Ride-hailing Impacts on Transit Ridership: Chicago Case Study

Helena Breuer
Graduate Research Assistant, Center for Sustainable Mobility, VTTI
3500 Transportation Research Plaza
Virginia Tech, Blacksburg, VA 24061
Email: lenab19@vt.edu

Jianhe Du
Senior Research Associate, Center for Sustainable Mobility, VTTI
3500 Transportation Research Plaza
Virginia Tech, Blacksburg, VA 24061
Email: jdu@vtti.vt.edu

Hesham Rakha, Ph.D., P.Eng. (Corresponding author)
Director, Center for Sustainable Mobility, VTTI
Samuel Reynolds Pritchard Professor of Engineering, Civil and Environmental Engineering
3500 Transportation Research Plaza
Virginia Tech, Blacksburg, VA 24061
Phone: (540) 231-1504, fax: (540) 231-1555, email: hrakha@vt.edu

ABSTRACT

Existing literature on the relationship between ride-hailing (RH) and transit services is limited to empirical studies that lack real-time spatial contexts. To fill this gap, we took a novel real-time geospatial analysis approach. With source data on ride-hailing trips in Chicago, Illinois, we computed real-time transit-equivalent trips for all 7,949,902 ride-hailing trips in June 2019; the sheer size of our sample is incomparable to the samples studied in existing literature. An existing Multinomial Nested Logit Model was used to determine the probability of a ride-hailer selecting a transit alternative to serve the specific O-D pair, P(Transit|CTA)[1]. We find that 31% of ride-hailing trips are replaceable, whereas 61% of trips are not replaceable. The remaining 8% lie within a buffer zone. We measured the robustness of this probability using a parametric sensitivity analysis and performed a two-tailed t-test. Our results indicate that of the four sensitivity parameters, the probability was most sensitive to the total travel time of a transit trip. The main contribution of our research is our thorough approach and fine-tuned series of real-time spatiotemporal analyses that investigate the replaceability of ride-hailing trips for public transit. The results and discussion intend to provide perspective derived from real trips and we anticipate that this paper will demonstrate the research benefits associated with the recording and release of ride-hailing data.

---

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

1   **TERMINOLOGY**

2   The following is an extensive list of terminology relevant to this paper, and their corresponding
3   contextual definitions.

4   ***First- and last-mile (FLM)***: this refers to the first and last leg of the transit trip that connect the
5   individual from their origin to the first transit stop, and/or from the last transit stop to their
6   destination.

7   ***In-vehicle travel time (IVTT):*** this is the portion of the total travel time, and accounts for all
8   time spent traveling inside the transit vehicle(s). In our analysis, we may refer to this as the
9   "transit time".

10 ***Not-replaced (NR) trip/group:*** this is the group containing all transit-equivalent trips with a
11 $P(Transit|CTA) \leq 0.45$. A transit-equivalent trip that has a probability in this range [0 ∪
12 0.45] is deemed to be inviable to the individual, and ultimately, does not compete with the RH
13 trip service.

14 ***Out-of-vehicle travel time (OVTT):*** this is a portion of the total travel time outside of the
15 vehicle(s), i.e. accessing, egressing, wait time, transfer walk time.

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

20 ***Replaced (R) group:*** this is the group containing all transit-equivalent trips with a
21 $P(Transit|CTA) \geq 0.55$. A transit-equivalent trip that has a probability in the range [0.55 ∪
22 1.0] is considered a viable mode of service for the specific O-D pair.

23 ***Ride-hailing (RH):*** this refers to the act of servicing a trip via a transportation network company
24 (TNC). Users must have an account with the respective TNC, and have the app downloaded onto
25 their smartphone. These trips are ordered using the TNC's app and require the user to input their
26 destination, whereas the origin is automatically determined using the smartphone's internal GIS
27 software. TNC trip fare pricing is dynamic and dependent on the surrounding demand. Ride-
28 hailing trips can be pooled or single passenger, refer to their definitions.

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

31 ***Sensitivity Condition:*** these are the percent-changes in the sensitivity variable used for the
32 analysis. For each variable, there existed 20 sensitivity conditions, ranging from -50% to +50%
33 in increments of 5%, where the 0% condition is the observed values and results.

34 ***Single-Passenger Trip, Ride-hailing:*** these are RH trips where a single person ordered the trip,
35 and there is one origin and destination. In some cases, these trips can have more than one
36 passenger.

37 ***TNC:*** Transportation Network Companies; these are the private businesses that offer ride-hailing
38 services. Examples include Uber and Lyft.

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

42 ***Transit-equivalent trip/CTA-equivalent:*** this refers to the routing solution from ArcGIS' Route
43 Analysis. It is important to consider that the output trip does not necessarily use transit. Under
44 certain conditions, the program determines that it is most efficiently serviced by walking, thus a
45 "transit-equivalent" trip does not imply the use of transit.

