Tristan Ford

Public Transit and Big Data: Reconciling the Opposing Demands of Mobility and Safety

2024-01-04

Abstract: The ongoing influx of big data in transportation is enhancing the standard transit agency's capacity to optimize its network of routes. This optimization encompasses crucial elements such as route design, stop locations, and service frequency resulting in increased mobility for transit users. Recent research, as cited in Verbavatz and Barthelemy (2019), highlights that augmenting public transit density ranks among the most effective strategies for reducing congestion and addressing other related factors. Additionally, evidence supports the notion that the presence of effective public transit correlates with improved road safety (Schepers et al., 2019).

Big data, beyond its role in optimizing transit networks, proves instrumental in monitoring road safety. The very mechanism enhancing mobility can be employed to monitor and estimate changes in road safety indicators. Proposing that both mobility and safety benefits may be realized from innovative policy and design considerations resultant from applications of big data, this study embarks on a literature review to explore this complex relationship. Furthermore, an empirical Bayes before-and-after analysis scrutinizes the road safety effects stemming from the implementation of RapidBus in Vancouver. Our findings, while not statistically significant, indicate an improvement in safety along the alignment of the R5 Hastings RapidBus route. These conclusions, drawn from observed collision statistics from ICBC, align with existing literature, highlighting the potential for big data to simultaneously enhance mobility and improve road safety through public transport innovations.

Contents

1 Introduction

In the realm of transportation, road safety and mobility emerge as inherently conflicting requirements. The essence of this dichotomy lies in the undeniable fact that if human mobility were unnecessary, road collisions would cease to exist. However, reality dictates otherwise, as individuals continually traverse their communities for work and leisure among other purposes. As professionals in transportation engineering and planning, our challenge is to harmonize these conflicting demands within our networks. The populace seeks travel options characterized by minimal friction and generalized cost, all while ensuring the paramount element of safety.

1.1 Public Transit and Road Safety

Road accidents stand as the primary cause of death for individuals aged 5-29, claiming an estimated 1.35 million lives annually, with 93% of these tragedies occurring in low- to middle-income countries (WHO, 2022). Beyond the human toll, road fatalities exact a considerable economic cost, averaging 3% of a country's gross domestic product each year. Recognizing the societal and economic advantages of enhancing road safety, it was proposed during the freeway era's inception in the 1930s to incorporate urban form considerations into transportation network development. However, the predominant focus on mobility led to the proliferation of auto-centric, sprawling networks in developed countries, making substantial safety improvements challenging (Schepers et al., 2019).

Emerging studies, such as Truong and Currie (2019), reveal that public transportation can be a catalyst for enhancing road safety. In their investigation of Melbourne, Australia, a mere 1% shift of commuters to buses yielded a remarkable 10% reduction in collisions. Notably, the benefits extend beyond safety, encompassing improvements in air quality, noise pollution, and congestion, all contributing to overall public health. Figure 1 underscores this correlation, depicting lower per capita fatalities in countries and cities with transit and active transportation-centric communities (Schepers et al., 2019). However, it is essential to acknowledge that some studies, including Vecino-Ortiz and Hyder (2015), caution against isolating public transportation investments as a panacea. While public transportation is generally safer for passengers, it can elevate risks for external road users, as illustrated in figure 2 (Schepers et al., 2019). For instance, research on bus rapid transit (BRT) projects reveal that while total collisions often decrease post-BRT implementation, issues such as heightened pedestrian collisions near stations persist (Vecino-Ortiz & Hyder, 2015).

