To What Extent Do Ride-Hailing Services Replace Public Transit? A Novel Geospatial, Real-Time Approach Using Ride-Hailing Trips in Chicago

Helena K. Breuer

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Civil Engineering

Hesham A. Rakha, Chair
Jianhe Du
Kevin P. Heaslip

August 28, 2020
Blacksburg, Virginia

Keywords: ride-hailing, ridesharing, public transit, Uber, Lyft, replacement, utility

To What Extent Do Ride-Hailing Services Replace Public Transit? A Novel Geospatial, Real-Time Approach Using Ride-Hailing Trips in Chicago

Helena K. Breuer

ABSTRACT

Existing literature on the relationship between ridehailing (RH) and transit services is limited to empirical studies that rely on self-reported answers and lack spatial and temporal contexts. To fill this gap, the research takes a novel approach that uses real-time geospatial analyzes. Using this approach, we estimate the extent to which ride-hailing services have contributed to the recent decline in public transit ridership.

With source data on ridehailing trips in Chicago, Illinois, we computed the real-time transit-equivalent trip for the 7,949,902 ridehailing trips in June 2019; the sheer size of this sample is incomparable to the samples studied in existing literature. An existing Multinomial Nested Logit Model was used to determine the probability of a ridehailer selecting a transit alternative to serve the specific origin-destination pair, $P(Transit|CTA)$[1].

The study found that 31% of RH trips are replaceable, 61% are not replaceable, and 8% lie within the buffer zone. We measured the robustness of this probability using a parametric sensitivity analysis, and performed a two-tailed t-test, with a 95% confidence interval. In combination with a Summation of Probabilities, the results indicate that the total travel time for a transit trip has the greatest influence on the probability of using transit, whereas the airport pass price has the least influence. Further, the walk time, number of stops in the origin and destination census tracts, and household income also have significant impacts on the probability of using transit. Lastly, we performed a Time Value Analysis to explore the cost and trip duration difference between RH trips and their transit-equivalent trips on the probability of switching to transit. The findings demonstrated that approximately 90% of RH trips taken had a transit-equivalent trip that was less expensive, but slower.

The main contribution of this study is its thorough approach and fine-tuned series of real-time spatial analyzes that investigate the replaceability of RH trips for public transit. The results and discussion intend to provide perspective derived from real trips and encourage public transit agencies to look into possible opportunities to collaborate with ridehailing companies. Moreover, the methodologies introduced can be used by transit agencies to internally evaluate opportunities and redundancies in services. Lastly, we hope that this effort provides proof of the research benefits associated with the recording and release of ridehailing data.

---

[1] This value defines the replaceability of the transit-equivalent trip, where the value ranging from 0 to 0.45 indicates the transit trip is not-replaceable (NR), and a value ranging from 0.55 to 1.0 indicates the transit trip is replaceable (R).

To What Extent Do Ride-Hailing Services Replace Public Transit? A Novel Geospatial, Real-Time Approach Using Ride-Hailing Trips in Chicago

Helena K. Breuer

GENERAL AUDIENCE ABSTRACT

Existing literature on the relationship between ridehailing (RH) and transit services is limited to empirical studies that rely on self-reported answers and lack spatial and temporal contexts. To fill this gap, the research takes a novel approach that uses real-time geospatial analyzes. Using this approach, we estimated the extent to which ride-hailing services have contributed to the recent decline in public transit ridership.

With source data on ridehailing trips in Chicago, Illinois, we computed the real-time transit-equivalent trip for the 7,949,902 ridehailing trips in June 2019; the sheer size of this sample is incomparable to the samples studied in existing literature. An existing Multinomial Nested Logit Model was used to determine the probability of a ridehailer selecting a transit alternative to serve the specific origin-destination pair, $P(Transit|CTA)$[2].

The study found that 31% of RH trips are replaceable, 61% are not replaceable, and 8% lie within the buffer zone. We measured the robustness of this probability using a parametric sensitivity analysis, and performed a two-tailed t-test, with a 95% confidence interval. In combination with a Summation of Probabilities, the results indicate that the total travel time for a transit trip has the greatest influence on the probability of using transit, whereas the airport pass price has the least influence. Further, the walk time, number of stops in the origin and destination census tracts, and household income also have significant impacts on the probability of using transit. Lastly, we performed a Time Value Analysis to explore the cost and trip duration difference between RH trips and their transit-equivalent trips on the probability of switching to transit. The findings demonstrated that approximately 90% of RH trips taken had a transit-equivalent trip that was less expensive, but slower.

The main contribution of this study is its thorough approach and fine-tuned series of real-time spatial analyzes that investigate the replaceability of RH trips for public transit. The results and discussion intend to provide perspective derived from real trips and encourage public transit agencies to look into possible opportunities to collaborate with ridehailing companies. Moreover, the methodologies introduced can be used by transit agencies to internally evaluate opportunities and redundancies in services. Lastly, we hope that this effort provides proof of the research benefits associated with the recording and release of ridehailing data.

[2] This value defines the replaceability of the transit-equivalent trip, where the value ranging from 0 to 0.45 indicates the transit trip is not-replaceable (NR), and a value ranging from 0.55 to 1.0 indicates the transit trip is replaceable (R).

Dedication

I would first like to thank Dr. Rakha for providing the opportunity to pursue a master's degree while performing research in CSM at VTTI. Moreover, I would like to thank Dr. Du, who served as a reference point for guidance for this project and furthered my experience with Matlab and ArcGIS. Secondly, I would like to thank all of my professors and mentors, especially Dr. Heaslip, Dr. Katz, and Dr. Scardina, who contributed to my personal and professional growth by believing in me and guiding and helping me during the completion of my master's degree. Lastly, I would like to thank my parents and my best friends for supporting my endeavors and being there for me when I needed them the most.

Acknowledgements

This effort was funded by the Urban Mobility and Equity Center (UMEC).

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AirPass: airport pass [price]

BF: base fare

CBD: central business district

CTA: Chicago Transit Authority

D: destination

FLM: first- and last-mile [arrangement]

GTFS: General Transit Feed Specification [dataset]

HHI: household income (average)

IVTT: in-vehicle travel time

LOS: level of service

MNL: multinomial nested logit [model]

NR: not replaced trip group

O: origin

O-D: origin-destination [pair]

OVTT: out-of-vehicle travel time

R: replaced trip group

RH: ride-hailing, ride-hailing services

SiT: stops in tract (number of transit stops per census tract)

TC: transfer cost

TNC: transportation network company

TTT: total travel time

WT: walk time

CHAPTER 1: INTRODUCTION

History of Ridehailing Services

The nascent ridehailing[3] (RH) market was first introduced to the United States in 2008 when Travis Kalanick and Garret Camp cofounded their company, *Uber*. Two years later, the company released its beta version and began services in San Francisco [1]. In 2012, an existing carpooling company, *Zimride*, launched a competing ridehailing service in San Francisco as well. By 2013, Zimride sold its carpooling business and renamed itself *Lyft*, to exclusively operate as a ridehailing service [2]. Over the next two years, competition heightened as Uber expanded to 60 cities across six continents, and Lyft announced its plan to expand to 24 more cities, totaling coverage of over 60 cities [2, 3]. As of January 2019, nearly a decade later, 36% of US adults have used or currently use ridehailing services [4].

Ridehailing is best defined as an app-based, on-demand transportation service that provides customers with door-to-door transportation for a single trip [5]. Through the company's smartphone app, a customer enters a specific pick-up and drop-off location (O-D pair). On the backend, the RH company's algorithm calculates an appropriate route and trip fare. It then selects the optimal driver to service the trip – this selection is based off of availability and range from the requested pick-up location. Once the algorithm selects a driver, the customer is notified of the estimated pick-up time, and vehicle/driver details. The company's drivers operate on their own schedule, independent from the company, and use their own vehicle.

Inherently, the novelty of ridehailing services brings about concern regarding its impact on the existing transportation network. Critics argue that ridehailing services have first taken ridership from a similar service, taxis [6]; the second service argued to be negatively impacted is public transit, which is explained in the next paragraph. While ridehailing and taxis share the concept of providing customers with private transportation, the novelty of ridehailing is attributed to its advantageous flexibility, real-time location data, availability (outside of cities) and ease of payment. The first notable difference is the ability for customers to view the trip fare and pay the fare through the app prior to placing an order. Once the trip is ordered, the cost of the trip cannot be changed regardless of any in-route deviations. This process is seamless and has increased convenience when compared to calling a taxi. Secondly, the app allows customers to plan a trip ahead or in the moment and view the location of the driver in real time. This allows for greater trip security and provides increased convenience for the rider's pre-trip agenda. Lastly, the rider has the opportunity to 'cancel' an order prior to pick-up. In contrast to the whole ridehailing process, traditional taxi's must be hailed curbside and the fare is unknown prior to the ride. These features of ridehailing services are advantageous in situations demanding flexibility or security, such as inclement weather or planned events.

Ridehailing and Public Transit

Coincidentally, when the ridehailing market began rapidly gaining traction through geographic expansion and increased acceptance in 2014, average public transit ridership in the United States began its decline. In the early twenty-first century, transit ridership in the United States experienced two periods of growth followed by decline (*Figure 1*).

[3] Within the existing literature, *ridehailing* is more commonly known as "ridesharing" but because it entails 'hailing' a ride which is not necessarily shared, *ridehailing* is most appropriate.

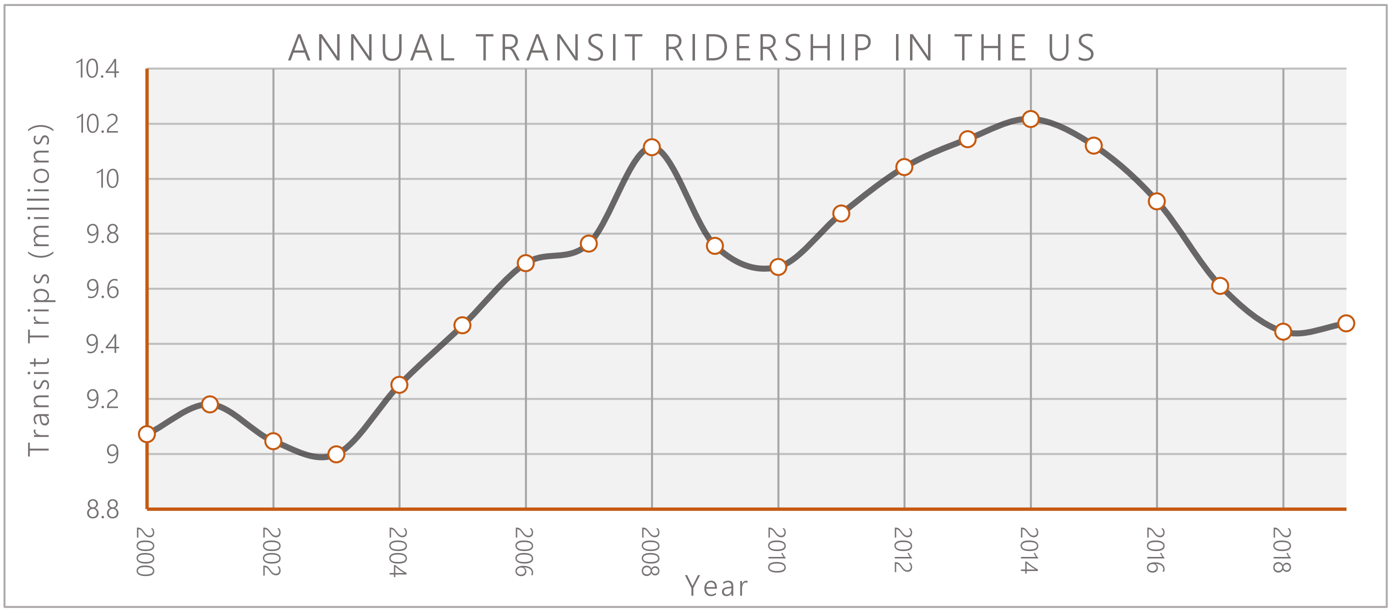


Figure 1 – Annual Public Transit Ridership in the United States from 2000-2019. Source Data: APTA Ridership by Mode and Quarter 1990-Present [7]. Annual ridership counts are the sum of bus, light rail, commuter rail, and heavy rail trips.

On average, from 2003 to 2008, public transit ridership in the United States increased by 2.58% each year. Following the 2008 economic recession, ridership levels took a downturn until 2010 when ridership began increasing again until 2014. Although, unlike the first period of growth, this growth rate was decreasing in magnitude each year until it plateaued in 2014. At this point, transit ridership began rapidly decreasing by losing more riders per year until 2019. While these statistics measure the national trend in mode-choice behavior, the trends within metropolitan transit agencies vary by year and mode. Nonetheless, continuous decline in ridership is significant and indicative of a disturbance to the market.

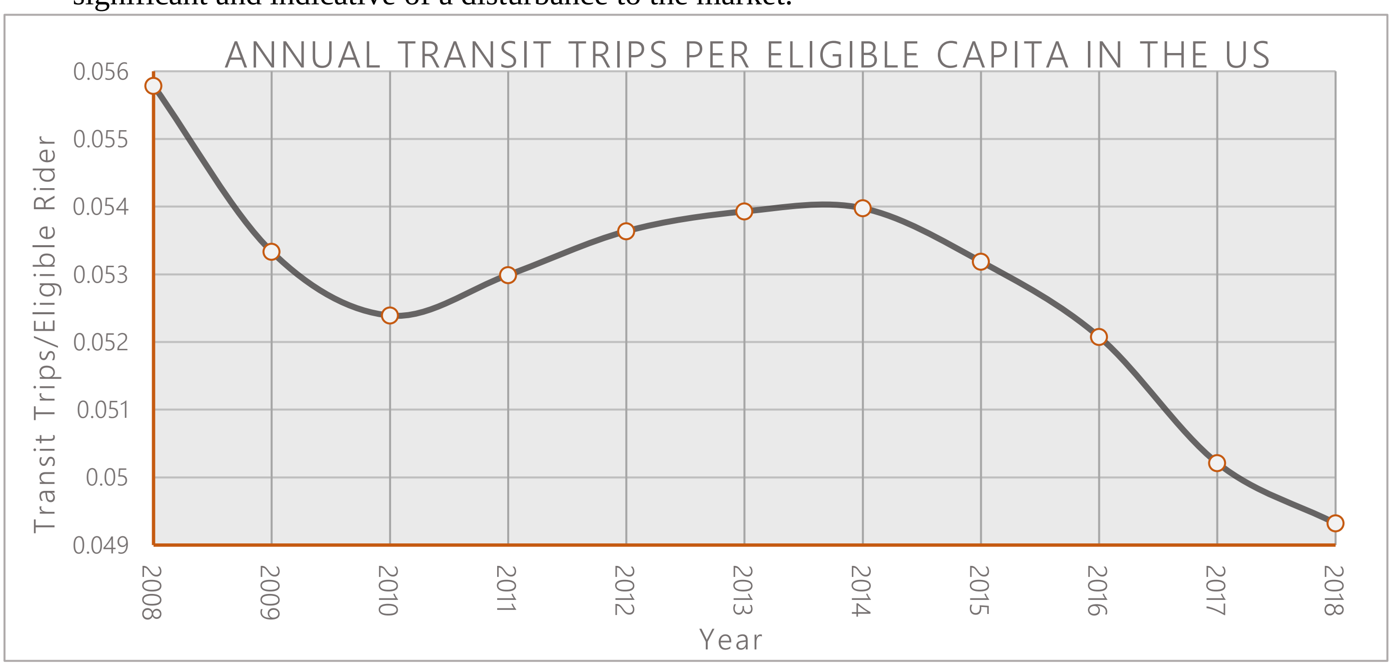


Figure 2 – Annual Transit Trips per Eligible Rider in the United States. Source Data: APTA Ridership by Mode and Quarter 1990-Present, and Population by Age from KFF

From 2008 to 2018, the population of eligible riders[4] increased by 5.63%. Yet, the annual transit trip per eligible rider decreased by 11.6%, as shown in Figure 2 above. More specifically, the transit ridership per eligible rider decreased by 8.6% from 2014 to 2018. Despite the steady growth of the US population, transit ridership does not reflect that.

Historically, declines in transit ridership can be a result of macroeconomic, geographic, and demographic changes in a region. The first period of ridership decline in the 21st century started in 2008 and was evidently a ramification of the economic recession. Yet, the cause(s) of the most recent decline is not as discernable. Further, this period exhibited a larger decline in magnitude and has spanned 5 years, as opposed to 2 years. So, what could have possibly caused a more crippling effect on transit ridership than the economic recession? Was the second decline in transit ridership a result of an emerging alternative mode, *ridehailing*?

The main obstacle in answering this question is deciphering if a ridehailer took a trip that was originally going to be serviced with transit. Under this condition, public transit would be considered "replaced" by ridehailing. However, if during the decision-making process, the individual did not consider public transit as a feasible mode of travel, then theoretically it was not in competition with the mode choices.

Based on spatial analysis, many ridehailing trips could be serviced by transit, with variance in travel times and access and egress distances. However, there is no guarantee that the trip experiences will be comparable; many ridehailers choose RH services because of its increased reliability, convenience, and cleanliness [9]. Ridehailing trips can be characterized by two sets of attributes: level of service (LOS) and individual preferences. LOS attributes are defined by quantifiable trip metrics such as trip length, duration, access and egress time, fare/cost, wait time, and walking time. Additionally, individualistic service metrics are likely to also vary, such as level of comfort, ease of payment process, cleanliness. [9]. These factors and preferences are all important in analyzing an individual's mode selection process. Large scale measurements of these preferences are not publicly available, hence, it has been challenging for researchers to probe the actual relationship between ridehailing services and public transit. In the existing literature, most researchers explore the relationship through empirical studies thus creating a gap in literature based on the source data.

In this study, ridehailing trip source data from the City of Chicago is used to explore the similarities and determine the trends in overlap between ridehailing trips and their equivalent transit trips. If O-D ridehailing pairs have a viable public transit equivalent trip with sufficient utility, we can infer that the Chicago Transit Authority (CTA) has an opportunity to gain ridership in these areas.

Thesis Contribution

This thesis will address the literature gap by identifying the replicability of transit by RH services through non-empirical based methods. In this study we use RH trip source data from the City of Chicago containing over 8,000,000 trips, hence the contributions of this research are expanded by the use of a relatively large dataset. We use a selection of ArcGIS spatial analyses

[4] Eligible riders: total US populations between age 18 and 64 [8] "Population by Age," 2008-2018 ed. Online: Kaiser Family Foundation, 2018.

and methodologies to deliver a real-time transit-equivalent route. These attributes of these trips are input into our selected utility model, from which we will calculate the corresponding probability that an original ride-hailer would select transit over RH. To explore the results in greater depth, we conducted a sensitivity analysis of the probability with respect to seven key LOS parameters. Secondly, we explore the opportunity cost of using RH services through a Time Value analysis. Lastly, we performed Summation of Probabilities to analyze the degree of influence each parameter has on $P(Transit|CTA)$.

Thesis Layout

This thesis contains six chapter, the current chapter being the first, the *Introduction*. In the next chapter, our literature review will be introduced and discussed. Chapter 3, *Methods*, is extensive as we introduce the datasets used, explain all programming procedures and logic, and introduce the utility model and explain how we adapted our output for the model's required input. In Chapter 4, *Results and Discussion*, we present our findings and their meanings in context of our study. In the final chapter, *Summary and Conclusions*, we will briefly review our study's contribution to the field, explain how our findings can be used by public transit agencies, and offer forward-thinking suggestions.

Terminology

The following is an extensive list of terminology relevant to this paper, and their corresponding contextual definitions. Following some definitions is their respective location within the paper.

Buffer Zone: this is the group of transit-equivalent trips that have a $P(Transit|CTA) = (0.45 \cup 0.55)$. Trips that lie within this zone are considered to be unreliable indicators of true mode-choice behaviors and are excluded from the analyzes.

First- and last-mile (FLM) arrangement: this refers to the first and last leg of the transit trip that connect the individual from their origin to the first transit stop, and from the last transit stop to their destination. This is commonly executed via walking and can be a deterrent to potential transit riders, especially the physically disabled. A more taxing FLM is associated with transit networks where the density of transit stops in the origin or destination zone is low.

In-vehicle travel time (IVTT): this is the portion of the total travel time, and accounts for all time spent traveling inside the transit vehicle(s). In this paper, IVTT may be referred to as the "transit time". For a trip that is executed by walking only, the IVTT equals zero.

Not-replaced (NR) trip/group: this is the group containing all transit-equivalent trips with a $P(Transit|CTA) \leq 0.45$. A transit-equivalent trip that has a probability in this range $[0 \cup 0.45]$ is deemed to be inviable to the individual, and ultimately, does not compete with the RH trip service. These transit-equivalent trips exhibit poor LOS attributes.

Out-of-vehicle travel time (OVTT): this is a portion of the total travel time outside of the vehicle(s), i.e. accessing, egressing, wait time, transfer walk time. For a trip that is executed by walking only, the OVTT equals the total travel time (TTT).

Pooled Trip, Ridehailing: these are ridehailing trips that combine two or more trips, such that passengers 'share' the ride. In some scenarios, all passengers meet at a specified location and are dropped off at a shared location. Whereas in other scenarios, passengers are picked up at their desired location and then dropped off in the most efficient order.

Replaced (R) group: this is the group containing all transit-equivalent trips with a $P(Transit|CTA) \geq 0.55$. A transit-equivalent trip that has a probability in the range [0.55 ∪ 1.0] is considered a viable mode of service for the specific O-D pair.

Ridehailing (RH): this refers to the act of servicing a trip via a transportation network company (TNC). Users must have an account with the respective TNC, and have the app downloaded onto their smartphone. These trips are ordered using the TNC's app and require the user to input their destination, whereas the origin is automatically determined using the smartphone's internal GIS software. TNC trip fare pricing is dynamic and dependent on the surrounding demand. Although, when ordering a trip, the displayed fare in present time becomes 'locked' and will not change even if the demand increases or decreases while the rider waits. Ridehailing trips can be pooled or single passenger, refer to their definitions.

Route: refers to the output from ArcGIS' Route Analysis: the transit-equivalent route for a given RH trip.

Sensitivity Condition: with reference to the section, *Sensitivity Analysis*, a sensitivity condition is defined as the percent-change in the sensitivity variable. For each variable, there existed 20 sensitivity conditions, ranging from -50% to +50% in increments of 5%, where the 0% condition is the observed values and results.

Sensitivity Variable: with reference to the section *Sensitivity Analysis*, a sensitivity variable is the variable that is modified and redefined for the 20 sensitivity conditions. All other variables, parameters, and constraints remain equal to their observed value. There are seven sensitivity variables for this analysis and are outlined in Chapter 3.

Single-Passenger Trip, Ridehailing: these are RH trips where there is one person who ordered the trip, and there is one origin and destination. In some cases, these trips can have more than one passenger. For example, a group of friends want to ride together so one person in the group orders the RH trip, and the remaining friends ride with this person. If the fare is split, payment transactions are not associated with the TNC company.

TNC: Transportation Network Companies; these are the private businesses that offer ridehailing services. Examples include Uber and Lyft.

Total travel time, transit (TTT): this is the time elapsed between the departure time at the origin and the arrival time at the destination for the transit-equivalent trip. This value accounts for OVTT and IVTT if applicable.

Transit-equivalent trip/CTA-equivalent: this is a trip classification and refers to the output from ArcGIS' Route Analysis. For an input RH trip with an O-D pair, Route Analysis will calculate the most efficient transit trip to serve the O-D pair. For the program, we input GTFS data that corresponded to the Chicago Transit Authority (CTA) only; there exists other transit services in Chicago, but the GTFS dataset was limited to CTA. Thus, any output trip that utilizes transit, is using CTA services. It is important to consider that the output trip does not necessarily use transit. Under certain conditions, the program determines that it is quicker for an individual to walk from the origin to destination, rather than using transit. Thus, a “transit-equivalent” trip does not imply the use of transit.

Trip duration, ridehailing: this is the total time between the pick-up time and drop-off time for a ridehailing trip.

Walk time (WT): this is the sum of time allocated to walking and is a portion of the OVTT. This value is output by the ArcGIS Route Analysis, and assumes a walking speed of 5 km/hr.

CHAPTER 2: REVIEW OF LITERATURE

The current body of research on ride-hailing is limited by its novelty and the lack of publicly available ride-hailing trip data. External research on the utility of RH and its impact is relatively untouched due to its introduction to the market in 2010. Moreover, ride-hailing services are privately owned, consequently, trip-specific data is exclusively withheld and unavailable for public research use. While there is no existing literature that definitively states how ride-hailing services impact public transit ridership, many stipulate a correlation between the two, and if ride-hailing is a contributor, it is likely not acting alone.