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

INTRODUCTION

4   The first ride-hailing[2] (RH) service came to US markets in 2008 when Travis Kalanick and
5   Garret Camp established their company, *Uber* [1]. Four years later, an existing carpooling
6   company, *Zimride*, launched a competing ride-hailing service in San Francisco later renaming
7   itself *Lyft*, to exclusively operate as a ride-hailing service [2]. Over the next two years,
8   competition heightened as Uber expanded to 60 cities across six continents, and Lyft announced
9   its plan to expand to 24 more cities, totaling coverage of over 60 cities [2, 3]. As of January
10   2019, nearly a decade later, 36% of US adults have used or currently use ride-hailing services
11   [4].
12   Ride-hailing is best defined as an on-demand, app-based, and real-time service that
13   provides customers with door-to-door transportation for a single trip [5]. Through the company's
14   smartphone app, a customer enters a specific pick-up and drop-off location (O-D pair). On the
15   backend, the RH company's algorithm calculates an appropriate route and trip fare. It then
16   selects the optimal driver to service the trip and notifies the customer of the estimated pick-up
17   time and vehicle/driver details.
18   Coincidentally, when the ride-hailing market began rapidly gaining traction through
19   geographic expansion and increased acceptance in 2014, average public transit ridership in the
20   United States began its decline. In the early twenty-first century, transit ridership in the United
21   States experienced two periods of growth followed by decline (Figure 1).
22

Annual Transit Ridership* 2000-2019
*Ridership for Bus, Light, Heavy, and Commuter Rail only
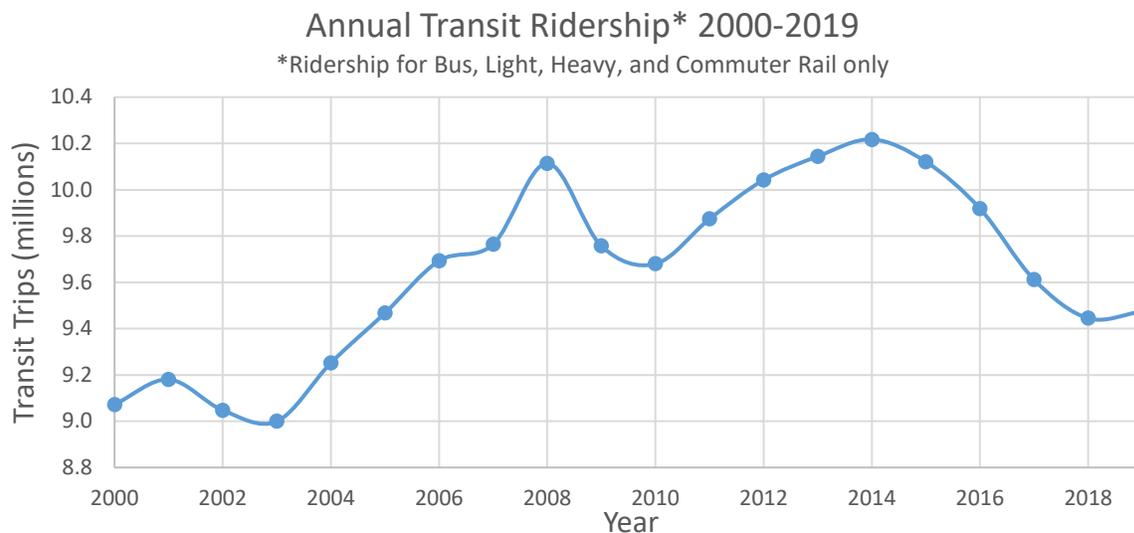
23
24   **Figure 1: Annual Public Transit Ridership in the United States from 2000-2019. Source Data:**
25   **APTA Ridership by Mode and Quarter 1990-Present [6]. Annual ridership counts are the sum of**
26   **bus, light rail, commuter rail, and heavy rail trips**
27
28   On average, from 2003 to 2008, public transit ridership in the United States increased by
29   2.58% each year. Following the 2008 economic recession, ridership levels took a downturn until

