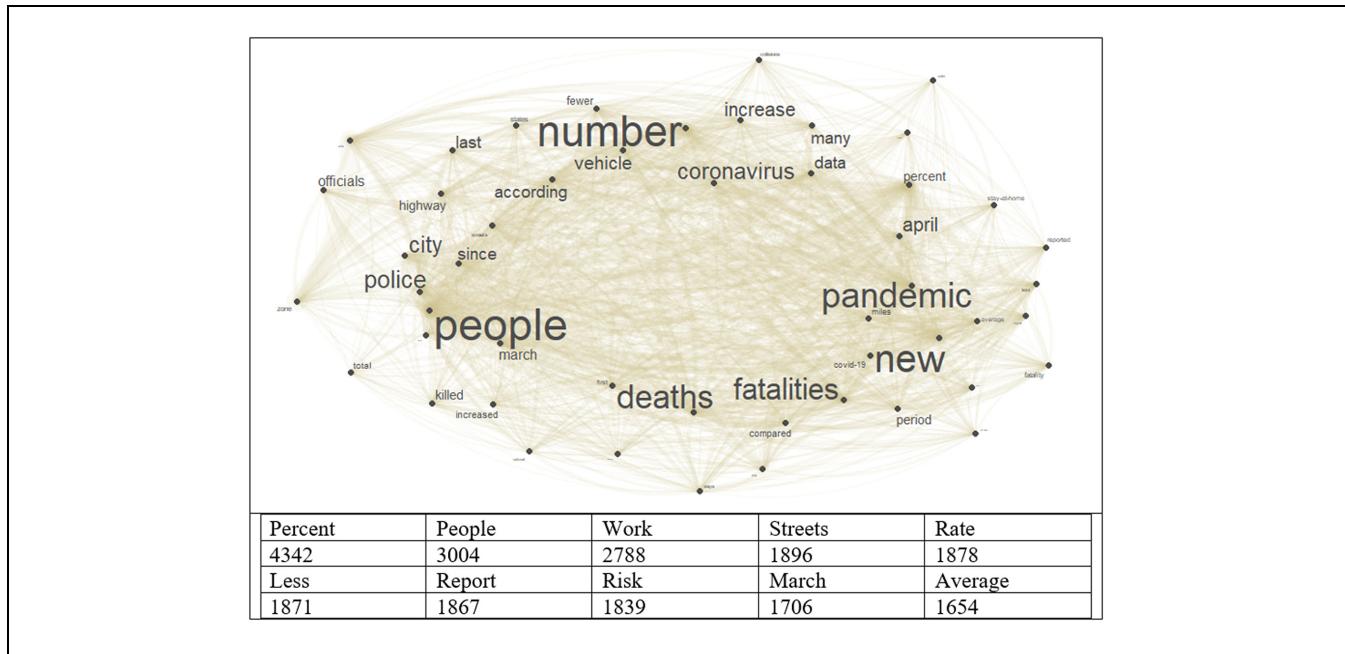


Figure 4. Text network from international news media text corpus in 2020.

counts of either three or two. Only three of the collocations have counts of three (“fatality rate,” “COVID-19 pandemic,” and “deaths fell”), and all the rest have counts of two. The collocations “April compared” and “partnership healthy cities” appear in the international news media text corpus, but not in the national or local corpora.

News Media in 2020. Figure 4 displays the text network for the international news media text corpus in 2020. “Percent,” “people,” “work,” “streets,” “rate,” “less,” “report,” “risk,” “march,” and “average” are listed as the key features. “Pandemic,” “people,” “percent,” “number,” and “data” appear to be the closest to the center of the image. Some words notably far from the center include “order,” “officials,” “national,” “decrease,” “fatal,” and “COVID-19.” The size of the words within this figure indicates node size. “People,” “number,” “pandemic,” “new,” “deaths,” and “fatalities” have the largest node sizes. Similar to the figures of text networks relating to the different news media, the node sizes of this figure seem to relate back to the number of deaths arising from the COVID-19 pandemic.

Table 9 lists the top 20 collocation measures of bigrams and trigrams of the 2020 news media text corpus, organized by count, length, lambda, and z-value. “Fatality rate” is the first collocation, with a z-value of 12.17. It also has the highest count. “Pedestrians cyclists” has the highest lambda value of 8.18. The last five collocations have significantly lower z-values than the top 15, each of them with z-values less than or equal to 1.48 (the

top 15 have z-values greater than or equal to 8.95); this is because of the length changing from two (for the top 15 collocations) to three (for the bottom five collocations). It is interesting to note that “pedestrians cyclists,” “drugs alcohol,” “fewer vehicles April,” and “risk April fewer” appear in Table 9. These seem to be related to the events of 2020, such as the COVID-19 lockdown and how there were fewer drivers on the road.

