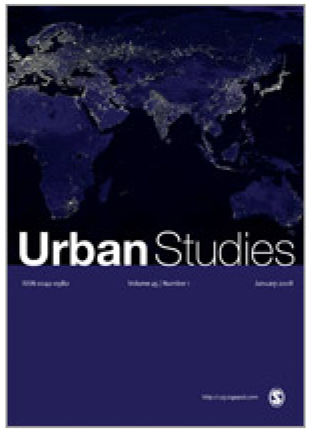

Measuring public transit transfer risk using high-resolution schedule and real-time bus location data

Journal:	*Urban Studies*
Manuscript ID	CUS-946-19-10.R1
Manuscript Type:	Special Issue
<b>Discipline: Please select a keyword from the following list that best describes the discipline used in your paper.:	Planning
World Region: Please select the region(s) that best reflect the focus of your paper. Names of individual countries, cities & economic groupings should appear in the title where appropriate.:	North America
Major Topic: Please identify up to 5 topics that best identify the subject of your article.:	Transport, Technology/Smart Cities, Planning, Method
You may add up to 2 further relevant keywords of your choosing below::	

Author's Response to Decision Letter for (CUS-946-19-10)

Measuring public transit transfer risk using high-resolution schedule and real-time bus location data

Dear Prof. O'Sullivan

Thank you for your response to our submitted manuscript 'Measuring public transit transfer risk using high-resolution schedule and real-time bus location data', (Manuscript ID CUS-946-19-10). We sincerely appreciate yours and the referees' thoughtful critiques. As you will see from our detailed revision report (see below), we were careful to respond to these critiques, revising the manuscript accordingly. Major changes include: i) situating the paper more explicitly within the context of urban big data; ii) reducing many of the equations, and; iii) streamlining the results, including the maps, for clarity.

In addition to the revision report below, the major changes in the manuscript are highlighted with red font color.

Again, we appreciate the review: we believe it has made the manuscript more focused, more accessible and better suited for the special issue.

We look forward to your response.

Best,

The Author

Revision report

We really appreciate the useful comments from the guest editor and reviewers. In the following paragraphs, we will thoroughly respond each comment and question.

Guest Editor Comments

GE-1. While clearly heading in the right direction, this paper will benefit considerably from further expositional clarity and additional content relating to the underlying themes of the Special Issue to which it is intended to contribute. The need for expositional clarity is raised by both reviewers, but more particularly by reviewer 1. In this vein, our view is that it is not clear the included math is adding much if anything at all – the underlying concepts seem rather intuitively obvious; is the math actually needed rather than just a slightly more considered verbal approach?

Response: Thank you for pointing this out. We incorporated formula 1 and 4 (previous draft) into the paragraph to avoid confusions. Per reviewer 1’s comments, we simplified the notations for formula 1 and formula 2 (current draft).

GE-2. In terms of diagrams, Figure 2 is confusing – again, is it actually needed?

Response: we decided to incorporate the Figure 2 (last draft) into the paragraph and remove the figure.

GE-3. Figure 3 is, as reviewer 2 notes, currently hard to read – and will not replicate well in grayscale. In overall terms, more attention to the quality and suitability of the graphics will greatly enhance the approachability of the paper.

Response: We changed the symbology of the map from graduated colors (circles with same size and different colors) to graduated symbols (circles with different sizes and same color). We also rework the maps so that they have smaller scale. Besides, to save words for the strict word limit and make the graph more intuitive, we decide to only keep the results for APC-GTFS data. We believe current version will replicate better in grayscale.

GE-4. The views of reviewer 2, on a need for the paper to speak more directly and substantively to an Urban Studies audience (and specifically to the special issue themes: Big Data as a tool for advancing our understanding of the urban, and to enhance urban well-being) also merit further emphasis. Attention to this matter has the potential to move the paper from a somewhat specialised contribution that will appeal to a limited cadre of transport experts, into something of much more general value and significance.

Response: We added many clarifications so that the paper is more suitable for the special issue theme. Generally, we focused on three parts in terms of emphasizing on the SI theme: the introduction, the literature review, and the conclusion.

In the introduction, we emphasize the introduction of big data and its implication, both opportunities and challenges, on the transfer studies. To leverage and overcome the complicated nature of big data, we thereby introduce the measures and their advantages over other traditional measures. We combine the big data and the new measures to demonstrate the main contribution of the paper: high-resolution, real-time, and computationally intensive measures of transfers with big data.

In the literature review, we start by introducing the history of transfer studies' data sources. Moreover, we adopt the suggestion of reviewer 2; we changed the "deliberate versus byproduct data" to "small versus big data" to avoid unclear definitions and classifications. In the second part, we focus on the development of transfer measures from non-real-time to real-time. We introduce the concept of temporal accuracy and highlight the real-time characteristic of the two measures we are about to introduce; we also connect this concept with later introduction of GTFS and APC. In the last paragraph of the literature review, we briefly compare the results of traditional non-real-time small data research, former big data real-time research, and this paper to emphasize the contribution of the paper.

In the conclusion, we adopt the reviewer 2's comment about re-engaging the SI theme. We revisit the vein of the paper from motivation ("real-time big data for transfers"), to the contribution (real-time notification and improved reliability for urban dwellers), then to the vision (more accurate and more abundant big data with ridership support). For more details about the conclusion, please refer to our responses to comment 3.9.

By doing this, we shifted the focus of this paper from "studying transfers" to the "application of transportation big data for transfers." The revised manuscript engages more with the big data sources and their application to solve complicated urban transportation questions. We believe the current draft is much more suitable for the SI theme and the broad audience of Urban Studies.

Reviewer 1

The paper makes a valuable contribution in public transport research by developing intuitive quality of transfer measures based on scheduled and real-time data. The paper is well written for the most part, however there are some unclear sentences. Some minor comments are listed below:

R1-1. Fig 1: x-axis label further away than the legend headers, makes figure a bit hard to read.

Response: We adjusted the graph according to the comment (see Figure 1 (current draft)). We indicated x and y to the corresponding label to avoid confusion; we also moved the x-axis label to the middle to keep it consistent with the y-axis label.

R1-2. The prime symbol used in the mathematical notation for "actual departure time" is quite difficult to notice. Can you please highlight that you use this symbol in text or alternatively use different symbols for clarity?

Response: We simplified the notations in the formula 1 (section 3.2, subsection transfer time penalties) and 2 (same section) in the current version. We now use uppercase T for the actual departure time and lowercase t for the scheduled departure time, which is much more obvious than prime symbols. Moreover, since all the time mentioned in the current formulas are now departure time, we removed the "departure" notations in the subscripts. All these adjustments should make the formulas simpler.

R1-3. Fig 2: this is not giving much additional information. It could be changed to depict all the scenarios on page 14. There should also be a scheduled arrival time for the figure to make more sense. Furthermore, the figure text could highlight that the blue line is the chosen option.

Response: As the reviewer 1 and the guest editor pointed out, the figure 2 (last draft) was confusing and did not give much additional information. Meanwhile, as the editor commented, the definition of the three types of transfers are inherently intuitive. Therefore, we remove figure 2 (last draft) and add some further explanation to section 3.2, subsection "transfers: the good, the bad, and the ugly" in the definition of each transfer type, especially preemptive transfers according to reviewer 2's comment.

R1-4. Explanations for nonobvious abbreviations could be repeated in analysis section for convenience.

Response: We add several explanations in section 4 and especially the start of each subsection. The relevant abbreviations include ATTP (average total time penalty), TR (transfer risk), GTFS (general transit feed specification), APC (automated passenger count), DBL (dedicated bus line), and COTA (Central Ohio Transit Authority).

R1-5. Are there large differences in frequency between weekdays and weekends? What are the impacts on the measures? (Similar effects as on the time of day comparison?)

Response: We added frequency correlation analysis in Figure 4 and Figure 5.

Weekdays
We added the correlation analysis between frequency and daily average measures in section 4.2 (current draft), paragraph 2. And yes, there are differences. COTA system have three schedules: weekdays, Saturdays, and Sundays. But the difference is rather small for trips that involves transfers. In terms of the transfer trips' headways, we do not observe a significant difference between the three schedules. However, this could be because the frequency difference between weekdays is too small and there are only seven weekdays and three schedules; therefore, the heterogeneous pattern within a day may be homogenized by this oversimplified aggregation.

Time of day
We add the correlation analysis between frequency and hourly average measures in section 4.2 (current draft), paragraph 3. Hourly correlation analysis has more abundant data points and shows the pattern within a day: for TR, frequency has no significant impact for both datasets; for ATTP, frequency has significant negative correlation with it for both datasets. This also suggests that: increasing frequency for both generating and receiving buses can help with reducing the total time penalty, however, it cannot help with the synchronization.

R1-6. Page 23, row 50: I think it is fair to assume zero delays for demonstrative purposes, however, it would be good to also highlight that it is a highly speculative assumption that there would be no delays on a BRT line. The only conclusion that can be drawn from this simulation is that improving punctuality even on one route will reduce ATTP.

Response: Yes, this is a fair point. We added some clarification in section 4.3, paragraph 2 to highlight the assumption is hypothetical. We also revised the conclusion from "DBL is an effective solution for reducing ATTP" to "improving punctuality even on one route will reduce ATTP". Besides the overall impact, the DBL simulation moreover pointed out the major differences between DBL's impact on the generating trips and receiving trips. Although this simulation is hypothetical, the simulation is one of the first to explore DBL's impact on transfers and it can demonstrate prospective results for future studies.

R1-7. Page 17, row 45: unclear sentence: "To investigate the spatial pattern of transfer risk, the first thing is spatial aggregation, since trip patterns (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns."

Response: We change the sentence to: "To investigate the spatial pattern of transfer risk, the first thing is to aggregate trips based on their generating stops, since trip patterns (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns."

Different transfer trips with the same transfer stop are not geographically distinguishable: we cannot distinguish these transfer trips on the map since they have the same geographic feature. For example, we have a transfer from route 1 to route 2 from stop A to stop B; we also have a transfer from route 3 to route 4 from stop A to stop B. Since their stops are both A and B, we cannot easily plot their geographic pattern individually on the map. Therefore, before we are investigating the geographic pattern of the transfers, we first need to aggregate trips based on their generating stop.

R1-8. Page 19, row 47: unclear sentence: "However, for the APC-GTFS dataset, we observe ATTP on Sundays is second lowest compared to Fridays, which is the lowest for original GTFS dataset."