---

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

1 2010 when ridership began increasing again until 2014. Although, unlike the first period of
2 growth, this growth rate was decreasing in magnitude each year until it plateaued in 2014, as
3 illustrated in Figure 2. At this point, transit ridership began rapidly decreasing by losing more
4 riders per year until 2019. While these statistics measure the national trend in mode-choice
5 behavior, the trends within metropolitan transit agencies vary by year and mode. Nonetheless,
6 continuous decline in ridership is significant and likely indicative of a disturbance to the market.
7
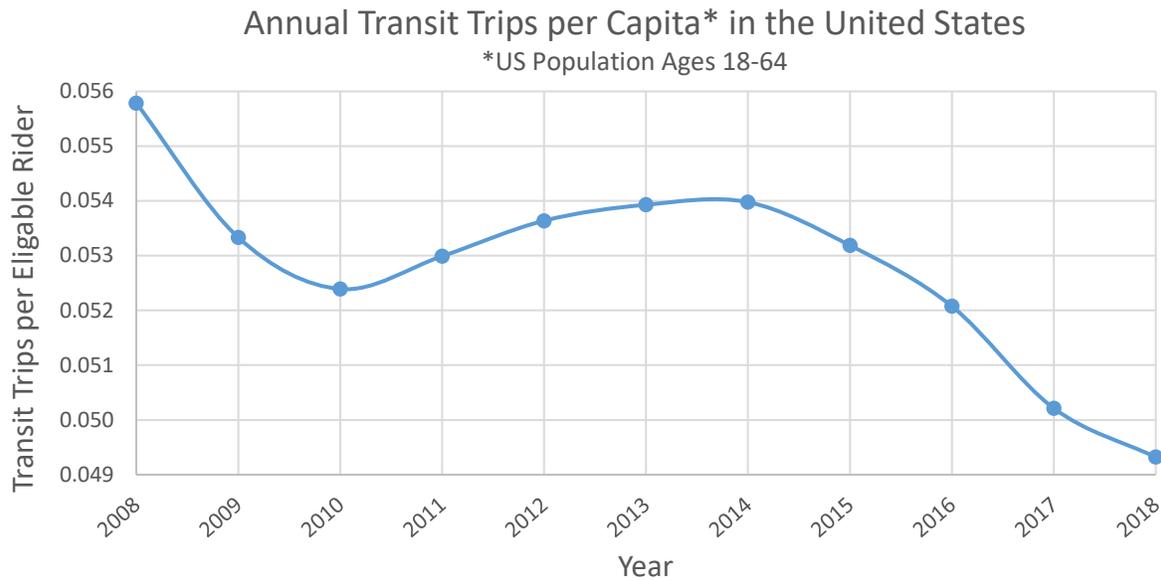
8
9 **Figure 2: Annual Transit Trips per Eligible Rider in the United States. Source Data: APTA**
10 **Ridership by Mode and Quarter 1990-Present, and Population by Age from KFF**
11

12   From 2008 to 2018, the population of eligible riders[3] increased by 5.63%. Yet, the annual
13 transit trip per eligible rider decreased by 11.6%, as shown in Figure 2. Moreover, the transit
14 ridership per eligible rider decreased by 8.6% from 2014 to 2018. Despite the steady growth of
15 the US population, transit ridership does not reflect that.
16   Historically, declines in transit ridership can be a result of macroeconomic, geographic,
17 and demographic changes in a region. The first period of ridership decline in the 21st century
18 started in 2008 and was evidently a ramification of the economic recession. Yet, the cause(s) of
19 the most recent decline is not as discernable. Further, this period exhibited a larger decline in
20 magnitude and has spanned 5 years, as opposed to 2 years. So, what could have possibly caused
21 a more crippling effect on transit ridership than the economic recession? Was the second decline
22 in transit ridership a result of an alternative mode, *ride-hailing*?
23   This paper will approach answering this question and will fulfill the literature gap by
24 identifying the replicability of transit by RH services through non-empirical based methods. In
25 our study we use ride-hailing trip source data from the City of Chicago containing over
26 8,000,000 trips, hence the contributions of our research are expanded by the use of a relatively
27 large dataset. We use spatial analyses and methodologies to deliver a real-time transit-equivalent
28 route. We then input trip characteristics into our selected utility model, from which we will
29 calculate the corresponding probability that an original ride-hailer would select transit over RH.

---

[3] Eligible riders: total US populations between age 18 and 64 7. *Population by Age.* 2018, Kaiser Family
Foundation: Online.

We will perform two analyses, the first exploring the replaceability of all RH trips as a whole, and second, a parametric sensitivity analysis of the P(Transit|CTA) with respect to four parameters.

LITERATURE REVIEW

It is important to review the *Terminology* section prior to reading this section, as much existing literature includes new terminology.

The current body of research on ride-hailing is limited by its novelty and the lack of publicly available ride-hailing trip data. Given that ride-hailing services were first introduced to the market in 2010, external research on its utility and impact is relatively untouched. Moreover, ride-hailing services are privately, 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 [8-11]. Some studies executed intercept surveys at points of interest [12], and one executing in-person interviews [13].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 will identify reoccurring themes and findings regarding the impact of ride-hailing services on public transit ridership. We will also highlight the methods used to obtain data. Lastly, we will determine gaps in the literature and how we will address these gaps in our study.

It is important to note that the relationship between ride-hailing and public transit encompasses one-to-many relationships. Bus and rail (light and heavy) both fall under 'public transit', although each mode services trips of differing purposes, rider demographics, and LOS metrics. Accordingly, most literature analyses each modality separately.