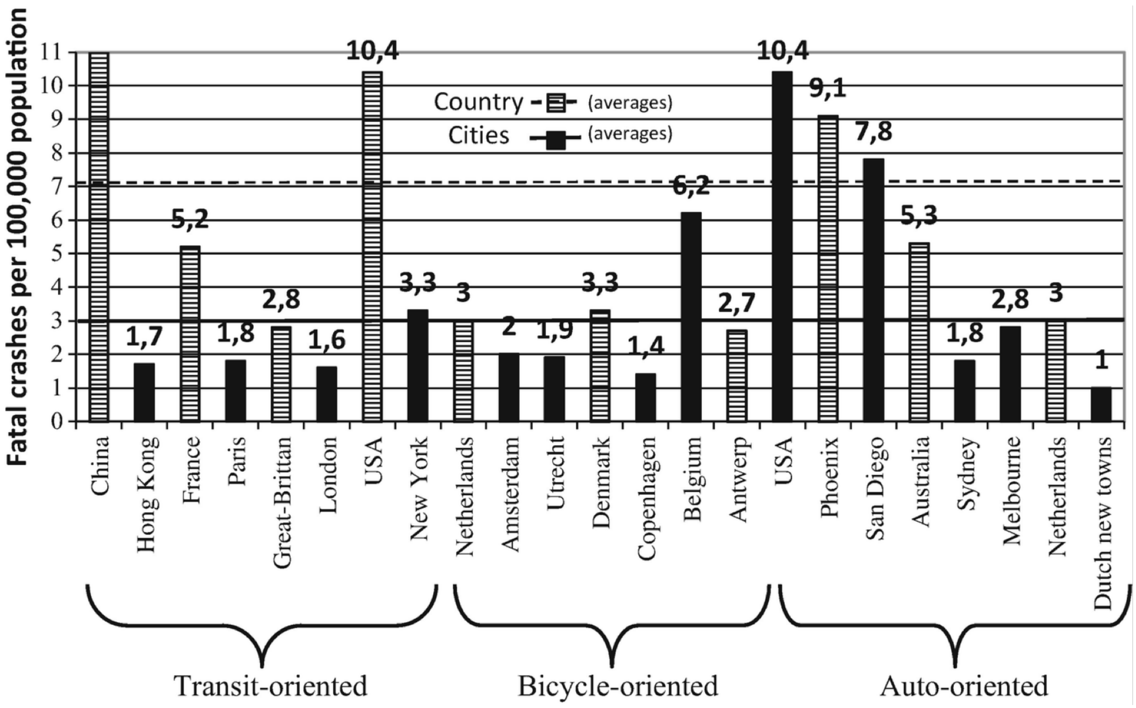


Figure 1: Road fatalities per capita in various countries and cities (Schepers et al., 2019).

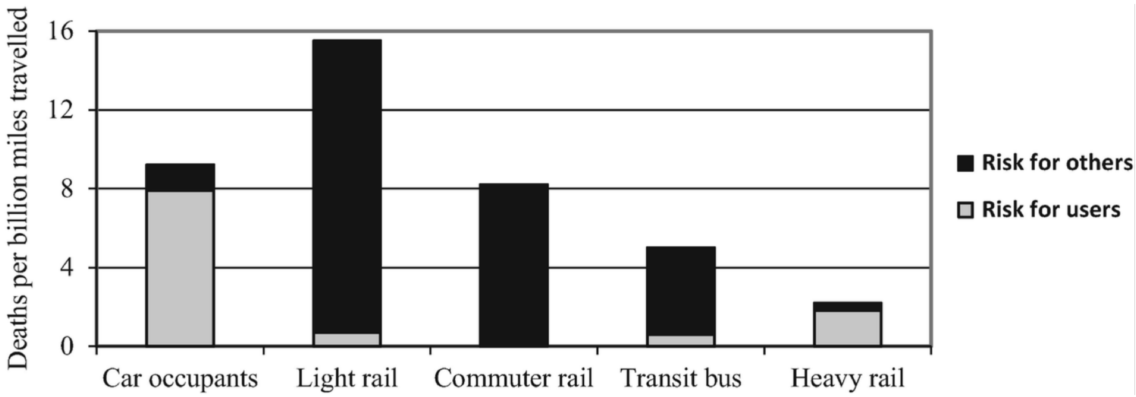


Figure 2: Safety risk for users and non-users of public transportation (Litman, 2015).

This nuanced landscape underscores the insufficiency of standalone public transportation investments in improving safety. However, the Safe System approach, a comprehensive methodology that substantially reduced fatalities in the Netherlands during the 1990s, offers a promising avenue for road safety enhancement. Cities in the Netherlands, designed with this approach in mind, exemplify low road fatalities and have been successfully applied in public transportation contexts (Schepers et al., 2019). By applying robust methodologies like the Safe System approach to public transport projects, we can anticipate significant strides in safety improvement.

1.2 Public Transit and Big Data

Having acknowledged the potential of public transportation in reducing road collisions, our focus extends to enhancing overall mobility. Litman (2015) underscores public transit's role as 'basic mobility,' providing essential services to equity-seeking groups and enhancing societal mobility. Public transportation, for example, presents a solution for increasing road capacity without the need for major infrastructure changes, promising improved mobility in our communities.

The historical practices of public transportation planning, constrained by limited resources and technologies, heavily relied on manually collected surveys and measurements. These traditional methods, while informative, may suffer from narrow temporal/spatial ranges and sample bias. Furthermore, these collection methods were typically designed with a specific policy question in mind (Welch & Widita, 2019). Transport planning was consequently bound to large, long-term infrastructure investments, aligning with aggregated trends that often overlooked temporal and spatial intricacies in the network.