News Media in 2021 (January to July 2021). The text network for the 2021 international news media text corpus is displayed in Figure 5. It has key features of “fatalities,” “fatal,” “city,” “fewer,” “increased,” “percent,” “killed,” “period,” “miles,” and “pedestrian.” Terms that appear closest to the center include “pandemic,” “people,” “data,” “fatalities,” and “deaths.” “Deaths,” “people,” “pandemic,” “data,” and “increase” have the largest nodes. Compared with Figure 4, the terms closer to the center: “pandemic,” “people,” “fatalities,” and “data” repeat. However, Figure 5 seems to have a greater focus on death, as it includes the terms “fatalities” and “deaths” close to the center, whereas Figure 4 only includes “fatalities.” As to key features, only the term “percent” is repeated in Figures 4 and 5. As mentioned above, Figure 5 once again seems to have a greater focus on fatalities. Similar to the node sizes in Figure 4, the node size in Figure 5 appears to relate back to the number of deaths from the pandemic. However, Figure 5 has the word “increase,” which may indicate discussions of a rising number of cases.

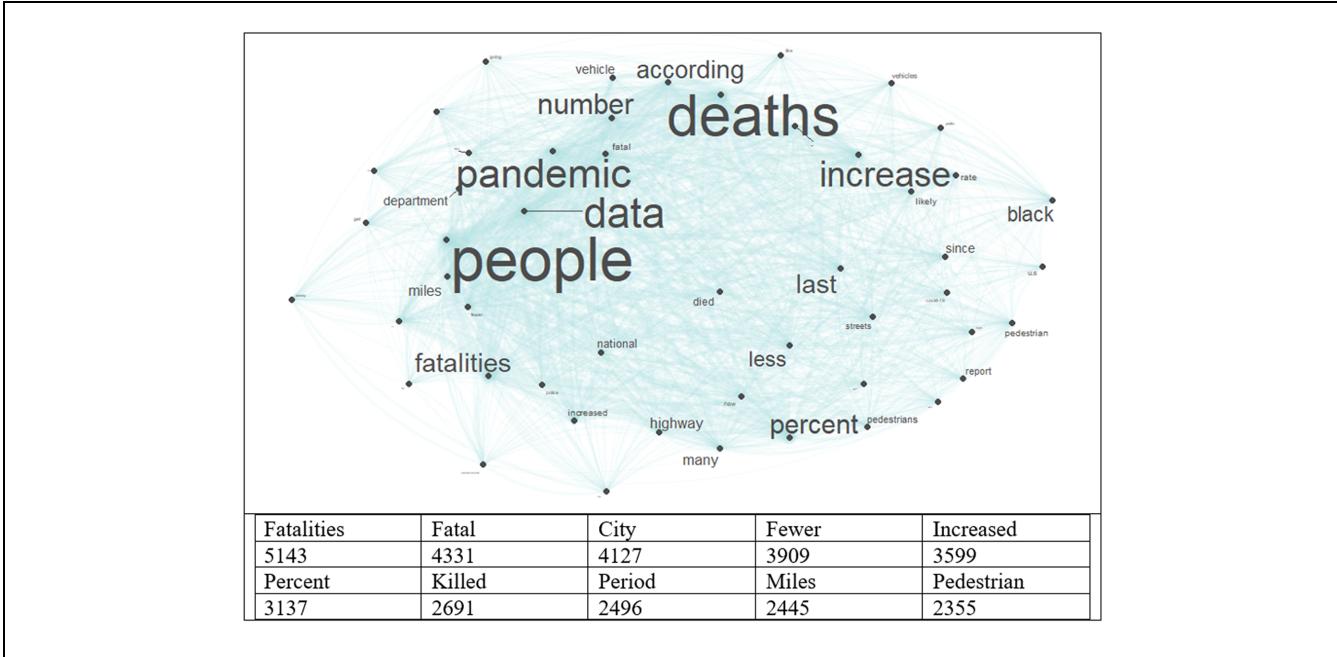


Figure 5. Text network from international news media text corpus in 2021.

Table 9. Collocation Measures of Bigrams and Trigrams of 2020 News Media Text Corpus

No.	Collocation	Count	Length	Lambda	z
1	Fatality rate	10	2	5.82	12.17
2	Stay-at-home order	9	2	6.01	11.80
3	York City	8	2	6.03	11.14
4	First half	6	2	6.00	10.12
5	Vehicle miles	7	2	4.83	10.06
6	Work zones	7	2	6.54	9.60
7	People died	9	2	4.91	9.41
8	National highway	6	2	4.74	9.36
9	Miles driven	6	2	6.54	9.33
10	Reduce risk	5	2	5.65	9.14
11	Pedestrian cyclist	5	2	8.18	9.12
12	Urban multi-vehicle	4	2	6.38	9.08
13	Two weeks	4	2	6.13	9.03
14	Number deaths	9	2	3.47	9.00
15	Drugs alcohol	4	2	6.64	8.95
16	Fewer vehicles April	2	3	3.78	1.48
17	Increased less miles	2	3	3.54	1.39
18	Factors increased fatalities	2	3	3.21	1.25
19	Risk April fewer	2	3	2.55	1.15
20	Multi-vehicle fewer vehicles	2	3	2.65	1.03