Response: We change the sentence to: "However, we observe Sundays have the lowest ATTP." To save words due to strict word limit, we decide to delete the results for original GTFS; current draft only show the results of APC-GTFS, which has higher temporal accuracy.

Reviewer 2

The paper presents two measures of transport network performance. It is clearly written, mostly easy to follow and quite technical in nature. The editor may want to take a view to which extent the paper fits the remit of this special issue and the Urban Studies audience more generally. In its current form, the paper seems to be more suited to a transportation-focused journal. Some suggestions:

R2-1. The authors could engage more thoroughly with the themes highlighted in the call of the SI. I imagine that an Urban Studies readership would be interested in a discussion of how big data can offer new understanding of urban transport systems. Of course, there are already numerous reviews on this question, but a more focused discussion in view of the SI and the particular specialism of the authors might add to the literature. This would also help readers appreciate the specific contribution of this paper.

Response: Thanks for pointing this out for us and we think this is a very good comment. To make the paper fits the SI theme better, we made several major adjustments in the current draft in following different sections.

* Introduction: we emphasize the introduction of big data and its implication, both opportunities and challenges, on the transfer studies. To leverage and overcome the complicated nature of big data, we thereby introduce the measures and their advantages over other traditional measures. We balance the equilibrium between the SI theme "big data" and the methodological "transfer measures"; moreover, we shift the focus from a transportation study for transfers to an application of transportation big data to understand transfers.

* Literature review: according to the comment: "a more focused discussion in view of the SI and the particular specialism of the authors", we rework the literature review. We made two major adjustments. First, we changed the "deliberate versus byproduct data" to "small versus big data" to avoid unclear definitions and classifications. By doing this, the paper is directly and explicitly connected with the SI theme of "urban big data". Second, we focus on the development of transfer measures from non-real-time to real-time. We introduce the concept of temporal accuracy in section 2.2, paragraph 1; then we highlight the real-time characteristic of the two measures we are about to introduce and the challenge and academic gaps. Combining the two aspects, we also demonstrate the core contribution of the new measures: high-resolution, real-time, and computationally intensive measures of transfers with big data.

* Conclusion: we adopt the reviewer 2's comment about re-engaging the SI theme. We reorganize the conclusion part so that the motivation (big data for transfer studies), the benefits (real-time notification and improved reliability for urban dwellers), the future (more accurate and more abundant big data with ridership support) can be revisited. We believe the can make the paper more accessible to the audience of Urban Studies and the special issue theme of the journal.

R2-2. For the benefit of an Urban Studies audience, I would suggest that the authors dedicate more space to a fuller discussion of the specialist literature, and highlight the different objectives of those studies they cite. What are the pros and cons of existing measures of evaluating transfer effectiveness? In which ways are the measures proposed by authors superior to existing ones?

Response: Thanks for pointing this out. We adjust numerous wordings and sequence in the literature review so that the objects and detail of the studies can be highlighted. First, we use italic font to highlight each study's object, as shown in section 2.2, subsection "Non-real-time measures". Second, we add more clarifications, as shown in the same paragraph.

We adjusted the literature review so that it fits the paper topic and the SI theme. To focus on the real-time nature of the proposed measures, we change the section 2.2 to "non-real-time measures" versus "real-time measure". We introduce the development of measures. Traditional studies used static measures that are built from non-real-time sources, such as schedules and non-volatile social factors. With the support of more big data sources and corresponding data supports, we now can demonstrate more real-time pattern and analysis.

"What are the pros and cons of existing measures of evaluating transfer effectiveness?"

In section 2.1 and 2.2, we discuss their pros and cons:

For non-real-time measures, though extremely useful in the designing and planning area, we can conclude that they are less effective to measure the actual real-time patterns since they only consider static qualities of transfer or transfer nodes.

For existing real-time measures, the lack of large-volume real-time big data is the major concern as shown in section 2, last paragraph.

Most importantly, very few studies consider the transfer's real-time performance with respect to delay and no study use actual real-time data sources to calculate the performance as shown in section 2, last paragraph. In this sense, TR (transfer risk) and ATTP (average total time penalty) are not measuring the same quality as the existing measures; therefore, the most important negative of the existing measures is simply that they are not measuring the real-time performance.

"In which ways are the measures proposed by authors superior to existing ones?"

Response: As section 1, paragraph 3 (introduction), section 2 (literature), and section 5, paragraph 3 – 5 (conclusion) point out, we can conclude following four major contributions and advantages compared with existing measures:

a) Compared with existing measures, TR and ATTP are the first measures that provide attainable solution to quantify the real-time performance of public transit transfers with respect to the schedules.

b) With this high-velocity and high-volume nature, the results of TR and ATTP are more detailed, more abundant, and more heterogeneous than the traditional measures. TR and ATTP can also be aggregated into different temporal and spatial scales; accordingly, TR and ATTP can provide more abundant and useful real-time information for ordinary passengers, compared with traditional measures dedicated to planning and designing.

c) As the literature review demonstrates in section 2, last paragraph, TR and ATTP are very easy to be implemented to a new public transit system with GTFS or APC-GTFS supports.

d) As the introduction in section 1, paragraph 3 and conclusion demonstrates in section 5, paragraph 3 and 5, TR and ATTP uses common metrics, namely probabilities and time. Compared with traditional composite scores, this feature can not only make them much easier to interpret and understand, but also can make intra-system and inter-system comparison much easier.

R2-3. How does the commonly made distinction between ‘deliberate data’ and ‘byproduct data’ apply to the datasets used by the authors? I certainly agree that smartcard data are ‘byproduct data’, but datasets such as APC are deliberately collected for the purpose of passenger counts. Similarly, GTFS data are more than just byproduct; they are purposively structured, standardised and documented. The authors should clarify the ways in which their work relates to the SI’s theme of ‘Big Data in the City’, and depending on their focus, offer a fuller discussion of ‘byproduct data’ potentially extending it to issues of bias and computational cost. Alternatively, perhaps an emphasis of ‘small data’ versus ‘Big data’ may be more appropriate for this particular paper.

Response: This is a very nice proposal. To avoid further confusions about how to classify “deliberate data” and “byproduct data” and keep the theme of the paper connected to the SI theme, we decided to change the two categories to “small data” versus “big data” as the reviewer 2 proposed. By doing this, we can clearly define GTFS and APC as typical examples of “big data in the city”.

As for the issues of bias and computational cost, we analyze their potential limitation in section 2.1, subsection “Big Data”, paragraph 3; since computational cost is a less important factor considering the advancement of computational technologies and Moore’s law for hardware, we choose to not include this part into the paragraph, also due to the tight word limit. Also, we already present several limits and potential biases in section 3.1, subsection “General Transit Feed Specification (GTFS) data”, paragraph 3 for GTFS data and section 3.1, subsection “Automated Passenger Count (APC) data”, paragraph 1 for APC data. Combining this two analyses on the data sources, we present a comprehensive discussion of the two relevant big data sources and their potential bias and computational cost. In the conclusion part in section 5, last paragraph, we moreover revisit the possible limitations and future direction of the transfer studies with big data: in general, more abundant and more high-resolution big data with ridership support can create more opportunities to understand public transit systems, especially transfers.

R2-4. There is a tension between the technical measurement and passengers' experience, which would warrant further discussion. To which extent are the components of transfer time penalties actually experienced by passengers? On high frequency services, ATP may actually not matter that much. And, if I understand correctly, transfer risk (which would be better called 'risk of transfer loss' or something) will increase for high frequency services. If a 'receiving' service runs every 5 minutes, the risk is inherently higher than if a service runs every 20 minutes, and yet the impact on the passengers' experience is higher in the latter case. Of course, this dimension is picked up by the other measures, but I do wonder how meaningful 'risk of transfer loss' is without considering frequency.

Response: This is a good observation. The best and perhaps the only way to analyze passengers' experience is to conduct surveys or interviews on the passengers per se. Since we did not do survey on passengers, we cannot have enough solid evidence to support a definite conclusion. This is also what we are conceptualizing in the conclusion part about the future research: more detailed and more abundant big data sources beyond real-time vehicle data. Although we do not have survey results, we can still discuss the questions by following aspects:

"To which extent are the components of transfer time penalties actually experienced by passengers?" We argue that the user will not experience the components at all, since the user will not see or perceive d_r (the delay of actual receiving bus (the second bus in the transfer)) and ATP (additional time penalty, the sum of receiving bus headways). For example, the nature of ATP determined that it sometimes cannot be physically perceived by the users at all: ATP can be negative. Instead, the user may experience TTP, since it is the "net time loss" for the users' transfer trip: user can easily calculate or perceive it by comparing with the current time when the receiving bus moves (T_r^n) and the time that the trip planner promised (t_r^0). However, since we do not have the data on user experience, we cannot conclude the exact threshold of the passengers' perceived transfer penalty.

"On high frequency services, ATP may actually not matter that much."
Yes, but this is partially right. On high frequency services, ATP, which is the sum of missed headways of receiving buses, is assumed to have lower value if the frequency increases. However, if the arrival time of the generating bus (the first bus in the transfer) does not change, the ATP will not change even if the headway is smaller, since it will simply miss more buses making the sum of headways stay same. But considering both generating and receiving buses, smaller headway will generally make it easier for transfers to synchronize, thus incur smaller time penalty.

"Transfer risk (which would be better called 'risk of transfer loss' or something) will increase for high frequency services. If a 'receiving' service runs every 5 minutes, the risk is inherently higher than if a service runs every 20 minutes, and yet the impact on the passengers' experience is higher in the latter case. Of course, this dimension is picked up by the other measures, but I do wonder how meaningful 'risk of transfer loss' is without considering frequency."

We added the correlation analyses between TR (transfer risk)/ATTP (average total time penalty) and frequency shown in Figure 5 (current draft). We can see that for both datasets, ATTP have significant negative correlation with frequency while TR does not have significant correlation with frequency. Therefore, we can conclude from the data that: among the same system in different hours, frequency/headway does not have a significant impact on transfer risk. This is also intuitive: if the frequency of both generating and receiving buses simultaneously increases, the synchronization result is likely to stay the same. Since frequency does not have significant impact on TR, we can conclude TR can be a good measure to gauge the desynchronization. Therefore, for the large difference of transfer risk between High Street (indicated by a red circle in Figure 2) and the downtown area (indicated by a green rectangle in Figure 2), we do not and cannot say the high risk is due to the high frequency.