This absence of trip data has led researchers to obtain empirical data through stated preference (SP) and revealed preference (RP) surveys [10-13]. Some studies executed intercept surveys at points of interest [6], and one executing in-person interviews [14].Yet to our knowledge, there exists no research on the relationship between ride-hailing and public transit that is uses source-data. Consequently, these empirical methods confine the spatial and temporal ranges, limiting the application and testing the integrity of the findings. Ultimately, this has led to conflicting arguments that have yet to be resolved. In the following literature review, we identify reoccurring themes and findings regarding the impact of ride-hailing services on public transit ridership. Additionally, we highlight the methods used to obtain data. Lastly, we determine gaps in the literature and how they will be addressed in this study.

It is important to note that one-to-many relationships are encompassed by the relationship between public transit and RH. Bus and rail (light and heavy) both fall under ‘public transit’, although trips of differing purposes, rider demographics, and LOS metrics are serviced by each mode. Accordingly, most literature analyzes each modality separately.

In general, the impact of ride-hailing on VMT and vehicle emissions, its relative safety, and its effect on mode selection are explored in the current literature. Yet, the latter of the three concerns is the least explored. Contreras and Paz presented three questions, one of which illustrates this concern, “have RHC’s [ride-hailing companies] had a negative or positive effect on transit ridership and/or revenue?” [9]. Answering this question requires empirical and source-data based research.

As stated previously, conflicting arguments have evolved from the lack of source-data based research. Considering that “public transit” encompasses many transit modes, positions tend to be unique per mode (bus, rail). Argued by the first position is that the perceived gains of ride-hailing services attract riders and thereby, substitutes transit. This is based on the significant difference between the gains, and marginal difference between the cost between public transit and ride-hailing. Thus, the cost differential is perceived to be worth the gain of ride-hailing, and thereby replacing public transit. Accordingly, it has been posed by critics that ride-hailing services contribute to the recent decline in public transit ridership. Whereas the second and opposing position argues that ride-hailing complements and reinforces the use of public transit by servicing the first- and/or last-mile (FLM) arrangement, and therefore induces revenue.

Most studies have explored mode choice behavior towards ride-hailing through observation-based research methods, such as SP, RP, and intercept surveys [11, 12, 15]. According to Clewlow and Mishra, ride-hailing services replaced 6% of bus trips and 3% of light rail trips.

Whereas ride-hailing acted complementary to commuter rail services, increasing ridership by 3%. Similarly, Graehler et al. found that the entry of a TNC decrease heavy rail and bus ridership by 1.3% and 1.7%, respectively [16].

Rayle et al. determined the primary reasons why individuals chose ride-hailing over the alternative of interest. In brief, users chose ride-hailing over the bus because it was faster and over rail because it was faster, easier to pay, and had less wait time [6].

Henao and Marshall worked as Uber drivers in Denver, Colorado to obtain observational data in real time via verbal, recorded interviews. Of the 311 passengers interviewed, only 5.5% of riders were using the ride-hailing service to get to or from a transit station [14]. This implies that 94.5% of ride-hailing trips do not service the FLM arrangement. However, the small sample size challenges the range of application and the question with a binary response option minimizes bias. Moreover, the utility of the surrounding transit network tests the application of this finding. The transit network in Denver has significantly less popularity than that of other US metropolitan cities. Hence, the percent of riders using RH for FLM arrangements is likely sensitive to transit network in question.

Nelson and Sadowsky used a difference in differences (DID) modeling by comparing transit ridership and operational metrics red before and after the entry of ride-hailing service(s). Their findings concluded that transit ridership increased following the entry of the first ride-hailing company, then decreased once the second company entered the regional market. The presence of the second company led to competition and increased affordability, allowing it to appeal to more people [17].

In 2016, APTA investigated the relationship between emerging modalities and public transit. The research areas included seven major US cities: Austin, Boston, Chicago, Los Angeles, San Francisco, Seattle, and Washington DC. Researchers executed in-depth interviews with transportation officials and surveying of network users. The most relevant finding is shared modes, i.e. ride-hailing services, are used most frequently for social trips during hours which public transit is not in operation or has reduced services. Hence, when transit operations are reduced, ride-hailing services supplement its decreased availability. Results from the survey show that 54% of respondents had used “ride-sourcing” (ride-hailing) to serve a recreational or social trip within the previous 3 months. Further, only 21% of respondents claimed to have used these services for commuting within the previous 3 months. However, this survey does not look at the trend in demand by trip type over a period of time. The percentage of respondents claiming to have used ride-sourcing for a specific purpose does not encapsulate the frequency of demand by type. For example, 21 out of 100 respondents could use ride-hailing services for commuting on a daily basis, whereas 74 out of 100 respondents only used ride-hailing once a week for recreational/social trips. The cumulative demand by trip purpose cannot be represented through a one-time survey [18].

While these methods are useful and highly qualitative, they assume an ideal condition that respondents are not biases. Hence, the results are vulnerable to many biases. The first, *hypothetical bias*, is the propensity of humans to view survey questions hypothetically to an extent that skews responses validity. Second, *strategic bias* is the tendency for a respondent to evaluate their hypothetical behavior such that it favors the response with greater perceived value. Lastly, *framing bias* is how the phrasing and wordage of a question influences its interpretation.

The overwhelming use of surveys and interviews serves as an opportunity to deploy a more quantitative study that focuses on individual trips and their corresponding LOS attributes. Until we can collectively concur upon the effect of ride-hailing, designers, planners, and politicians cannot make sound decisions. We hope to contribute to the field by pioneering new methods and approaches for analysis the impact of RH. The use of source data-based research will not only result in greater clarity and insight but will illuminate gray areas with more intensity. From this, empirical studies should be refined to focus on investigating these ambiguous regions and identifying their sources.

CHAPTER 3: METHODS

The primary goal of this thesis is to answer the research question using trip-based data rather than empirical data. Further, to avoid biases in the results from using proprietary data, we chose to use publicly available data. We searched for a dataset including individual spatial and temporal trip characteristics to increase the representativeness of the conclusions. Unfortunately, because all TNCs are privately owned, the availability of trip data is extremely limited. We exhausted many avenues of resources and discovered the only publicly available dataset containing individual trip attributes, provided by the City of Chicago's Online Data Portal. This dataset is titled "Transportation Network Providers (TNP) – Trips" and contains records of all RH trips within the city limits of Chicago, Illinois from November 2018 to the present day [19]. The dataset is further explained in the subsection, *Data* We chose to use this as the only RH trip dataset, thus the study area spans the city of Chicago. In the following subsection, *Area of Study*, we introduce relevant characteristics of the city of Chicago.

Area of Study

Geography and Demographics of Chicago

Per the U.S. Census, the population of Chicago was estimated to as 2,705,994 persons in July 2018. The city spans 227.63 square miles and contains 801 census tracts, according to the 2010 Census. As of 2010, 21.2% of the population is 18-years or younger and 12% of the population is 65+ years [20]. From 2014-2018, there was an average of 1,056,118 households with a median income of $57,238 [21]. As of 2015, 26.5% of households do not own a vehicle, where the average vehicles owned per household is 1.11 [20].

Public Transit in Chicago

Chicago Transit Authority (CTA) is the second largest transit agency in the U.S. as of 2018 [22]. CTA runs and operates bus and rapid transit (rail) services within the city and the 35 surrounding suburbs. There are 1,864 buses that run 129 routes and 1,429 rail cars that serve 145 stations [23]. Additionally, CTA operates certain routes and lines during early morning and late-night hours, and some operate all hours of the day.

Ridehailing Services in Chicago

Historical data on the services present during the study period (June 2019) is unavailable at this time. However, as of January 2020, three personal-car ridehailing services operate in the City of Chicago: Uber, Lyft, and Via [24].

Data

TNP (Transportation Network Providers) – Trips Dataset

This dataset served as the source data for ridehailing trips and was obtained from the City of Chicago's online data portal. The dataset contains 129 million unique TNC trips that span from November 2018 to the present day and is aggregated by the month [19]. Given the expansive size we chose to only study one month, June 2019. This month was selected because it does not contain any nationally recognized holidays that could hinder the representativeness of the results. All RH trips with a start time on or after June 1, 2019 12:00:00 AM and before July 1, 2019 12:00:00 AM are included in this data set. The dataset contains 21 fields per trip, including a unique identifier, trip start and end time, pick-up and drop-off longitudinal and latitudinal coordinates, pick-up and drop-off census tract ID, trip fare, and if the ride was authorized as "shared" through the respective TNC app. A full list of the dataset attributes can be found in Appendix A.

Public Transit Data (General Transit Feed Specification (GTFS) Dataset)
To perform a public transit network analysis in ArcGIS, it requires the GTFS dataset corresponding the area of interest, as an input. GTFS is a publicly available data feed hosting real-time and fixed components of transit agencies' schedules. This data is uploaded by the responsible agency and is readily available through an online database published by OpenMobilityData [25]. For each transit agency, there exist many subsets of data, spanning approximately two-month periods. The dataset holds the corresponding schedules, routes, stops, and transfers for the time period. GTFS serves as an open-source data feed that can be used by public and private entities. With respect to this paper, this dataset will be integrated into ArcGIS such that the Network Analysis program can identify the corresponding transit route under spatial and temporal conditions.

Street Centerlines
To create a network geodatabase in ArcGIS, the user must have an existing feature layer consisting of the roadway centerlines. Hence, a SHP file of all street centerlines within the limits of the city of Chicago was obtained from the City of Chicago Data Portal [26].

Census Tract Boundaries
A shape file of all census tract boundaries was downloaded from Chicago's Data Portal. This file was used to estimate the census tracts containing the origin and destination, for trips with corresponding null values in the raw dataset [27].

Preliminary Analyses

The first preliminary data analysis aggregates ridehailing trips by calendar day to provide a visual representation of the demand by type of day (Monday-Sunday). Figure 3 below is a bar graph depicting the results of this analysis.

For purposes of clarity, the legend in Figure 3 is explained below. Further, this legend applies to Figures 4 and 5. With reference to the legend in Figure 3, the bars shaded blue corresponds to the volume of trips on the weekend (Saturday and Sundays), whereas the bars shaded orange corresponds to weekday trip demand. For each bar, it is split into two different color intensities, a darker and lighter section. The darker portion of each bar represents the number of trips classified as "pooled" (shared) by the ridehailing service. The lighter portion of each bar represents the volume of single occupancy trips.

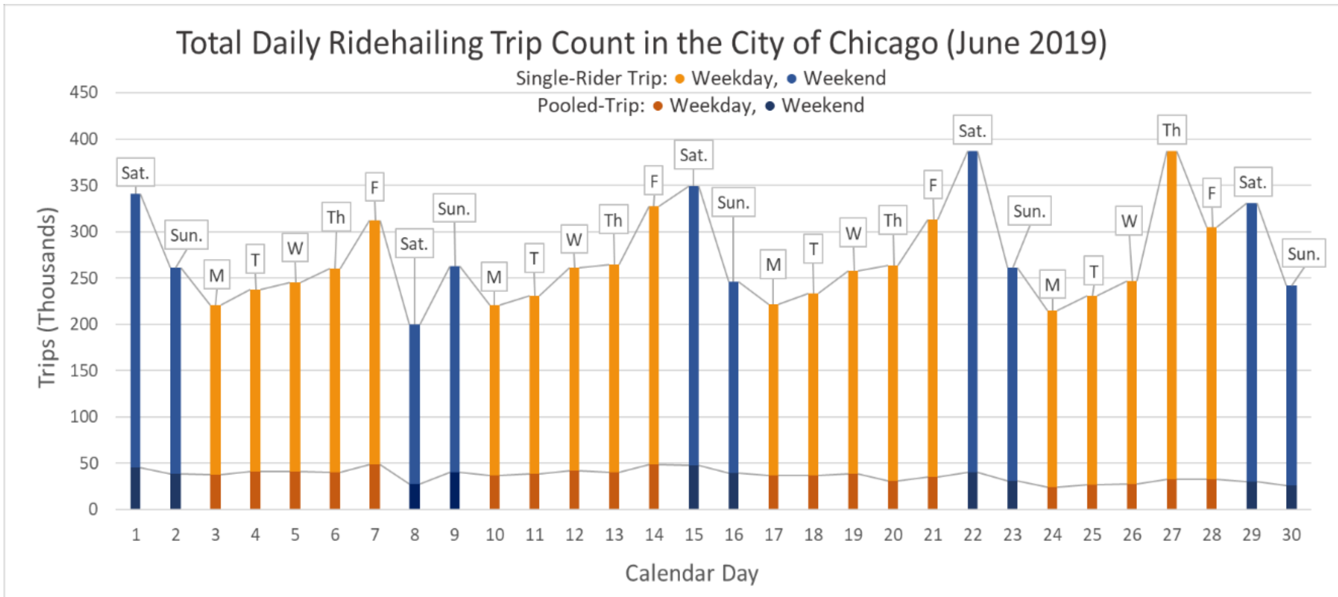


Figure 3 – Daily Ridehailing Trip Counts in the City of Chicago during June 2019

First, this figure demonstrates that single occupancy ride-hailing trips have significantly greater demand than pooled trips. When compared to driving alone, this modality has a higher contribution towards congestion because its utility is comparably low due to deadheading mileage.

Referring to the temporal trend of demand during the weekdays, it is clear that there is an upward trend in demand from Monday to Friday. Let us assume that there exists a baseline demand for RH commuting trips. With the traditional work week spanning Monday to Friday, we can infer that each workday will have this baseline demand. However, as shown in Figure 3, there exists growth in demand from Monday to Friday. Hence, in addition to the baseline volume of trips, there is a volume of trips that are not work-related, or are work-related trips taken by individuals who do not regularly commute via RH.

The next two figures show the distribution of trips by starting hour; weekday trips are depicted in Figure 4 and weekend trips are depicted in Figure 5. For each hour, the bar height represents the number of trips in June 2019 starting during the corresponding hour's period[5]. The purpose of these two figures is to compare the temporal trend of demand between weekday and weekend trips.

When comparing between Figures 4 and 5, it should be noted that the range of the y-axes are different – Figure 4 has a greater range, spanning twice that of Figure 5. The majority of this difference can be attributed to the numbers of days spanned per subset of data.

[5] A hour's "period" spans the 60-minutes following the hour. For example, if the 8:00 bar has a height of 5,000 trips, then there were 5,000 RH trips started betweed 8:00:00 and 8:59:59.

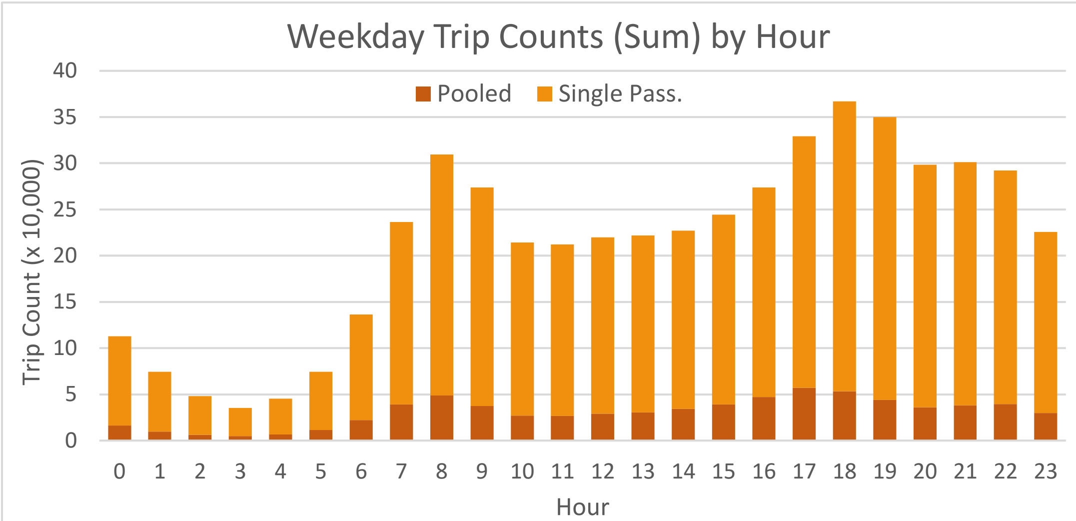


Figure 4 – Weekday Trip Counts (Sum) by Hour

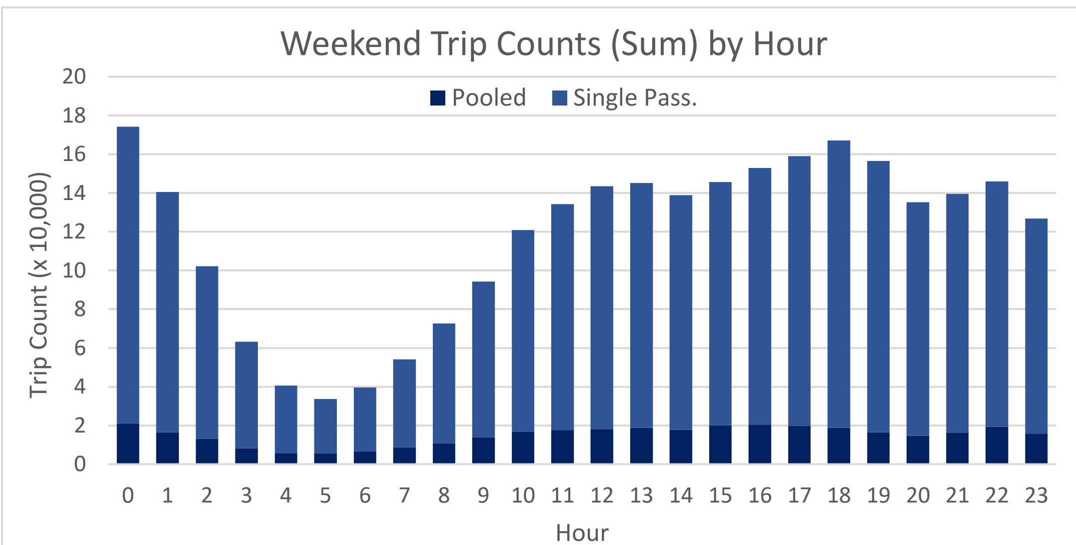


Figure 5 – Weekend Trip Counts (Sum) by Hour

Let us consider travel behavior during the hours 7 and 19 (7:00 AM – 7:00 PM). For weekday trips, two demand peaks exist at hours 8 and 16, whereas for weekend trips peak periods are not as distinct. This is a result of a more even distribution of demand between hours 11 and 22 (11:00 AM – 10:00 PM). The more level demand on weekends during this period is likely a result of a shift in trip purposes. Work trips are commonly serviced on a predictable schedule due to traditional 8-5/9-5 jobs. On weekends, people tend to allocate their time for social outings, leisure activities, and shopping. These activities have the opportunity to occur at any hour, as opposed to outside of working hours. Due to the nature of these activities, their duration is less predictable and can span a greater range of time. Inherently, with social and leisurely activities, the demand for parking (short-term and overnight) increases. This demand evolves into competition when the parking supply is limited. Consequently, in densely developed regions and cities, such as Chicago, parking availability is low. Overall, the growing population of drivers influences congestion and competition for parking. In conditions where this is of concern, RH services become a more attractive option. Search time, access and egress walking time, and the parking fare are all eliminated with the operational structure of RH. Additionally, being a passenger, as opposed to being a driver, allows for the redirection of attention and energy to activities that originally could not have been performed while driving.

Outside of 7:00 AM – 7:00 PM on weekends, the global peak exists at hour 0 (12:00 AM). We can assume that this spike in demand is for social-based purposes [6, 18]. RH services provide transportation for people who cannot legally drive due to alcohol consumption. The consumption of alcohol in combination with a demand for travel, yields increased demand for modes that do not require personal-auto use.

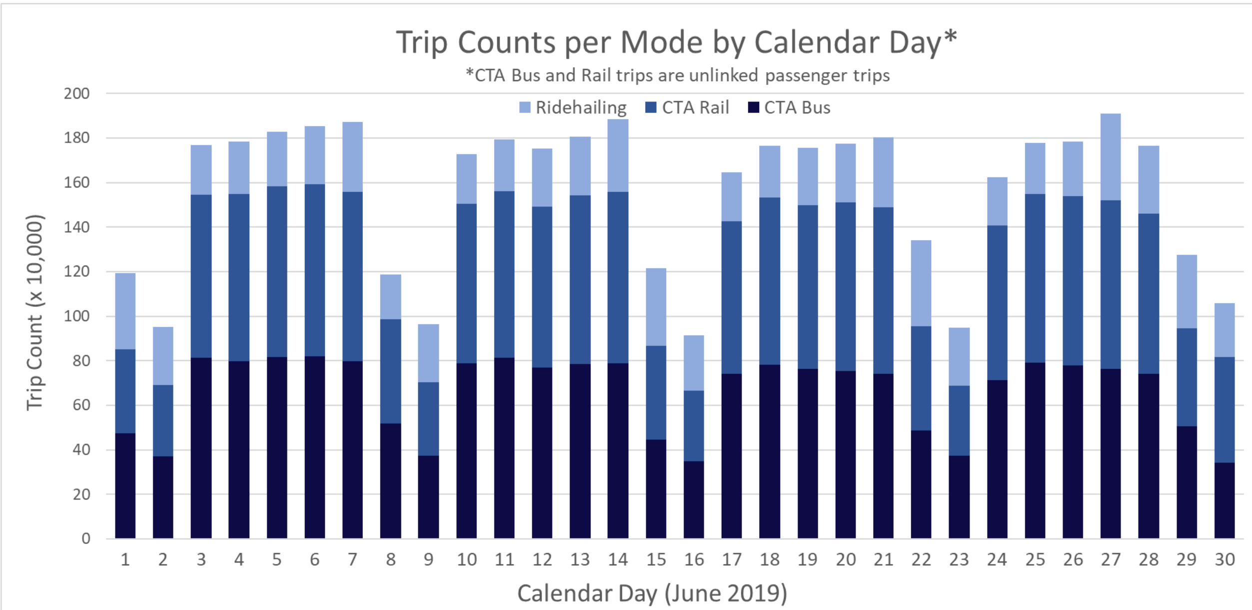


Figure 6 – Trip Counts per Mode by Calendar Day

We then transition to a preliminary comparison of trip per mode (CTA bus and rail, and ridehailing). The stacked column bar chart above (Figure 6**Error! Reference source not found.**) shows the total number of trips aggregated by day, where each bar is composed of the volumes of trips by mode (RH, bus, and rail). When interpreting this figure, it is important to consider that the number of transit trips are unlinked. As an example, if a rider took the bus from O_i to D_i with 2 transfers, this one trip (O-D) is subdivided and classified as three separate trips: (1) O_i to $station_A$, (2) $station_A$ to $station_B$, and (3) $station_B$ to D_i. Whereas a ridehailing trip from O_i to D_i would count as one trip. Thus, if the average number of transfers is greater than zero, then the proportion of transit trips to RH trips would be overestimated. However, the trend in percent-share by mode remains significant. The volume of transit trips on weekdays is significantly higher than that of weekends, which can be attributed to commuting trips. Moreover, it appears the average percent-makeup of the volume of transit trips is shared evenly between bus and rail. Excluding June 8th and 9th, the volume of RH trips appears to increase from Thursday to Saturday. Further, the proportion of RH trips to transit trips is greatest on Friday, Saturday, and Sunday. This can likely be attributed to an increase in transit disutility due to a significant decrease in transit frequency and in-operation lines/routes. Hence, longer wait times and decrease in serviceability yield a favoring towards RH services.

Data Processing

Data processing was completed in three steps, with the ultimate output being the probability of a rider choosing public transit. This probability is derived from a multinomial nested logit (MNL) model based on the Chicago's travel behaviors in 2015 [13]. This model and its relevancy are descriptively explained later in this section.

As a brief overview, the first two steps were performed in the program, *ArcGIS*, using two separate tools: (1) Route Analyst and (2) Spatial Join. These two steps are novel, in that GTFS data and source data are combined to compute the time-conscious transit-equivalent route. The output of these two steps, per RH trip, were a transit-equivalent trip and the number of transfers required to complete the trip. For the third step, results from the route analysis were input into our Matlab code to continue processing procedures and to compute $P(Transit|CTA)$.

STEP 1: Transit-Equivalent Trip Generation

For the first step, the Route Analyst tool within the ArcGIS Network Analysis toolbox was employed to determine the transit-equivalent route for each ride-hailing trip. This tool processes a set of trips containing 2+ stops (per trip), and outputs the most efficient route given a specified travel mode (driving, public transit, or walking) a specified impedance (travel time, walk time, or travel cost). Prior to running the tool, the *travel mode* was set to 'public transit' and the *impedance* to 'walk time'. These conditions make the solver utilize public transit when possible and minimize the walking time to, from, and between stops.