In general, the current literature seeks to explore the impact of ride-hailing on VMT and vehicle emissions, its relative safety, and its effect on mode selection. 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?" [14]. Answering this question requires empirical and source-data based research.

As stated previously, the lack of source-data based research has led to conflicting arguments. Considering that "public transit" encompasses many transit modes, positions tend to be unique per mode (bus, rail). The first position argues 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, these critics pose 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 [9, 10, 15].

1   According to Clewlow and Mishra, ride-hailing services replaced 6% of bus trips and 3% of light
2   rail trips. Whereas ride-hailing acted complementary to commuter rail services, increasing
3   ridership by 3%. Similarly, Graehler et al. found that the entry of a TNC decrease heavy rail and
4   bus ridership by 1.3% and 1.7%, respectively [16].
5   Rayle et al. determined the primary reasons why individuals chose ride-hailing over the
6   alternative of interest. In brief, users chose ride-hailing over the bus because it was faster and
7   over rail because it was faster, easier to pay, and had less wait time [12].
8   Henao and Marshall worked as Uber drivers to obtain observational data in real time.
9   Verbal and recorded interviews were used during ride-hailing trips. Of the 311 passengers
10   interviewed, only 5.5% of riders were using the ride-hailing service to get to or from a transit
11   station [13]. This implies that 94.5% of ride-hailing trips do not service the FLM arrangement.
12   However, the small sample size challenges the range of application and the question with a
13   binary response option minimizes bias.
14   Nelson and Sadowsky used a difference in differences (DID) modeling by comparing
15   transit ridership and operational metrics red before and after the entry of ride-hailing service(s).
16   Their findings concluded that transit ridership increased following the entry of the first ride-
17   hailing company, then decreased once the second company entered the regional market. The
18   presence of the second company led to competition and increased affordability, allowing it to
19   appeal to more people [17].
20   In 2016, APTA investigated the relationship between emerging modalities and public
21   transit. The research areas included seven major US cities: Austin, Boston, Chicago, Los
22   Angeles, San Francisco, Seattle, and Washington DC. Researchers executed in-depth interviews
23   with transportation officials and surveying of network users. The most relevant finding is shared
24   modes, i.e. ride-hailing services, are used most frequently for social trips during hours which
25   public transit is not in operation or has reduced services. Hence, when transit operations are
26   reduced, ride-hailing services supplement its decreased availability. Results from the survey
27   show that 54% of respondents had used "ride-sourcing" (ride-hailing) to serve a recreational or
28   social trip within the previous 3 months. Further, only 21% of respondents claimed to have used
29   these services for commuting within the previous 3 months. However, this survey does not look
30   at the trend in demand by trip type over a period of time. The percentage of respondents claiming
31   to have used ride-sourcing for a specific purpose does not encapsulate the frequency of demand
32   by type. For example, 21 out of 100 respondents could use ride-hailing services for commuting
33   on a daily basis, whereas 74 out of 100 respondents only used ride-hailing once a week for
34   recreational/social trips. The cumulative demand by trip purpose cannot be represented through a
35   one-time survey [18].
36   While these methods are useful and highly qualitative, they assume an ideal condition
37   that respondents are not biases. Hence, the results are vulnerable to many biases. The first,
38   *hypothetical bias*, is the propensity of humans to view survey questions hypothetically to an
39   extent that skews responses validity. Second, *strategic bias* is the tendency for a respondent to
40   evaluate their hypothetical behavior such that it favors the response with greater perceived value.
41   Lastly, *framing bias* is how the phrasing and wordage of a question influences its interpretation.
42   The overwhelming use of surveys and interviews serves as an opportunity to deploy a
43   more quantitative study that focuses on individual trips and their corresponding LOS attributes.
44   Until we can collectively concur upon the effect of ride-hailing, designers, planners, and
45   politicians cannot make sound decisions. We hope to contribute to the field by pioneering new
46   methods and approaches to analyze the impact of RH on public transit ridershiop. Our

1   publication of source data-based research will not only result in greater clarity and insight but
2   will illuminate gray areas with more intensity. From this, empirical studies should be refined to
3   focus on investigating these ambiguous regions and identifying their sources.

4   **METHODS**

5   **Geography and Demographics of Chicago**
6   Per the U.S. Census, the population of Chicago was estimated to have been 2,705,994 persons in
7   July 2018 across 801 census tracts. From 2014-2018, there was an average of 1,056,118
8   households with a median income of $57,238 [19]. As of 2015, 26.5% of households do not own
9   a vehicle, where the average vehicles owned per household is 1.11 [20].