The top 20 collocation measures of bigrams and trigrams of a 2021 news media text corpus, organized by count, length, lambda, and z-values, are covered in Table 10. The first collocation is “national highway,” with a z-value of 15.48. Notably, it has a z-value that is 2.51 higher than the second collocation (“COVID-19

pandemic”), which has a z-value of 12.97. “People died” has the highest count, and “behind wheel” has the highest lambda value. Compared with the top 15 collocations, the last five have significantly lower z-values, each having z-values less than or equal to 1.94, while the top 15 have z-values greater than or equal to 9.71. This is

Table 10. Collocation Measures of Bigrams and Trigrams of 2021 News Media Text Corpus

No.	Collocation	Count	Length	Lambda	z
1	National highway	17	2	5.92	15.48
2	COVID-19 pandemic	14	2	4.85	12.97
3	People died	18	2	5.71	12.82
4	Fatality rate	9	2	5.99	12.43
5	Across country	8	2	6.69	12.16
6	Highway administration	16	2	7.83	11.14
7	Percent increase	11	2	3.73	10.64
8	Behind wheel	9	2	9.41	10.58
9	Stay home	7	2	7.35	10.57
10	Americans drove	6	2	7.94	10.48
11	Law enforcement	7	2	7.12	10.38
12	Communities color	5	2	7.28	10.20
13	Miles driven	7	2	5.61	10.10
14	Stay-at-home order	6	2	7.94	9.85
15	Preliminary data	7	2	4.79	9.71
16	Early pandemic last	2	3	4.94	1.94
17	Percent drop fatal	2	3	4.70	1.85
18	Despite fewer pandemic	2	3	4.65	1.82
19	Pedestrian percent likely	2	3	4.63	1.82
20	Increase reported fatalities	2	3	3.64	1.65

because of the word length changing from two to three. Some interesting collocations appear in Table 10, including “behind wheel,” “law enforcement,” and “communities color.” Compared with 2020, these collocations appear to be diversified with the inclusion of topics such as people of color and law enforcement.

Topic Modeling Results

This study then created visualizations for these LDA themes using LDAvis, a free visualization tool for topic modeling (38). Additionally information was compiled from the dataset (e.g., signature counts and petition creation dates) and Google Trends to interpret the topic. Based on the corpora size, 20 topics were developed for national news media content and 10 topics were developed for local and international news media contents separately. The illustrations (Figures 6 to 8) and explanations provided in this study are related to relative locations of the topic models and explanation on the words presented in topic 1. The researchers provided interactive tool URLs for additional exploration. Discussion on other topics is not provided in this study.

Topics From National News Media. This study labeled each topic according to the LDAvis visualization results to facilitate interpretation and additional analysis. For the topics from national news media (Figure 6), the words in each topic were sorted in descending order based on the estimated term frequency within the selected topic (red bars), which informs topic words that are highly relevant

to the specific topic. A term’s relevance to the topic is determined by the weight parameter λ . $\lambda = 0.6$ was used for the topic words, and the width of the blue bar indicates the “corpus-wide frequencies of each term” and the width of the red bar represents “the topic-specific frequencies of each term.” After extracting the 30 topics, labels are selected from the top 10 topic words displayed by LDAvis (relevance parameter $\lambda = 0.6$) and also by considering semantic meaningfulness. The top 10 relevant terms are shown to the left of Figure 6, including “traffic,” “drivers,” “crashes,” “safety,” “pandemic,” “people,” “roads,” “miles,” “data,” and “increase.” Topic 1 in the national news media topics completely overlaps with topics 12 and 19 and partially overlaps with topics 6, 7, 11, and 14. Additionally, topics 2 and 3 and topics 3 and 4 overlap with each other. The remaining topics are completely independent. The interactive tool can be found at: https://subashis.github.io/pages/lDavis/spd_cr_cvd/national.