CTA's GTFS data for June 2019 was then loaded into the program using the GTFS toolkit. These consisted of files defining the transit network's geometric and temporal structure. This dataset provides the means to determine a RH trip's (O-D pair and start time) public transit alternative using CTA only.

The trip data was imported as text files, where each calendar day had a separate text file containing all the trips with the corresponding start date-time. For each day, trips were aggregated by start times using 15-minute intervals, i.e. 12:00:00 AM, 12:15:00 AM, etc. For each RH trip, the pick-up location was defined as the first stop and the drop-off location was defined as the second and final stop[6]. With reference to the GTFS dataset and the output transit network, the route analyst tool then output the most efficient transit-equivalent route per RH trip. The output from this tool provided LOS metrics of the transit-equivalent trip, such as the TTT, WT, and the start and end times. Route analyst also outputs an ArcGIS layer that contains polylines spatializing all routes. However, these polylines do not contain any characteristics regarding which transit lines and stops were used. When this layer overlaid the transit network layers, each route would visually intersect the transit network. Considering this, we manipulated the spatialized data using ArcGIS' analytical tools to identify all transit network elements intersected (i.e. transit stops, transit lines, transit stations).

[6] It is important to note that all pick-up and drop-off locations in the dataset were the coordinates of the centroid of the census tract the location lies within.

STEP 2: Transfer Count Estimation
In the second step, data were first processed in ArcGIS and lastly in Matlab. For Part 1 of this step, we used the *Spatial Join* function to calculate how many "stop connectors" were contained in each route. This term and its corresponding step are explained in the proceeding paragraphs. The output of the spatial join served as input for Part 2, the calculation of transfer count per route using Equation 1.

As a result of integrating the GTFS dataset, feature layers were created separately containing transit stops and lines that are respectively contained in the network layers, *Stops* and *LineVariantElements*, and their counterparts, *Stops on Streets* and *Stop Connectors*. Attributes in the Stops and Stops on Streets layers are represented as points, whereas attributes in the *LineVariantElements*, *Stop Connectors*, and *Streets* layers are represented as polyline elements (Figure 7).

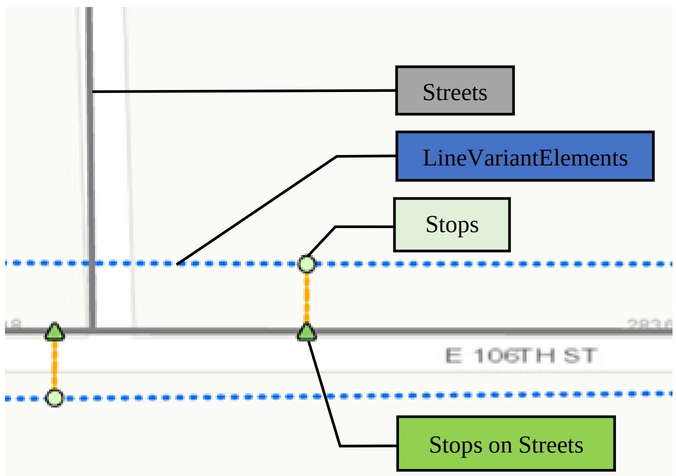


Figure 7 – GTFS Network Dataset Layers

The transit stops (Stops) are spatially offset from the street centerlines (Streets) because the GTFS transit lines (LineVariantElements) do not spatially overlap the streets for modeling purposes. When route analysis is performed, the transit-equivalent route will overlap the streets when not using transit (OVTT) i.e. when walking to/from transit and will overlap the transit routes when using transit (IVTT).

Transit stops are reflected onto the street it is offset from, generating a second element that is stored in a new layer called "Stops on Streets". Thus, for each transit stop there are two corresponding points (1) on the transit network (2) on the street centerlines. After the second point is created, a polyline element is generated that connects the two points. These polylines are called 'stop connectors.' Refer to Appendix B for a map of the GTFS-integrated Chicago (CTA) transit network developed in ArcGIS.

When boarding or deboarding transit, the route intersects both points and overlaps the corresponding stop connector. Thus, it can be assumed that if a route contains (overlaps) a stop connector, the rider is either accessing or egressing a transit service. To encourage program efficiency, we used a non-visual program, Matlab, to execute this calculation. The table of transit-equivalent trip characteristics and results from the spatial join were then imported into Matlab, and the equation below was used for each trip:

$$n_{transfers} = \left(\frac{number\ of\ overlapping\ stop\ connectors}{2}\right) - 1 \qquad (1)$$

STEP 3: Probability Estimation
To estimate the probability of a rider selecting public transit to service their O-D pair, a utility model and logit transformation formulae are required. However, the scope and limited timeline of our project made it impossible to develop a utility model. Hence, we transitioned our efforts to finding an appropriate utility model and logit transformation in existing literature.
Moreover, we sought out a model that contained obtainable input values and that was derived from a sample with similar demographics and travel behaviors. We reviewed many models based on the nature of the research, the study area, and modalities modeled, and then compared our methods and dataset against these models to determine which existing utility model was most suitable.

The extensive literature review resulted in a selection of a multinomial nested logit (MNL) model developed by Javanmardi et al. [13]. The basis for development of their mode choice model was a revealed preference survey. Traditionally, mode choice models are developed from TAZ (traffic analysis zones) level data that uses average travel times. However, Javanmardi et al. used Google Maps API (Application Programming Interface) and RTA's Goroo TripPlanner to obtain personal trip data that better represents individual travel behavior. Such data included point-to-point travel times, and feasible alternatives and their LOS attributes [13]. Overall, this MNL model is used to measure variance in mode choice behaviors regarding alternative transportation, with increased accuracy from RP surveying.

Coincidentally, this model was developed using trip data from the same area of study as our project, Chicago, Illinois. Thus, this model allowed for increased representativeness of travel behaviors to a greater degree of accuracy. Lastly, the study year (2015) of their research is appropriate in that ridehailing was introduced to Chicago before that time, thus their model should capture any evolution of mode choice behavior and preferences towards or against alternative transportation.

The model's formulae are represented by the equations below (Equations 2-6). Given that the model was used for a range of modes, the subscripts were modified to align with the variables in this paper. The constants, coefficient values, and variables are mode-specific and were provided in the paper corresponding to the model [13].

Utility of Transit, $U_{Transit}$

$$\begin{aligned} U_{Transit} = 2.93 - 1.04TT - 0.13TC - 0.17n_t - 0.77HHI \\ + 0.45wrktrp - 0.39d_A - 0.23d_E \end{aligned} \qquad (2)$$

Probability of Selecting Transit, $P_{Transit}$

$$P_{Transit} = \frac{e^{U_{transit}}}{1 + e^{U_{transit}}} \tag{3}$$

Utility of CTA, U_{CTA}

$$U_{CTA} = -0.39TT - 0.059TC - 0.33n_t + 0.022n_{Stop,O} + 0.0089n_{Stop,D} + 0.77shptrp + 1.78wrktrp - 0.46HHI \tag{4}$$

Probability of Selecting CTA, P_{CTA}

$$P_{CTA} = \frac{e^{U_{CTA}}}{1 + e^{U_{CTA}}} \tag{5}$$

Probability of Selecting CTA given Transit Selection, $P(Transit|CTA)$

$$P(Transit|CTA) = P_{Transit} \times P_{CTA} \tag{6}$$

The equations were executed in the respective order per trip, with Equation 6 outputting the final probability used in the analyses.

Table 1 outlines the input attributes in the above formulae, their corresponding definition, and their availability with respect to the RH trip dataset.

While this derivation of this model exhibited strong similarities to this study's characteristics, it did contain several caveats. The attributes of the RH trip dataset used in this paper did not completely satisfy all required input of the model, therefore missing values were generalized, estimated, or calculated.. The determination of these missing values required multiple assumptions. Following the Table 1, each assumption-based variable, and its calculation process(es) is explained in greater detail.

Table 1 is located on the next page for formatting purposes.

Table 1 – Utility Model Input Variables

Source	Variable	Definition
Output from spatial analysis	TT	Total travel time (hr); wait time + walk time + transit time
Calculation; assumption-based	TC	Total travel cost (USD); total cost of fare for transit trip
Spatial Analysis	n_t	Number of transfers (transfers);
US Census Bureau[7]	HHI	Household income (10^{-5} USD);
Calculation; assumption-based	wrktrp	Purpose, Work trip (1/0); if trip purpose is for work, wrktrp = 1. Assumed if trip start day = weekday, and start hour in 5 a.m. – 7 p.m., trip purpose was for work
Calculation; assumption-based	d_A	Access distance (km); walking distance from origin to first transit stop (pickup)
Calculation; assumption-based	d_E	Egress distance (km); walking distance from last transit stop (drop-off) to destination
Spatial Analysis	$n_{Stop,O}$	Number of transit stops in origin zone (stops); total number of transit stops within census tract containing origin
Spatial Analysis	$n_{Stop,D}$	Number of transit stops in destination zone (stops); total number of transit stops within census tract containing destination
Calculation; assumption-based	destCBD	Purpose, Destination in CBD during rush hour (1/0); if trip destination is within the geographic boundary of the Chicago CBD, and started during rush hour, *destCBD* = 1.

Household Income; *HHI*: given the privatization of the ride-hailing dataset, we were unable to access the socioeconomic characteristics of each individual ride-hailer. To compromise, we defined the HHI for a rider using a dataset containing the average HHI per census tract. We defined the HHI to equal the average HHI of the origin, if the trip was executed on a weekday between 5:00 AM and 12:00 PM, or of the destination, if the trip was executed on a weekday between 12:00 PM and 7:00 PM. For all trips outside this boundary, the HHI was defined as the average between the origin and destination HHI.

Four variables outlined below are a function of two census tracts, containing the origin and destination. These variables depend on the census tract IDs, as the ID is used to index data from related tables. The source dataset contained the GPS coordinates for the origin and destination census tracts, and the corresponding tract IDs. However, a subset of trips exhibited null values for the tract IDs. Thus, to remedy this, we estimated the corresponding tract IDs via a minimum distance program in Matlab. First, we obtained a SHP file of the geographic boundaries for all census tracts from the Chicago Data Portal. Once imported into ArcGIS, geometric calculations were performed to output each tract's centroid in GPS coordinates (latitude, longitude). This output table was imported into Matlab as a matrix. Using each trip origin and destination latitude

[7] Due to confidentiality of trips, we did not have access to demographics of ridehailers. Therefore, we assumed the income for each rider to be the 2018 average household income (HHI) of all persons in City of Chicago from the US Census Bureau [20] *QuickFacts Chicago city, Illinois*, United States Census Bureau.

and longitude, the distance between each tract and O-D coordinates was calculated. The census tract ID corresponding to the smallest distance value was selected and replaced the null value for the origin or destination tract value.

Number of Transit Stops per O-D Zone *nStopOrigin; nStopDest*: these two attributes were a function of the O-D census tracts per trip and were calculated as the number of transit stops in the corresponding origin or destination zone (census tract). The number of transit stops per census tract were calculated using a spatial join in ArcGIS, where the output was an 801x2 table with each census tract ID and the corresponding number of transit stops within the tract boundary. Indexing was then used to retrieve and append the number of transit stops per census tract to the master table.

Saturday/Sunday Classification: To test true for the attributes below (*nStopOrigin*, *nStopDest*, *destCBD*), a trip could not have been serviced on a Saturday or Sunday. Therefore, to classify a 'weekend' (Saturday or Sunday) trip, we composed a vector of June 2019 calendar days corresponding to each pair of Saturdays and Sundays. If a trip's start calendar day was identified as a weekend day, then it results in a *false* value. Therefore, any trip taken on 06-[1,2,8,9,15,16,22,23,29,30] -2019 tested false for *destCBD* and *wrktrp*.

Destination within CBD at Rush Hour; *destCBD*: this binary attribute indicates if the destination of a trip lies within the central business district and was serviced during rush hour. To determine this value, we first determined which census tracts lied within Chicago's CBD. Using a SHP file containing the geographic boundary of the CBD, we spatially joined it with the aforementioned census tract centroids layer using the "completely within" condition [27]. The output of this procedure was a table of 20 census tracts, their IDs, and GPS coordinates. The second clause of this test was to determine if the trip started during the peak period. We assumed there to be two, 3-hour peak periods (AM and PM). The AM peak period occurred between 6:00:00 and 9:00:00 AM, and the PM peak period occurred between 16:00:00 and 19:00:00 PM. For a trip to test "true" (destCBD = 1), First, we determined if the destination census tract was a member of the CBD census tracts array. If true, the trip start hour was then tested against the two peak periods. If the trip start hour had a value in the following vector, [6,7,8,9,16,17,18,19], then *destCBD* = 1, otherwise, *destCBD* = 0.

Trip Purpose: Work; *wrktrp*: similar to *destCBD*, this binary attribute indicates if the trip was a commute to or from work. As stated previously, due to the anonymity of the dataset, we did not have individual details on trip characteristics such as the trip purpose. To accommodate for this, we made a conservative assumption that all trips with a start time between 5:00:00 and 19:00:00 were work-related/commuting trips. This ensures the exclusion of any social or leisurely trips taken on weekends and/or during late-night hours.

Following the analysis of the observed data (*Replaceability Analysis*) is a sensitivity analysis of $P(Transit|CTA)$. Seven key attributes of the utility model were selected and used to analyze and determine its influence on the probability value. The results from this analyzes can be further used by transit agencies to identify how altering services and modifying operations could either increase or decrease ridership.

Analyses

In this section, we will introduce the three analyses used and their respective methods. Below is a list of the three analyses and their relevancies.

1. Replaceability Analysis: per RH trip, identify the $P(Transit|CTA)$ of the corresponding transit-equivalent trip. These results will be summarized by group type (R/NR).
2. Time Value Analysis: compare the travel times and total fare/cost for each RH trip and its transit-equivalent trip.
3. Sensitivity Analysis: explore the sensitivity of $P(Transit|CTA)$ with respect to seven variables in the utility model, which are introduced later in this section.

Replaceability of a Transit Trip

Ultimately, to determine if a ridehailing trip "replaced" its transit-equivalent trip, we computed the probability of using CTA, given the selection of transit. The magnitude of these probabilities indicates the viability of public transit serving a specified trip and depends on how favorable the trip's LOS attributes and trip-specific characteristics are to the rider. We chose to classify a transit-equivalent trip by its replaceability, categorized by two groups: replaced (R) trips and not-replaced (NR) trips. Initially We assumed the threshold value distinguishing a trip being "replaced" (R) or "not replaced" (NR), to be 0.5. Thus, all trips with a $P(Transit|CTA) \; < \; 0.5$, were assumed to not have a viable public transit-equivalent trip and was deemed "not replaced" (NR) by public transit. Whereas all trips with a $P(Transit|CTA) \; \geq \; 0.5$, were assumed to have a reasonable and competitive public transit-equivalent trip and was classified as "replaced" (R) by public transit. Although, following the first sensitivity analysis trial, we found that trips with $P(Transit|CTA)$ close to 0.5 switch between the R and NR groups. These volumes of trips were considered fuzzy and unreliable indicators of true mode-choice modeling behavior. Thus, we chose to implement a buffer, where trips with $P(Transit|CTA) \; = \; (0.45 - 0.55)$ grouped into the *Buffer Zone* and are removed and excluded from the summary statistics. This modification is represented by the conditional statement below.

For an individual ridehailing trip, T,

$$replaceability\ group_T = \begin{cases} Not\ Replaced\ (NR), & 0 \; \leq P(Transit|CTA) \leq 0.45 \\ Replaced\ (R), & 0.55 \; \leq P(Transit|CTA) \leq 1.0 \\ Buffer\ Zone, & 0.45 < P(Transit|CTA) < 0.55 \end{cases}$$

Following the grouping of each trip, statistical measurements were calculated for each group and undivided dataset as a whole. These values are introduced and discussed in the *Results* section.

Time Value Analysis

The most notable differences between the RH trip and transit-equivalent trip appeared to exist in the cost and travel times per trip. To further explore this relationship, we compared the magnitude and sign of each difference between both parameters. For a single trip, there are four possible outcomes (I-IV) which are outlined below, in Figure 8.

		Total Travel Time (TTT)	
	RH_i - $Transit_i$	+	-
Fare	+	IV	III
	-	I	II

Figure 8 – TTT vs. Fare Outcome Scenarios

$$f(x) = \begin{cases} I, & TTT: \quad Transit > RH \\ & Cost: \quad Transit < RH \\ II, & TTT: \quad Transit < RH \\ & Cost: \quad Transit < RH \\ III, & TTT: \quad Transit < RH \\ & Cost: \quad Transit > RH \\ IV, & TTT: \quad Transit > RH \\ x, & Cost: \quad Transit > RH \end{cases}$$

Verbal explanations for each outcome are explained below:

I. These are trips where, in comparison to the RH trip, the transit-equivalent trip was less expensive, but slower.
II. These are trips where, in comparison to the RH trip, the transit-equivalent trip was less expensive **and** faster.
III. These are the trips where, in comparison to the RH trip, the transit trip was quicker, but more expensive.
IV. These are the trips where, in comparison to the RH trip, the transit-equivalent trip was more expensive **and** slower.

Sensitivity Analysis

To explore how the attributes of a trip affect the probability of choosing CTA, we conducted a parametric sensitivity analysis of $P(Transit|CTA)$ with respect to the following decision variables:

1. Transit stops per census tract (SiT)
2. Base fare (BF)
3. Transfer cost (TC)
4. Household income (HHI)

5. Total travel time (TTT)
6. Walk time (WT)
7. Airport pass price (Airpass)

The $P(Transit|CTA)$ was recalculated under a set of percentage-change conditions, for each decision variable. Per variable, there were a total of 20 trials, where the observed value of the variable was incrementally adjusted in increments of 5%, ranging from -50% to +50%. Given that each variable was tested independently, there were a total of 140 trials. Variables 2, 3, 4, and 7 were fixed values defined at the beginning of the program. Whereas, for variables 1, 5, and 7, their original values are unique per trip, thus the new value is dependent on the trip attributes and is not a fixed value. The algorithm was run under the new condition, and a new $P(Transit|CTA)$ was output for all trips, and per group (R/NR).

Assuming that both groups share the same standard deviation, we can estimate σ by calculating the pooled standard deviation, s_p, with the equation below. The pooled standard deviation for the observed and sensitivity condition data sets, for group R or NR, is:

$$s_p(group, sensitivity\ condition) = \sqrt{\frac{[(n_{observed} - 1) * s_{observed}^2] + [(n_i - 1) * s_i^2]}{(n_{observed} + n_i) - 2}} \tag{7}$$

Where,

$n_{observed}$ = the number of trips in the observed group
$s_{observed}$ = standard deviation of the observed group
n_i = number of trips in the sensitivity group
s_i = standard deviation of the sensitivity group

It should be noted that $n_{observed}$ and $s_{observed}$ are fixed values under all sensitivity conditions. These values are shown in Table 3 in the *Results and Discussion* section.

To measure the level of influence and statistical relationship of each decision variable and the $P(Transit|CTA)$, we performed a two-tailed pooled t-test. Considering there is no overlap between the observed and sensitivity condition data, the two-tailed test was most suitable. A t-test was performed for each group (trip type), replaced (R) and not replaced (NR). Per group and under each sensitivity condition (decision variable and percentage-change), the mean $P(Transit|CTA)$ was compared between the observed and sensitivity data sets. The relationship between the t-statistic and the critical value indicate whether we accept or reject the null hypotheses stated below:

$H_{Null\,(R)}$ = The $\bar{P}(Transit|CTA)$ of the observed R group is not statistically different from the $\bar{P}(Transit|CTA)$ of the sensitivity R group.

$H_{Null\,(NR)}$ = The $\bar{P}(Transit|CTA)$ of the observed NR group is not statistically different from the $\bar{P}(Transit|CTA)$ of the sensitivity NR group.

If the t-statistic is greater than the critical value, then we reject the null hypothesis and refer to the alternative hypothesis. The alternative hypothesis opposes the null by concluding that there is a statistically significant difference between the observed and the sensitivity condition data.

Meaning, the influence of the decision variable on the $P(Transit|CTA)$ is expected to have an effect on the whole population, similar to the effect of the sensitivity condition.

The following equation was used to compute the t-statistic per variable and group for each sensitivity condition (Equation 8):

$$t_i = \left| \frac{\bar{P}_i - \bar{P}_{observed}}{\sqrt{\left(\frac{{s_i}^2}{n_i}\right) + \left(\frac{{s_{observed}}^2}{n_{observed}}\right)}} \right| \tag{8}$$

Where,

i = sample trips of group g (R or NR), decision variable *var*, and percent-change condition $\%\Delta$.

P_i = mean $P(Transit|CTA)$

$P_{observed}$ = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-change condition $\%\Delta$.

s_i= mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-change condition $\%\Delta$.

$s_{observed}$ = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-change condition $\%\Delta$.

n_i= sample size for group g (R/NR), decision variable *var*, and percentage-change condition $\%\Delta$.

$n_{observed}$ = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-change condition $\%\Delta$.

In the next chapter, *Results and Discussion*, the outcome of the three aforesaid analyses will be presented.

CHAPTER 4: RESULTS & DISCUSSION

The following section introduces a trip-classification system, results from the utility model and replaceability analysis i.e. $P(Transit|CTA)$ calculation, sensitivity analysis, and time value analysis. Before presenting the results, we will introduce a classification system of transit-equivalent trips which was developed in attempt to further group trips based on modality used. The ArcGIS programs output the "transit-equivalent" trip for each RH O-D pair, although after spatially analyzing the results, the solution did not necessarily use transit. Hence, we created a system that distinguishes trips based on modalities used, which is explained in *Classification of Transit-Equivalent Trips*. Following the aforementioned section, we will present the findings in the following order:

I. Replaceability Analysis
II. Time Value Analysis
III. Sensitivity Analysis

The results are best represented by visuals as there are many contributing factors that must be known for accurate analysis. Given the extensivity of the results, t-test results and additional supporting figures are located in the appendices.

Before proceeding, the following should be taken into consideration. As mentioned in the *Data* section, there were 8,136,461 RH trips in the raw dataset. Upon preparing the data to be input into ArcGIS, we found that 186,559 trips did not have geographic coordinates of their origin and/or destination. Given the coordinates are required input for the Route Analysis tool, we removed them from the dataset for analysis. The remaining trips, 7,949,902, were processed and all results are representative of that sub-selection of the raw dataset.

Classification of Transit-Equivalent Trips

For clarification, a RH trip's "transit-equivalent trip" is its alternative solution to its O-D pair using transit and/or walking and is output from ArcGIS Route Analyst. All transit-equivalent trips contain trip legs of walking, but not all transit-equivalent trips utilize transit. We refer to the transit-equivalent trips that use transit, as *transit-utilized* trips. For these trips, a walking distance is executed during the FLM arrangement and between transfers (if applicable). We refer to the transit-equivalent trips that do **not** utilize transit as *walk-only* trips; within this class are two sub-groups detailed later. These are trips that can be most efficiently serviced by only walking rather than using transit. Examples could include trips where the access and egress time required to use transit is comparable to the direct walk time from origin to destination. Below, we provide definitions per class, and introduced the two sub-classes for walk-only trips. Following these definitions, is Table 2 summarizing the mean travel times per group.

1) Transit-Utilized Trips: output trip that a rider uses at least one form of transit (bus and/or rail) to complete. These trips exhibit a $TTT > WT$. The difference between the TTT and walk time is the transit travel time. The census tract of the origin and destination must be different.
2) Walk-Only Trips: output trip where the rider does not utilize transit and is assumed to complete via walking. Thus, the transit travel time is non-existent such that $TTT = WT$.

a. Between-Tracts Trips: trips where the origin and destination census tracts are different, therefore the $TTT > 0$. These trips are distinguishable from the transit-utilized trips in that the total travel time equals the walk time.

$$TTT = Walk\ Time + Transit\ Travel\ Time \tag{9}$$

Considering Equation 9, if TTT equals the walk time, then the transit travel time must equal TTT-walk time (0).

b. Within-Tract Trips: walk-only trips where the origin and destination lie within the same census tract, thus the GPS coordinates of the origin and destination are identical. Considering the algorithm used, the route is calculated based on the spatial difference between the origin and destination census tracts. Therefore, if they have the same geographic coordinates, no spatial separation exists between the origin and destination, and the program outputs a travel distance of zero[8]. All travel time values are a function of this distance, so the travel times will equal zero.