10   **Public Transit in Chicago**
11   Chicago Transportation Authority (CTA) is the second largest transit agency in the U.S. as of
12   2018 [21]. The Chicago Transit Authority (CTA) runs and operates bus and rapid transit (rail)
13   services within the city and the 35 surrounding suburbs. There are 1,864 buses that run 129
14   routes and 1,429 rail cars that serve 145 stations [22]. Additionally, CTA operates certain routes
15   and lines during early morning and late-night hours, and some operate all hours of the day.

16   **Ride Hailing Services in Chicago**
17   Historical data on the services present during our study period is unavailable at this time.
18   However, as of January 2020, three personal-car ride-hailing services operate in the City of
19   Chicago: Uber, Lyft, and Via [23].

20   **Data**
21   *TNP (Transportation Network Providers) – Trips Dataset*
22   This dataset served as our primary source data for ride-hailing trips and was obtained from the
23   City of Chicago's online data portal. The dataset contains 129 million unique TNC trips that span
24   from November 2018 to the present day and is aggregated by the month [24]. We chose to only
25   study one month, June 2019 because it does not contain any nationally recognized holidays that
26   could hinder the representativeness of the results. Our selection from this dataset contains all
27   TNC trips that have a trip start time on or after June 1, 2019 12:00:00 AM and before July 1,
28   2019 12:00:00 AM. The dataset contains 23 fields per trip, including a unique identifier, trip start
29   and end time, pick-up and drop-off longitudinal and latitudinal coordinates, pick-up and drop-off
30   census tract ID, trip fare, and if the ride was authorized as "shared" through the respective TNC
31   app.
32   *Public Transit Data (General Transit Feed Specification (GTFS) Dataset)*
33   To perform a public transit network analysis in ArcGIS, it requires the GTFS dataset
34   corresponding the area of interest. GTFS datasets are publicly available from OpenMobilityData,
35   which contains archived real-time and fixed components of transit agencies' schedules [25]. The
36   dataset holds the corresponding schedules, routes, stops, and transfers for the time period. With
37   respect to this paper, this dataset will be integrated into ArcGIS such that the Network Analysis
38   program can identify the corresponding transit route under spatial and temporal conditions.
39
40   **Preliminary Data Analyses**
41   In Figure 3 below, is a preliminary analysis of the dataset in terms of trip passenger count and
42   day type, weekday (M-F) or weekend (Saturday and Sunday). Over the month of June 2019,