Meanwhile, as reviewer 2 commented, TR is only one dimension, which focuses on the synchronization performance on the system side; ATTP is more appropriate for users' experience, since all that user cares is time penalty, instead of a percentage. The future research can measure the perceived ATTP and compare it with actual ATTP.

R2-5 Similarly, I couldn't follow why 'ugly, pre-emptive' transfers would be experienced negatively. 'Pre-emptive' suggests that passengers pre-emptively transfer to avoid risk elsewhere; but in this particular context, they simply get on an earlier vehicle without necessarily being aware of this. Except for potential crowding, they may not be any different from 'good' transfers.

Response: The reason why we called it "ugly" is not because preemptive transfers will be experienced negatively, and we added corresponding clarification in the paper about this in section 3.3 (current draft), subsection "Transfers: the good, the bad, and the ugly", paragraph "The ugly".

Just like the comment says, compared to the schedule, a preemptive transfer may result in positive, or zero, or negative TTP. This is exactly the reason why we define transfer risk as the proportion of missed transfers alone, instead of missed transfer and preemptive transfers together. We acknowledged that a preemptive transfer can be as "good" as or even better than a normal transfer. However, we have to point out that preemptive transfers are different from the normal transfers. With the metaphor of "ugly", we intended to describe its unpredictable and unsustainable nature.

R2-6: P.16, last sentence, "Only those combinations with … unique routes are selected". I couldn't follow what unique routes means in this context. Does this mean that origin-destination pairs with multiple transfer possibilities have been excluded? More explanation would be helpful.

Response: We added some clarification in the section 3.3, paragraph 3. The idea is: for multiple transfers with the same route combinations and the same generating stop (the stop where the user gets off from the first bus), we will only keep the one with closest walking distance to remove some redundancy. This is because the passengers will only walk to the closest stop when she/he gets off the first bus to finish a transfer. For example, if a passenger gets off the first bus at stop A, and she/he has several options (stop B, C, D…) to catch the same second bus. If stop B is the closest one, there is really no point for the passenger to walk further to catch the same bus.

R2-7. Figure 3 is hard to read. The authors may want to explore alternative types of visualisations, e.g. heat maps or contour maps.

Response: This is a very good point. We adjusted the figure in following several aspects:

We changed the extent of the map to the center of the city of Columbus to keep consistent with the geographic pattern described in section 4.1, subsection “spatial patterns”, paragraph 3. The center of city is the area with most ridership in the Central Ohio Transit Authority (COTA) bus system. By doing this, we can not only cover most interesting patterns, but also the symbols are much clearer to see.

We did not use heat map or contour map and there are several reasons. First, the underlying data are bus stops; interpolation of their values beyond the bus stops themselves is not valid. Second, each stop is extremely close to each other while their ATTP / TR difference can be very large; this would be even more confusing. Instead, we changed the visualization type from graduated colors (circles with same size and different colors) to graduated symbols (circles with same color and different sizes). The figure also replicates better in grayscale in terms of the editor’s concern Combined with a larger map scale, we believe it is easier for readers to distinguish the major patterns in Figure 2.

R2-8. P.21 first paragraph, I wonder if statistical tests should be added to assess differences on days with rain or a football match or in the DBL scenario. By the look of the values, I don’t see how they indicate ‘considerable impact’ as the authors conclude. The authors could also focus on connections with larger differences or on those with higher passenger counts.

Response: Thanks for pointing this out; we apologize for our neglect. We added corresponding statistical tests to each sections and the results are:

* Football. We did the Kolmogorov–Smirnov (KS) test to see whether the two distributions are the same. For ATTP (average total time penalty), the p-value for original GTFS is 0.036 and the p-value for APC-GTFS is 0.53; for TR (transfer risk), the p-value for original GTFS is smaller than 〖10〗^(-5) and for APC-GTFS is 0.65. Therefore, for original GTFS, the difference between football game days and ordinary days is significant and for APC-GTFS it is not.

* Precipitation. The results of KS test are all not significant. For ATTP, the p-value for original GTFS is 0.46 and the p-value for APC-GTFS is 0.40; for TR, the p-value for original GTFS is 0.27 and for APC-GTFS is 0.88.
Even if the football days are significantly different from ordinary days for original GTFS data, the results show that the two results in the last draft are not statistically significant under most scenarios. Therefore, as reviewer 2 pointed out, we also would like to put more attention to geographic and temporal patterns; the causality between the introduced measures and other social and environmental factors is not the major research question of this paper. Meanwhile, also because of the strict words limit, we decided to delete it to save space for other more important issues.

* Dedicated bus lane (DBL). The KS test shows very strong significant results for the difference between two scenarios. For ATTP, the p-value for original GTFS is 0.006 and the p-value for APC-GTFS is 0.005. We also add the analysis results to section 4.3 (current draft), paragraph 2.

However, as we mention above, since we decide to delete original GTFS results to save space, we will not show these test results in the paper.

R2-9. In view again of the theme of the SI, I would expect authors to re-engage with the theme of Big Data in the conclusions. This may include a discussion of what new insights we may gain with regard to urban dwellers' experience of their cities.

Response: To emphasize the SI theme "Big Data as a tool for advancing our understanding of the urban, and to enhance urban well-being.", we made several adjustments so that the paper can fit better. We reorganize the conclusion part so that the motivation (big data for transfer studies), the contribution (real-time notification and improved reliability for urban dwellers), the vision (more accurate and more abundant big data with ridership support) can be revisited:

Motivation: Instead of starting with the transfers, we start the conclusion from the perspective of big data sources as shown in section 5, paragraph 1. By doing this, we want to draw more attention to the application of urban and transportation big data in this paper; we can also make the paper more accessible for broader audience of the journal beyond transportation.

Contribution: We also reorganize the concluded advantages/utilities of the new measures in section 5, paragraph 3 – 5 so that we can have a comprehensive discussion of their potential impact. We added more content on urban dweller and passengers' experiences in section 5, paragraph 3; we also merge the management level and administration level to moreover emphasize the urban dwellers' experience.

Vision: In the section 5, last paragraph, we moreover revisit the development of big data and possible improvements in the future.

Measuring public transit transfer risk using high-resolution schedule and real-time bus location data

Abstract

The emergence of urban Big Data creates new opportunities for deeper understanding of transportation within cities, revealing patterns and dynamics that were previously hidden. Public transit agencies are collecting and publishing high-resolution schedule and real-time vehicle location data to help users schedule trips and navigate the system. We can use these data to generate new insights into public transit delays, a major source of user dissatisfaction. Leveraging open General Transit Feed Specification (GTFS) and administrative Automatic Passenger Counter (APC) data, we develop two measures to assess the risk of missing bus route transfers and the consequent time penalties due to delays. *Transfer Risk* (TR) measures the empirical probability of missed transfers and *Average Total Time Penalty* (ATTP) shows overall time loss compared to the schedule. We apply these measures to data from the Central Ohio Transit Authority (COTA), a public transit agency serving the Columbus, Ohio, USA metropolitan area. We aggregate, visualize, and analyze these measures at different spatial and temporal resolutions, revealing patterns that demonstrate the heterogeneous impacts of bus delays. We also simulate the impacts of dedicated bus lanes reducing transfer risk and time penalties. Results demonstrate the effectiveness of measures based on high-resolution schedule and real-time vehicle location data to assess the impacts of delays and guide planning and decision making that can improve on-time performance.

Keywords: Public transit; transfer risk; General Transit Feed Specification (GTFS) data; Automatic Passenger Counter (APC) data

1. Introduction

Transfers between routes are often necessary in public transit systems. The expansion of city footprints can make direct routes difficult to provide (Knoppers and Muller, 1995). Configuring systems that include route transfers allows public transit providers to cover more space ~~and time~~ with fewer vehicles (Walker, 2012). However, transit delays, defined as a positive deviation of a transit vehicle's actual arrival time from the scheduled time, are inevitable due to traffic, equipment malfunctions, external events and other circumstances. Transit delays causing users to miss intended transfers between routes imposes significant time penalties on users, making the system less functional and desirable.

Until ~~the~~ recently, scientific analysis of transit transfers has been limited, focusing on the users' experience and transfer node design (Guo and Wilson, 2004, 2011; Sun et al., 2007, 2010), with limited investigation into real-time behavior and system performance (Jang, 2010; Nesheli and Ceder, 2015; Nishiuchi et al., 2015). However, the context for scientific understanding of public transit is changing due to the emergence of location-aware and wireless communication technologies. These technologies allow public transit agencies to capture real-time data on vehicle locations across the entire transit system. Many agencies are publishing these data along with schedule data to enable public transit webpages and apps that make navigating the system easier for users. These data can also be leveraged to enable new scientific insights into public transit dynamics, including public transit transfers and the impacts caused by delays.

In this paper, we develop measures for the evaluation of ~~the~~ transfer performance in public transit systems for high-resolution schedule and real-time vehicle location datasets. *Transfer Risk* (TR) measures the empirical probability of missing transfers based on historical data, while the *Average Total Time Penalty* (ATTP) shows overall time loss compared to the schedule. Unlike many former composite scores, TR and ATTP are ~~more~~ intuitive to compute and understand ~~for its direct and precise nature~~; the measures are also easier to aggregate into different levels of spatial and temporal resolution and expand to other systems with real-time support. We implement these measures using high-resolution schedule and real-time vehicle location data from the Central Ohio Transit Authority (COTA) bus system in Columbus, Ohio, including both open data published in the General Transit Feed Specification (GTFS) format and administrative data derived from Automated Passenger Counters (APCs). We explore the patterns of TR and ATTP for different spatial and temporal resolution, and simulate the impact of dedicated bus lanes on TR and ATTP. The results demonstrate the ability of TR and ATTP measures to leverage large-volume big data and assess the impacts of delays on transfers; the measures can not only guide planning and decision making to improve on-time performance, but also provide important information for ordinary users about transfers' empirical performance.

In the next section of this paper, we review previous research on transfers from the perspectives of the data utilized and the analysis conducted. In the subsequent section we introduce our data sources and methods. Following this, we show the results of the spatial and temporal analyses. We conclude this paper with a discussion of the strengths and limitations of this study, and steps for future research.