Topics From Local News Media. As before, the topic words were sorted in descending order based on the estimated term frequency within the selected topic (red bars in Figure 7), which identifies topic words that are highly relevant to the specific topic. The topic terms depicted in Figure 7 are the result of employing $\lambda = 0.6$, an ideal value recommended in the literature (38). The breadth of the blue bar shows the “corpus-wide frequency of each term,” while the width of the red bar indicates “the topic-specific frequency of each term.” In Figure 7, the red bars for “year” and “traffic” are nearly entirely red, with little

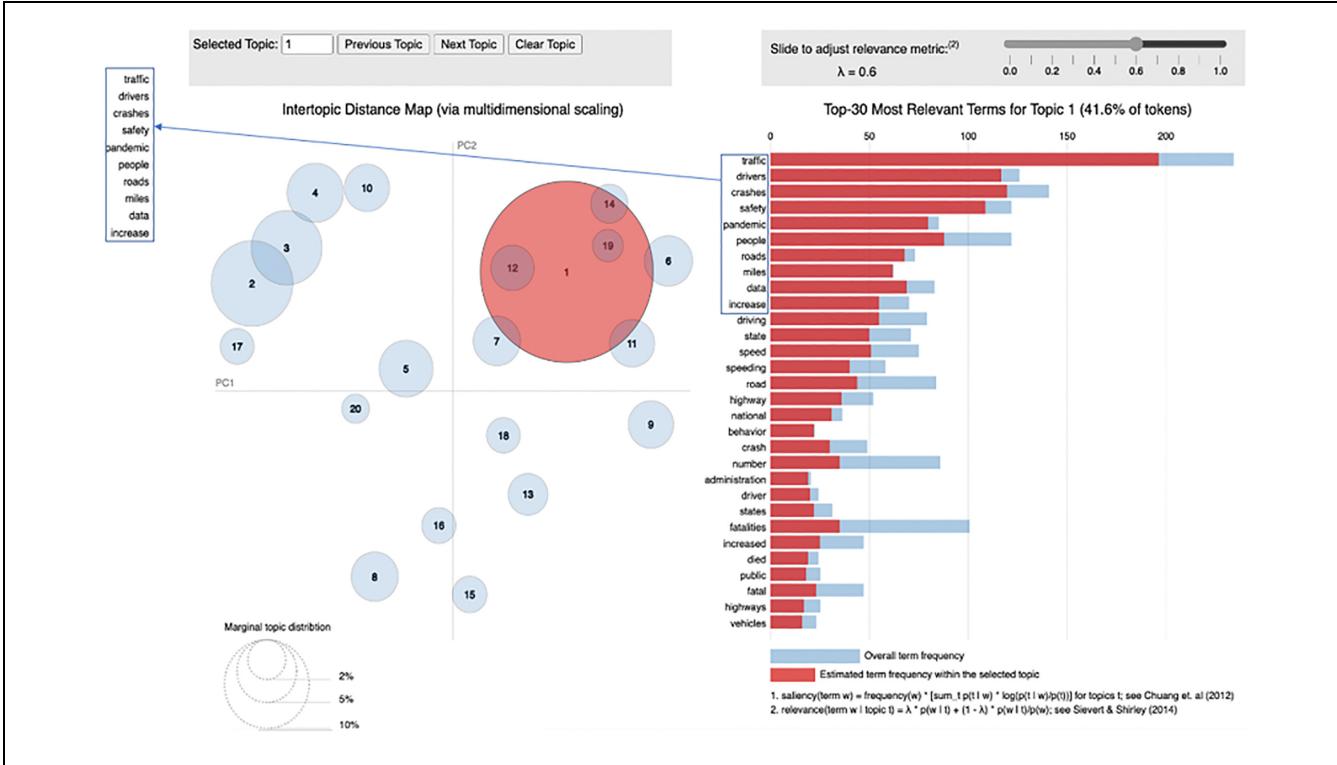


Figure 6. LDAvis results acquired using $\lambda = 0.6$ (national media).