Table 2 – Counts and Mean Travel Times per Trip Classification

Class		Trip Count	Transit Travel Time	Walk Time	Total Travel Time
		trips	*minutes*	*minutes*	*minutes*
Transit-utilized	1	7,143,648	19.31	17.39	36.70
Between-tracts	2A	507,300	0	18.41	18.41
Within-tract	2B	298,954	0	0	0
All Trips		7,949,902	17.35	16.80	34.15

Each class of trips was further analyzed to identify commonalities and differences. In doing so, we discovered a source of error in the walk-only (Classes 2A and 2B) trips. For all between-tract trips, ArcGIS' Route Analysis tool was unable to determine a serviceable transit route. This is a function of the O-D pair and its start time. For example, the trip may not be serviceable due to a transit route's operating hours. Therefore, the trip's solution is an alternative walking route. With reference to the utility model equations (Equations 2-6), the probability of choosing CTA is a function of non-zero variables such as HHI, number of transit stops in the origin and destination census tracts, and trip purposes (work, destination in CBD). Therefore, without using transit, the calculated utility values will always be non-zero values. Hence, through the logit model transformation, the calculated probabilities for walk-only trips will be greater than zero. This non-zero value exhibits error by contradicting the lack of transit serviceability of the trip. Theoretically, if the trip cannot be serviced by public transit, then the probability of selecting it as a mode is zero. As previously mentioned, the transit route is a function of the start time and trip O-D pair. Considering the O-D pairs are redefined by the centroids of the census tracts of the origin and destination, there lies a possibility for a transit solution to exist if the exact origin and destination locations were used. Thus, we employed a walking distance threshold of determine if

[8] Although this does not reflect the trip distance, it does indicate trips of relatively short distances considering there are 801 census tracts that span the city limits of Chicago. Attention should be focused on the volume of within-tract trips – this is the same volume of trips that RH served. can assume from this output that the destination is within walking distance of the origin, for a capable user. Although the physical workload or environmental constraints (such as safety) could challenge this assumption.

a trip's $P(Transit|CTA)$ should be reassigned a value of zero. Per the Pedestrian Safety Guide for Transit Agencies, people are willing to traverse 0.25-0.50 miles to access/egress transit [28]. we assumed a threshold value of 0.75 miles which is computed as the mean of this range multiplied by two to account for access and egress distances. For any trip with a walking distance greater than 0.75 miles, the $P(Transit|CTA)$ was reassigned a value of zero.

Further analysis of trips by classification is included in the next section, following the introduction of the $P(Transit|CTA)$s.

Replaceability Analysis (Probability of Selecting CTA)

The $P(Transit|CTA)$ estimation is a function of the abovementioned procedures and their respective outputs. We then developed a program in Matlab that first calculated any unknown required input, and lastly would calculate the $P(Transit|CTA)$. The results are shared below and are categorized based on their replaceability. For ease of recall, the group for a trip, T, is categorized by the following conditional:

$$replaceability\ group_T = \begin{cases} Not\ Replaced\ (NR), & 0 \leq P(Transit|CTA) \leq 0.45 \\ Replaced\ (R), & 0.55 \leq P(Transit|CTA) \leq 1.0 \\ Buffer\ Zone, & 0.45 < P(Transit|CTA) < 0.55 \end{cases}$$

Of the 7,949,902 trips output from the route analysis, approximately 8% (794,464 trips) had a probability lying within the buffer zone. As mentioned in Chapter 3, these trips are excluded from all analyses. Proceeding, all findings and conclusions are representative of R and NR groups only. Specific trip counts and standard deviations of these two groups are in the table below (Table 3).

Table 3 – Trip Count and Standard Deviation per Trip Group

Group	$n_{observed}$	$S_{observed}$
R	2,465,504	0.0775
NR	4,837,590	0.1261

Altogether, these two groups host 7,156,438 trips, equating to 90% of all trips. A summary of fundamental parameter means per group is shown in Table 4. Following, we explore the degree of skew for travel times, transfer count, and fare in Table 5.

Table 4 – Count and Means per Trip Group Type (Observed Data)

Group	Trip Count	$P(Transit|CTA)$	WT	TTT	Transfer Count	Fare	Std. Dev.
	trips		*min*	*min*	*transfers*	*$*	
R	2,465,504	0.684	13.22	27.30	0.61	2.48	0.0849
NR	4,837,590	0.221	18.66	37.16	0.93	2.61	0.1237

Referring to the trip counts in Table 4, approximately 31% of the all trips are replaced and 61% are not replaced, with 8% of trips lying in the buffer zone. The standard deviations for both groups are comparable, although the magnitude of σ_{NR} is slightly greater, which can be attributed to the larger sample size.

To obtain greater insight into the statistical meaning of the mean, each was compared to its corresponding median. Considering that the quantitative relationship between the mean and median is not fully indicative of the skew, we calculated Pearson's Second Coefficient[9] for all respective group-parameter pairs. The table below shows the results from these calculations (Table 5).

Table 5 – Skewness of Parameters per Trip Group Type (Observed Data)

Group	$P(Transit\|CTA)$			TTT (min)			Walk Time (min)			Transfer Count			Fare ($)		
	Mean	Med.	Skew	Mean	Med.	Skew	Mean	Med.	Skew	Mean	Med.	Skew	Mean	Med.	Skew
R	0.684	0.680	0.131	27.30	25.58	60.95	13.22	12.41	28.62	0.61	0	21.45	2.48	2.60	-4.24
NR	0.221	0.235	-0.340	37.16	33.35	92.49	18.66	17.02	39.85	0.93	1.0	-1.70	2.61	2.60	0.24

Six out of the ten group-variable mean pairs exhibited positive skews ranging from 0.24 to 92.49. For these pairs, we can conclude that more than 50% of the trips in their respective groups are below the mean. Consequently, there exists a volume of trips of greater magnitude at a statistically significant distance above the median. The volume of these values and their magnitudes directly influence the degree of difference and skew of the data. For example, referring to the (NR, TTT) pair, there is a significant positive skew, with a coefficient value of 92.49. This indicates that there was a notable volume of trips in the RH dataset that had transit alternatives, but the LOS of the trip was too low to be competitive. These trips represent O-D pairs that have a significant demand, but do not have a viable transit alternative, thus RH is complementary. Moreover, this behavior warrants an opportunity for CTA to implement services for these trips.

Referencing the skews for the $P(Transit|CTA)$, each group exhibits opposing signs. The negative skew for group NR indicates that there exists a greater volume of trips above the median (0.235), with $P(Transit|CTA)$ values approaching the upper bound, 0.45. Reversely, for group R, the positive skew implies there is a greater volume of trips with $P(Transit|CTA)$ values below the median (0.680) approaching the lower bound value, 0.55. Recall that 10% of all trips fall within the buffer zone, i.e. $P(Transit|CTA)$ ranges from 0.45 to 0.55. Given this percentage of trips exist about the center ($P = 0.5$), it is anticipated there to exist skews pulling groups R and NR towards this value.

Referring to the mean TTT values, the value for R group is 27% shorter than that of group NR, which was nonetheless expected – a trip is more attractive, and thus more replaceable, if the TTT is minimized. More interestingly, the mean WT is approximately half of the TTT for each group. This implies that the mean transit travel time for each group is approximately equal to the mean walk time. Thus, the degree of replaceability may not influence the temporal structure of a transit trip. Group NR had the higher mean value for number of transfers, at 0.93 transfers. This could be explained by its classification as a 'not-replaced' trip, which implies the transit-equivalent trip is not viable. One primary reason a transit trip may not compete well due to lack of connectivity; poor connectivity is indicated by greater wait times and access/egress distances, and increased IVTT due to the indirectness of the route.

[9] Pearson's Second Coefficient of Skew is calculate with the following formula: $\frac{3*(mean-median)}{std.\ deviation}$

It should be clarified that there is a zero-dollar fare associated with walking trips. Thus, the mean fare values for transit-utilized trips is greater. Further, the mean transfer count is greater than 0, hence for transit-utilized trips the mean fare will be greater than $2.35 to account for the transfer cost. This relationship can be visualized by the histogram below (Figure 9).

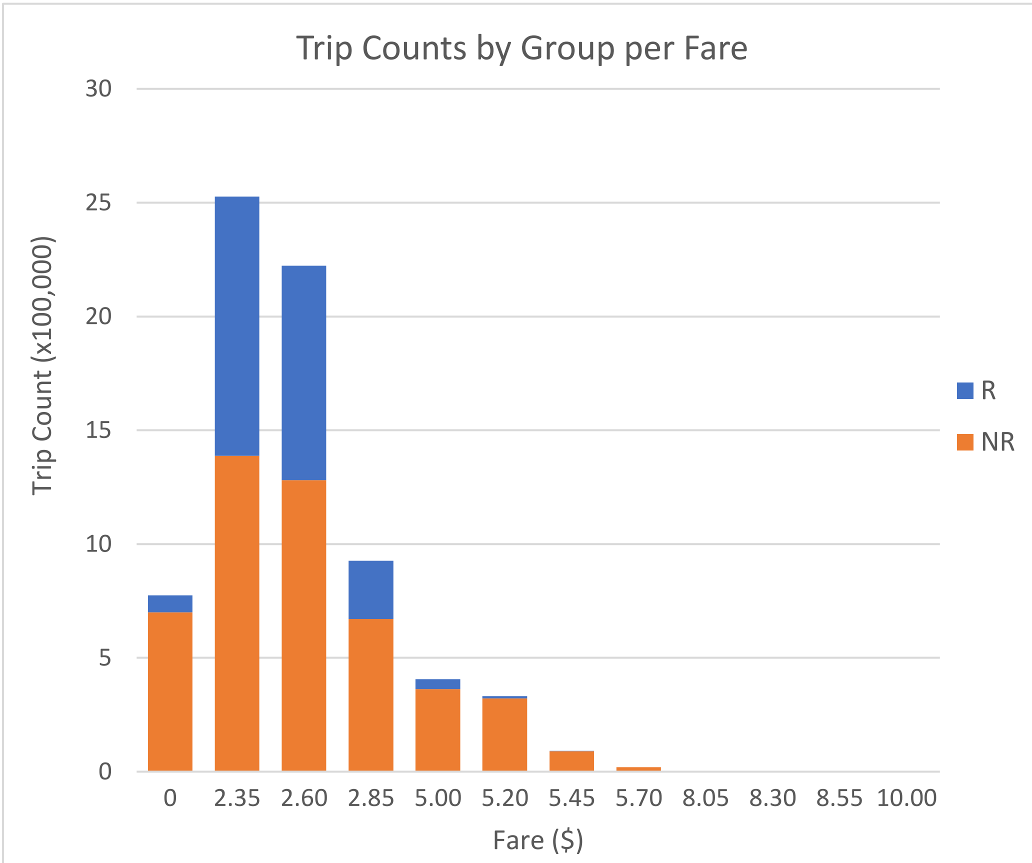


Figure 9 – Histogram of Trip Counts by Fare

Given that the 'fare' for a walk-only trip is $0, it should not be interpreted as a "free" transit trip, rather it is a trip that is most efficiently serviced by walking, which has no fare cost. Moreover, this zero-dollar fare does not fully encapsulate all expenses for walk-only trips, as it excludes cost of time, preferences, and needs.

The replaceability of a trip is dependent upon a combination of factors. First, a trip's replaceability is contingent on the persons' capabilities of physical exertion. Under certain circumstances, the replaceability of the RH trip may be incomparable due to the physical capabilities of the rider. For example, an elderly, disabled person may need to traverse two blocks to get to the grocery store. Under ArcGIS' Route Analysis program it will likely output that walking is most efficient, although given the user's conditions, walking is not an option. Secondly, the replaceability is influenced by the user's safety, which is dependent upon the perception of the route's surrounding physical environment(s). For example, consider a walk-only trip that requires the person to walk along a busy road with limited pedestrian infrastructure or one that requires an individual to walk in inclement weather conditions. In these scenarios, the surrounding environment may have greater impact on mode-choice decisions, and likely will influence the user's preference to favor personal safety. Moreover, when personal safety is of concern, people act conservatively to mitigate hazard opportunities from occurring. All of these conditions, concerns, and exceptions cannot be explicitly accounted for in our model. Therefore, when examining the proceeding results, these points should be taken into consideration.

In the following table is a matrix of trip counts for walk-only trips by their transit-equivalent classification (2A = between-tract, 2B = within-tract), and replaceability group. From Table 2, out of the 7,949,902 trips, 806,254 are classified as walk-only. With 219,903 trips lying in the buffer zone, the remaining 586,351 walk-only trips are depicted in the table below (Table 6). Assuming a normal distribution, this portion of trips in the buffer zone to the R and NR trip volume is expected.

Table 6 – Walk-Only Trip Counts by Classification and Group

Group	Between-Tract	Within-Tract	*Sum*
R	174,121	172,752	*346,873*
NR	198,610	40,868	*239,478*
Sum	*213,620*	*372,731*	*586,351*

In comparison to the original RH trip, the alternative being a walk-only does not necessarily cost the person greater time. The selection of RH at a greater cost accounts for the advantages of and opportunities that come with RH. These opportunities are preferential, and may be exhibited by the RH trip, or exhibited by its alternatives and is consequently is avoided through RH. As stated in the literature review, examples include level of convenience, comfort, and cleanliness [9]. This extends to trips that utilize transit; we can assume the cost difference between a RH trip and its transit-equivalent trip represents the difference between the environmental conditions and individual's preferences per alternative. This cost differential is further explored in the next subsection, *Time Value Analysis*.

The replaceability of a trip has many implications, one of which regards the LOS and operations of a network, *congestion*. RH trips that have a transit-utilized solution have an alternative delay that is different from RH trips that replace walk-only trips. For transit utilized trips, there is an overall decrease in delay per person. Essentially the demand shifts modes to transit and becomes pooled. The delay from the original RH trip is eliminated and delay is added from the transit services, but the delay per person is significantly lower for transit. These delays are inherent to the transit system, but vary in magnitude depending on type (bus, rail), service region (suburbs, city), and the hour and day. For bus services, an increase in passengers consequently results in increased delays from dwell times at access and egress points and vehicle entry and exit delays that are a result of demand. Buses will run with zero utility (no passengers), hence there is a baseline level of congestion added to the network since every bus is one more vehicle in the network. Although, as the utility increases (passenger volume increases), the magnitude of delay contributed by the bus increases. Moreover, this magnitude is a function of the existing LOS of its route. A bus at 50% capacity in the suburbs will incur less delay than that of the same bus in the city. For rail services, the contributing congestion is less sensitive to an increase in passengers. Considering the systematic structure of rail services, vehicles enter stations and execute stops with or without demand. As opposed to buses, the vehicle entry and exit delay is inherently a part of the baseline level of congestion. However, rail delays are influenced by dwell times as an increase in passenger volumes warrants an increase in time for riders to enter and exit the vehicle. Nonetheless, these confounding delays are expected from a growth of public transit ridership. However, it is assumed that an increase in public transit ridership implies a decrease in alternative mode demand, such as personal vehicles.

In opposition to transit-utilized trips, the replacement of a walk-only trip implies the elimination of delay induced from RH, and transferred to the pedestrian. Given the nature of pedestrians, they do not contribute high volumes of delay to transportation networks. First, pedestrians do not use the network like vehicles do. Other than pedestrian crosswalks, there is essentially no shared right of way (ROW). Secondly, pedestrians occupy significantly less ground area. Hence, an increase in pedestrian volumes does not imply the same delay as an increase in vehicle volumes. For example, the delay incurred by 10 RH trips (10 individual riders) is likely greater than that of 10 pedestrians. Thus, transfer of demand (replacement) of walk-only trips by RH services implies the addition of delay to the transportation network.

These are important relationships that must be considered when analyzing the quantitative output and conclusions of this study.

Time Value Analysis

All observed trips were classified into one of four classes (**Error! Reference source not found.**). To classify a trip, the TTT and fare for the RH trip and its transit alternative were compared. For each class, there is a scenario differing the relationship between the TTT_{RH} from $TTT_{Transit}$ and $fare_{RH}$ and $fare_{Transit}$. For purposes of convenience we have restated the criterion for all four classes (I-IV) below. Following the class definitions is a summary table containing descriptive statistics per class (Table 7).

I. These are trips where, in comparison to the RH trip, the transit-equivalent trip was less expensive, but slower.
II. These are trips where, in comparison to the RH trip, the transit-equivalent trip was less expensive **and** faster.
III. These are the trips where, in comparison to the RH trip, the transit trip was quicker, but more expensive.
IV. These are the trips where, in comparison to the RH trip, the transit-equivalent trip was more expensive **and** slower.

Table 7 – Count, Mean, and Median per Cost-Travel Time Class

	Class I	Class II	Class III	Class IV
Fare	*Transit < RH*	*Transit < RH*	*Transit > RH*	*Transit > RH*
TTT	*Transit > RH*	*Transit < RH*	*Transit < RH*	*Transit > RH*
Trip Count	7,174,581	39,374	732,833	3,114
% Total	90.25%	0.5%	9.22%	0.03%
$\overline{P(Transit \mid CTA)}$	0.4519	0.5432	0.6179	0.6643
$P(\widetilde{Transit} \mid CTA)$	0.4104	0.6152	0.6399	0.7216

These results align with the existing conclusions that RH trips are perceived to be faster than the transit alternative. Class I and II trips exhibit this condition yielding a conclusion that 99.47% of

RH trips are quicker than their transit alternative. This finding quantitatively supports the stated preference for RH (versus transit) because of the decrease in travel time. Of this percentage of trips, 90.25% (Class I) exhibited RH trip fare greater than that of the transit alternative. With reference to the economic concept, *opportunity cost*, for Class I trips, the riders' chose to pay more (cost) for a quicker trip in return (opportunity). Although, given that mode-choice decisions are multifaceted, in this scenario, the explicit cost difference only accounts for the difference in travel times. Moreover, there are likely implicit costs associated with each decision, so the relationship between these two variables may not be as clear.

The confidence in the output from this time value analysis can be challenged by the nature of the datasets used. Recall that the O-D geographic coordinates are that of the corresponding census tracts that contain the O or D. This variance between the experimental and actual geographic location may impact the accuracy of the transit trip and its path. Consequently, the TTT and fare will have some error built in.

Sensitivity Analysis

Results from the parametric sensitivity analyzes are extensive, so the results will be summarized and discussed. In addition to the figures provided in this section, there are more detailed supporting figures provided in the appendices. The supporting figures show the trend in $P(Transit|CTA)$ and population size for R and NR groups separately, per sensitivity condition.

For each trip, the variable of interest was adjusted and rerun through the program. All intermediate variables dependent upon the decision variable were also recalculated in the program. The new $P(Transit|CTA)$ was computed, and the trip was recategorized based on its magnitude. Thus, it is important to consider that for each percent-change condition, the sample for R and NR trip groups will vary in size. Considering $P(Transit|CTA) = [0 \cup 1]$, a change in a group's sample size indicates that the difference in trips has transferred to either the buffer zone or the opposing trip group (R/NR). Moreover, means of both groups will fluctuate with the sensitivity condition.

Results of T-Test

For the seven aforementioned sensitivity parameters, t-tests were conducted per group and condition, hence a total of 288 t-tests were executed. We fail to reject the null hypothesis for 17 scenarios (Variable, %Δ, Group), all of which were exhibited by TC and AirPass values. The conditions of the rejected scenarios are listed below by variable, percent-change, and group.

i. TC
    a. ± 5% (R)
    b. ± 10% (R)
    c. ± 5% (NR)
ii. Airpass
    a. + 5% (R)
    b. - 5% to - 50% (R)

From these results, we can conclude that for each scenario, an adjustment of the variable by the corresponding percent change will not yield a significant change in $P(Transit|CTA)$ for the stated group. The remaining 271 scenarios exhibited t-values greater than the critical value. This implies that with 5% error, we can assume the adjustment of each of their corresponding parameters will have a statistically meaningful impact on the mean probability. A table of these values per condition and variable are in Appendix C.

Overview of Sensitivity Test Results

In the following section, we introduce a plot that provides the means to compare between each variable's influence on the probability of a person selecting public transit (Figure 1010). This figure displays the sum of all trip probabilities across groups R and NR. On the primary y-axis, is the summation of probabilities in millions. The maximum value for one condition is the number of trips in groups R and NR multiplied by the maximum probability, 1. Although, this value could only be obtained if every trip had a $P(Transit|CTA) = 1$. On the x-axis is the sensitivity condition, percent-difference of the variable from the observed value. There are 20 sensitivity conditions, and 1 observed value (at 0% change), thus for each sensitivity variable, there are 21 points plotted where each share the observed value. Hence, all connecting lines intersect at Sensitivity Condition = 0%.

Secondly, we introduce stacked bar charts per sensitivity variable that depicts the overall weighted mean P per sensitivity-condition, with the volumetric distribution of trips between groups R and NR (Figures 11-17). On the primary y-axis is the total weighted mean probability, which is a measure of the sum of group R and NR's contribution to the probability using the following equation:

$$P_{weighted} = P_R\left(\frac{n_R}{n_T}\right) + P_{NR}\left(\frac{n_{NR}}{n_T}\right) \quad (10)$$

The data labels (percentages) within each bar corresponds to the percentage of total trips (n_T) that each group contains. Per the legend, the blue portion of the stacked bar corresponds to the replaced trips, whereas the orange portion corresponds to the not replaced trips. It should be clarified that these percentages are independent from the portion heights of the stacked bars. In some scenarios, the height of the bar may be increasing as the percentage decreases. When analyzing these figures, the subset of trips should be considered. For each condition, the total sample size only includes trips where $0 \leq P \leq 0.45$ or $0.55 \leq P \leq 1$, meaning that all trips in the buffer zone are excluded. To provide greater insight into the behavior within each trip group, we provide two figures per variable in Appendix D; each combination graph compares the sample size and mean probability for the R and NR groups separately, by sensitivity condition. These figures are located in the appendices out of consideration for the paper's length, although these visuals are important to reference when analyzing the findings.

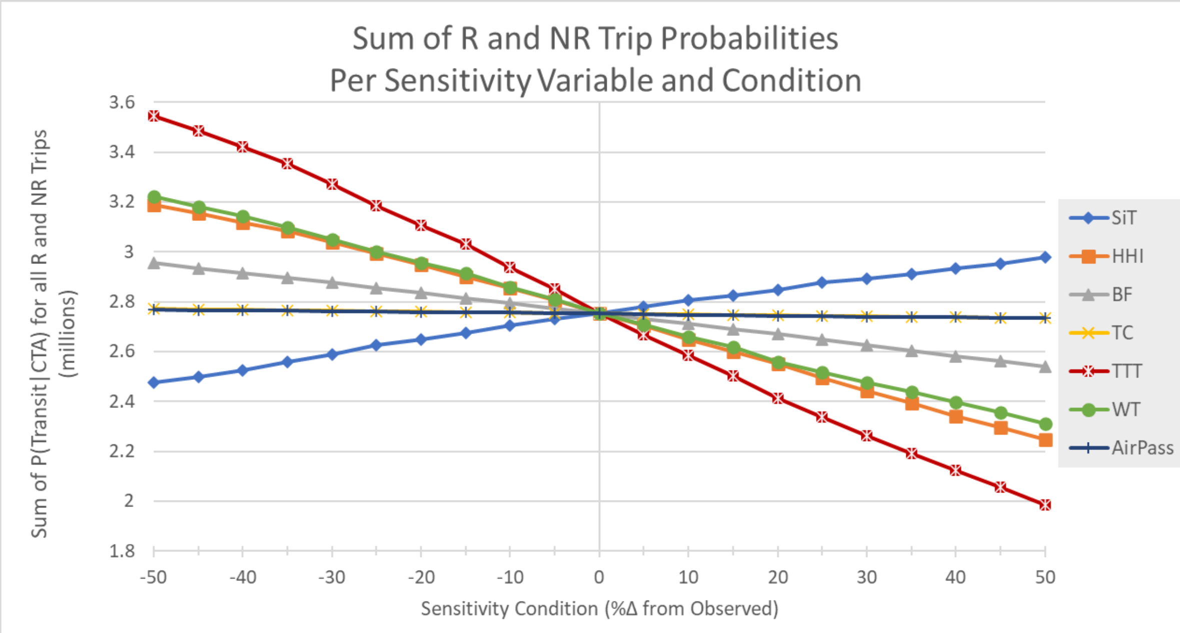


Figure 10 – Sum of R and NR Trip Probabilities per Sensitivity Variable and Condition

In Figure 10, the slopes illustrate $P(Transit|CTA)$'s sensitivity to each variable. An increase a slope's steepness implies an increase in $P(Transit|CTA)$ sensitivity. Six variables (HHI, BF, TC, TTT, WT, AirPass) exhibit a negative relationship with $P(Transit|CTA)$ whereas SiT exhibits a positive relationship with $P(Transit|CTA)$. Each of these seven directionalities (positive or negative) were expected, however the magnitude of each slope was unknown. It can be concluded that TTT has the greatest influence on $P(Transit|CTA)$, whereas TC and AirPass have the least influence.