In the era of technological advancements, particularly the Internet of Things (IoT) and connected systems, the emergence of 'big data' has revolutionized transportation. Figure 3 illustrates the rising prominence of big data since 2012, encapsulated by the definition from Zannat and Choudhury (2019) as any data beyond the confines of an Excel spreadsheet. In the transportation context, big data is often passively generated and without specific questions in mind. Connected vehicle sensors, mobile phones, smart cards, and automated systems contribute to a wealth of transportation information, enabling the extraction of dynamic trends and insights beyond the scope of traditional methods (Zannat & Choudhury, 2019). However, it's crucial to note the privacy concerns associated with such data sources, emphasizing the importance of caution and due process.

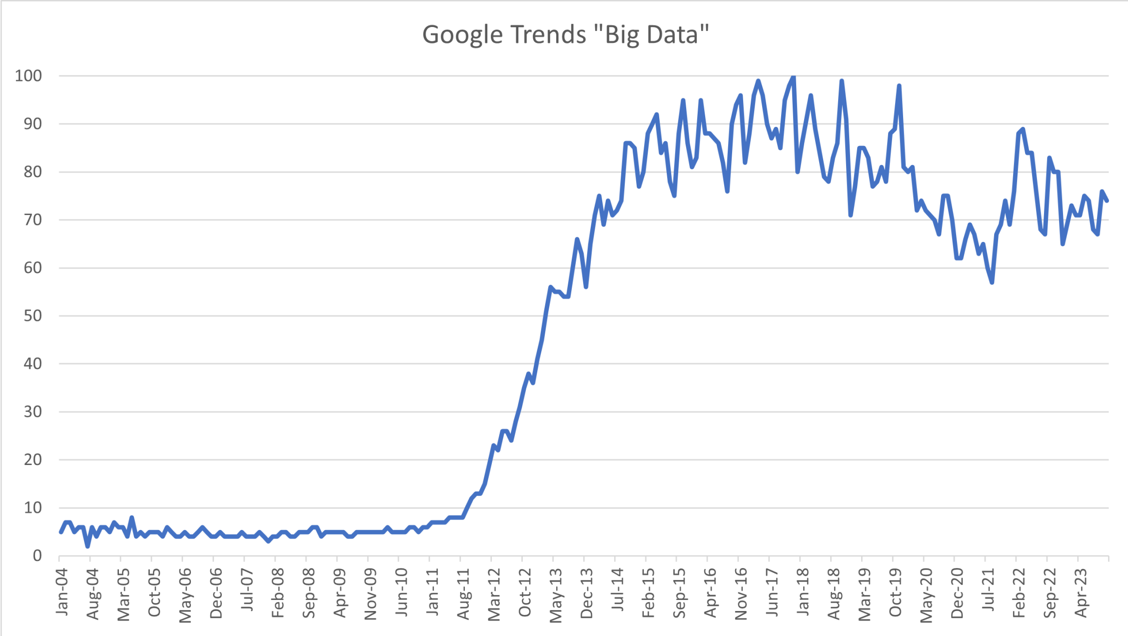


Figure 3: Google Trends search relevance for 'big data.'

The application of big data in transportation studies, particularly since 2013, has catalyzed improvements across various public transportation aspects. These include service/performance measures, accessibility, user behaviors, demand estimation, mode choice, facility and data management, and health and safety (Welch & Widita, 2019). Noteworthy advancements include the estimation of origin-destination matrices, enabling service improvements for transit agencies through systems like automatic vehicle location (AVL) and automated passenger counters (APC). Predictive maintenance techniques, fueled by big data, have enhanced maintenance schemes, subsequently reducing operational costs (Welch & Widita, 2019). Despite this progress, a gap exists in research directly assessing road safety impacts concerning public transit services, emphasizing the need for further exploration (Welch & Widita, 2019; Zannat & Choudhury, 2019). The continued integration and analysis of big data in public transportation stand as imperative steps toward optimizing safety, efficiency, and overall effectiveness in urban mobility.

1.3 Research Objective

Having summarized the intricate connections among big data, public transportation, and road safety, a noticeable gap in the literature pertains to a scarcity of studies examining the direct impact of public transportation services on road safety. To bridge this gap, our research considers a case study in which public transport influences road safety. Our investigation centers on the East Hastings corridor in Vancouver, BC, where the introduction of rapid transit service in 2020 serves as an example of transit network improvement. Employing an empirical Bayes before-and-after assessment, we aim to discern the effect of the R5 Hastings RapidBus implementation on road collisions along this corridor. Our hypothesis posits a reduction in collisions subsequent to the introduction of RapidBus. In this way, we may conclude that enhancements to our transit networks via applications of big data may serve as a means to improve both mobility and safety concurrently.