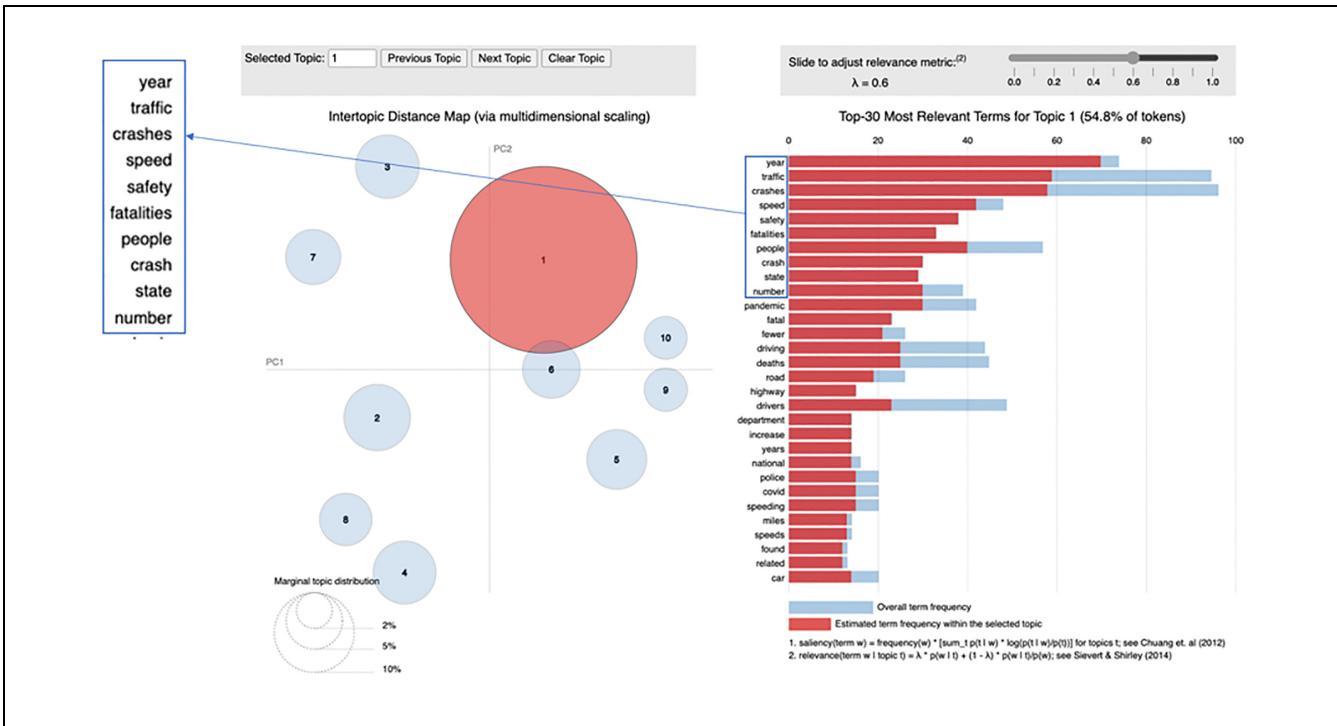


Figure 7. LDAvis results acquired using $\lambda = 0.6$ (local media).



Figure 8. LDAvis results acquired using $\lambda = 0.6$ (international media).

blue bar visible, indicating that these variables are highly reflective of the local new media outcomes. When using $\lambda = 0.6$ in Figure 7, these two terms represent the local news media outcomes as the first and second most significant terms. After extracting the 30 subjects, labels are chosen from the top 10 topic terms revealed by LDAvis (relevance parameter $\lambda = 0.6$) while taking semantic significance into account. The top 10 relevant terms for the local results are shown on the left of Figure 7, containing terms such as “year,” “traffic,” “crashes,” “speed,” “safety,” “fatalities,” “people,” “crash,” “state,” and “number.”

The sizes of the circles (on the left side of Figure 7, which depicts the global topic view) “are proportional to the relative prevalence of the topics in the corpus” (38). According to the circle size of Figure 7, topic 1 is frequent in around 20% of the corpus, whereas topic 8 is prevalent in approximately 2% of the corpus. According to a study conducted by Hagen (39), the largest and smallest themes are frequently difficult to interpret because they contain a variety of topics. Additionally, the distance between topics represents their semantic distance. Topic 1 is demonstrated to be largely independent, despite some overlap with topic 6. The remaining topics are independent. For the usability evaluation, this study developed software with interactive features (snapshots

of the package are in Figures 6, 7, and 8). The LDAvis results demonstrate the contextual richness of topic modeling results by revealing the topological position of the subjects and their relationships to other closely related topics. In addition, the red bar on the right indicates the relevance of each term inside the topic. The interactive tool can be found at: https://subashish.github.io/pages/l davis/spd_cr_cvd/local/.

Topics From International News Media. For the international topics (Figure 8), $\lambda = 0.6$ was again used for the topic words. As before, the topic words were sorted in descending order based on the estimated term frequency within the selected topic (red bars), which identifies topic words that are highly relevant to the specific topic. After extracting the 30 topics, labels are selected from the top six topic words displayed by LDAvis (relevance parameter $\lambda = 0.6$) and also by considering semantic meaningfulness.

The top six relevant terms are shown on the left of Figure 3, being “traffic,” “pandemic,” “people,” “road,” “increase,” and “drivers.” In the international topics, topic 1 is completely independent. Topics 4, 9, and 6 overlap, and topics 3 and 5 overlap. The remaining topics are independent. The interactive tool can be found at: https://subashish.github.io/pages/l davis/spd_cr_cvd/international/.

Key Findings

Reviewing the news reports as well as performing a series of text mining techniques on textual datasets helps researchers to understand international, national, and local concerns and safety trends across the world.