2. Background

This literature review covers two dimensions of the development of measuring and analyzing public transit transfers: data and measures. We first discuss two types of data sources: *deliberate* or *small data* collected purposefully for analysis and *big data* generated automatically as a byproduct of operations (Ceder, 2016). Following this, we discuss existing measures that uses these data for ~~the~~ transfer measurement and their different tasks and benefits.

2.1. From small data to big data

Small data. Traditionally, studies of public transit transfer properties and behaviors use data collected deliberately for specific research questions, often using dedicated and survey instruments and location tracking via Global Positioning System (GPS) receivers. We call these datasets *small data* not only due to ~~its~~ smaller volume but also since it is often collected for a specific purpose. While small data are valuable and easy to interpret, there are several issues that limit their usefulness. First and most importantly, small data are often focused samples that do not provide blanket coverage of the entire system, making it difficult to see broader temporal and spatial patterns. Since most small data are purposely created, they are expensive and time-consuming to collect. For example, Guo and Wilson (2011) created and maintained special purpose station inventory and field survey databases. This requires substantial time and resources, often for a relatively small volume of data.

Another issue is the lack of universal standards and definitions, making comparisons difficult. Different transfer studies have varying definitions of transfers and their data (Guo and Wilson, 2004), limiting comparability. In addition, sampling frames are fragile, meaning that data collected for one set of questions cannot easily be repurposed for other questions (Miller and Goodchild, 2015).

An example of small data is stated preference (SP) data, used widely to support mode choice models (Guo and Wilson, 2011). Although many transfer assessment studies use SP data, the choice dimension is typically small, meaning that SP data may not be able to capture the full diversity of transfer situations (Bovy and Stern, 2012). Other semi-quantitative data collecting methods, such as on-board questionnaires, can also lack precision and reliability. The result of these imprecise data sources is that most studies provide a partial assessment of the system since it is difficult to have a detailed assessment at high spatial or temporal resolutions across the entire system (Guo and Wilson, 2011).

Big data. In the past, detailed, real-time performance data about public transit was difficult to acquire (Dessouky et al., 1999). However, this has changed due to the development of new data collection and sharing technologies. The widespread application of new information and communication technologies (ICTs) provide the technical support for what is often labeled *big data* (Hilbert, 2016). The definition of big data is diverse; a commonly accepted definition encompasses the "three Vs": large volume, high variety, fast velocity (Chen et al., 2014). However, as Miller and Goodchild (2015) argue, in many applications, especially in urban science, the unique and valuable characteristic of Big Data is ubiquity: its widespread coverage and availability, often as a byproduct of digitally-enable operations and activities.

In public transit, inexpensive GPS receivers and wireless communication allow widespread tracking of vehicle locations in real-time. These data are collected automatically on an ongoing basis by public transit authorities, meaning they are readily available without additional and prohibitive cost or effort. Meanwhile, the World Wide Web combined with data services allow sharing schedule and real-time vehicle location data. This technology revolution allows the possibility of more detailed investigation of transfer performance across an entire transit system.

However, big data also has limitations. As suggested by the characteristic "high variety," these data are often heterogeneous in terms of structure, quality and support: diverse data sources, lack of metadata, and lack of quality control all make big data challenging from a scientific perspective (Miller and Goodchild, 2015). Accordingly, standard protocols for sharing transit schedule and real-time data, such as General Transit Feed Specification (GTFS) (Google Developers, 2016, 2018), to help solve these problems.

2.2. From non-real-time measures to real-time measures

Non-real-time measures. Along with the data source's development from small data to big data, we also witness the progress of ~~the~~ measures from non-real-time measures to real-time measures. Non-real-time measures have a relatively low *temporal accuracy*; similar to Firmani et al. (2016)'s definition, we define it as: how accurate is the measure's recorded time compared to the actual time.

For example, many non-real-time measures do not measure ~~the~~ actual performance. Instead, they try to gauge static features, like *transfer nodes' design* and *transfer* *connectivity*. For example, Guo & Wilson (2011) assess transfer cost based on both users' and operators' perspective; they develop an index that measures each transfer node's effectiveness based on average time and economic cost per capita and apply it to the London Underground system. Hadas & Ranjitkar (2012) combine transfer connectivity and travel time to represent the effectiveness of transfers. Although the non-real-time measures have been proven to be extremely useful to assess the static qualities of ~~the~~ system design, the results based on only schedules cannot represent the actual performance.

Some non-real-time measures also use second-hand sources like stated preference surveys; their temporal accuracy is also ~~considerably~~ low since the surveys are usually conducted long after the actual trips. For example, many research analyze users' perceptions and attitudes about transfers (Guo and Wilson, 2004; Liu et al., 1997). These studies focus on measuring *transfer penalties*, namely how much and why people prefer not to take transfer trips; these penalties encompass a broad range of factors such as walking time, number of transfers in a single trip, waiting time, ticket fare, and other environmental factors.

Real-time measures. Even before the emergence of big data, on-board questionnaires can be considered as a form of real-time measure. However, it is until automated big data that we can really develop real-time measures based on first-hand information with *high velocity* and *large volume*. For example, Nishiuchi et al. (2015) used data envelop analysis to derive measures to evaluate the efficiency of user transfers between transportation systems. The measure concentrates on the transfer stations' commuting efficiency using users' smart card real-time data.

Similarly, ~~due to the lack of big data,~~ little research assesses transfers based on real-time performance and the risk of missing transfers due to delays. Progress in data availability, real-time monitoring and other smart city technologies are making this topic an area of active investigation: researchers can now conduct more detailed analysis and develop more precise measures and models of public transit transfer performance (Hadas and Ranjitkar, 2012; Kujala et al., 2018).

From this review, we can see that big data and real-time measures are the future direction of transfer studies. This paper contributes to this literature by developing measures of transfer risk and transfer time penalties using high-resolution real-time big data sources. The new measures are one of the first to consider transfers' real-time performance due to delay and the first to use bus systems' actual real-time big data to calculate it~~, which was largely overlooked by existing measures~~.

~~Moreover, the~~ new measures can demonstrate detailed patterns for any geographic and temporal resolution. While ~~the~~ small data sources can only present a homogeneous average pattern (Guo and Wilson, 2011), smart card data can provide temporal pattern during different hours (Nishiuchi et al., 2015), ~~the new measures with big data can not only measure temporal patterns of~~ a single specific trip, but also daily hours, weekdays, and months.

3. Methodology

This section discusses the methodology in our study. We first describe our data sources; then we define public transit transfers from a space-time perspective and conceptualize the impact of vehicle delays as a problem in transfer synchronization. Then, we discuss the methods involved in transfer risk measurement and analysis.

3.1. Data sources

In this paper, we leverage two datasets for measures and analytics.

General Transit Feed Specification (GTFS) data. General Transit Feed Specification (GTFS) is a combination of two data standards: *GTFS static* and *GTFS real-time expansion*. GTFS static reports the schedule data for a public transportation system. GTFS static is now the *de facto* standard for public transportation schedules and associated geographic data (Google Developers, 2016). Public transit system administrations are encouraged to share their GTFS static publicly: by 2010, almost 85% of transit miles traveled in the U.S were covered by open data published by transit authorities (Antrim and Barbeau, 2013).

GTFS real-time expansion provides frequently updated vehicle location data. GTFS real-time includes two components: vehicles' location and the trip updates, which contains vehicles' arrival and departure time at every sequential stop. Moreover, the temporal resolution can be as high as 1 min (Kujala et al., 2018). GTFS overcomes the disadvantages of both traditional data and unclean big data: it is high-volume, frequently updated, publicly accessible, standardized formats, and covers the entire public transit system. These features make GTFS a good big data source to conduct real-time measures.

However, despite all the valuable features, GTFS data has limitations. In particular, its temporal accuracy is lower than other sources such as automated passenger count data (discussed below). This is because GTFS real-time data feed is updated based on a fixed temporal interval, not based on the actual events of a public transit vehicle entering and leaving a designated stop. Consequently, the actual arrival/departure time may be different from the times recorded in the GTFS data.

Automated Passenger Count (APC) data. Automated Passenger Counting (APC) data are generated by devices that are installed on ~~the~~ vehicles to track and report transit ridership (Chu, 2010; Transit Wiki, 2019). These data often contain arrival time and departure time at each stop. A major advantage of APC data compared to GTFS data is the arrival and departure time is recorded at the stop as the events occur instead of updated according to a specified temporal interval. However, APC data are not open. As administrative data, APC data are not available for the public and transit mobile applications. Moreover, APC data does not have widespread coverage of the whole system. ~~Typically~~, a subset of public transit vehicles are installed with APC devices rather than blanket coverage as with GTFS data.

To leverage the positive features of both data sources, we merge ~~the~~ raw APC data and ~~the~~ GTFS data into a new combined dataset. We enumerated all GTFS trips while querying the APC database: if the corresponding trip is in the APC database, we override the record in the GTFS database to take advantage of the higher temporal accuracy of the APC data. Below, we will provide results based on the merged APC-GTFS dataset.

3.2. Transfer synchronization, desynchronization and time penalties

Public transit transfers link *generating trip* and *receiving trip*. A user first boards a bus to start the generating trip, then alights to catch the next bus to start the receiving trip. The transfer itself can be characterized as: i) a street-crossing transfer; ii) a sidewalk-based transfer, and; iii) a non-walk transfer at the same stop (Hadas and Ranjitkar, 2012). Based on this categorization, we can generalize the transfers as: i) *non-walking transfer*, which does not require ~~a~~ walking for the transfer, and; ii) *walking transfer*, which requires walking from the generating trip's stop (which we label the *generating stop*) to receiving trip's stop (the *receiving stop*).

Synchronization and desynchronization. We further conceptualize transfer as a process of synchronization ~~among~~: i) the generating trip brings passengers to the generating stop; ii) ~~transition of users~~ to receiving stop; iii) the receiving trip picks up passengers at receiving stop.

Transit delays can result in inconsistent arrival and departure times hence the desynchronization of scheduled generating trip and receiving trip. For each transfer, we can measure the time penalty when the receiving bus is leaving; this is the time point when the desynchronization happens. Due to desynchronization between the generating and receiving trip at the receiving stop, the actual transfer can differ from the schedule according to the relative temporal order of the two trips arrival/departure time. Figure 1 illustrates this process using a time-space diagram.