For the number of transit stops in the pick-up/drop-off area (SiT), we predicted that an increase in transit stops would cause a positive shift in the $P(Transit|CTA)$s for all trips. Inherently, a greater volume of transit stops implies greater service, increased frequencies, and decreased wait times. Hence, an increase in SiT would yield shorter TTT and WT, which in return, would increase the attractiveness of the transit alternative.

While there are six variables that share the same negative directionality, only four of those significantly influence $P(Transit|CTA)$. Positive changes in TTT, HHI, WT, and BF values all yield decreases in $P(Transit|CTA)$, with the listed order corresponding to their level of influence. An adjustment in transfer costs (observed value of \$0.25) and the airport pass prices (observed value of \$5.00) yield minimal change in $P(Transit|CTA)$. Thus, we can conclude that altering these variable's values will not significantly impact CTA ridership. This is further supported by summary of the t-test in *Results of T-Test* on page 33.

The weak sensitivity of $P(Transit|CTA)$ to the transfer cost can be reasoned by the proportion of the transfer cost to the total trip fare. Consider a transit trip with one transfer; the total trip fare would equal $2.60 (base fare of $2.35 + transfer cost ($0.25). The transfer cost is roughly 10% of the total trip fare, thus any adjustment in the transfer cost would yield negligible changes in the total trip fare. Moreover, the average number of transfers between both groups is less than 1 (Table 4). Hence, the average transit trip will can be described by the aforementioned scenario. Figure 11 below further supports this claim by displaying the volumetric distribution of trips and their weighted means per condition. The distribution and weighted means at the -50% and +50% sensitivity conditions are nearly identical to the values under the observed environment. Hence, an increase in TC would little impact on ridership behaviors and levels. To increase revenue for transit agencies without disrupting existing behaviors, an increase in cost per transfer could be considered.

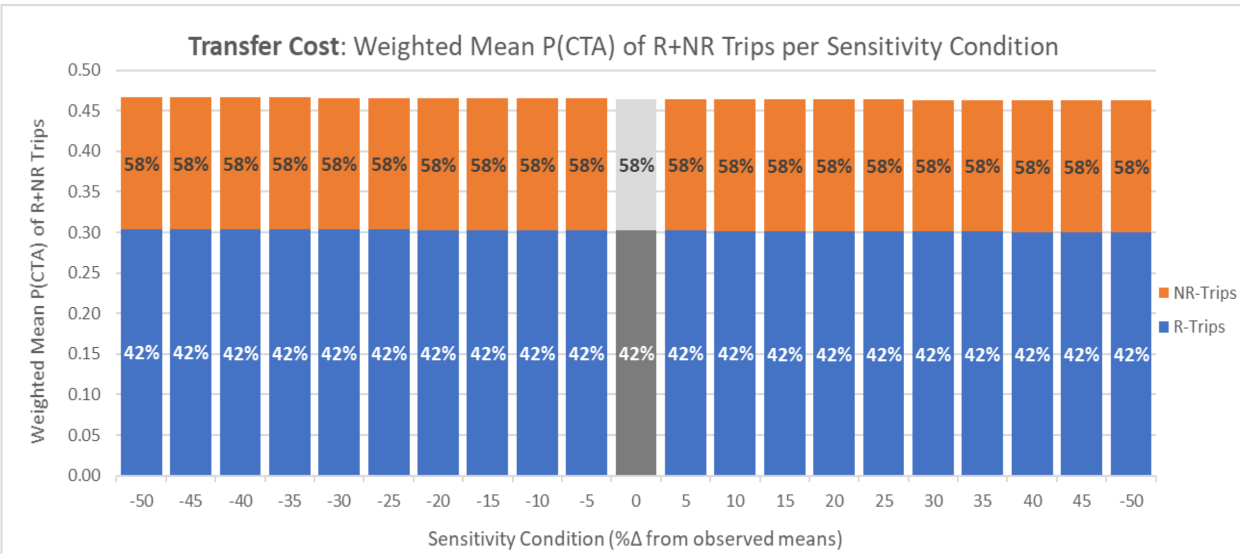


Figure 11 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Transfer Cost (TC)

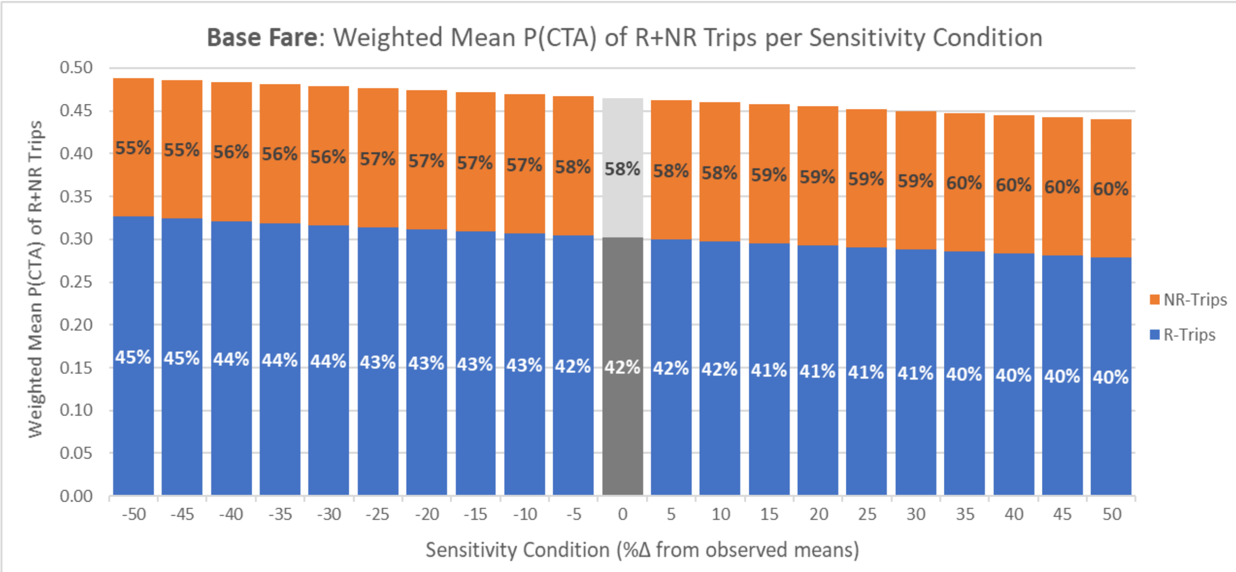


Figure 12 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Base Fare

The probability's sensitivity under varying base fare conditions is more distinct than that of the transfer cost (Figure 12). This difference can be attributed to the price, \$2.35 as opposed to \$0.25. Secondly, recall that for both groups R and NR, the average number of transfers was 0.61 and 0.93 respectively, meaning that overall, the average transit trip has approximately one transfer. A transit trip with one transfer equals the base fare (\$2.35) plus the cost of one transfer (\$0.25), totaled to \$2.60, meaning the value the base fare has greater weight. Thus, an equivalent percent-change in base fare, as opposed to transfer cost, would impact $P(Transit|CTA)$ to a greater degree.

Recall that the airport pass price is the cost of a one-way 'ticket' to the airport via CTA. The adjustment of this pass price yields minimal deviation in $P(Transit|CTA)$ shown by the flat slope in Figure 10. In Figure 13, the distribution of R and NR trips is shown per sensitivity condition. 42-43% of trips to-or-from the airport are considered replaced, whereas 57-58% are considered not replaced. The deviation from the observed distribution is indicative of the small percentage of trip to or from the airport.

In Chicago, CTA's blue line, "L", is a bus rapid transit (BRT) service between Chicago O'Hare International Airport and the Forest Park Terminal located west of the downtown Chicago. Transit agencies often provide services to and from the airport that are more direct than other modes, such as RH or driving which incur additional delay from entering and maneuvering airport grounds. Therefore, transit services exhibit great utility when arriving at or departing the airport; in comparison, transit takes passengers closer to the access point, mitigating lost time in queues. This is highly beneficial, especially since trips to the airport tend to be constrained by time (flight departures and arrivals). Even considering these factors challenging the utility of non-transit modes, the demand to-and-from the airport via RH is still considerable. It can be assumed that there exist balancing preferences towards RH to serve the remaining first or last mile of the trip. In other words, this arrangement connecting the last stop to home, or home to first stop, perceived to be so taxing in terms of time and/or workload, that RH appeals to this disutility. Given that Forest Park Terminal is not centrally located downtown, nor does it provide viable service to north and south Chicago, taking transit to or from Chicago O'Hare is likely justified by the longer travel times, and number of transfers. This is a primed opportunity for CTA to expand its services to outlying regions, with competitive modes such as demand response transit (DRT), transit shuttles, or subsidized alternatives. This also serves as an opportunity for transit agencies to collaborate with RH companies to service the FLM via RH, and the main leg to the airport via CTA.

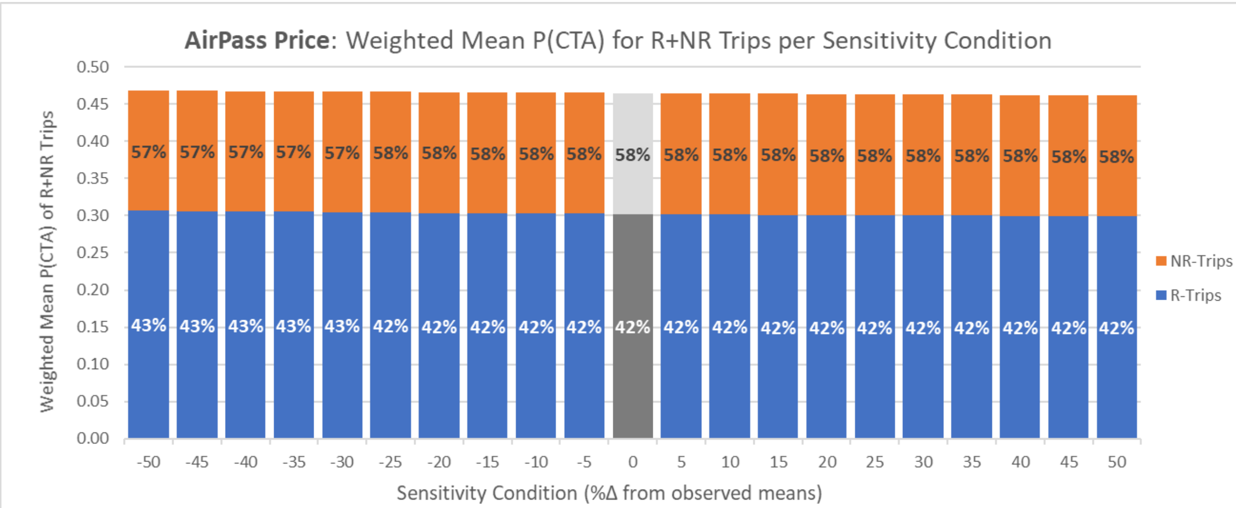


Figure 13 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Airport Pass Price (AirPass)

Next, are the variables associated to travel time: WT and TTT (Figures 14 and 15). For the negative sensitivity (from -50% to 0%) conditions, the TTT exhibits a slightly steeper slope than that of the WT. Whereas for the positive sensitivity (0% to 50%) conditions, the TTT and WT slope are nearly identical, as they overlap in the plot. The difference in slopes for the negative sensitivity conditions can be explained by TTT's formula and the nesting of WT in its value.

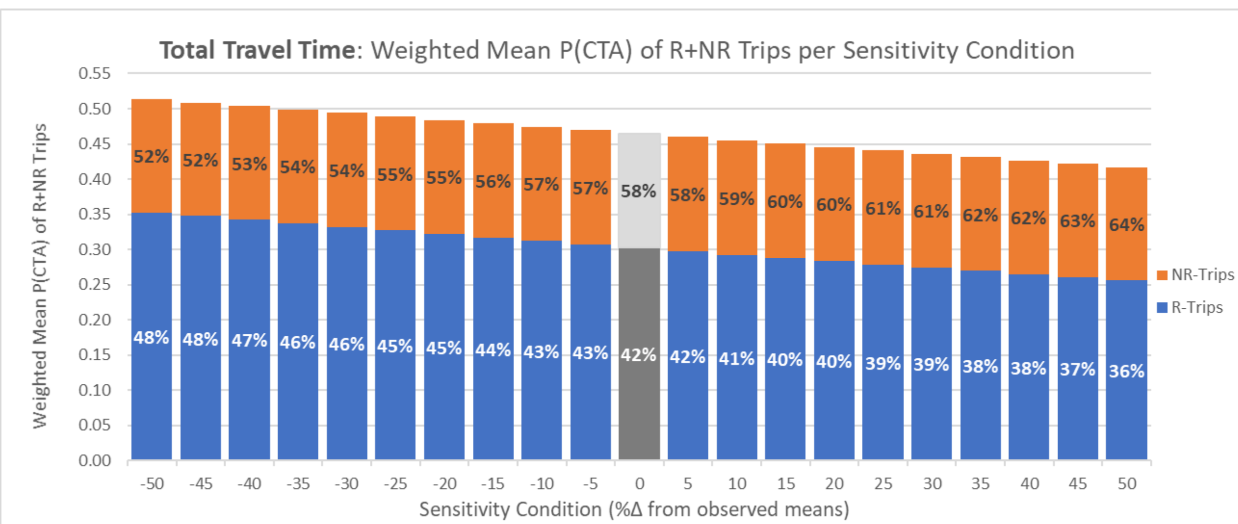


Figure 14 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Total Travel Time (TTT)

Recall that the TTT is the sum of the IVTT, wait time, and walk time. Meaning, a 50% decrease in TTT includes a 50% decrease in IVTT and wait time in addition to a decrease in walk time. Whereas a 50% decrease in WT does not include the reduction in IVTT and walk time. When the new TTT is calculated, it uses the observed wait time and IVTT, but changes the WT. While these differences exist, they are relatively small in comparison to their distribution patterns and weighted probability values. This can be seen by the similarity in trip volumes and percentages between Figures 14 and 15. Moreover, the overall weighted mean $P(Transit|CTA)$ and the two probabilities it is composed of, are almost identical for every sensitivity condition.

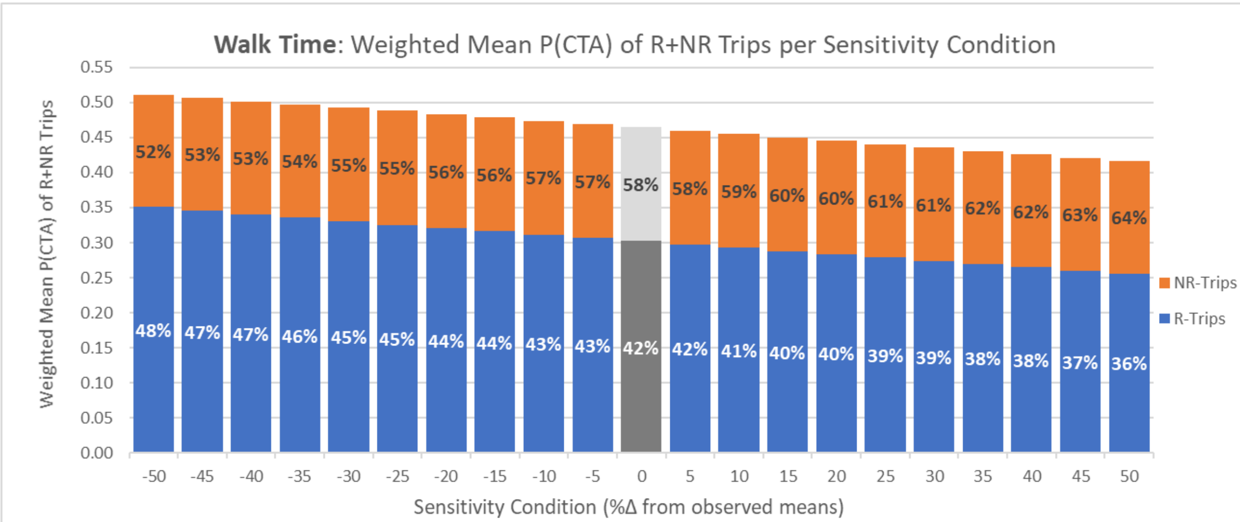


Figure 15 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Walk Time (WT)

The next variable, average household income (HHI), has less impact on $P(Transit|CTA)$ per Figure 16. Although we must account for its static behavior and derivation. The utility model called for the input household income to be rider specific. In existing literature, it was determined that ridehailers exhibit demographic characteristics that are at variance with the average American. Ridehailers were found to be more educated and be of a higher income class.

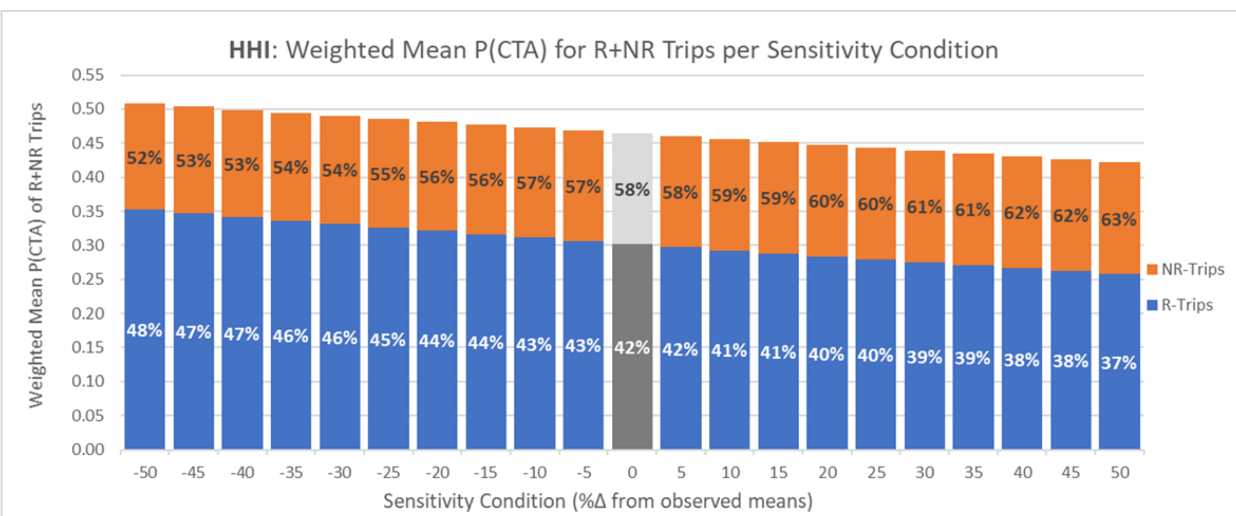


Figure 16 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Average Household Income (HHI)

Therefore, the use of the average HHI may undervalue that of the average ridehailer and the accuracy of these results could be challenged. Nonetheless, the trend and behavior HHI has on $P(Transit|CTA)$ is transposable.

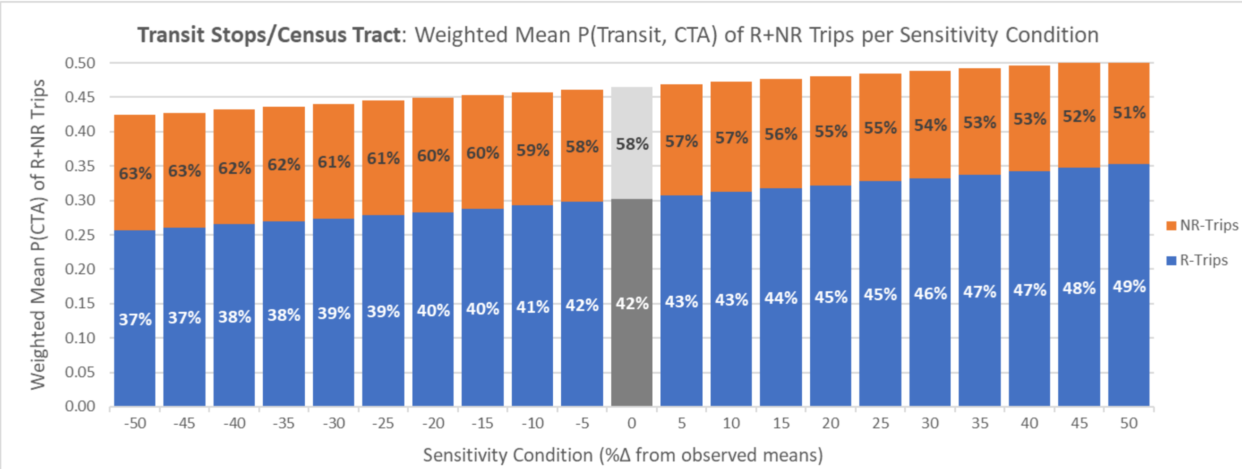


Figure 17 – Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Transit Stops per Census Tract

The transit stops per census tract (SiT) was the only variable where the was a positive correlation between the $P(Transit|CTA)$ and the sensitivity condition. Similar to the other six variables, this positive relationship was predicted. The number of transit stops in a network has many implications on operations and ridership. An increase in transit stops implies an increase in route LOS. As the distance between consecutive stops is decreased, the average access and egress distance decreases. Moreover, as the accessibility of transit services increase, the volume of serviceable patrons increases, and an increase in frequency is more likely. Although there exist caveats with the more extreme positive sensitivity conditions. With reference to its source, these are likely not captured by the utility model. The addition of transit stops to an existing route must be optimized to account for the consequence: add lost time. At every transit stop, delay is incurred in the operational timeline when approaching, operating at, and exiting the stop.
The first being the dwell time, which is the amount of time a transit vehicle waits at the stop. Embedded in the dwell time is boarding time – the time required for all approaching individuals to enter the vehicle and their ridership be validated. This can quickly accumulate during peak period hours when there are large platoons of approaching riders, and there is discontinuity in payment forms. Additionally, this is a consequence of increased ridership. The second source is called the 're-entry' delay; this is the time required for the driver to merge into oncoming traffic. For every additional transit stop, one re-entry delay is incurred per cycle. The summation of these delays per stop and per cycle can adversely affect the travel time between stops, and the TTT of each rider. In summary, the addition of transit stops increases the accessibility and consequently, utility. Although designing addition to increase ridership must strategically consider implications it has on the existing travel times and LOS attributes.

Chapter 5: Summary and Conclusions

The impact of RH services on the recent decline in public transit ridership has not been widely explored. The current body of research is constrained to empirical studies that vary in methodologies used and study relatively small samples. These studies analyze preferences and experiences explained by users, but do not explicitly include trip records and their attributes. Thus, when aggregated, the conclusions yield a variety of results and implications, yielding conclusions that cannot be widely agreed upon.

Moreover, sample sizes in existing literature have been restricted by the framework of empirical methods. To our knowledge, there are no studies that explore the research question using a massive dataset containing individual trips. Further, our findings are derived from a 30-day study period covering four full weeks. This allows behavioral results and trends to be represented with greater confidence.

Lastly, our approach to exploring the research question is resourceful and novel. We define the replaceability of a RH trip by a series of spatial and mathematical analyses. First, the real-time transit equivalent trip was computed using the GTFS-integrated ArcGIS Route Analysis. Then, the probability of choosing transit over all other alternatives defined if the transit-equivalent trip was a viable option and replaced by RH or if it was incomparable. If a trip was deemed the latter, the use of ridehailing supplemented the unpractical transit services.

Our findings indicate that 31% of ride-hailing trips were executed where the transit alternative exhibited a competitive utility, with respect to travel times, fare/expenses, and workload. Over the month of June, the total revenue lost from trips replaced by ride-hailing is estimated to be $6,114,450[10]. If we assume the percentage of replaced trips and trip counts for each month can be represented by June 2019, then the total loss in fare revenue over one year would be approximately 73 million dollars. Further, the ramifications of the demand transfer to RH services is not fully represented by the loss in revenue. As such, public transit agencies should employ strategies to increase transit utility such that a significant portion of this estimate can be recovered.

As summarized in the introduction, the RH decision making process is highly complex, situational, and the output is variable. The utility model used in this paper accounted for travel time, walking distance, fare, trip purpose, distribution of transit stops, transfer count and cost, and household income. While this model includes LOS attributes and household income, many situational factors are not considered. These unexplored factors serve as an opportunity to obtain a deeper understanding of the mode choice decision making process. One example would be to determine if a relationship exists between ridehailing trip demand and inclement weather conditions.