- Comparing the traffic safety-related news coverage during the pandemic situation in 2020 to 2021 reveals that safety concerns were quite different in these two years. Particularly, in 2020 news coverage frequently reported on the “empty streets” and drivers’ responses to the quieter roads. Although a few articles highlighted the surge in car crashes with a higher probability of “fatality” because of “reckless driving,” hardly any report pointed out other indirect impacts of shelter-in-place ordinances. On the other hand, in 2021, accepting the impact of stay-at-home orders on more severe crashes, most discussed the responses to the new situation and even proposed methods to control the rise in fatal crashes, such as more serious law enforcement actions.
- Reviewing local news shows that crashes related to “pedestrians” and “bicyclists” were covered more often than in national and international coverages. In local news, every fatality, and, more specifically, each crash, is important. Local news agencies report the details related to the crashes and interview police and local government officials to investigate the rise in fatal crashes. Moreover, it seems local news found it more helpful to alert residents about the higher rate of severe crashes during the pandemic, since people might still consider emptier roads to be safer as well.
- National news media, unlike local and international coverage, specifically uncovered the inequitable distribution of traffic incidents among community groups; for instance, they particularly mentioned that Black people were involved in a higher rate of crashes or the specific concerns about elderly commuters or low-income populations. Moreover, “alcohol/drug” related crashes are the ones that national news mostly warned about. Some reports have considered that assumptions about lower “police” presence and “law enforcement” on road networks made drivers attempt to break the law. There is a need to investigate the potential reason for such driving behavior during the lockdowns.
- International news media, on the other hand, mostly reported the basic trends and the “new normal” situation to increase “public awareness” through the news. It is also interesting that traffic

crash fatalities occurred at different rates around the world.

- Compared with 2020, the word groups in 2021 appear to be more diversified with the inclusion of topics such as people of color and law enforcement.
- This study developed three topic modeling-related interactive tools which can be used to understand the trends of different topics by media types. The words in each topic and relevance of locations of the topics can be further explored for difference and similarities of news media reporting on this public health concern.

Conclusion

News is the reporting of selected information on recent happenings, where the selection process is made based on the level of newsworthiness and how it interests the public. News is a reliable source used by people for discovering current events and it can also help academics to explore general public perceptions toward recent happenings, as news coverage of events is likely to affect public opinion toward the topics and form people’s point of view about the events. Moreover, news coverage intensely explores social trends and mostly offers legitimate solutions for recent events. Therefore, analyzing news articles seems a critical first step in investigating social phenomena and recent trends, as they are important to understanding human perception and finding potential solutions for a certain issue. However, the information era has provided an extremely high number of articles offered to the public. Moreover, textual articles often “lack metadata such as topic, semantic structure, and labels” (40) and frequently include noise. This makes textual news articles a very difficult information source to use for analysis and research. News articles are content-rich documents for investigating qualitative information using descriptive text mining. However, not many studies, specifically in the field of transportation, have utilized this approach to investigate news articles.

The COVID-19 pandemic has affected lives from every imaginable perspective. It also provided some opportunities to explore the possible impacts from different viewpoints. Understanding the impacts of the pandemic could help mitigate the negative effects of the ongoing pandemic and help prepare for the next one. This paper explains how news media have shaped the mass-media coverage of “quieter roads, over speeding, and more crashes.” This study posits that when examining impediments of speed-crash-volume from media coverage, research must critically examine the intent of the content published in news media. This study collected 57 relevant news articles from various sources (local,

national, and international) by web scraping. A manual content analysis was performed. Additionally, text mining was performed by applying text network analysis. The results show that news coverage was mostly about the surge in more reckless driving and fatal crashes, although different rates were reported from around the world. Local news often provides more fine-grained coverage of traffic crashes, particularly pedestrian and bicycle-related crashes. In 2021, news agencies mostly shifted from just reporting the crash number trends to informing about the indirect consequences of the “new normal” for the lives of disadvantaged populations. In addition, this study developed three interactive tools for the topic models developed from textual contents by media type. These tools can assist the people or agencies of interest in exploring the overall contexts of the sudden increase of speeding-related traffic crashes and crash associated severities and injuries.

The current study has some limitations. First, the data collection was limited. Additional news articles can provide additional information. Future studies can explore this issue by including more relevant news media sources. Second, the content analysis can be extended to qualitative analysis using advanced tools such as NVIVO. Third, the text analysis was limited to text network and LDA, a commonly used topic model. Future studies can explore advanced topic modeling (e.g., structural topic model) and language model (e.g., Bidirectional Encoder Representations from Transformers [BERT]) to provide latent insights from unstructured textual contents.

Acknowledgments

The authors would like to thank three anonymous reviewers who provided critical and rigorous comments. The current version of the paper is much improved as a result of the suggested changes that were made in responding to the queries of the reviewers.