- **Insert Figure 1**

Due to the desynchronization, the actual receiving bus can be different from the scheduled receiving bus. We can conceptualize the schedule of all buses running on the same route as an array of trips (a *trip sequence array*). We assume the passenger will always take the first available bus. If the generating bus is sufficiently late, the passenger will miss the scheduled bus and need to take a later scheduled bus. Likewise, if the receiving buses are sufficiently late, the passenger can catch an earlier receiving bus in the trip sequence array. The *desynchronization degree* (DD), an integer variable, measures a transfer's desynchronization in the trip sequence array: it represents the order number of the actual bus before/after the scheduled bus. For example, if the actual bus is the *n-th* bus after the scheduled bus, the DD is *n*; if the actual bus is the *n-th* bus before the scheduled bus, the DD is -*n*; if the actual bus is the scheduled bus, then the DD is 0.

Transfer time penalties. We calculate two types of potential time penalties for each transfer. The first is *total time penalty* (TTP):

$$TTP = T_r^{\mathrm{n}} - t_r^0 \tag{1}$$

where: T_r^{n} is the actual departure time of actual receiving bus (DD = n) and t_r^0 is the scheduled departure time of scheduled receiving bus (DD = 0). TTP represents the total time loss compared to the schedule at the receiving stop. The value shows the total delay when the receiving trip starts, which encompass both the generating bus and receiving bus time loss. However, since the synchronization process is involved with two vehicles, it is also important to determine the corresponding time loss caused by each bus. For example, a large TTP could be because of the receiving bus's large delay but the synchronization is not disturbed; on the other hand, a large TTP could be also because of the first bus's delay, which results in desynchronization and thus long waiting time.

To quantify ~~this~~ two different types of delay, we decompose TTP as follows:

$$TTP = T_r^{\mathrm{n}} - t_r^0 = \left(T_r^n - t_r^n\right) + \left(t_r^n - t_r^0\right) = d_r + \left(t_r^n - t_r^0\right) \\ = d_r + ATP \tag{2}$$

where: t_r^n is the scheduled departure time of the actual receiving bus (DD = n), t_r^0 is the scheduled departure time of the scheduled receiving bus (DD = 0), d_r is the delay of the actual receiving bus at the receiving stop. The second part of the decomposition $\left(t_r^n - t_r^0\right)$ is defined as *additional time penalty* (ATP), which represents the time cost caused by the transfer desynchronization.

The two parts of TTP's decomposition, ATP and d_r, correspond to the time penalty caused by desynchronization and normal delay of the actual receiving bus. The value of ATP depends on the passenger's actual arrival time at receiving stop and the receiving buses' schedules. If the passenger's actual arrival time is before the scheduled bus's departure time (DD = 0), there will be no additional time penalty; if the actual arrival time is after the *nth* bus's departure time (DD = n), which can be noted as *n-th* receiving bus, then there is an additional time penalty which is worth sum of *n* receiving buses' headways.

Beyond a single transfer's time penalty, we can expand the measure to a collection of transfers. The collection can have different spatiotemporal definitions depending on different purposes, such as transfers between two routes during an hour every day, or transfers at a stop during a year. We can measure the *average total transfer time penalty* (ATTP) for a collection of transfers C:

$$ATTP_C = E_C(TTP_i) = \frac{1}{|C|}\sum_{i=1}^{|C|} TTP_i \tag{3}$$

where: TTP_i is the measurement of transfer i's total time penalty.

Transfers: The good, the bad, and the ugly. We classify all transfers into three types according to their real-time synchronization performance. We can distinguish them by the receiving bus's desynchronization degree.

- *The good: normal transfers* (DD = 0). A passenger getting on a normal transfer will catch the same bus as the scheduled transfer. Under this circumstance, ATP = 0, which means there is no additional time penalty, while the performance can be still different from the schedule due to the normal delay of the receiving trip.
- *The bad: missed transfers* (DD > 0). The passenger will take a bus after the scheduled bus, hence will suffer from additional time penalty other than normal delay. Under this circumstance, ATP > 0. The missed transfers have several scenarios: 1) generating trip is delayed that the user cannot catch the scheduled receiving bus; 2) the scheduled receiving bus is out of service; 3) the scheduled receiving bus is severely delayed after another receiving bus. Scenario 1 is the most common circumstances. For scenario 2, if the scheduled receiving trip is no longer running, the passenger must take the next bus. Likewise, for scenario 3, a severely

delayed bus can be caught up by another bus on the same route scheduled after it. It is natural for users to take the closest bus despite the buses being out of sequence.

- *The ugly: preemptive transfers* (DD < 0). During a preemptive transfer, instead of the scheduled bus, the user will get on a bus which should have arrived earlier than the passenger at the receiving stop. This is due to delays in the receiving buses. The passenger will naturally take the nearest bus regardless of the schedule. The ATP's value can be negative, zero or positive, however, a negative ATP will not necessarily suggest a better performance since the TTP can be positive meanwhile. In fact, a preemptive transfer's TTP does not guarantee to be worse than a normal transfer; it may achieve better, same, or worse performance depending during the synchronization process. With some license, we refer to these as "ugly" since they are ~~they are~~ unpredictable with respect to impact.

Accordingly, we measure each transfer with a binary value t that represents whether it is a missed transfer. Based on the assessment of single transfer, we define *transfer risk*: it is the proportion of missed transfers in a collection of transfers, based on the empirical schedule and real-time vehicle location data; we can interpret this as an empirical probability of a missed transfer in this collection. Mathematically, in a collection C, the transfer risk is:

$$TR_C = E_C(t) = \frac{1}{|C|}\sum_{i=1}^{|C|} t_i \tag{4}$$

$$t_i = \begin{cases} 1, if\ DD_i > 0 \\ 0, \text{otherwise} \end{cases}$$

where: C is the collection of transfers and t_i is the binary measurement variable of each transfer.

3.3. Determining valid transfers

There are four policies for transfer scheduling in public transit systems: 1) Unscheduled transfers; 2) Scheduled transfers without vehicles waiting; 3) Single holding strategy that lower frequency vehicles wait for higher frequency vehicles; 4) Double holding transfer that both vehicles hold for transfers (Ceder, 2016; Knoppers and Muller, 1995). Many transit authorities, especially those that rely on buses, use an unscheduled transfer policy, meaning there are few explicitly scheduled transfers in the GTFS static data. Moreover, in reality, transit users' transfer behavior and transit real-time apps will not strictly follow the scheduled transfers.

Consequently, we have to search empirically for possible transfers from the GTFS static data. Theoretically, any two trips at two stops which are proximal enough for users to access can be regarded as a valid transfer. This can be refined with passenger data that shows actual transfers; this is likely to be a subset of the valid transfers. However, the danger with this approach is we may miss a potential transfer if it did not occur in the data.

Based on the data structure in the GTFS data, we define three levels of aggregation: *stop*, *route*, and *trip*. Every trip is run according to a fixed schedule by a bus at a specific time. Trips with a same schedule can be aggregated into a route, and some routes can be bound to a stop. To find transfer schedule from GTFS schedule, we developed a hierarchical searching algorithm in the Python and MongoDB environment. Using the algorithm, we derived all possible stops combinations, route combinations, and GTFS trip combinations. Only those combinations with near distance (Euclidean distance < 100 meters) and unique routes are selected for the transfer schedule: if there are multiple transfers with the same route combination and same generating stop, we ~~will~~ only keep the one with closest walking distance to remove some redundancy, since passengers are most likely to walk to closest stop for a transfer.

4. Analysis

We conducted a case study using data from Central Ohio Transit Authority (COTA) bus system in Columbus, Ohio from February 2018 to January 2019. We acquired the General Transit Feed Specification (GTFS) schedule and real-time data via the COTA public application programming interface. COTA shared the Automated Passenger Count (APC) dataset from May 2018 to January 2019. When merging the APC and GTFS datasets, 45.06% of the total records were matched on average, meaning that roughly half of the GTFS data was updated with the more accurate APC data. Based on the GTFS alone, the average transfer risk (TR) over the study period is 7.14% ($\sigma = 25.75\%$) and the average total time penalty (ATTP) is 3.74 ($\sigma = 12.97$) minutes; based on the merged APC-GTFS, the average TR is 8.55% ($\sigma = 27.96\%$) and the ATTP is 4.57 ($\sigma = 15.44$) minutes. Although the mean value is relatively small, ~~however,~~ the standard deviation is substantially large, which suggests the temporal and spatial variation is large. This suggest that GTFS data alone underestimates transfer risk and time penalties, although it provides reasonable estimates.

We archived the data using a MongoDB database. The GTFS real-time data, APC data, and their auxiliary databases total nearly one terabyte. Due to large database size, we optimized and parallelized our code to deal with the computational burden. We also developed different summary measures based on varying spatial or temporal aggregations.

4.1. Spatial patterns

To investigate the spatial pattern of transfer risk, the first thing is spatial aggregation, since *trip patterns* (each vehicle trip; the finest level of resolution) are too specific and not representative of broader patterns. We can aggregate in different ways. Naturally, *route patterns* are useful, which aggregate the trip combinations based on their route schedules. *Stop patterns* are also useful since the quality of transfers between stops is assessed and stop combinations are geographically distinguishable, making it especially crucial for visualization. We concentrate on stop patterns in our analyses.

Figure 2 shows the spatial pattern of the TR and ATTP. It shows some differences between TR and ATTP's spatial distribution, especially on High Street (the major north-south thoroughfare in Columbus, indicated by a ~~red circle in Figure 2), and the northern downtown area (indicated by a green rectangle~~ in Figure 2). Stops among High Street has relatively higher transfer risk but also have relatively lower average total time penalty. This is likely due to traffic and other disturbances on this route elevating the risk, although headway between buses is short meaning the time penalty is small. Similarly, the high ATTP clusters on some roads in downtown area and some peripheral roads that do not have higher transfer risk. Although the transfer risk is low, the time delay can be high, especially for downtown, due to longer headways.

- **Insert Figure 2**

4.2. Temporal patterns

We now examine aggregate temporal patterns of transfer risk and time penalties. Figure 3 provides the monthly trends of transfer risk (TR) and average total time penalty (ATTP) for both datasets. July, December, and January show an overall low time penalty pattern. This can be due to better overall traffic conditions during summer and holiday season vacation. August is the worst month to take a transfer; this may be due to the start of an academic year in a city with a massive university campus near the city center.