It is worth noting that the empirical Bayes approach in this context does not leverage big data nor does it apply the industry standard approach of collision prediction models. We are implicitly assuming a linear relationship between crash frequency and traffic volumes which has been shown to have many defects (Sawalha & Sayed, 1999). Nevertheless, we continue with this approach due to limited data availability and recognize the importance of continuing this type of research utilizing more robust datasets and analysis methods. This research contributes to the field in three key ways:

1. **Consolidation and Summarization**: By consolidating and summarizing the intricate relationships between big data, public transportation, and road safety, our work provides a comprehensive overview of these interconnected elements.

2. **Evidence for Improved Road Safety**: Through the case study of the East Hastings corridor, we aim to furnish empirical evidence supporting the notion that public transit investments, particularly the introduction of RapidBus, can lead to improved road safety.

3. **Importance of Continued Research**: Beyond the immediate findings, our research underscores the importance of ongoing investigations into enhancing public transit, and emphasizes the pivotal role of big data in fostering safety and mobility within urban environments.

The remainder of this paper is structured as follows. In section 2, we summarize the case study including a background of RapidBus projects, including the R5, that were put into service in Vancouver, BC. We further outline the data used in this study, and demonstrate the methodology, analysis and results from the empirical Bayes assessment. Finally in section 3, we conclude with comments and suggestions for future work. Overall, we aim to show that the analysis of big data and application of new technologies can allow effective improvements to public transportation and thus simultaneously improve mobility and safety.

2 Case Study

2.1 R5 Hastings RapidBus

In early 2020, Bus Rapid Transit (BRT) was introduced across multiple corridors in Metro Vancouver, marking a significant enhancement over basic bus-operated or local transit services. This transit innovation, implemented by TransLink, boasts several improvements, including more widely spaced stops, all-door boarding, enhanced station amenities, and transit priority measures. These BRT routes, characterized by high frequency and capacity, can efficiently accommodate up to 12,000 passengers per hour during peak periods and operate up to 20% faster than local services (TransLink, 2023).

One notable addition to Metro Vancouver's transit landscape is the R5 Hastings RapidBus, which commenced operations on January 6, 2020. This route spans from Simon Fraser University to the Burrard SkyTrain station in Downtown Vancouver, as illustrated in figure 5. The R5 Hastings RapidBus succeeded the 95 B-Line, which was implemented in 2016 to support the launch of the Evergreen Line SkyTrain Extension. Unfortunately, the launch of this RapidBus service, alongside others, coincided with the onset of the Covid-19 pandemic, resulting in a substantial reduction in transit ridership. Despite this challenge, figure 4 reveals a notable increase of approximately 5% in the ratio of rapid to local ridership post-implementation, demonstrating resilience in the face of overall ridership decreases.

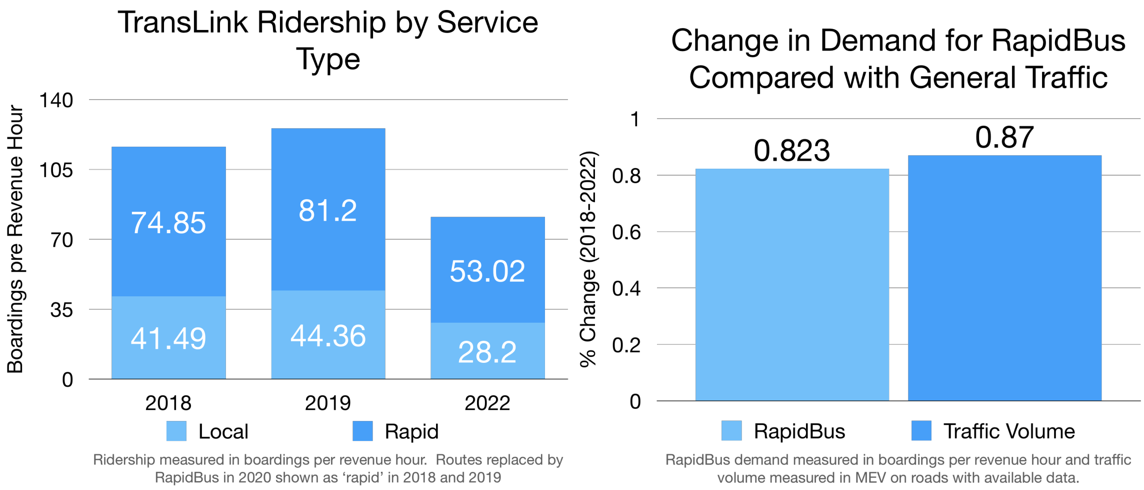


Figure 4: (Left) Average boardings per revenue hour for TransLink local and rapid transit routes. In 2018 and 2019, routes which were replaced by rapid service in 2020 are shown as 'rapid.' Source: https://www.translink.ca/plans-and-projects/strategies-plans-and-guidelines/managing-the-transit-network. (Right) Comparison of demand for rapid transit with general traffic volume.