1   there were 8,136,461 ride-hailing trips serviced in the City of Chicago. This figure demonstrates
2   that single occupancy ride-hailing trips have significantly greater demand than pooled trips.
3   When compared to driving alone, this modality has a higher contribution towards congestion
4   because its utility is comparably low due to deadheading mileage.
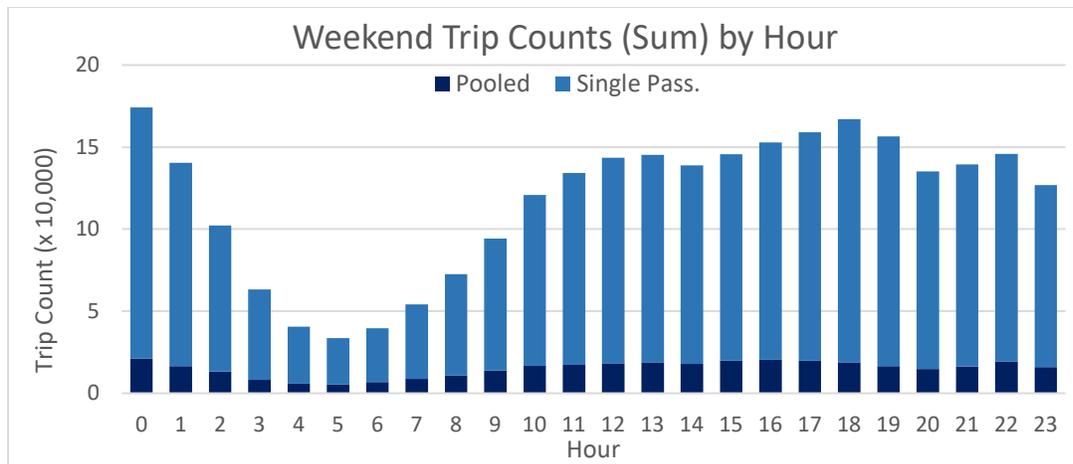5
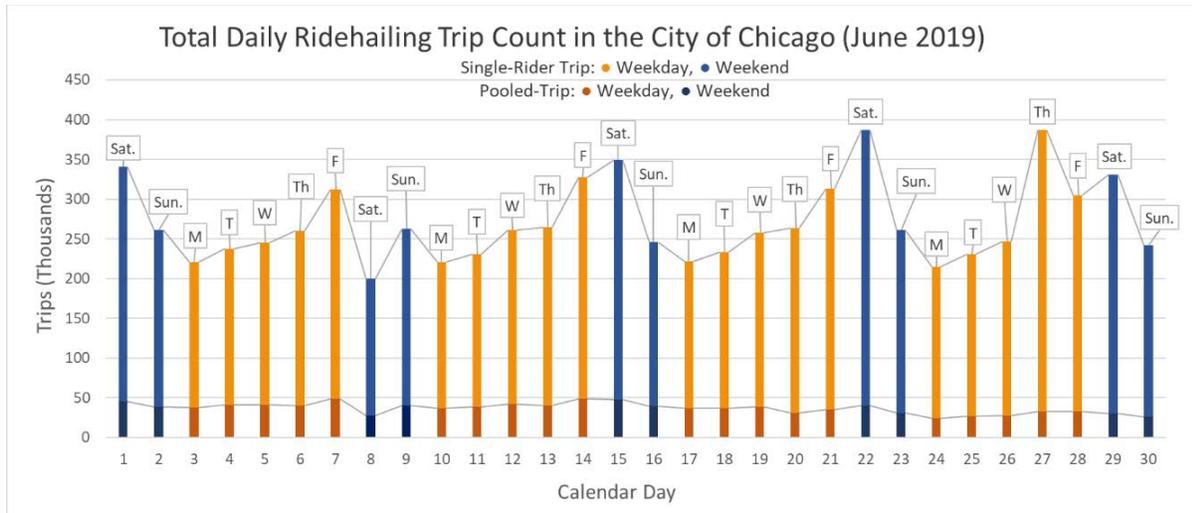
6
7   **Figure 3: Daily Ride-hailing Trip Counts in the City of Chicago during June 2019**
8
9   The next two figures show the distribution of trips by starting hour, by type of day;
10   weekday trips (Monday-Friday) are depicted in Figure 4 and weekend trips (Saturday – Sunday)
11   are depicted in Figure 5.
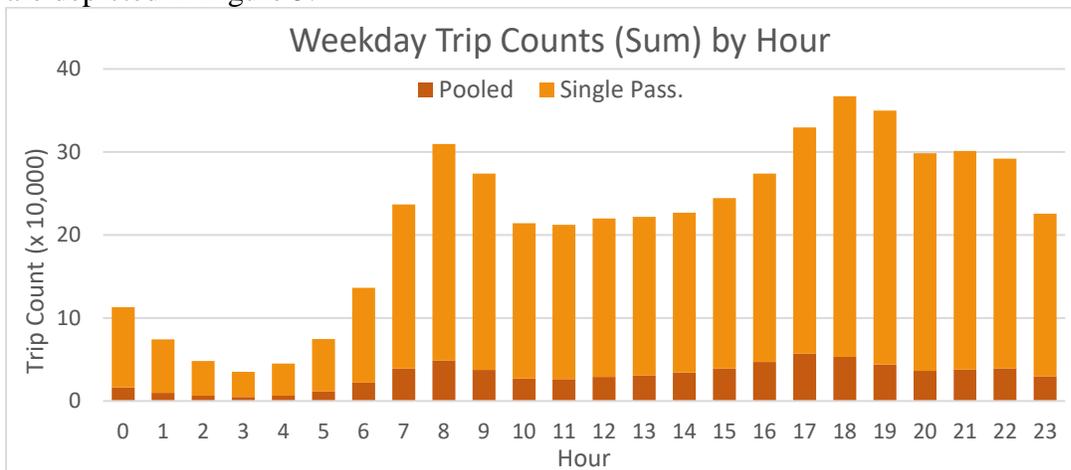
12
13   **Figure 4: Weekday Trip Counts (Sum) by Hour**
14

1
2 **Figure 5: Weekend Trip Counts (Sum) by Hour**
3 For weekday trips, two demand peaks exist at hours 8 and 16, whereas for weekend trips peak
4 periods are not as distinct. On weekends, people tend to participate in social and leisure activities
5 that are not confined to the traditional 9-5 work schedule. Hence, travel demand is more evenly
6 distributed throughout daylight hours.