- **Insert Figure 3**

Figure 4 provides the aggregate trends by day of the week and frequency. We can see the overall TR and ATTP peak on Friday; Wednesday, Thursday, and Friday exhibit higher levels of risk and time penalties, likely due to the overall traffic pattern in this city. Both measures are relatively low on weekends, as would be expected due to lower traffic congestion. TR and ATTP are relatively low on Mondays, possibly due to flexible working schedule and long weekends for some residents, leading to less commuting. However, we observe Sundays have the lowest ATTP. Intuitively, frequency can be a significant factor accounting for the measures. We conducted a Pearson correlation analyses, ATTP and TR and found no significant correlation with daily frequency: p-values are 0.38 for ATTP and 0.118 for TR.

- **Insert Figure 4**

Figure 5 illustrates the hourly trend and there are three major time periods when transfer risk and penalties are high: mornings (8:00–10:00), afternoon (17:00–19:00), and night hours (22:00–24:00). High risk and penalties during the morning and afternoon periods can be explained by overall traffic pattern during these busy hours. However, nighttime with lower traffic also displays high transfer risk and high total time penalty. At night, as the transfer risk increases and service frequency decreases, the time penalties are higher due to sparser scheduled service. In terms of frequency impact, according to the Pearson correlation analyses between each measure and hourly frequency shown in Figure 6, ATTP has significant negative correlation with the frequency, while TR has no significant correlation with the frequency.

- **Insert Figures 5 and 6**

4.3. Simulating the impacts of dedicated bus lanes

Dedicated bus lanes (DBL) can provide benefits for a bus system by reducing delays due to automobile traffic. Without the disturbance of traffic congestion, bus rapid transit systems with separated DBL can achieve rail-like performance (Li et al., 2009). We simulated the impact of DBL on delays, transfer risk and time penalties using the methods in this paper.

We selected the COTA (Central Ohio Transit Authority) bus route No.2 as the target, which has the most transfers and most ridership in the system. We simulate the impact of a DBL by assume all the buses running on this route will behave according to the GTFS static schedule data after DBL is in effect (i.e., no delay). This assumption is hypothetical, however, the results can be an upper bound on ~~the~~ actual DBL performance. We analyze TR and ATTP's changing trend before and after applying the assumption and the difference's spatial and temporal pattern. Across all stops on the route, the DBL will save 1.72 minutes ($\sigma =$ 10.09 minutes) and Kolmogorov–Smirnov (KS) test shows the two scenarios have significantly different distributions (p-value = 0.005). Therefore, although the average time savings is modest, the impacts are statistically significant and highly differentiated across stops.

- **Insert Figure 7**

Also, we calculated different impacts on the generating trips and receiving trips. We categorized all affected transfers into two classes: transfers with generating trip on the

DBL (*DBL-generating transfers*) and transfers with receiving trip on the DBL (*DBL-receiving transfers*). DBL will save DBL-generating transfers 2.25 minutes and 5.25% transfer risk while only save DBL-receiving transfers 0.32 minutes and increase 9.03% transfer risk. The KS tests between the two types of transfers show significant differences for both measures (p-value $<10^{-14}$). This suggest that the DBL will eliminate delays for all transfers thus decrease all transfers' total time penalty universally; but will simultaneously decrease DBL-generating transfers' risk while increasing DBL-receiving transfers' risk, however, it will not necessarily enlarge its time penalty. Based on this simulation, we conclude that improving punctuality via a DBL can reduce ATTP, and DBL-generating transfers will benefit more than DBL-receiving transfers.

5. Conclusion

Big data creates an unprecedented opportunity for more and deeper understanding of ~~the~~ urban public transit systems and the study of transfers. However, due to the lack of attainable big data sources, few studies focus on the transfers' on-time performance in the real-time context. Based on high-resolution GTFS and APC real-time and static data of huge volume, we developed transfer risk (TR) and average total time penalty (ATTP) to assess ~~the~~ transfers~~'~~ performance. TR and ATTP indicate the systematic quality of transfers and corresponding potential time cost. These measures provide important information for ~~the~~ transit system planners and administrators concerning ~~the~~ transfers~~'~~ feasibility, quality, and user experience. The spatial and temporal analysis show similar pattern like overall traffic and transit system delay, while it also shows some unique patterns, such as high time penalty during the nighttime due to larger headway. Additionally, we simulated dedicated bus routes' impact on ~~the~~ transfer performance. It suggests even a single route DBL can reduce ATTP, especially for DBL-generating transfers.

With the support of big data, the TR and ATTP measures are a further step towards sustaining ~~a~~ smarter public transit system. Compared with existing indexes and measuring systems, the spectrum of the proposed measures' audience is broad: besides academic and administrating purposes, ordinary passengers and open source developers can also be potential users. Thanks to ~~the~~ high-resolution public transit big data, we can calculate corresponding performance based on specific transfers as well as overall broad patterns:

- At the application level, urban dwellers can query each transfer's performance in their real-time transit apps and react correspondingly. Current mainstream transit apps do not show empirical risk and average time loss, especially for transfers which users have no control. If a proposed transfer's empirical performance is shown when the apps plan the trip, urban dwellers can avoid high risk routes. This is similar to airlines apps showing the on-time performance of air routes. Unlike some composite indexes that are hard to conceptualize, TR and ATTP are ~~all~~ intuitive since they use common metrics, namely probabilities and time.
- At the management level, administrators can check the high risk and high time penalty areas and respond. With support of real-time data and the measures, ~~the~~ transit authorities can make operational changes such as adding additional buses, enforcing bus's time table to reduce transfer risk, and planning flexible time table adjustment accordingly. City planners can analyze ~~the~~ spatiotemporal patterns of risk and time penalties. The patterns of proposed measures can demonstrate important information about ~~the~~ roads, transit system design and other transport and non-transport factors.

- At the policy-making level, policy makers can compare different public transit systems' transfer real-time performance across the US. Due to the high reusability and expandability of the indexes and the system, they can be easily implemented and applied to any transit system with published GTFS scheduled and real-time data without major modification. The common metrics also make intra-system and inter-system comparison much easier.

Future research direction can concentrate on the application of both smart and human sensors, generating abundant and high-resolution big data for analysis. In this paper, we compared datasets of different temporal accuracy, nevertheless, we do not have a good answer for how spatial accuracy will influence the results and how the overall impact of inaccuracy can be decomposed into the two factors. It may be useful to utilize ~~a~~ third-party data to calibrate ~~the~~ GTFS data, so that GTFS data can achieve higher accuracy. Volunteered data from humans is also a possible strategy for collecting data. Based on more precise and abundant data, there are more possibilities for more scientific planning, improvement, and knowledge derivation of transfer activities and the transit system. Moreover, we do not consider population and ridership factors; with transfer ridership data, we can incorporate these factors into the system.

References

Antrim A and Barbeau SJ (2013) The Many Uses of GTFS Data–Opening the Door to Transit and Multimodal Applications. *Location-Aware Information Systems Laboratory at the University of South Florida* 4.

Bovy PH and Stern E (2012) *Route Choice: Wayfinding in Transport Networks: Wayfinding in Transport Networks*. Springer Science & Business Media.

Ceder A (2016) *Public Transit Planning and Operation: Modeling, Practice and Behavior, Second Edition*. *Public Transit Planning and Operation: Modeling, Practice and Behavior, Second Edition*. CRC press.

Chen M, Mao S and Liu Y (2014) Big data: A survey. *Mobile networks and applications* 19(2). Springer: 171–209.

Chu X (2010) *A Guidebook for Using Automatic Passenger Counter Data for National Transit Database (NTD) Reporting*. National Center for Transit Research (US).

Dessouky M, Hall R, Nowroozi A, et al. (1999) Bus Dispatching at Timed Transfer Transit Stations Using Bus Tracking Technology. *Transportation Research Part C: Emerging Technologies* 7(4). Elsevier: 187–208.

Firmani D, Mecella M, Scannapieco M, et al. (2016) On the Meaningfulness of "Big Data Quality". *Data Science and Engineering* 1(1). Springer: 6–20.

Google Developers (2016) GTFS Static Overview | Static Transit | Google Developers. Available at: https://developers.google.com/transit/gtfs/ (accessed 8 March 2018).

Google Developers (2018) Trip Updates. Available at: https://developers.google.com/transit/gtfs-realtime/guides/trip-updates (accessed 8 April 2019).

Guo Z and Wilson NHM (2004) Assessment of the Transfer Penalty for Transit Trips: Geographic Information System-based Disaggregate Modeling Approach.

Transportation Research Record 1872(1872). Transportation Research Board of the National Academies: 10–18. DOI: 10.3141/1872-02.

Guo Z and Wilson NHM (2011) Assessing the Cost of Transfer Inconvenience in Public Transport Systems: a Case Study of the London Underground. *Transportation Research Part A: Policy and Practice* 45(2). Pergamon: 91–104.

Hadas Y and Ranjitkar P (2012) Modeling Public-transit Connectivity With Spatial Quality-of-transfer Measurements. *Journal of Transport Geography* 22. Elsevier: 137–147. DOI: 10.1016/j.jtrangeo.2011.12.003.

Hilbert M (2016) Big Data for Development: A Review of Promises and Challenges. *Development Policy Review* 34(1). Wiley Online Library: 135–174.

Jang W (2010) Travel Time and Transfer Analysis Using Transit Smart Card Data. *Transportation Research Record* (2144). Transportation Research Board of the National Academies: 142–149.

Knoppers P and Muller T (1995) Optimized Transfer Opportunities in Public Transport. *Transportation Science* 29(1). INFORMS: 101–105.

Kujala R, Weckström C, Mladenović MN, et al. (2018) Travel Times and Transfers in Public Transport: Comprehensive Accessibility Analysis Based on Pareto-optimal Journeys. *Computers, Environment and Urban Systems* 67. Elsevier: 41–54.

Li JQ, Song M, Li M, et al. (2009) Planning for Bus Rapid Transit in Single Dedicated Bus Lane. *Transportation Research Record* (2111). Transportation Research Board of the National Academies: 76–82.

Liu R, Pendyala RM and Polzin S (1997) Improved Reliability of Public Transportation Using Real-time Transfer Synchronization. *Transportation Research Record* (1607). Transportation Research Board of the National Academies: 74–80.

Miller HJ and Goodchild MF (2015) Data-driven Geography. *GeoJournal* 80(4). Springer: 449–461.