2.2 Data Sources

To conduct a comprehensive empirical Bayes before-and-after study, transportation data encompassing traffic volumes and crash frequencies at both treatment and reference sites before and after the implementation period is required.

Traffic volumes were procured from the City of Vancouver's permanent vehicle counts open data portal, reflecting information as of the April 7, 2022 update. The relevant webpage can be accessed at https://maps.

vancouver.ca/portal/apps/sites/#/vanmap/maps/f1df4c0b64384a27a27d0c9d97586aa0/explore. Concurrently, crash frequencies were obtained from the ICBC statistics open data portal, incorporating data as of the September 25, 2023 update, accessible through https://public.tableau.com/app/profile/icbc/viz/ICBCReportedCrashes/ICBCReportedCrashes-LowerMainland#1. Note that all data and information for this study was accessed in November 2023. Additionally, it is imperative to acknowledge several limitations concerning the data and its suitability for this study.

1. **Sparse Distribution of Traffic Volume Sensors**: The traffic volumes are derived from permanent sensors, sparsely distributed across the road network. Consequently, there is a limitation in terms of the number of sites and years with available data.

2. **Single-Direction Traffic Volume Data**: In certain instances, traffic volumes were provided for a single direction. While this configuration remains unaltered in our analysis, it is acknowledged that the ideal scenario would involve data for both directions to more accurately assess the ratio of crashes to volume before and after.

3. **Road Name Level Crash Frequencies**: Crash frequencies were documented at the road name level, potentially resulting in higher counts for longer roads. To rectify this skew, we have incorporated estimated road length in the denominator of our crash rate measure.

These considerations highlight the need for caution in interpreting the results and underscore the importance of addressing these limitations in our data analysis and subsequent conclusions.

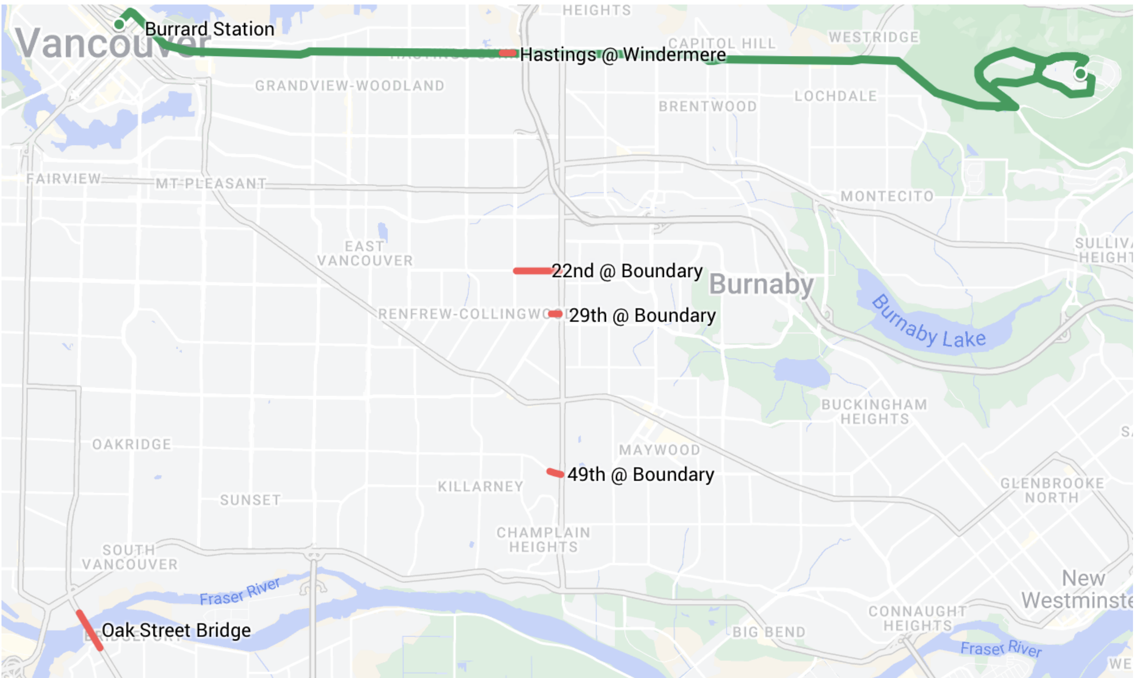


Figure 5: Map of Vancouver showing the R5 Hastings alignment (green) and traffic count locations (red).

2.3 Methodology

The empirical Bayes before-and-after assessment approach is designed to address three confounding factors that naive statistical methods may overlook.