Author Contributions

The authors confirm the contribution to the paper as follows: study conception and design: S. Das; data collection: S. Das; analysis and interpretation of results: S. Das; draft manuscript preparation: S. Das, S. Sarkar. Both authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Subasish Das  <https://orcid.org/0000-0002-1671-2753>
Sobhan Sarkar  <https://orcid.org/0000-0001-6040-6051>

References

1. Mohler, G., A. L. Bertozzi, J. Carter, M. B. Short, D. Sledge, G. E. Tita, C. D. Uchida, and P. J. Brantingham. Impact of Social Distancing During COVID-19 Pandemic on Crime in Los Angeles and Indianapolis. *Journal of Criminal Justice*, Vol. 68, 2020, p. 101692.
2. COVID-19 - Mobility Trends Reports. *Apple*. <https://www.apple.com/covid19/mobility>. Accessed October 25, 2020.
3. Zuo, F., J. Wang, J. Gao, K. Ozbay, X. J. Ban, Y. Shen, H. Yang, and S. Iyer. An Interactive Data Visualization and Analytics Tool to Evaluate Mobility and Sociability Trends During COVID-19. *arXiv Preprint arXiv:2006.14882*, 2020.
4. NHTSA. Fatality Analysis Reporting System (FARS). <https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars>.
5. Inada, H., L. Ashraf, and S. Campbell. COVID-19 Lockdown and Fatal Motor Vehicle Collisions due to Speed-Related Traffic Violations in Japan: A Time-Series Study. *Injury Prevention*, Vol. 27, No. 1, 2021, pp. 98–100.
6. Katrakazas, C., E. Michelarakis, M. Sekadakis, and G. Yannis. A Descriptive Analysis of the Effect of the COVID-19 Pandemic on Driving Behavior and Road Safety. *Transportation Research Interdisciplinary Perspectives*, Vol. 7, 2020, p. 100186.
7. Gao, J., J. Wang, Z. Bian, S. D. Bernardes, Y. Chen, A. Bhattacharyya, S. S. M. Thambiran, K. Ozbay, and S. Iyer. The Effects of the COVID-19 Pandemic on Transportation Systems in New York City and Seattle, USA. *arXiv Preprint arXiv:2010.01170*, 2020.
8. Aarts, L., and I. Van Schagen. Driving Speed and the Risk of Road Crashes: A Review. *Accident Analysis & Prevention*, Vol. 38, No. 2, 2006, pp. 215–224.
9. Qureshi, A. I., W. Huang, S. Khan, I. Lobanova, F. Siddiq, C. R. Gomez, and M. F. K. Suri. Mandated Societal Lockdown and Road Traffic Accidents. *Accident Analysis & Prevention*, Vol. 146, 2020, p. 105747.
10. Doucette, M. L., A. Tucker, M. E. Auguste, A. Watkins, C. Green, F. E. Pereira, K. T. Borup, D. Shapiro, and G. Lapidus. Initial Impact of COVID-19's Stay-at-Home Order on Motor Vehicle Traffic and Crash Patterns in Connecticut: An Interrupted Time Series Analysis. *Injury Prevention*, Vol. 27, No. 1, 2021, pp. 3–9.
11. Pishue, B. *COVID-19 Effect on Collisions on Interstates and Highways in the US*. INRIX, Kirkland, WA, 2020.
12. Shilling, F., and D. Waetjen. Impact of COVID19 Mitigation on California Traffic Crashes. University of California, Davis, CA, 2020.
13. Qureshi, A., W. Huang, S. Khan, I. Lobanova, F. Siddiq, C. Gomez, and M. Suri. Mandated Societal Lockdown and Road Traffic Accidents. *Accident Analysis & Prevention*, Vol. 146, 2020.
14. Dong, X., K. Xie, and H. Yang. How Did COVID-19 Impact Driving Behaviors and Crash Severity? A