7 **Data Processing**
8 Data processing was completed in three steps, with the ultimate output being the probability of a
9 rider choosing public transit. This probability is derived from a multinomial nested logit (MNL)
10 model based on the Chicago's travel behaviors in 2015 [11]. This model is descriptively
11 explained, and its relevancy is introduced later in this section.
12   As a brief overview, the first two step were performed in the program, *ArcGIS*, using two
13 separate tools: (1) Route Analyst and (2) Spatial Join. These two steps are novel, in that we
14 combine GTFS data and source data to compute the time-conscious transit-equivalent route. The
15 output of these two steps, per RH trip, were a transit-equivalent trip and the number of transfers
16 required to complete the trip. For the third and final step, we used an existing utility model to
17 compute the P(Transit|CTA), the probability of choosing to service a trip using CTA.
18   Considering the scope of our research, we chose to conduct a thorough literature review
19 to find an existing utility model with obtainable input values, and that was derived from a sample
20 with similar demographics and travel behaviors. We compared our methodology and dataset
21 against these models to determine which existing utility model was most suitable. We selected
22 the multinomial nested logit (MNL) model developed by Javanmardi et al. [11]. Their model was
23 developed from a revealed preference survey using Google Maps API and RTA's Goroo
24 TripPlanner trip data. The purpose of their MNL model is to measure the varying mode choice
25 behaviors regarding alternative transportation, with increased accuracy from RP surveys. This
26 model was developed from trips executed within the city limits of Chicago. Given the model
27 shares the same study region of this paper, it is the most appropriate option in that it will capture
28 the travel behavior with a greater degree of accuracy. Lastly, the study year (2015) of their
29 research is appropriate in that ride-hailing was introduced to Chicago before that time, thus their
30 model should capture any evolution of mode choice behavior and preferences towards or against
31 alternative transportation.
32   The model's formulae are represented by the equations below (Equations 1-6). Given that
33 the model was used for a range of modes, we modified its subscripts to align with the variables

1 we used. The constant and coefficient values, and variables are mode-specific and were provided
2 [11].
3 **Utility of Transit, $U_{Transit}$**

$$U_{Transit} = 2.93 - 1.04TT - 0.13TC - 0.17n_t - 0.77HHI + 0.45wrktrp \\ - 0.39d_A - 023d_E \tag{1}$$

4 **Probability of Selecting Transit, $P_{Transit}$**

$$P_{Transit} = \frac{e^{U_{transit}}}{1 + e^{U_{transit}}} \tag{2}$$

5 **Utility of CTA, U_{CTA}**

$$U_{CTA} = -0.39TT - 0.059TC - 0.33n_t + 0.022n_{Stop,O} + 0.0089n_{Stop,D} \\ + 0.77shptrp + 1.78wrktrp - 0.46HHI \tag{3}$$

6 **Probability of Selecting CTA, P_{CTA}**

$$P_{CTA} = \frac{e^{U_{CTA}}}{1 + e^{U_{CTA}}} \tag{4}$$

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

$$P_{CTA|Transit} = P_{Transit} \times P_{CTA} \tag{5}$$

8   The equations were executed in the respective order per trip, with Equation 6 outputting
9 the final probability used in the analyses.
10   While this derivation of this model exhibited strong similarities to our study
11 characteristics, it did contain several caveats. Our dataset did not completely satisfy all required
12 input of the model, therefore we either generalized, estimated, or calculated these parameters. In
13 doing so, assumptions were made. Table 1 outlines the input attributes, their corresponding
14 definition, and their availability with respect to our dataset. Following the table, each
15 assumption-based variable, and its calculation process(es) is explained in greater detail.
16
17 **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	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	$dest_{CBD}$	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, $dest_{CBD} = 1$.

18

1   **Household Income** *(HHI):* given the privatization of the ride-hailing dataset, we were not
2   granted access to socioeconomic characteristics of the individual ride-hailer. However, we
3   computed the HHI for a rider using a dataset containing the average HHI per census tract. We
4   defined the HHI to equal the average HHI of the origin, if the trip was executed on a weekday
5   between 5:00 AM and 12:00 PM, or of the destination, if the trip was executed on a weekday
6   between 12:00 PM and 7:00 PM. For all trips outside this boundary, the HHI was defined as the
7   average between the origin and destination HHI.
8   **Census Tracts of Origins and Destinations**: The ride-hailing dataset contained two fields for
9   the origin and destination tract values, although a group of trips would had null origin or
10 destination tract values but would have the geographic coordinates of the origin and destination.
11 To account for this, we estimated the corresponding tract IDs via a minimum distance program in
12 Matlab. First, we obtained the geographic boundaries of the census tracts from the Chicago Data
13 Portal. We imported this SHP file into ArcGIS and performed geometry calculations to calculate
14 the tract's centroid and to output the corresponding latitudes and longitudes. This output table
15 was imported into Matlab as a matrix. Using each trip origin and destination latitude and
16 longitude, we calculated the distance between each tract and O-D coordinates. The census tract
17 ID corresponding to the smallest distance value was selected and replaced the null value for the
18 origin or destination tract value.
19 **Number of Transit Stops per O-D Zone** *nStopOrigin; nStopDest*: these two attributes were a
20 function of the O-D census tracts for the trip and were calculated as the number of transit stops in
21 the corresponding origin or destination zone (census tract). For each trip, we indexed the number
22 of stops in the O-D tracts from this table and appended these values to the final table for the
23 utility calculation.
24 **Saturday/Sunday Classification**: To test true for the attributes below (*destCBD, wrktrp*), a trip
25 could not have been serviced on a Saturday or Sunday. Therefore, to classify a 'weekend'
26 (Saturday or Sunday) trip, we composed a vector of June 2019 calendar days corresponding to
27 each pair of Saturdays and Sundays. If a trip's start calendar day was identified as a weekend
28 day, then it results in a *false* value.
29 **Destination within CBD at Rush Hour (*destCBD*)**: Due to the anonymity of the dataset, we did
30 not have individual details on trip characteristics such as the trip purpose. We assumed the AM
31 and PM peak period to occur between 6:00:00 and 9:00:00 AM and 16:00:00 and 19:00:00 PM,
32 respectively. For a trip to test "true" (destCBD = 1), we first determined if the destination census
33 tract was a member of the CBD census tracts array. If true, the trip start hour was then tested
34 against the two peak periods, otherwise, *destCBD* = 0.
35 **Trip Purpose: Work (*wrktrp*)**: like *destCBD*, we assumed that a trip was deemed as a "to-or-
36 from work" commute (wrktrp = 1) if the start time lied between 5:00:00 and 19:00:00
37 on a weekday.