1. **History**: Factors other than the treatment might contribute to observed changes in the dependent variable. In our case study, the effects of the Covid-19 pandemic fall into this category.

2. **Maturation**: Processes influencing the occurrence of the dependent variable, such as crashes, may naturally evolve over time. People can become more accustomed to certain environments, and the occurrence of the independent variable may change independently of the implemented countermeasure.

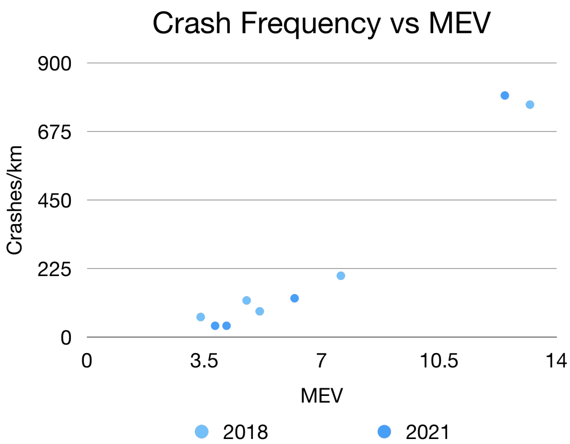
3. **Regression to the mean**: Following extreme events, there is a tendency for occurrences to become less extreme. Therefore, targeting locations with high crash occurrences in one period may result in observing a significant decrease in the following period, potentially due to natural fluctuations rather than the countermeasure's impact.

To address historical effects, we introduce *reference sites* alongside our *treatment site*. These *reference sites* closely resemble the *treatment site* in terms of configuration and traffic volumes, with options limited by available data. Additionally, we use data from 2018/2021 for the before/after periods to mitigate the impact of traffic behavior maturation influenced by the Covid-19 pandemic, which began in early 2020. The empirical Bayes before-and-after methodology accounts for regression to the mean artifacts, and we provide a concise explanation of this method below.

In road safety, it is widely acknowledged that collisions are rare and random occurrences. Traditional statistical estimation techniques based on the Gaussian (normal) distribution are ill-suited for identifying 'outliers' in a population of road sites where collisions are better approximated by a Poisson distribution (Oppe, 1982). The Poisson distribution, which expresses the number of collisions among similar sites, is given by the following formula, where n is the number of collisions, and a is the expected collision frequency.

$$p(n) = \frac{e^{-a} a^n}{n!}$$

Moreover, we allow a to be a gamma-distributed random variable, a reasonable assumption in practice (Wright et al., 1988). This flexibility, intrinsic to Bayesian analyses, involves estimating the distribution using historical data from both the *reference sites* and the collision history of our *treatment site* to refine parameter estimation.



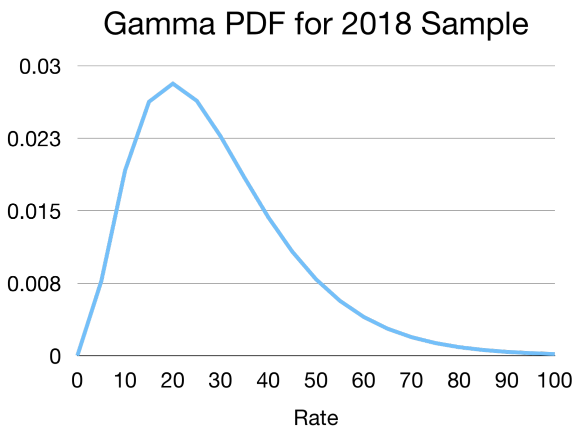


Figure 6: (Left) Scatter plot of traffic volume and crash data. (Right) Gamma distribution fit to the 2018 sample data. Note we do not have a sufficient sample size to estimate a goodness of fit to this data.

Site	Year	Million Entering Vehicles (MEV)	Crashes/km	Rate
22nd @ Boundary	2018	5.15	85	16.4
29th @ Boundary	2018	3.39	66	19.4
49th @ Boundary	2018	4.76	120	25.2
Hastings @ Windermere	2018	7.57	201	26.6
Oak Street Bridge	2018	13.21	763	57.8
22nd @ Boundary	2021	4.16	37	8.9
29th @ Boundary	2021	3.82	37	9.7
Hastings @ Windermere	2021	6.19	127	20.6
Oak Street Bridge	2021	12.46	793	63.6

Table 1: Traffic volumes and crash data for all sites.