Nesheli MM and Ceder A (2015) Improved reliability of public transportation using real-time transfer synchronization. *Transportation Research Part C: Emerging Technologies* 60. Pergamon: 525–539.

Nishiuchi H, Todoroki T and Kishi Y (2015) A Fundamental Study on Evaluation of Public Transport Transfer Nodes by Data Envelop Analysis Approach Using Smart Card Data. *Transportation Research Procedia* 6. Elsevier: 391–401.

Sun L, Rong J, Ren F, et al. (2007) Evaluation of Passenger Transfer Efficiency of an Urban Public Transportation Terminal. In: *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2007, pp. 436–441. IEEE.

Sun L, Rong J and Yao L (2010) Measuring Transfer Efficiency of Urban Public Transportation Terminals by Data Envelopment Analysis. *Journal of Urban Planning and Development* 136(4). American Society of Civil Engineers: 314–319.

Transit Wiki (2019) Automated Passenger Counter. Available at: https://www.transitwiki.org/TransitWiki/index.php/Automated_passenger_counter (accessed 10 February 2019).

Walker J (2012) *Human Transit: How Clearer Thinking about Public Transit Can Enrich*

Our Communities and Our Lives. *Human Transit: How Clearer Thinking About Public Transit can Enrich our Communities and our Lives*. Island Press.

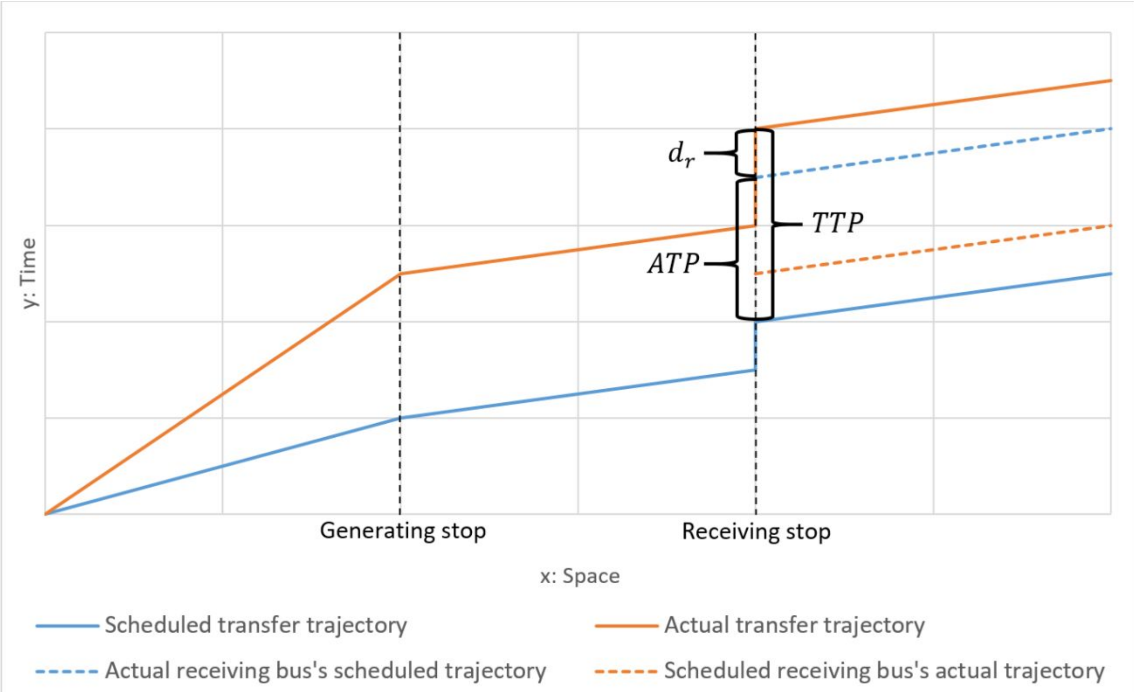

Figure 1: Time-space diagram of a delayed transfer and the corresponding scheduled transfer

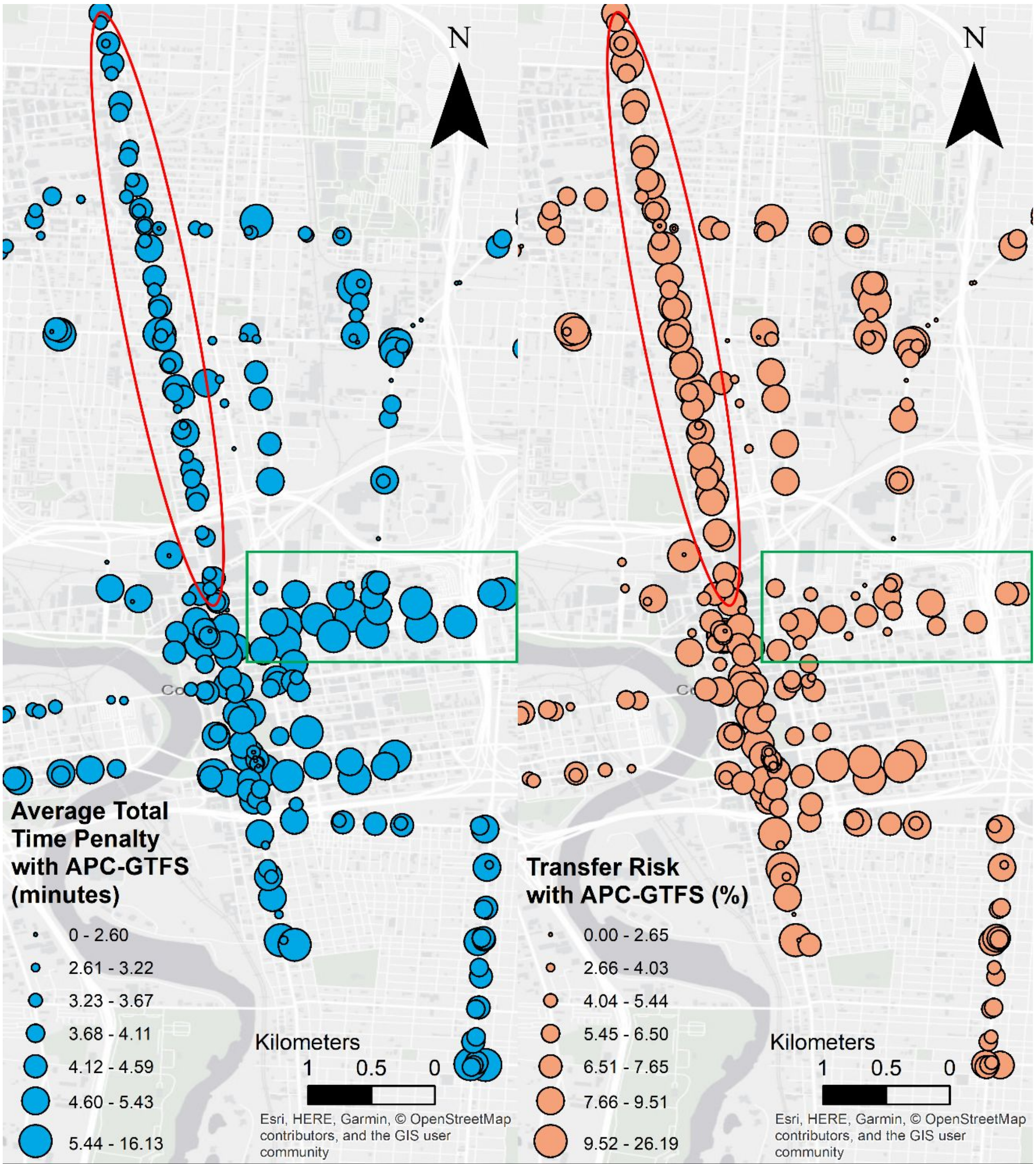

Figure 2: Spatial pattern of ATTP and TR (in quantile classification)

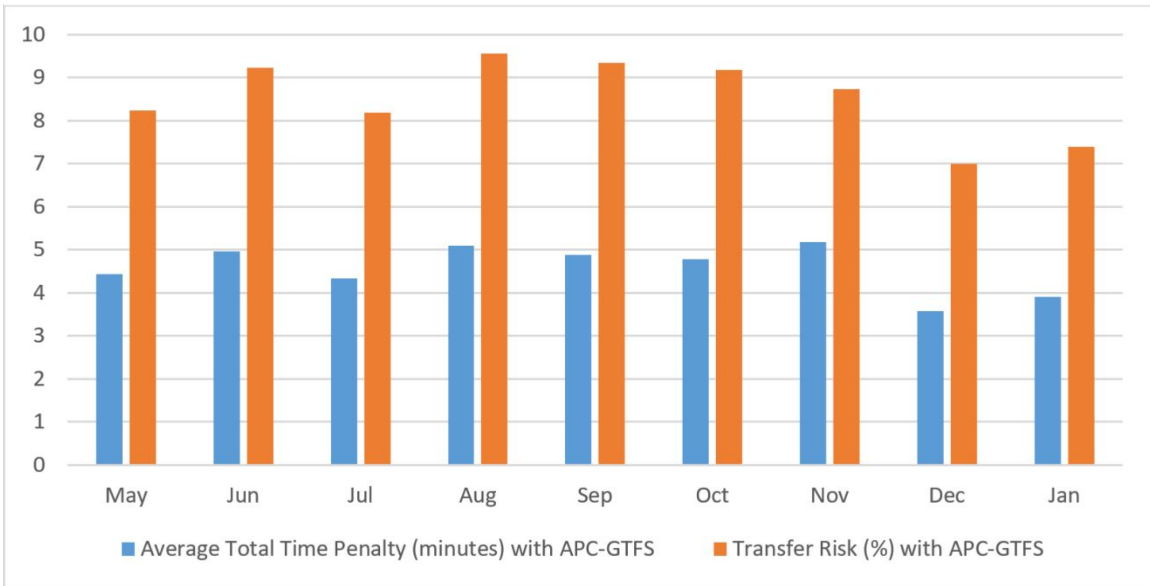

Figure 3: Overall monthly TR and ATTP trends.