Next, further analyses of the replaced trip groups should be executed. Replaced trips have transit-equivalent trips that are comparable to the RH trip, in terms of LOS attributes. However, the selection of RH can be attributed to personal preferences and perceptions towards ridehailing

[10] Estimated by multiplying the average fare ($2.26) of the replaced trips by the number of replaceable trips.

that outweigh the utility of transit. Future research should focus on studying mode-choice behavior to thoroughly understand the conditions in which a person selects RH instead of transit services. Regarding NR trips, transit agencies should turn inwards and evaluate services, or the lack thereof, in the corresponding origin and destination zones.

Considering our research was limited to the city of Chicago, the continuation of this study in different cities and suburbs will yield more representative conclusions. Moreover, it will allow for the identification of behavioral trends and geo-specific characteristics that influence RH ridership.

Given the scope of this project, we were unable to further explore the behavior of pooled trips. The dataset used provides indicators of a pooled trip and the number of passengers pooled in one trip. Future research should focus on modeling pooled trips, and their differentiation from single-occupancy RH trips. Inherently, pooled riders exhibit the willingness to compensate travel time, privacy, and walking time for a reduced travel cost. In most circumstances, when selecting transit over RH, riders are willing to have a greater travel time for a smaller fare. This opportunity cost perspective parallels with ridehailers selecting pooled over single occupancy. Given the similarities, the selection of pooled RH over transit and pooled riders, should be investigated.

Publicly available RH trip data will likely maintain its anonymity by recording origins and destinations as their census tract centroids. Given it is unlikely for the precision to increase, studies that are macroscopic and encompass all attributes types (temporal, spatial, monetary) should be executed. However, the use of our methodologies and approach is only doable for regions that mandate the submission of all RH trips. Like the City of Chicago, government agencies across the United States should require TNCs to report all RH trips with trip attributes that include spatial and temporal parameters of the origin and destination. Recording and releasing this data will enable institutions to publish research that will provide a greater understanding of how RH impacts the transportation network and economy.

References

[1] "The history of Uber." Uber. https://www.uber.com/newsroom/history/ (accessed June 6, 2020.

[2] A. Greiner, M. McFarland, I. Sherman, and J. Tse. "A History of Lyft, From Fuzzy Pink Mustaches to Global Ride Share Giant." CNN. https://www.cnn.com/interactive/2019/03/business/lyft-history/index.html (accessed June 11, 2020, 2020).

[3] S. Melendez, "How Uber Conquered the World in 2013," vol. 2020, ed: Fast Company, 2014.

[4] J. Jiang, "More Americans are using ride-hailing apps," in *FactTank* vol. 2020, ed. Online: Pew Research Center, 2019.

[5] "Ride-hailing Phenomenon Under Hood." HEREmonility. https://mobility.here.com/learn/smart-mobility/ride-hailing-phenomenon-under-hood (accessed 2020).

[6] L. Rayle, D. Dai, N. Chan, R. Cervero, and S. Shaheen, "Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco," *Transport Policy*, vol. 45, pp. 168-178, 2016, doi: 10.1016/j.tranpol.2015.10.004.

[7] *APTA Ridership by Mode and Quarter 1990-Present*.

[8] "Population by Age," 2008-2018 ed. Online: Kaiser Family Foundation, 2018.

[9] S. D. Contreras and A. Paz, "The effects of ride-hailing companies on the taxicab industry in Las Vegas, Nevada," *Transportation Research Part A: Policy and Practice,* vol. 115, pp. 63-70, 2018, doi: 10.1016/j.tra.2017.11.008.

[10] J. Lo and S. Morseman, "The Perfect uberPOOL: A Case Study on Trade-Offs," *Ethnographic Praxis in Industry Conference Proceedings,* vol. 2018, no. 1, pp. 195-223, 2018, doi: 10.1111/1559-8918.2018.01204.

[11] S. Y. Amirkiaee and N. Evangelopoulos, "Why do people rideshare? An experimental study," *Transportation Research Part F: Traffic Psychology and Behaviour,* vol. 55, pp. 9-24, 2018, doi: 10.1016/j.trf.2018.02.025.

[12] R. R. Clewlow and G. S. Mishra, "Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States," UC Davis Institute for Transportation Studies, Research Report 2017. [Online]. Available: http://www.reginaclewlow.com/pubs/2017_UCD-ITS-RR-17-07.pdf

[13] M. Javanmardi, M. F. Langerudi, R. S. Anbarani, and A. K. Mohammadian, "Mode Choice Modeling Using Personalized Travel Time and Cost Data," University of South Florida, Tampa, Flordia, 2117-9060-02-C, May 2015 2015.

[14] A. Henao and W. E. Marshall, "The impact of ride-hailing on vehicle miles traveled," *Transportation,* 2018, doi: 10.1007/s11116-018-9923-2.

[15] M. Stiglic, N. Agatz, M. Savelsbergh, and M. Gradisar, "Enhancing urban mobility: Integrating ride-sharing and public transit," *Computers & Operations Research,* vol. 90, pp. 12-21, 2018, doi: 10.1016/j.cor.2017.08.016.

[16] M. Graehler, R. Mucci, and G. Erhardt, "Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes?," presented at the TRB 2019 Annual Meeting, Washington, D.C., 2018.

[17] N. Sadowsky and E. Nelson, "The Impact of Ride-Hailing Services on Public Transportation Use: A Discontinuity Regression Analysis," Bowdoin College, Working Paper, May 26, 2017, 2017.
[18] C. Murphy, "Shared Mobility and the Transformation of Public Transit," Shared-Use Mobility Center (SUMC), Research Analysis TCRP Project J-11, Task 21, 2016.
[19] *Transportation Network Providers - Trips*, City of Chicago.
[20] *QuickFacts Chicago city, Illinois*, United States Census Bureau.
[21] N.A. "DATA USA: Chicago, IL." Deloitte. https://datausa.io/profile/geo/chicago-il (accessed May 25, 2020, 2020).
[22] M. Hughes-Cromwick, "2018 Public Transportation Fact Book," American Public Transportation Association, December 2018 2018.
[23] "Facts at a glance." Chicago Transit Authority (CTA). (accessed May 25, 2020, 2020).
[24] "Transportation Network Providers (Ride-Hail)." City of Chicago. https://www.chicago.gov/city/en/depts/bacp/supp_info/transportation-network-providers.html (accessed June 11, 2020, 2020).
[25] *CTA GTFS*, OpenMobilityData, May 16, 2019 - July 31, 2019.
[26] *Street Center Lines*, City of Chicago.
[27] *Boundaries - Census Tracts - 2010*, City of Chicago.
[28] D. Nabors, R. Schneider, D. Leven, K. Lieberman, and C. Mitchell, "Pedestrian Safety Guide for Transit Agencies," FHWA, Online, February 2008 2008.

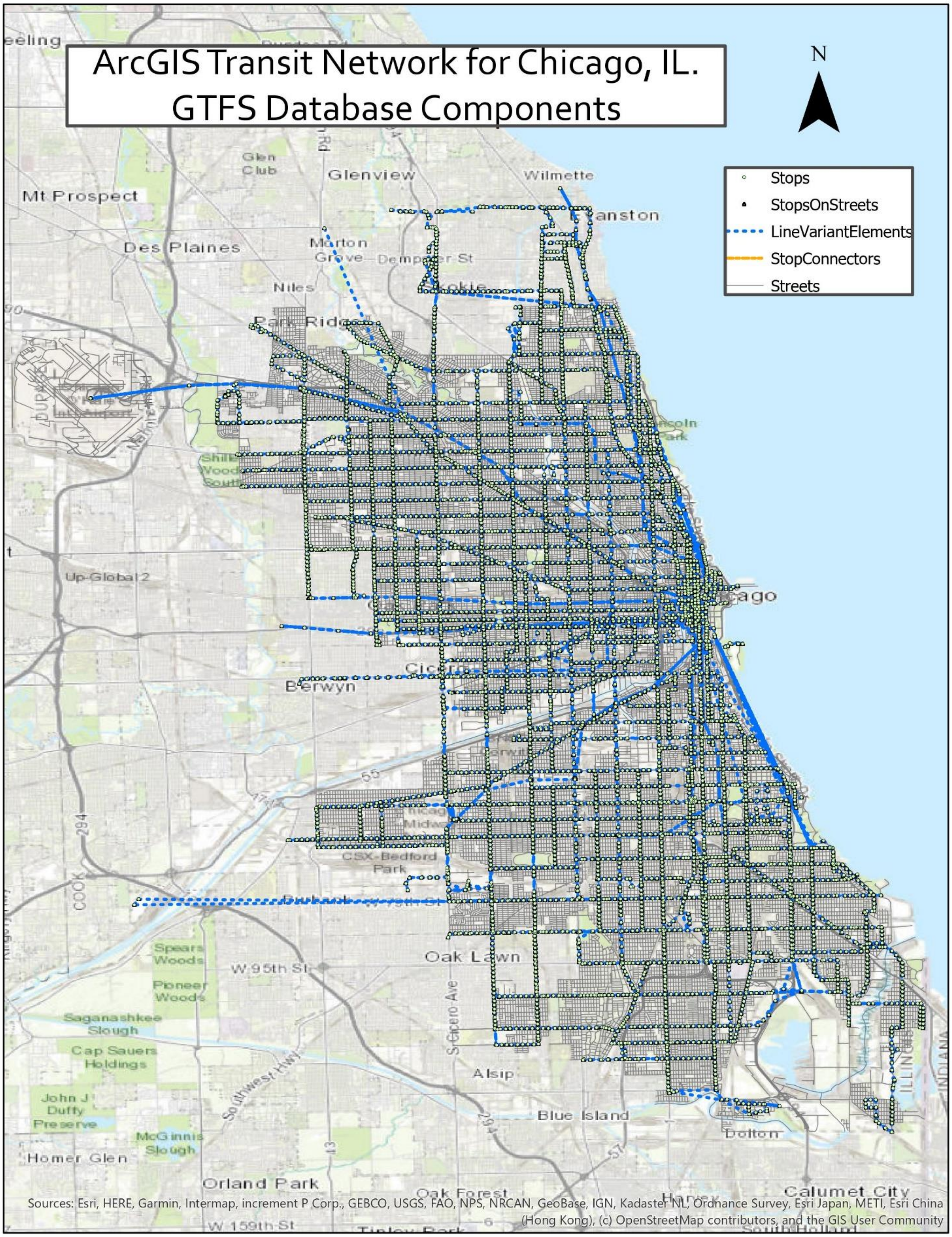
Appendices

Appendix A: TNC Dataset Description

This is the source dataset that contained all ridehailing trips within Chicago from June 1-June 30, 2019. The city's ordinances require for all trips to be descriptively reported. For each ridehailing trio, the following was recorded and is accessible in this dataset. The bolded items were used in our study.

1. Trip ID
2. Trip Start Timestamp
3. Trip End Timestamp
4. Trip Seconds
5. Trip Miles
6. Pickup Census Tract
7. Dropoff Census Tract
8. Pickup Community Area
9. Dropoff Community Area
10. Fare
11. Tip
12. Additional Charges
13. Trip Total
14. Shared Trip Authorized
15. Trips Pooled
16. Pickup Centroid Latitude
17. Pickup Centroid Longitude
18. Pickup Centroid Location
19. Dropoff Centroid Latitude
20. Dropoff Centroid Longitude
21. Dropoff Centroid Location

Appendix B: ArcGIS Transit Network Map with GTFS Components



Appendix C: T-Test Results for Sensitivity Analysis

C.1- Transit Stops per Census Tract (SiT)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
SiT	R	5	0.6853	2495025	0.0061	0.0804	21.232	Yes
SiT	NR	5	0.2222	4809856	0.0161	0.1264	20.706	Yes
SiT	R	10	0.6868	2525322	0.0062	0.0772	44.065	Yes
SiT	NR	10	0.2239	4779661	0.0163	0.1264	41.197	Yes
SiT	R	15	0.6883	2558169	0.0062	0.0774	65.628	Yes
SiT	NR	15	0.2250	4735662	0.0165	0.1268	55.374	Yes
SiT	R	20	0.6896	2586860	0.0063	0.0777	83.617	Yes
SiT	NR	20	0.2262	4706984	0.0166	0.1270	69.189	Yes
SiT	R	25	0.6912	2618863	0.0064	0.0780	107.414	Yes
SiT	NR	25	0.2281	4670298	0.0169	0.1274	92.251	Yes
SiT	R	30	0.6926	2642486	0.0064	0.0782	126.769	Yes
SiT	NR	30	0.2291	4637077	0.0170	0.1278	104.046	Yes
SiT	R	35	0.6938	2668359	0.0065	0.0784	145.108	Yes
SiT	NR	35	0.2305	4606489	0.0172	0.1281	119.437	Yes
SiT	R	40	0.6951	2696837	0.0065	0.0786	163.663	Yes
SiT	NR	40	0.2317	4569306	0.0174	0.1285	133.102	Yes
SiT	R	45	0.6963	2725702	0.0066	0.0788	180.931	Yes
SiT	NR	45	0.2329	4532542	0.0176	0.1288	147.420	Yes
SiT	R	50	0.6975	2759675	0.0067	0.0789	199.042	Yes
SiT	NR	50	0.2343	4492632	0.0178	0.1292	163.327	Yes
SiT	R	-5	0.6823	2438043	0.0059	0.0820	20.347	Yes
SiT	NR	-5	0.2195	4873169	0.0158	0.1267	12.916	Yes
SiT	R	-10	0.6806	2405911	0.0059	0.0772	44.856	Yes
SiT	NR	-10	0.2176	4902156	0.0156	0.1253	36.351	Yes
SiT	R	-15	0.6788	2373237	0.0058	0.0770	70.783	Yes
SiT	NR	-15	0.2158	4935580	0.0154	0.1249	58.272	Yes
SiT	R	-20	0.6774	2336696	0.0056	0.0767	91.535	Yes
SiT	NR	-20	0.2145	4970717	0.0152	0.1245	75.396	Yes
SiT	R	-25	0.6758	2308129	0.0056	0.0762	114.161	Yes
SiT	NR	-25	0.2131	4998027	0.0151	0.1242	93.668	Yes
SiT	R	-30	0.6742	2264330	0.0054	0.0761	137.116	Yes
SiT	NR	-30	0.2108	5030303	0.0148	0.1239	122.950	Yes
SiT	R	-35	0.6724	2231023	0.0053	0.0755	163.332	Yes
SiT	NR	-35	0.2089	5059346	0.0146	0.1234	147.634	Yes
SiT	R	-40	0.6703	2198227	0.0052	0.0752	193.005	Yes
SiT	NR	-40	0.2068	5085604	0.0144	0.1230	175.403	Yes
SiT	R	-45	0.6683	2167613	0.0051	0.0748	221.981	Yes
SiT	NR	-45	0.2050	5113631	0.0142	0.1225	198.951	Yes
SiT	R	-50	0.6665	2143729	0.0049	0.0744	248.691	Yes
SiT	NR	-50	0.2037	5137202	0.0141	0.1222	217.399	Yes

Appendix C: T-test Results for Sensitivity Analysis
C.2- Household Income (HHI)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
HHI	R	5	0.6808	2408085	0.0059	0.0776	42.680	Yes
HHI	NR	5	0.2179	4886602	0.0157	0.1254	32.058	Yes
HHI	R	10	0.6781	2345069	0.0058	0.0774	80.714	Yes
HHI	NR	10	0.2151	4929292	0.0155	0.1252	67.958	Yes
HHI	R	15	0.6754	2283041	0.0057	0.0770	119.126	Yes
HHI	NR	15	0.2124	4975941	0.0154	0.1247	101.731	Yes
HHI	R	20	0.6727	2220503	0.0056	0.0766	156.470	Yes
HHI	NR	20	0.2100	5028041	0.0153	0.1244	132.453	Yes
HHI	R	25	0.6704	2152209	0.0054	0.0763	188.184	Yes
HHI	NR	25	0.2074	5074954	0.0152	0.1242	166.246	Yes
HHI	R	30	0.6680	2085683	0.0053	0.0759	220.380	Yes
HHI	NR	30	0.2048	5122517	0.0151	0.1240	199.882	Yes
HHI	R	35	0.6657	2019579	0.0052	0.0756	251.423	Yes
HHI	NR	35	0.2024	5173659	0.0150	0.1237	231.483	Yes
HHI	R	40	0.6635	1953736	0.0051	0.0753	281.228	Yes
HHI	NR	40	0.2001	5227001	0.0149	0.1236	261.790	Yes
HHI	R	45	0.6614	1887490	0.0050	0.0750	308.813	Yes
HHI	NR	45	0.1979	5283259	0.0149	0.1234	290.387	Yes
HHI	R	50	0.6593	1822124	0.0049	0.0747	335.484	Yes
HHI	NR	50	0.1959	5342197	0.0149	0.1234	317.428	Yes
HHI	R	-5	0.6868	2522353.00000	0.0061	0.0684	49.192	Yes
HHI	NR	-5	0.2236	4799066.00000	0.0162	0.1272	37.716	Yes
HHI	R	-10	0.6899	2577295.00000	0.0062	0.0772	88.897	Yes
HHI	NR	-10	0.2263	4752580.00000	0.0164	0.1266	71.286	Yes
HHI	R	-15	0.6932	2628223.00000	0.0064	0.0776	136.475	Yes
HHI	NR	-15	0.2290	4705527.00000	0.0166	0.1270	103.919	Yes
HHI	R	-20	0.6963	2680427.00000	0.0065	0.0780	182.537	Yes
HHI	NR	-20	0.2319	4660504.00000	0.0169	0.1274	137.490	Yes
HHI	R	-25	0.6997	2728886.00000	0.0066	0.0784	230.420	Yes
HHI	NR	-25	0.2349	4618853.00000	0.0172	0.1279	172.699	Yes
HHI	R	-30	0.7030	2776000.00000	0.0067	0.0789	278.446	Yes
HHI	NR	-30	0.2378	4576431.00000	0.0175	0.1285	207.047	Yes
HHI	R	-35	0.7066	2816994.00000	0.0068	0.0793	329.813	Yes
HHI	NR	-35	0.2408	4532805.00000	0.0179	0.1291	240.361	Yes
HHI	R	-40	0.7103	2855191.00000	0.0069	0.0797	382.425	Yes
HHI	NR	-40	0.2433	4480779.00000	0.0182	0.1298	268.396	Yes
HHI	R	-45	0.7138	2894857.00000	0.0070	0.0801	432.773	Yes
HHI	NR	-45	0.2457	4425159.00000	0.0185	0.1304	294.145	Yes
HHI	R	-50	0.7171	2936793.00000	0.0072	0.0805	479.626	Yes
HHI	NR	-50	0.2478	4363087.00000	0.0188	0.1310	315.636	Yes

Appendix C: T-test Results for Sensitivity Analysis
C.3- Base Fare (BF)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
BF	R	5	0.6824	2443011	0.0060	0.0776	19.103	Yes
BF	NR	5	0.2193	4854155	0.0158	0.1254	14.372	Yes
BF	R	10	0.6811	2419966	0.0059	0.0772	37.976	Yes
BF	NR	10	0.2182	4871577	0.0158	0.1255	28.223	Yes
BF	R	15	0.6798	2396195	0.0059	0.0771	56.346	Yes
BF	NR	15	0.2171	4888079	0.0157	0.1254	42.667	Yes
BF	R	20	0.6785	2372581	0.0058	0.0769	74.964	Yes
BF	NR	20	0.2160	4905349	0.0156	0.1252	56.689	Yes
BF	R	25	0.6773	2347405	0.0058	0.0768	92.419	Yes
BF	NR	25	0.2148	4921357	0.0155	0.1251	71.463	Yes
BF	R	30	0.6760	2322877	0.0057	0.0766	110.511	Yes
BF	NR	30	0.2137	4939065	0.0155	0.1249	85.259	Yes
BF	R	35	0.6748	2297106	0.0057	0.0765	127.645	Yes
BF	NR	35	0.2126	4956561	0.0154	0.1248	99.162	Yes
BF	R	40	0.6736	2271684	0.0056	0.0763	145.149	Yes
BF	NR	40	0.2115	4974259	0.0154	0.1247	112.960	Yes
BF	R	45	0.6723	2247343	0.0056	0.0762	163.581	Yes
BF	NR	45	0.2104	4992291	0.0153	0.1246	126.567	Yes
BF	R	50	0.6711	2221355	0.0055	0.0760	180.671	Yes
BF	NR	50	0.2094	5010727	0.0153	0.1245	139.931	Yes
BF	R	-5	0.6851	2487428	0.0061	0.0736	20.521	Yes
BF	NR	-5	0.2217	4821075	0.0160	0.1258	14.358	Yes
BF	R	-10	0.6865	2509844	0.0061	0.0772	38.697	Yes
BF	NR	-10	0.2228	4803942	0.0161	0.1260	28.293	Yes
BF	R	-15	0.6878	2532383	0.0062	0.0773	57.560	Yes
BF	NR	-15	0.2240	4788141	0.0161	0.1262	42.978	Yes
BF	R	-20	0.6891	2554739	0.0062	0.0775	76.431	Yes
BF	NR	-20	0.2252	4771586	0.0162	0.1263	57.166	Yes
BF	R	-25	0.6904	2576762	0.0063	0.0777	95.425	Yes
BF	NR	-25	0.2264	4755773	0.0163	0.1265	71.744	Yes
BF	R	-30	0.6917	2599248	0.0063	0.0778	113.930	Yes
BF	NR	-30	0.2275	4738851	0.0164	0.1267	85.611	Yes
BF	R	-35	0.6930	2619848	0.0063	0.0780	133.699	Yes
BF	NR	-35	0.2287	4722688	0.0165	0.1268	99.875	Yes
BF	R	-40	0.6944	2641085	0.0064	0.0782	152.894	Yes
BF	NR	-40	0.2297	4703201	0.0166	0.1270	112.130	Yes
BF	R	-45	0.6956	2663710	0.0065	0.0783	170.858	Yes
BF	NR	-45	0.2308	4685046	0.0167	0.1272	125.136	Yes
BF	R	-50	0.6968	2686780	0.0065	0.0785	188.265	Yes
BF	NR	-50	0.2319	4667249	0.0167	0.1273	138.255	Yes

Appendix C: T-test Results for Sensitivity Analysis
C.4- Transfer Cost (TC)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
TC	R	5	0.6837	2462878	0.0060	0.0774	0.543	No
TC	NR	5	0.2204	4839298	0.0159	0.1253	1.054	No
TC	R	10	0.6837	2460108	0.0060	0.0772	0.970	No
TC	NR	10	0.2203	4841171	0.0159	0.1258	2.004	Yes
TC	R	15	0.6837	2457473	0.0060	0.0772	1.495	Yes
TC	NR	15	0.2203	4843135	0.0159	0.1258	2.905	Yes
TC	R	20	0.6836	2454963	0.0060	0.0772	2.114	Yes
TC	NR	20	0.2202	4845277	0.0159	0.1258	3.703	Yes
TC	R	25	0.6836	2452335	0.0060	0.0773	2.635	Yes
TC	NR	25	0.2201	4847099	0.0159	0.1258	4.690	Yes
TC	R	30	0.6836	2449618	0.0060	0.0773	3.081	Yes
TC	NR	30	0.2200	4848936	0.0159	0.1258	5.668	Yes
TC	R	35	0.6835	2446649	0.0060	0.0773	3.324	Yes
TC	NR	35	0.2200	4850828	0.0159	0.1258	6.615	Yes
TC	R	40	0.6835	2443904	0.0060	0.0773	3.736	Yes
TC	NR	40	0.2199	4853057	0.0159	0.1258	7.366	Yes
TC	R	45	0.6835	2441385	0.0060	0.0773	4.320	Yes
TC	NR	45	0.2198	4855062	0.0160	0.1258	8.248	Yes
TC	R	50	0.6835	2438066	0.0060	0.0773	4.268	Yes
TC	NR	50	0.2198	4856992	0.0160	0.1259	9.176	Yes
TC	R	-5	0.6838	2468088	0.0060	0.0770	0.583	No
TC	NR	-5	0.2206	4835583	0.0159	0.1260	0.872	No
TC	R	-10	0.6839	2470599	0.0060	0.0772	1.226	No
TC	NR	-10	0.2207	4833767	0.0159	0.1258	1.858	Yes
TC	R	-15	0.6839	2473642	0.0060	0.0772	1.461	Yes
TC	NR	-15	0.2207	4832003	0.0159	0.1257	2.873	Yes
TC	R	-20	0.6839	2476350	0.0060	0.0772	1.964	Yes
TC	NR	-20	0.2208	4830291	0.0159	0.1257	3.918	Yes
TC	R	-25	0.6839	2479164	0.0060	0.0772	2.390	Yes
TC	NR	-25	0.2209	4828349	0.0159	0.1257	4.826	Yes
TC	R	-30	0.6840	2481692	0.0060	0.0772	3.044	Yes
TC	NR	-30	0.2210	4826483	0.0159	0.1257	5.779	Yes
TC	R	-35	0.6840	2484472	0.0060	0.0771	3.507	Yes
TC	NR	-35	0.2211	4824794	0.0159	0.1257	6.835	Yes
TC	R	-40	0.6841	2487194	0.0060	0.0771	4.021	Yes
TC	NR	-40	0.2211	4822843	0.0158	0.1257	7.736	Yes
TC	R	-45	0.6841	2490188	0.0060	0.0771	4.330	Yes
TC	NR	-45	0.2212	4821064	0.0158	0.1257	8.738	Yes
TC	R	-50	0.6841	2492930	0.0060	0.0771	4.840	Yes
TC	NR	-50	0.2213	4819360	0.0158	0.1257	9.783	Yes