Utilizing the data presented in table 1, we can compute the Odds Ratio to estimate the reduction in collisions attributed to the implementation of RapidBus. This measure serves to rectify the previously mentioned confounding factors by assessing the change at the treatment site in comparison to the change at the reference sites between the before and after periods. The effectiveness of the treatment can be quantified as $\text{OR} - 1$, where a negative value denotes a decrease, and a positive value signifies an increase in the observed quantity. Additionally, the standard error and t-statistic can be expressed as detailed below.

$$\text{OR} = \frac{A/C}{B/D}, \quad \sigma_{\text{OR}} = \sqrt{A^{-1} + B^{-1} + C^{-1} + D^{-1}}, \quad t = \frac{\ln \text{OR}}{\sigma_{\text{OR}}}$$

Where, $A =$ the number of collisions in the reference group in the before period.
$B =$ the empirical Bayes estimate of the number of collisions that would have occurred at the treatment site had RapidBus not been implemented.
$C =$ the number of collisions in the reference group in the after period.
$D =$ the number of collisions at the treatment site in the after period.

In computing B, our initial step involves estimating the parameters for the assumed Gamma prior distribution. We employ the method of moments estimates (MME), which sets the mean and variance of the Gamma distribution equal to those of the sample. The prior distribution is visually represented in figure 6, and the parameter calculation proceeds as follows: n denotes the number of sites, $\bar{\lambda}$ represents the average collision rate, N_i stands for the number of collisions, and V_i represents the traffic volume for site i.

$$\bar{\lambda} = \frac{1}{n}\sum_{i=1}^{n}\frac{N_i}{V_i}, \quad s_{\bar{\lambda}}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{N_i}{V_i} - \bar{\lambda}\right)$$

$$\beta = \frac{\bar{V}\bar{\lambda}}{\bar{V}s_{\bar{\lambda}}^2 - \bar{\lambda}}, \quad \alpha = \beta\bar{\lambda}$$

Here, $\bar{V}$ stands for the harmonic mean of the normalized traffic volumes. Subsequently, we can derive parameter estimates for the posterior distribution at our treatment site, denoted as $\beta_t = \beta + V_t$ and $\alpha_t = \alpha + N_t$, where t signifies the treatment site. Ultimately, we can compute $B = \bar{\lambda}_t V_t$ using the average collision rate from the posterior of our treatment site and the traffic volume observed during the after period.

2.4 Results

Following the steps in section 2.3, we find $A = 913$, $B = 165$, $C = 867$, and $D = 127$ which yields an estimated reduction in collisions of 19% for the treatment site of Hastings @ Windermere.

$$\text{OR} \approx 0.81, \quad \sigma_{\text{OR}} \approx 0.13, \quad t \approx -1.62$$

It is crucial to note that our sample size is insufficient to generate a reliable estimate for the goodness of fit concerning the Gamma distribution. Furthermore, the obtained result lacks statistical significance at a high confidence interval. As a result, it is advisable to approach these findings with skepticism, and additional investigations are imperative to either substantiate or refute the null hypothesis that RapidBus did not influence crash occurrence along its alignment.

3 Discussion

Let us revisit our research objective and contributions outlined in section 1.3 and assess the implications of this study. Our primary goal was to examine the impact of RapidBus implementation on road safety along the R5 Hastings alignment in Vancouver, BC. Traditionally, justifying transit improvements has been challenging due to manual data gathering and analysis processes, resulting in limited and costly implementations. The integration of big data into transit planning has revolutionized the design and operation of public transit projects, such as RapidBus, potentially allowing for concurrent enhancements in mobility and safety.

> Employing an empirical Bayes before-and-after assessment, we aim to discern the effect of the R5 Hastings RapidBus implementation on road collisions along this corridor. Our hypothesis posits a reduction in collisions subsequent to the introduction of RapidBus.

Regrettably, we cannot assert with certainty that RapidBus unequivocally enhanced road safety along its East Hastings alignment. While we observed indications of reduced vehicle crashes, these results lack statistical significance at a high confidence level. Moreover, certain factors were overlooked, such as non-vehicle incidents or the spatial concentration of accidents around stations.

> **Consolidation and Summarization**: By consolidating and summarizing the intricate relationships between big data, public transportation, and road safety, our work provides a comprehensive overview of these interconnected elements.

Our success lies in unravelling the intricate connection between big data, public transport, mobility, and safety through a comprehensive literature review, summarizing current trends, and presenting results. Operational improvements derived from transit fleet data analysis have boosted system efficiency, enabling additional service with minimal cost increases and thereby enhancing mobility. Additionally, we reviewed several papers indicating that public transit can reduce vehicle collisions and mitigate risks for passengers. However, these claims are accompanied by caveats such as ignorance of non-vehicle incidents.

> **Evidence for Improved Road Safety**: Through the case study of the East Hastings corridor, we aim to furnish empirical evidence supporting the notion that public transit investments, particularly the introduction of RapidBus, can lead to improved road safety.