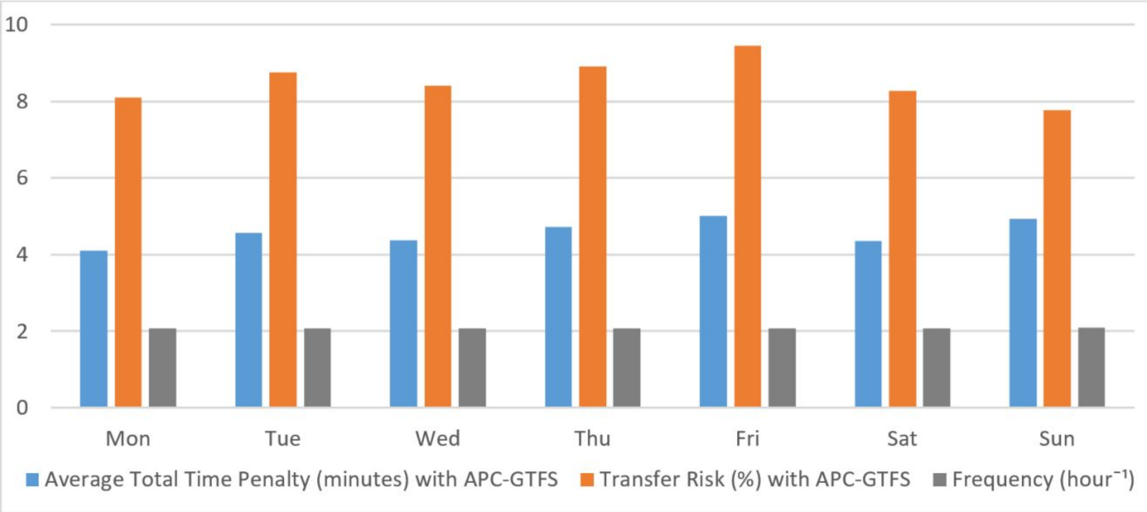

Figure 4: Overall weekday TR and ATTP trends and daily frequency.

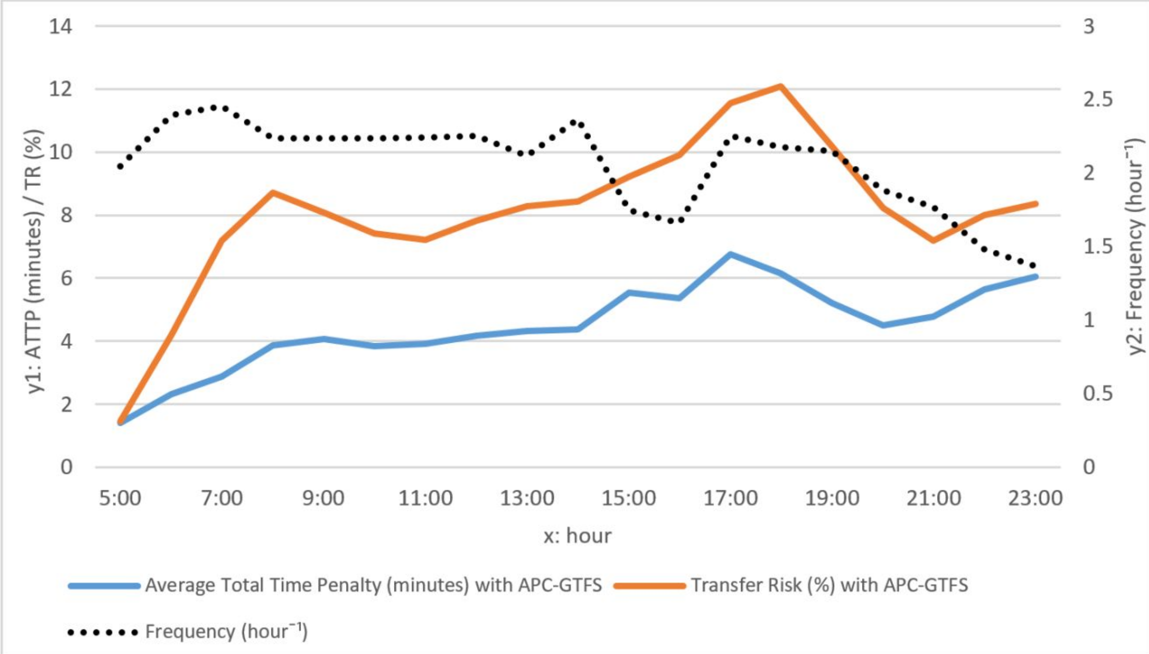

Figure 5: Overall hourly TR and ATTP trends.

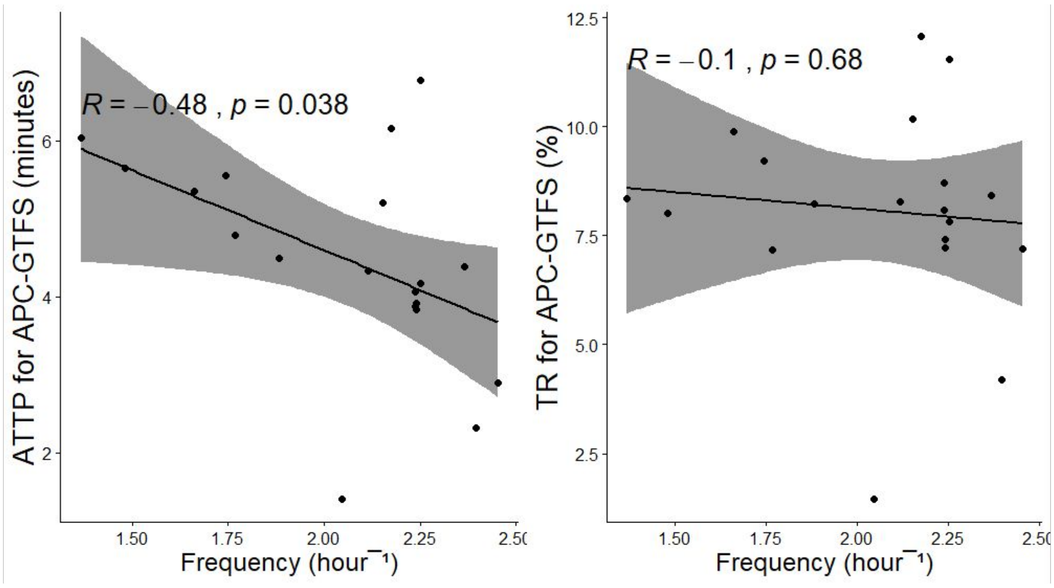

Figure 6: Scatter plots of ATTP (left side) and TR (right side) versus frequency.

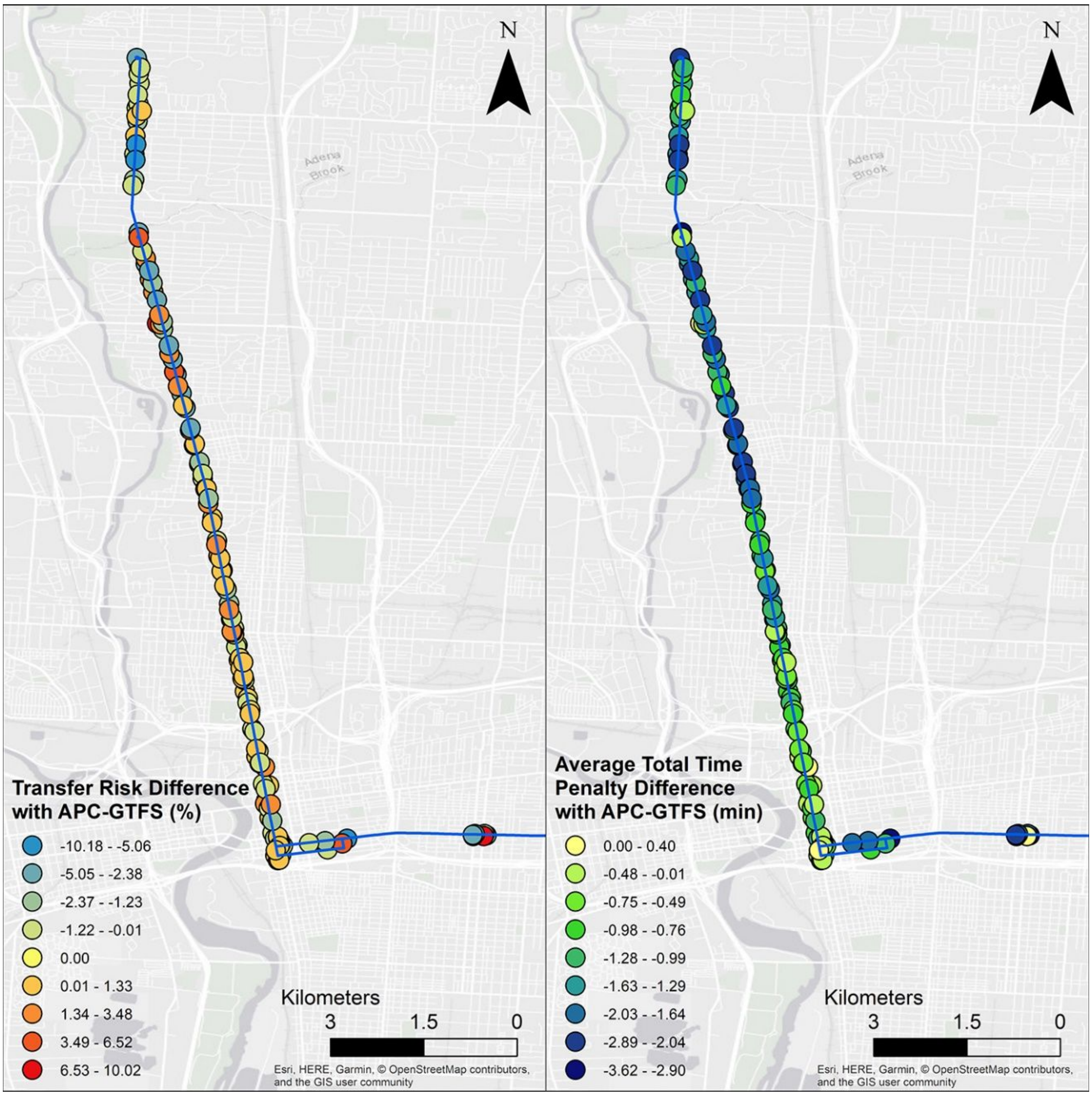

Figure 7: TR and ATTP difference after simulated implementation of a dedicated bus lane.