Appendix C: T-test Results for Sensitivity Analysis
C.5- Airport Pass Price (Airpass)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
Airpass	R	5	0.6838	2462635	0.0060	0.0832	0.529	No
Airpass	NR	5	0.2203	4840440	0.0159	0.1197	2.885	Yes
Airpass	R	10	0.6840	2457211	0.0060	0.0773	3.305	Yes
Airpass	NR	10	0.2201	4843881	0.0159	0.1257	5.152	Yes
Airpass	R	15	0.6840	2455356	0.0060	0.0772	3.385	Yes
Airpass	NR	15	0.2199	4847229	0.0159	0.1257	7.624	Yes
Airpass	R	20	0.6841	2452422	0.0060	0.0772	4.374	Yes
Airpass	NR	20	0.2197	4850314	0.0159	0.1257	10.254	Yes
Airpass	R	25	0.6841	2450841	0.0060	0.0772	4.409	Yes
Airpass	NR	25	0.2195	4853522	0.0159	0.1257	12.817	Yes
Airpass	R	30	0.6841	2449203	0.0060	0.0772	4.557	Yes
Airpass	NR	30	0.2193	4856732	0.0159	0.1257	15.380	Yes
Airpass	R	35	0.6841	2447734	0.0060	0.0772	4.619	Yes
Airpass	NR	35	0.2191	4861133	0.0159	0.1257	17.265	Yes
Airpass	R	40	0.6842	2445550	0.0060	0.0772	5.314	Yes
Airpass	NR	40	0.2189	4863855	0.0159	0.1258	20.137	Yes
Airpass	R	45	0.6842	2443079	0.0060	0.0772	6.311	Yes
Airpass	NR	45	0.2186	4866652	0.0159	0.1258	22.959	Yes
Airpass	R	50	0.6843	2441081	0.0060	0.0772	7.044	Yes
Airpass	NR	50	0.2184	4868870	0.0159	0.1258	26.105	Yes
Airpass	R	-5	0.6837	2468804	0.0060	0.0770	0.782	No
Airpass	NR	-5	0.2207	4834097	0.0159	0.1261	2.351	Yes
Airpass	R	-10	0.6837	2471243	0.0060	0.0772	0.786	No
Airpass	NR	-10	0.2209	4830867	0.0159	0.1258	4.849	Yes
Airpass	R	-15	0.6837	2474040	0.0060	0.0772	0.979	No
Airpass	NR	-15	0.2211	4826979	0.0159	0.1258	6.943	Yes
Airpass	R	-20	0.6838	2475912	0.0060	0.0772	0.385	No
Airpass	NR	-20	0.2211	4820255	0.0159	0.1258	7.311	Yes
Airpass	R	-25	0.6838	2477990	0.0060	0.0771	0.113	No
Airpass	NR	-25	0.2213	4816551	0.0159	0.1257	9.407	Yes
Airpass	R	-30	0.6838	2481005	0.0060	0.0771	0.047	No
Airpass	NR	-30	0.2214	4811558	0.0159	0.1258	10.708	Yes
Airpass	R	-35	0.6838	2484339	0.0060	0.0771	0.357	No
Airpass	NR	-35	0.2215	4806175	0.0159	0.1257	11.721	Yes
Airpass	R	-40	0.6837	2487690	0.0060	0.0771	0.561	No
Airpass	NR	-40	0.2215	4800445	0.0158	0.1257	12.470	Yes
Airpass	R	-45	0.6838	2490328	0.0060	0.0771	0.112	No
Airpass	NR	-45	0.2216	4795339	0.0158	0.1257	13.514	Yes
Airpass	R	-50	0.6838	2494030	0.0060	0.0771	0.387	No
Airpass	NR	-50	0.2217	4789826	0.0158	0.1257	14.260	Yes

Appendix C: T-test Results for Sensitivity Analysis
C.6- Total Travel Time (TTT)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
TTT	R	5	0.6804	2364525	0.0059	0.0861	42.548	Yes
TTT	NR	5	0.2151	4921913	0.0159	0.1273	65.870	Yes
TTT	R	10	0.6772	2265643	0.0059	0.0777	91.732	Yes
TTT	NR	10	0.2099	5001648	0.0159	0.1253	133.044	Yes
TTT	R	15	0.6745	2162121	0.0058	0.0776	128.975	Yes
TTT	NR	15	0.2050	5088479	0.0160	0.1253	194.990	Yes
TTT	R	20	0.6712	2049177	0.0056	0.0774	171.333	Yes
TTT	NR	20	0.1996	5191848	0.0161	0.1253	263.687	Yes
TTT	R	25	0.6687	1951227	0.0056	0.0769	204.680	Yes
TTT	NR	25	0.1954	5278065	0.0162	0.1257	316.657	Yes
TTT	R	30	0.6665	1852785	0.0055	0.0767	231.717	Yes
TTT	NR	30	0.1915	5364485	0.0162	0.1258	367.701	Yes
TTT	R	35	0.6643	1759145	0.0054	0.0765	258.261	Yes
TTT	NR	35	0.1878	5452893	0.0163	0.1259	416.146	Yes
TTT	R	40	0.6624	1665139	0.0053	0.0763	278.874	Yes
TTT	NR	40	0.1842	5542000	0.0164	0.1261	463.279	Yes
TTT	R	45	0.6604	1575359	0.0052	0.0761	301.164	Yes
TTT	NR	45	0.1804	5630910	0.0164	0.1262	512.839	Yes
TTT	R	50	0.6586	1482437	0.0051	0.0760	318.617	Yes
TTT	NR	50	0.1763	5724512	0.0165	0.1263	566.080	Yes
TTT	R	-5	0.6875	2573128	0.0061	0.0661	62.470	Yes
TTT	NR	-5	0.2275	4753748	0.0158	0.1333	81.408	Yes
TTT	R	-10	0.6912	2672780	0.0062	0.0768	108.865	Yes
TTT	NR	-10	0.2337	4670034	0.0158	0.1261	161.045	Yes
TTT	R	-15	0.6952	2772131	0.0063	0.0771	168.623	Yes
TTT	NR	-15	0.2404	4585664	0.0156	0.1260	242.313	Yes
TTT	R	-20	0.6992	2857939	0.0063	0.0775	228.794	Yes
TTT	NR	-20	0.2458	4502970	0.0156	0.1258	307.213	Yes
TTT	R	-25	0.7034	2943271	0.0064	0.0778	292.331	Yes
TTT	NR	-25	0.2524	4423434	0.0155	0.1257	385.541	Yes
TTT	R	-30	0.7077	3033133	0.0065	0.0780	357.355	Yes
TTT	NR	-30	0.2594	4331803	0.0153	0.1256	468.471	Yes
TTT	R	-35	0.7119	3123318	0.0066	0.0783	421.456	Yes
TTT	NR	-35	0.2666	4239614	0.0150	0.1252	553.227	Yes
TTT	R	-40	0.7162	3199891	0.0067	0.0788	485.792	Yes
TTT	NR	-40	0.2726	4148712	0.0148	0.1247	624.465	Yes
TTT	R	-45	0.7209	3267041	0.0067	0.0792	555.579	Yes
TTT	NR	-45	0.2787	4057333	0.0147	0.1244	695.136	Yes
TTT	R	-50	0.7254	3331646	0.0068	0.0794	623.560	Yes
TTT	NR	-50	0.2843	3974717	0.0146	0.1240	759.213	Yes

Appendix C: T-test Results for Sensitivity Analysis

C.7- Walk Time (WT)

Variable	Group	Condition	Sample Mean P	Sample Size	Variance	Standard Error	t	Reject Null?
-	R/NR	%Δ	-	Trips	-	-	-	Yes/No
Observed	R	0	0.6838	2465504	0.0060	-	-	-
Observed	NR	0	0.2205	4837540	0.0159	-	-	-
WT	R	5	0.6817	2414906	0.0060	0.0875	25.771	Yes
WT	NR	5	0.2173	4882494	0.0160	0.1175	41.819	Yes
WT	R	10	0.6801	2359991	0.0059	0.0775	52.680	Yes
WT	NR	10	0.2143	4923712	0.0160	0.1256	77.292	Yes
WT	R	15	0.6781	2312289	0.0059	0.0772	80.843	Yes
WT	NR	15	0.2113	4966393	0.0160	0.1257	115.004	Yes
WT	R	20	0.6760	2246414	0.0058	0.0773	109.698	Yes
WT	NR	20	0.2075	5021655	0.0161	0.1256	162.845	Yes
WT	R	25	0.6744	2194689	0.0057	0.0769	131.927	Yes
WT	NR	25	0.2051	5063841	0.0161	0.1259	192.846	Yes
WT	R	30	0.6729	2142018	0.0057	0.0767	151.339	Yes
WT	NR	30	0.2029	5107312	0.0162	0.1260	220.471	Yes
WT	R	35	0.6715	2089980	0.0056	0.0766	170.147	Yes
WT	NR	35	0.2007	5150211	0.0162	0.1260	247.570	Yes
WT	R	40	0.6702	2038566	0.0056	0.0765	188.053	Yes
WT	NR	40	0.1987	5194013	0.0162	0.1261	274.148	Yes
WT	R	45	0.6689	1983881	0.0055	0.0764	204.243	Yes
WT	NR	45	0.1963	5237490	0.0162	0.1261	304.356	Yes
WT	R	50	0.6674	1930468	0.0055	0.0762	224.095	Yes
WT	NR	50	0.1935	5285123	0.0163	0.1261	340.064	Yes
WT	R	-5	0.6859	2525332	0.0061	0.0709	34.084	Yes
WT	NR	-5	0.2250	4788457	0.0157	0.1297	53.716	Yes
WT	R	-10	0.6880	2579411	0.0061	0.0770	60.916	Yes
WT	NR	-10	0.2287	4742580	0.0157	0.1258	101.374	Yes
WT	R	-15	0.6902	2635158	0.0062	0.0772	93.217	Yes
WT	NR	-15	0.2331	4698168	0.0155	0.1256	154.583	Yes
WT	R	-20	0.6922	2680209	0.0063	0.0775	123.670	Yes
WT	NR	-20	0.2362	4658046	0.0155	0.1253	193.251	Yes
WT	R	-25	0.6944	2726679	0.0063	0.0777	155.961	Yes
WT	NR	-25	0.2403	4617247	0.0154	0.1252	243.145	Yes
WT	R	-30	0.6967	2779116	0.0064	0.0778	189.169	Yes
WT	NR	-30	0.2446	4559922	0.0151	0.1251	294.911	Yes
WT	R	-35	0.6990	2828944	0.0064	0.0780	224.551	Yes
WT	NR	-35	0.2491	4506906	0.0148	0.1245	351.382	Yes
WT	R	-40	0.7013	2870907	0.0065	0.0783	257.717	Yes
WT	NR	-40	0.2530	4461531	0.0146	0.1240	398.853	Yes
WT	R	-45	0.7035	2912681	0.0065	0.0785	290.345	Yes
WT	NR	-45	0.2568	4415482	0.0144	0.1236	446.608	Yes
WT	R	-50	0.7055	2953952	0.0066	0.0787	320.012	Yes
WT	NR	-50	0.2600	4372348	0.0143	0.1232	486.327	Yes

Appendix D: Supplemental Figures

The following figures seven pairs of are combination graphs displaying the probability as points with a trend line, and the sample size (trips) of the respective group as bar elements. The first figure plots the results for the replaced trip group, and the second figure plots the results for the not-replaced trip group. The $P(Transit|CTA)$ values reference the secondary axis, to the right of the graph. Similarly, the sample size values correspond to the primary axis, located to the left of the graph. Per figure, there is a total of 21 scenarios plotted, with one being the observed scenario. The $P(Transit|CTA)$ and sample size are plotted in grey and correspond to the x-axis value of 0.

For formatting purposes, the figures begin on the next page.

I. Transit Stops per Census Tract (SiT)

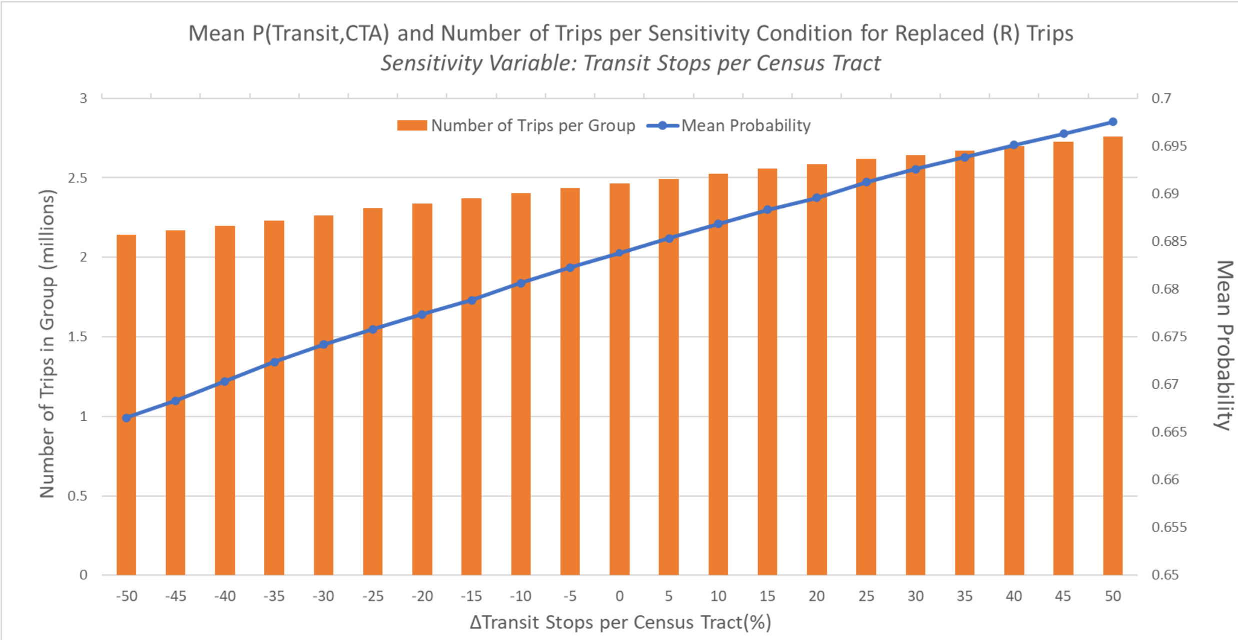


Figure D1 – Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Transit Stops per Census Tract

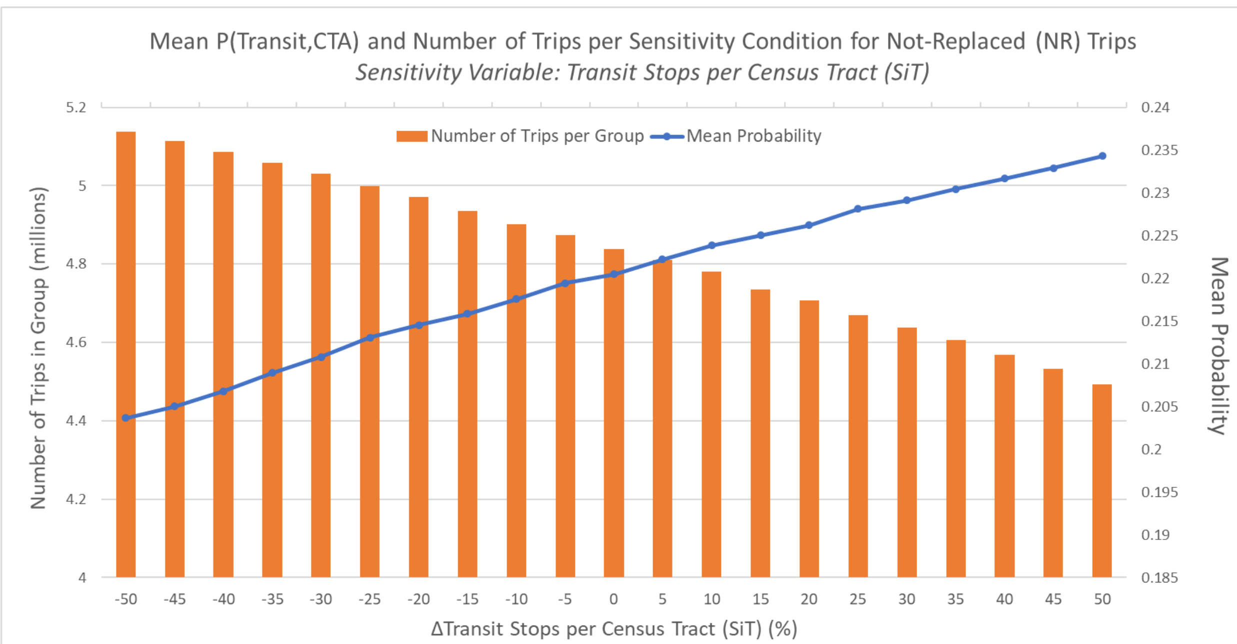


Figure D2 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Transit Stops per Census Tract