Despite the lack of statistical significance, our analysis of road safety along the R5 Hastings RapidBus alignment aligns with broader road safety studies for public transit. Specifically, our results suggest a correlation between RapidBus implementation and an improvement in road safety.

> **Importance of Continued Research**: Beyond the immediate findings, our research underscores the importance of ongoing investigations into enhancing public transit, and emphasizes the pivotal role of big data in fostering safety and mobility within urban environments.

In line with Vecino-Ortiz and Hyder (2015), we underscore the necessity for further research into the effects of public transit on road safety, encompassing both vehicle and non-vehicle collisions. Simultaneously, we advocate for the integration of big data and modern safety assessment techniques, such as considering real-time sensor data and crash surrogates. By doing so, we can persist in advancing mobility and safety by investing in public transport and executing projects with sound technical methodologies.

3.1 Limitations

This research is subject to several inherent limitations. Primarily, the restricted availability of data led to a shallow sample size, preventing the calculation of goodness of fit statistics for our assumed Gamma distribution that models the data. Additionally, the use of inconsistent data sources introduced uncertainty when comparing the data. For instance, traffic volumes and vehicle crashes were independently observed and measured over varying road lengths. Furthermore, the available data does not encompass non-vehicle crashes and incidents, potentially resulting in an overestimation of road safety. Lastly, the consistency between the treatment and comparison sites is questionable and stems from the limitations in data availability.

3.2 Conclusion and Recommendations

Despite the numerous limitations, this research yields valuable insights. We indeed discovered evidence suggesting improved road safety correlated with the presence of RapidBus along East Hastings in Vancouver, BC, aligning with existing research. Our literature review underscored the connection between innovative applications of big data and the optimization of transit networks. Additionally, we highlighted the applicability of robust design approaches, such as the Safe System approach, to enhance the efficacy and safety impacts of public transport ventures. Hence, this research substantiates the idea that through leveraging big data, transit agencies may endeavor in enhancing both mobility and road safety through projects like RapidBus.

In conclusion, we propose several extensions and improvements to this study to further validate these findings. Firstly, leveraging a more robust dataset could furnish sound estimates for road safety improvements. Incorporating big data, like connected vehicle sensor data, enables the analysis of safety levels concerning public transit, surpassing the capabilities of standard statistical methods such as the empirical Bayes used in this study. A focused investigation into non-vehicle incidents is also recommended to comprehensively formalize the safety impact of public transit. Finally, future studies should extend their analyses to different environments (e.g., various road facilities) and strategies (e.g., Safe System approach) to discern which combinations of these factors yield the most substantial safety improvements.

4 Acknowledgements

All analysis, inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not reflect the opinions, position or policies of ICBC, City of Vancouver, and/or TransLink.

References

Litman, T. (2015). *Evaluating public transit benefits and costs*. Victoria transport policy institute Victoria, BC, Canada.

Oppe, S. (1982). *Detection and analysis of black spots with even small accident figures*. Institute for Road Safety Research SWOV, the Netherlands.

Sawalha, Z., & Sayed, T. (1999). Accident prediction and safety planning models for urban arterial roadways. *Report for the Insurance Corporation of British Columbia by the Department of Civil Engineering, University of British Columbia, Vancouver, Canada*.

Schepers, P., Lovegrove, G., & Helbich, M. (2019). Urban form and road safety: Public and active transport enable high levels of road safety. *Integrating human health into urban and transport planning: a framework*, 383–408.

TransLink. (2023). Rapidbus projects. https://www.translink.ca/plans-and-projects/projects/bus-projects/rapidbus-projects

Truong, L. T., & Currie, G. (2019). Macroscopic road safety impacts of public transport: A case study of melbourne, australia. *Accident Analysis & Prevention*, *132*, 105270.

Vecino-Ortiz, A. I., & Hyder, A. A. (2015). Road safety effects of bus rapid transit (brt) systems: A call for evidence. *Journal of urban health*, *92*, 940–946.

Verbavatz, V., & Barthelemy, M. (2019). Critical factors for mitigating car traffic in cities. *PLoS one*, *14*(7), e0219559.

Welch, T. F., & Widita, A. (2019). Big data in public transportation: A review of sources and methods. *Transport reviews*, *39*(6), 795–818.

WHO, W. H. O. (2022, June). Road safety. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

Wright, C., Abbess, C., & Jarrett, D. (1988). Estimating the regression-to-mean effect associated with road accident black spot treatment: Towards a more realistic approach. *Accident Analysis & Prevention*, *20*(3), 199–214.

Zannat, K. E., & Choudhury, C. F. (2019). Emerging big data sources for public transport planning: A systematic review on current state of art and future research directions. *Journal of the Indian Institute of Science*, *99*(4), 601–619.