- Multigroup Structural Equation Modeling. *Accident Analysis & Prevention*, Vol. 172, 2022, p. 106687.
15. Hosseini, P., M. Jalayer, S. Das, and H. Zhou. Identifying Wrong-Way Driving (WWD) Crashes in Police Reports Using Text Mining Techniques. Presented at 100th Annual Meeting of the Transportation Research Board, Washington, D.C., 2021.
16. Das, S., X. Sun, and A. Dutta. Text Mining and Topic Modeling of Compendiums of Papers From Transportation Research Board Annual Meetings. *Transportation Research Record: Journal of the Transportation Research Board*, 2016. 2552: 48–56.
17. Wen, T., Q. Gao, Y. Chen, and K. H. Cheong. Exploring the Vulnerability of Transportation Networks by Entropy: A Case Study of Asia–Europe Maritime Transportation Network. *Reliability Engineering & System Safety*, Vol. 226, 2022, p. 108578. <https://doi.org/10.1016/j.ress.2022.108578>.
18. Das, S., K. K. Dixon, X. Sun, A. Dutta, and M. Zupanich. Trends in Transportation Research: Exploring Content Analysis in Topics. *Transportation Research Record: Journal of the Transportation Research Board*, 2017. 2614: 27–38.
19. Das, S., A. Dutta, G. Medina, L. Minjares-Kyle, and Z. Elgart. Extracting Patterns From Twitter to Promote Biking. *IATSS Research*, Vol. 43, No. 1, 2019, pp. 51–59.
20. Rose, R. L., T. G. Puranik, D. N. Mavris, and A. H. Rao. Application of Structural Topic Modeling to Aviation Safety Data. *Reliability Engineering & System Safety*, Vol. 224, 2022, p. 108522. <https://doi.org/10.1016/j.ress.2022.108522>.
21. Das, S., A. Dutta, T. Lindheimer, M. Jalayer, and Z. Elgart. YouTube as a Source of Information in Understanding Autonomous Vehicle Consumers: Natural Language Processing Study. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 242–253.
22. Das, S. Exploratory Analysis of Unmanned Aircraft Sightings Using Text Mining. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 291–300.
23. Kutela, B., S. Das, and B. Dadashova. Mining Patterns of Autonomous Vehicle Crashes Involving Vulnerable Road Users to Understand the Associated Factors. *Accident Analysis & Prevention*, Vol. 165, 2022, p. 106473.
24. Das, S. Understanding Fatal Crash Reporting Patterns in Bangladeshi Online Media Using Text Mining. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 960–971.
25. Wang, Y., Z. He, and J. Hu. Traffic Information Mining From Social Media Based on the MC-LSTM-Conv Model. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 2, 2022, pp. 1132–1144. <https://doi.org/10.1109/TITS.2020.3021096>.
26. Wu, L., K. Ye, and M. Hastak. A Comparison Study on Environmental Policies for Expressway Construction Projects Between China and the US: A Tiered Analysis Approach. *Journal of Environmental Management*, Vol. 305, 2022, p. 114298.
27. Das, S., A. Dutta, and I. Tsapakis. Topic Models From Crash Narrative Reports of Motorcycle Crash Causation Study. *Transportation Research Record: Journal of the Transportation Research Board*, 2021. 2675: 449–462.
28. Luo, S., and S. Y. He. Understanding Gender Difference in Perceptions Toward Transit Services Across Space and Time: A Social Media Mining Approach. *Transport Policy*, Vol. 111, 2021, pp. 63–73.
29. Das, S., and M. Theel. Pandemic and Transportation Research: Bibliometric Analysis and Topic Modeling. 2021. Presented at 100th Annual Meeting of the Transportation Research Board, Washington, D.C., 2021.
30. Das, S., A. Dutta, and K. Fitzpatrick. Technological Perception on Autonomous Vehicles: Perspectives of the Non-Motorists. *Technology Analysis & Strategic Management*, Vol. 32, No. 11, 2020, pp. 1335–1352.
31. Sadri, A. M., S. Hasan, S. V. Ukkusuri, and M. Cebrian. Crisis Communication Patterns in Social Media During Hurricane Sandy. *Transportation Research Record: Journal of the Transportation Research Board*, 2018. 2672: 125–137.
32. Mimno, D. Package “Mallet”. 2013. <https://cran.r-project.org/web/packages/mallet/mallet.pdf>. Accessed October 9, 2015.
33. Blei, D. M. Probabilistic Topic Models. *Communications of the ACM*, Vol. 55, No. 4, 2012, pp. 77–84.
34. Rampin, R., V. Steeves, and S. De Mott. *Taguette Zenodo*. Version 0.9.2. 2020. <https://app.taguette.org/>. Accessed September 4, 2021.
35. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. November 17, 2018. 2020. <https://www.r-project.org/>.
36. Benoit, K., K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo. quanteda: An R Package for the Quantitative Analysis of Textual Data. *Journal of Open Source Software*, Vol. 3, No. 30, 2018, p. 774. <https://doi.org/10.21105/joss.00774>.
37. Csardi, G. Igraph. Network Analysis and Visualization. 2020. <https://igraph.org/>
38. Sievert, C., and K. E. Shirley. LDAsvis: A Method for Visualizing and Interpreting Topics. *Proc., Workshop on Interactive Language Learning, Visualization, and Interfaces*, Association for Computational Linguistics, Baltimore, MD, 2014, pp. 63–70.
39. Hagen, L. *Topic Modeling for e-Petition Analysis: Interpreting Petitioners’ Policy Priorities*. PhD dissertation. State University of New York, Albany, 2016.
40. De Fortuny, E. J., T. De Smedt, D. Martens, and W. Daelemans. Media Coverage in Times of Political Crisis: A Text Mining Approach. *Expert Systems With Applications*, Vol. 39, No. 14, 2012, pp. 11616–11622.