II. Household Income (HHI)

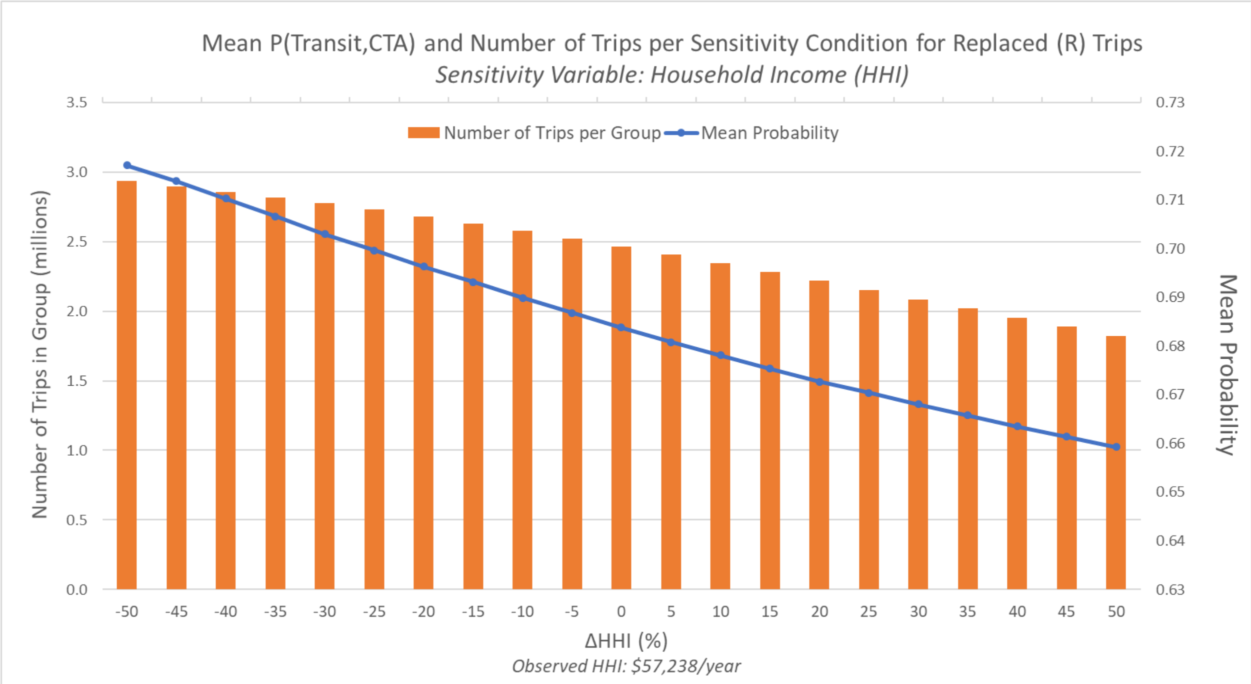


Figure D3 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Household Income

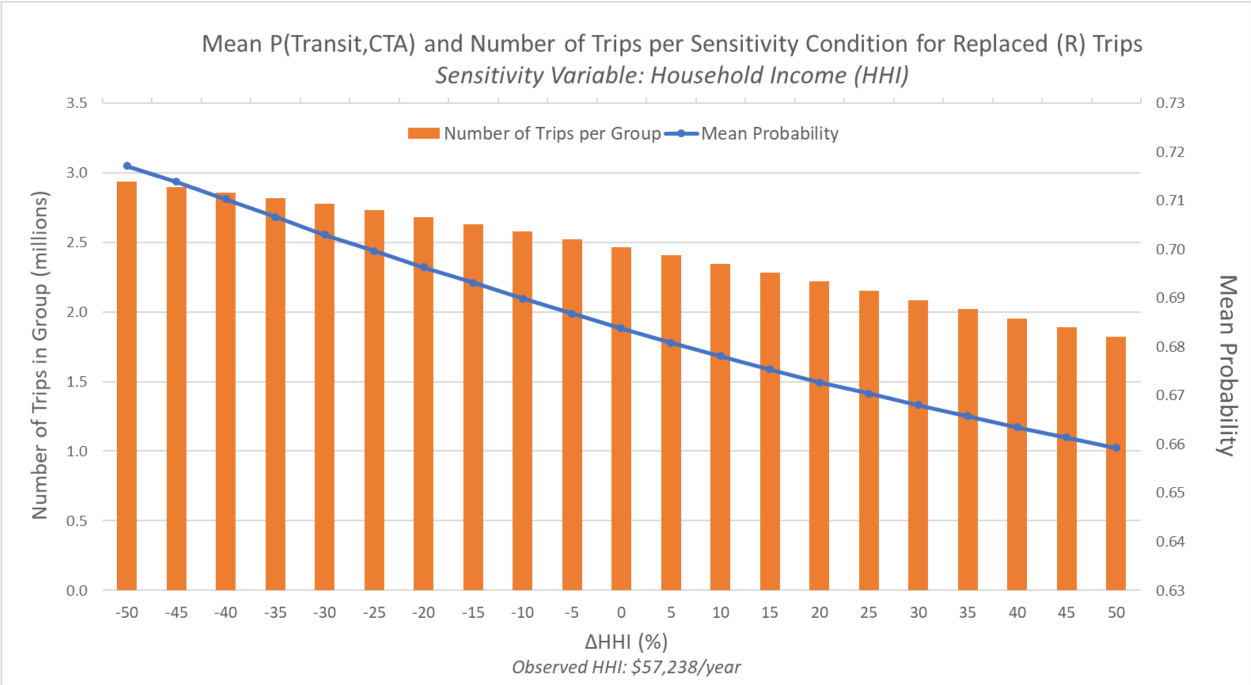


Figure D4 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Household Income

III. Total Travel Time (TTT)

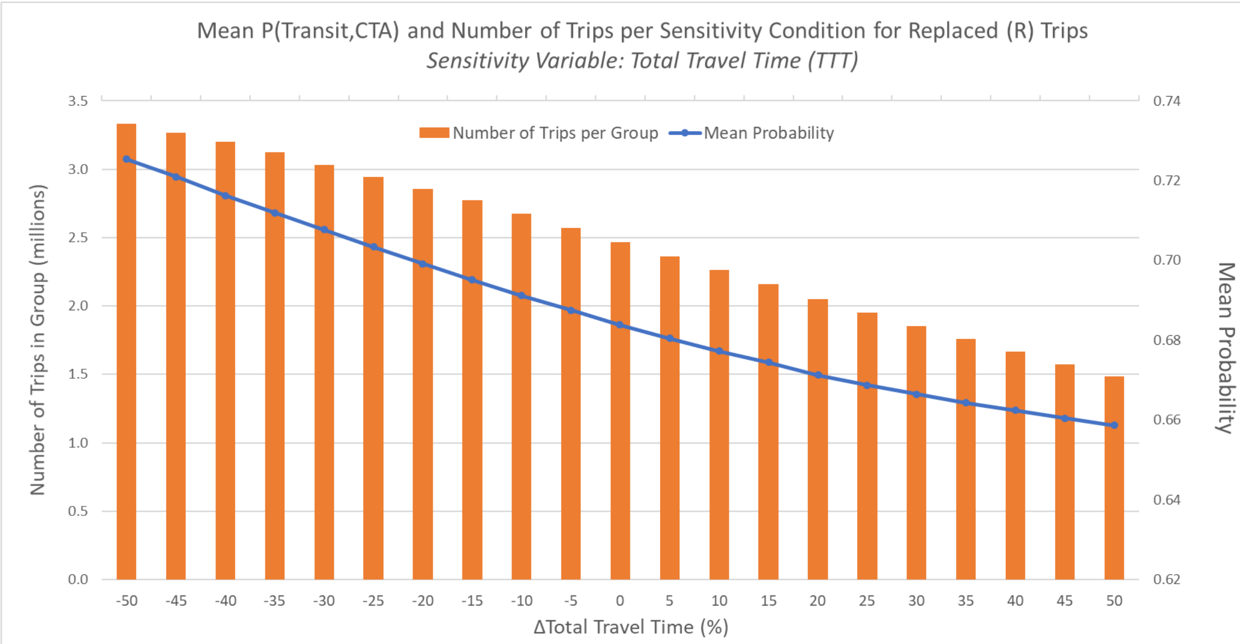


Figure D5 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Total Travel Time

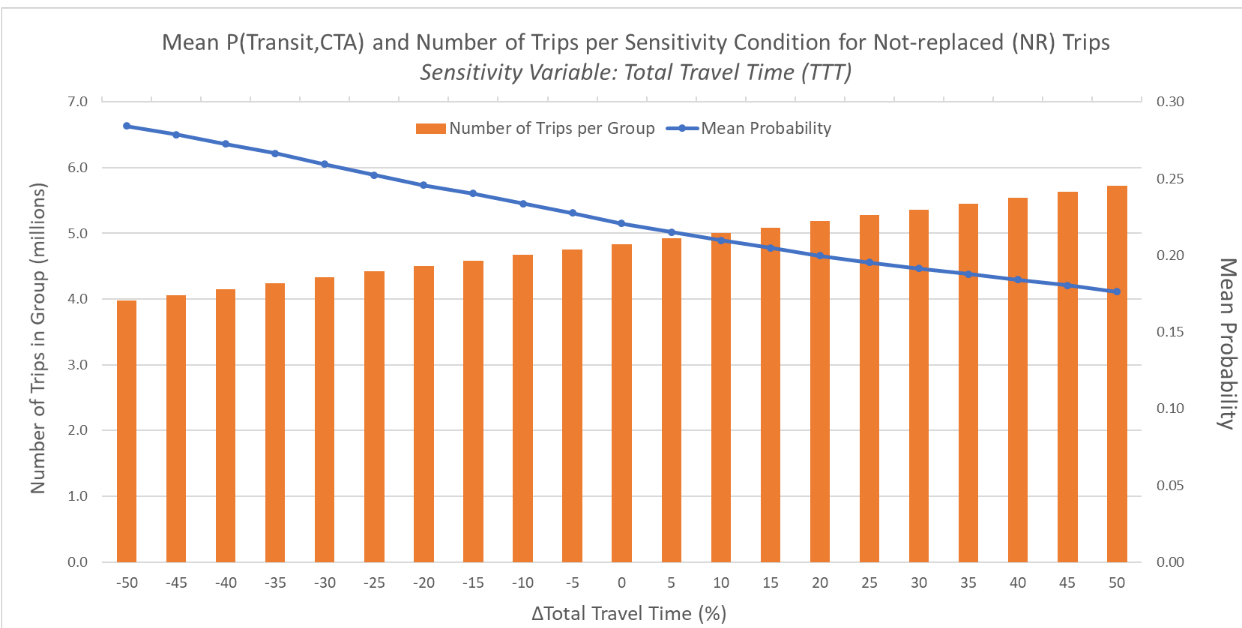


Figure D6- Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Total Travel Time

IV. Walk Time (WT)

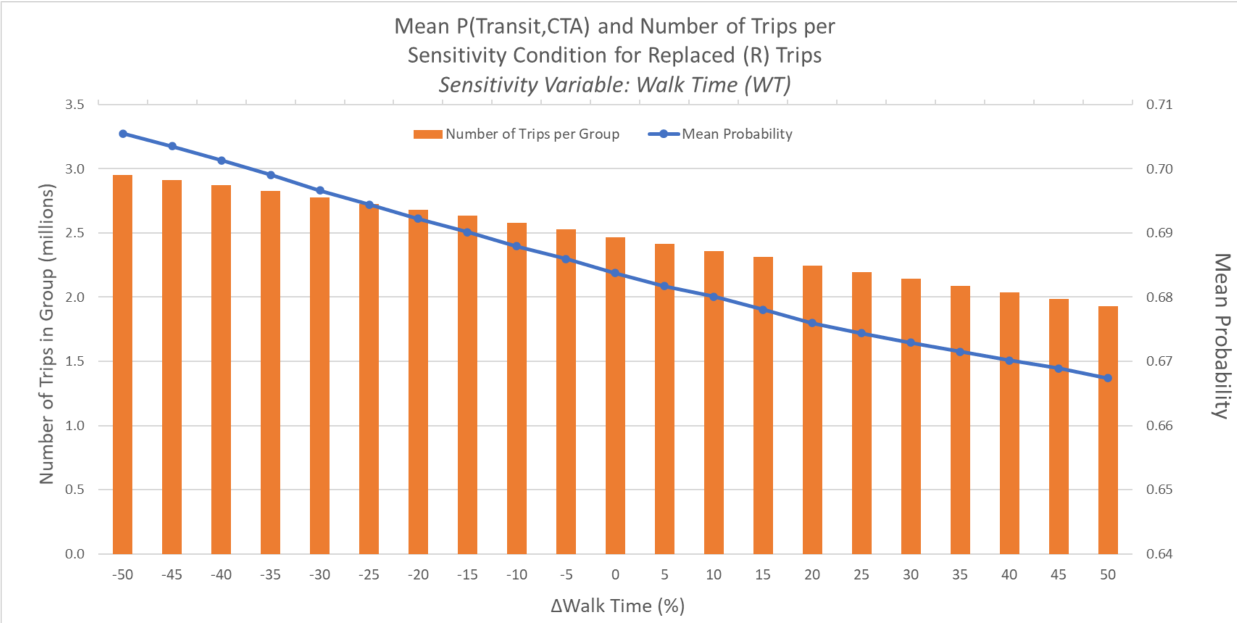


Figure D7 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Walk Time

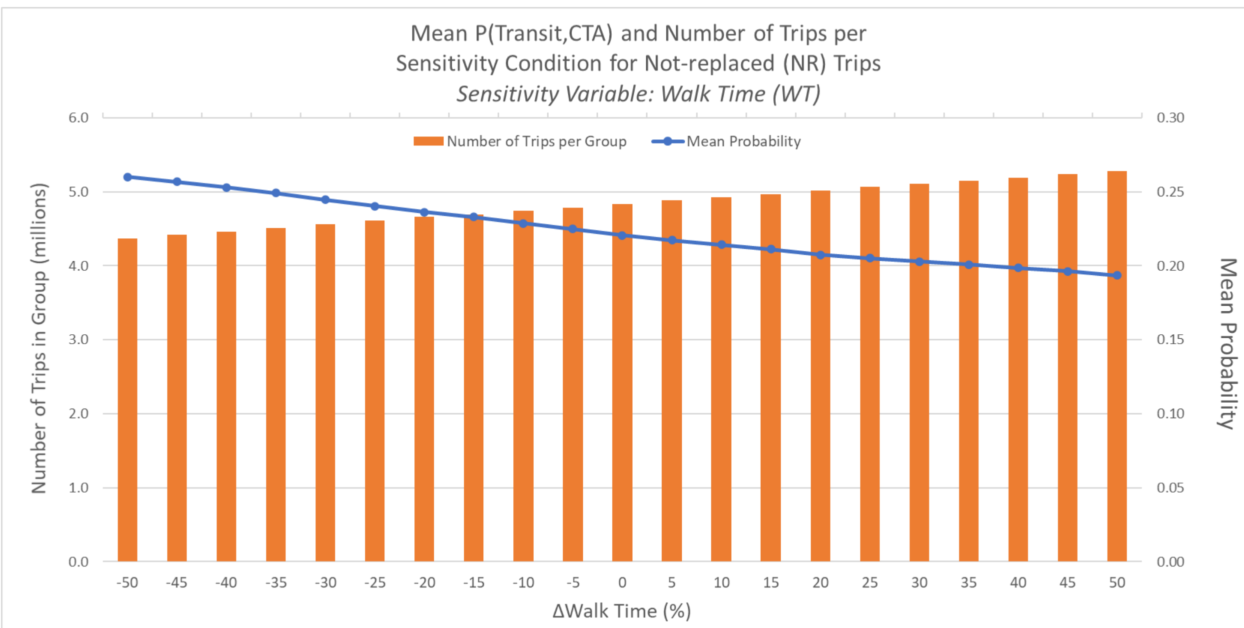


Figure D8 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Walk Time

V. Base Fare

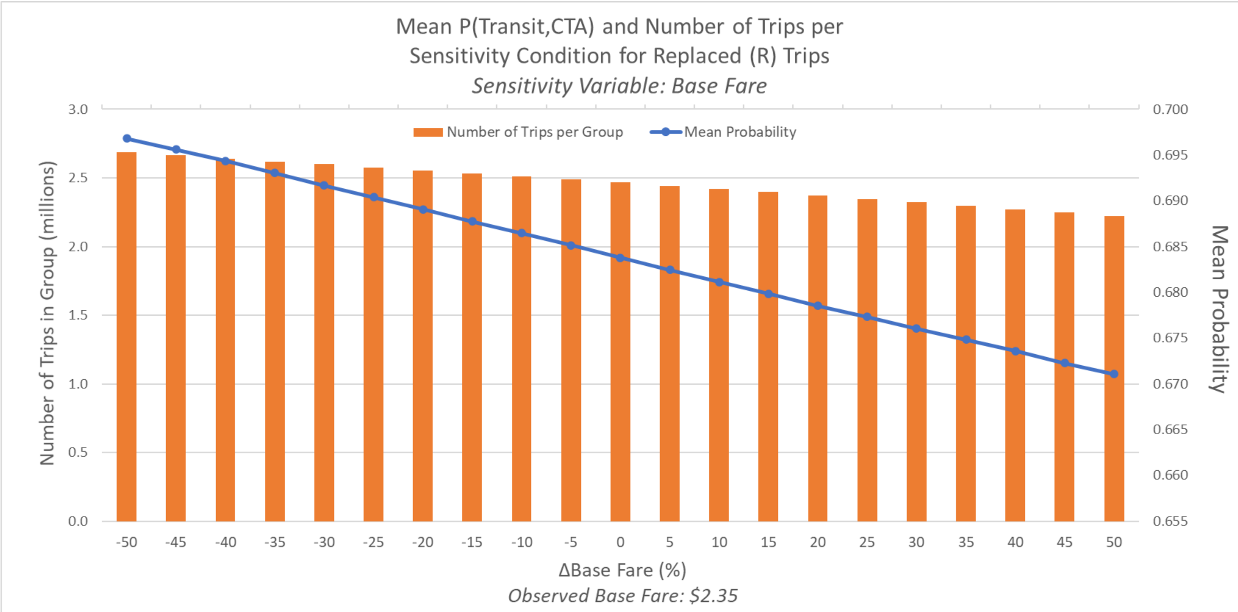


Figure D9 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Base Fare

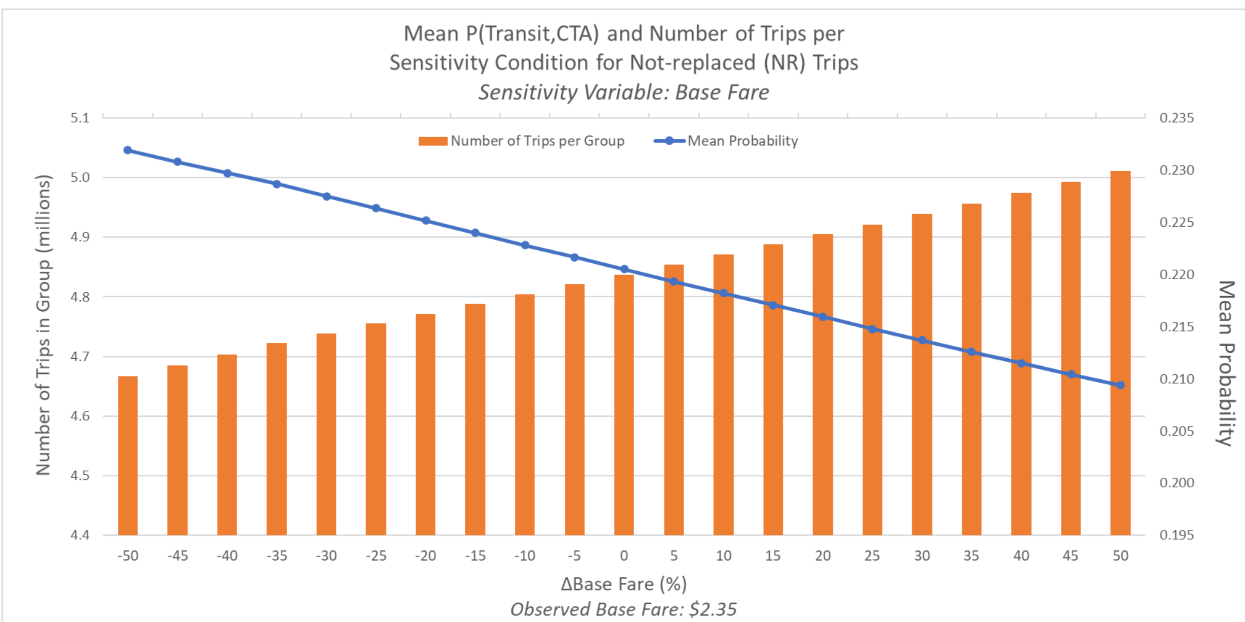


Figure D10 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Base Fare

VI. Transfer Cost

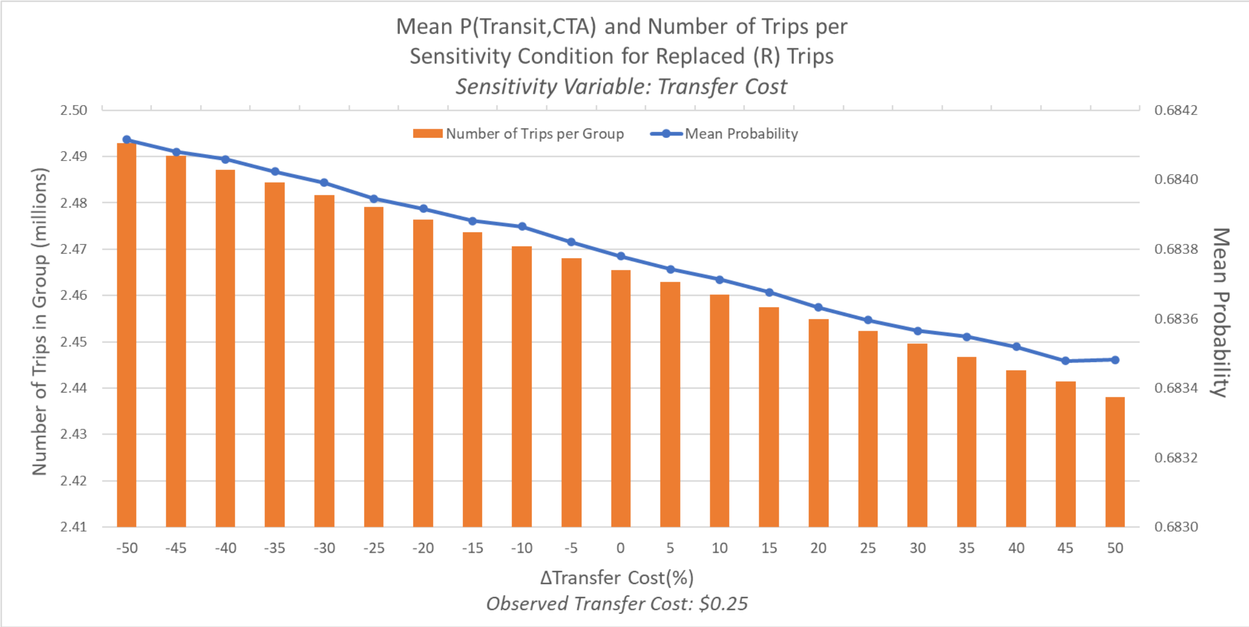


Figure D11 - Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Transfer Cost

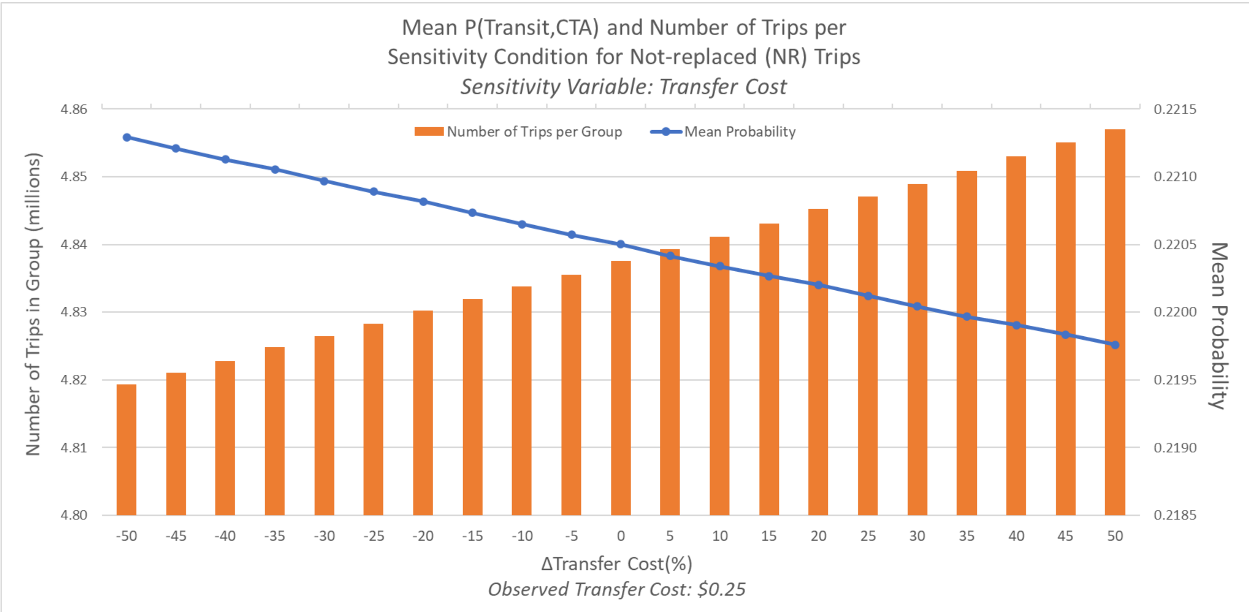


Figure D12- Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Transfer Cost

VII. Airport Pass Price

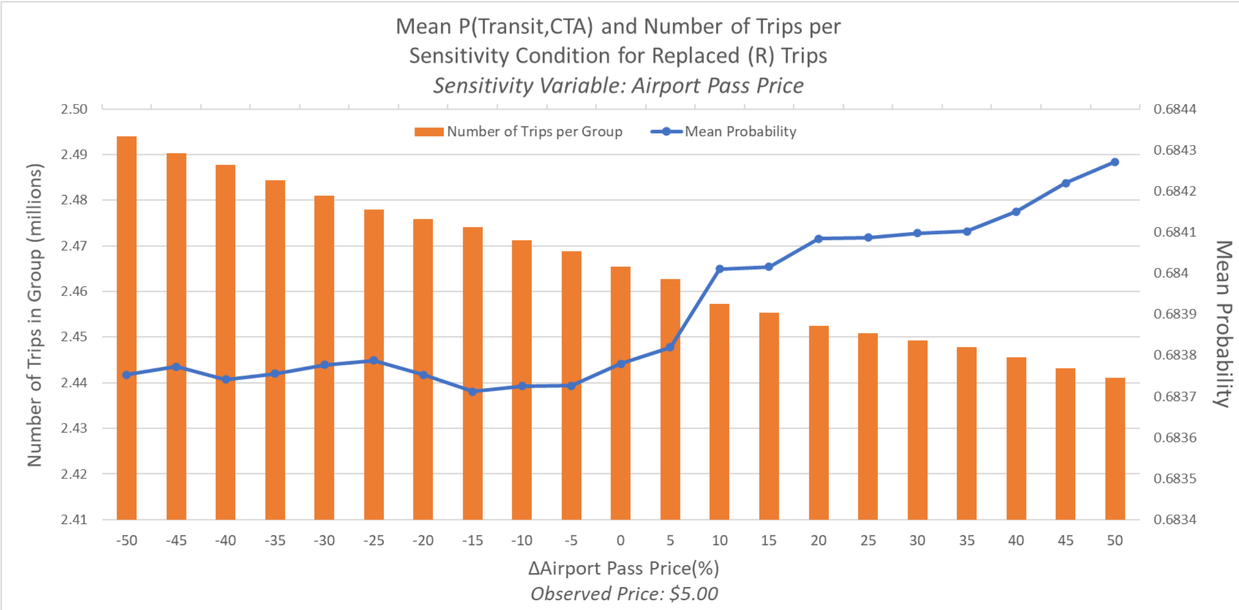


Figure D13- Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Replaced Trips, for Sensitivity Variable: Airport Pass Price

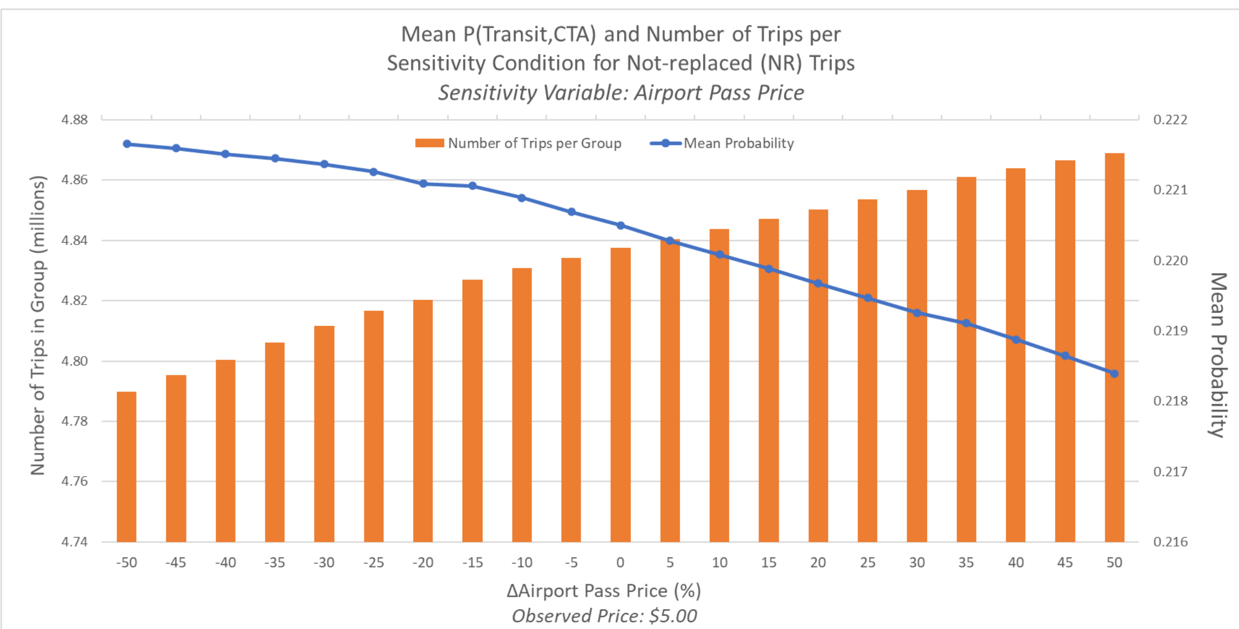


Figure D14- Mean Probability of Selecting CTA and Sample Size per Sensitivity Condition for Not-replaced Trips, for Sensitivity Variable: Airport Pass Price