38 ## Analysis
39 ### *Replaceability of a Transit Trip*
40 Ultimately, to determine if a ride-hailing trip "replaced" its transit-equivalent trip, we compute
41 the probability of using CTA. The magnitude of these probabilities indicates the viability of
42 public transit serving a specified trip and depends on how favorable the trip's LOS attributes and
43 trip-specific characteristics are to the rider. We chose to classify a transit-equivalent trip by its
44 replaceability, categorized by two groups: replaced (R) trips and not-replaced (NR) trips. We
45 initially assumed the threshold value distinguishing a trip being "replaced" (R) or "not replaced"
46 (NR), to be 0.5. Where all trips with a P(Transit|CTA) < 0.5 were classified as not replaced and

1   all trips with a P(Transit|CTA)>=0.5 were classified as replaced. However, following the first
2   sensitivity analysis trial, we determined that trips with P(Transit|CTA) close to 0.5 switch
3   between the R- and NR-groups. These trips are fuzzy and are not reliable indicators of true
4   mode-choice modeling behavior. Thus, we chose to implement a buffer, where trips with
5   P(Transit|CTA) = (0.45-0.55) are removed and excluded from the summary statistics. This
6   modification is represented by the conditional statement below.
7   For an individual ride-hailing trip, T,

8
$$replaceability\ group_T = \begin{cases} Not\ Replaced\ (NR), & 0 \le P(Transit|CTA) \le 0.45 \\ Replaced\ (R), & 0.55 \le P(Transit|CTA) \le 1.0 \\ Buffer\ Zone, & 0.45 < P(Transit|CTA) < 0.55 \end{cases}$$

9   ***Sensitivity of P(Transit|CTA)***
10   We then chose to conduct a parametric sensitivity analysis of P(Transit|CTA) with respect to the
11   following decision variables:
12   1. Transit stops per census tract (StopsinTract)
13   2. Household income (HHI)
14   3. Total travel time (TTT)
15   4. Walk time (WT)
16   The $P(Transit|CTA)$ was recalculated under 20 sensitivity conditions, for each decision
17   variable. Per variable, its observed value was adjusted in increments of 5%, ranging from -50%
18   to +50%. Given that each variable was tested independently, there were a total of 80 trials. For
19   each sensitivity condition, the algorithm was executed with the adjusted parameter and a new
20   P(Transit|CTA) was output for all trips and averaged per group (R/NR).
21   Assuming that both groups share the same standard deviation, we can estimate σ by
22   calculating the pooled standard deviation, s_p, with the equation below. The pooled standard
23   deviation for the observed and sensitivity condition data sets, for group R or NR, is:

$$s_p(group, sensitivity\ condition)$$
$$= \sqrt{\frac{[(n_{observed} - 1) * s^2_{observed}] + [(n_i - 1) * s^2_i]}{(n_{observed} + n_i) - 2}} \qquad (6)$$

24   Where,
25   $n_{observed}$ = the number of trips in the observed group
26   $s_{observed}$ = standard deviation of the observed group
27   n_i = number of trips in the sensitivity group
28   s_i = standard deviation of the sensitivity group
29   It should be noted that $n_{observed}$ and $s_{observed}$ are fixed values under all sensitivity
30   conditions.
31   To measure the level of influence and statistical relationship of each decision variable
32   and the $P(Transit|CTA)$, we performed a two-tailed pooled t-test. Considering there is no
33   overlap between the observed and sensitivity condition data, the two-tailed test was most
34   suitable. We performed the t-test per variable, sensitivity condition, and trip group (R and NR).
35   The relationship between the t-statistic and the critical value indicate whether we accept or reject
36   the null hypotheses stated below:
37
38   $H_{Null\ (R)}$ = The $\bar{P}(Transit|CTA)$ of the observed R group is not statistically
39   different from the $\bar{P}(Transit|CTA)$ of the sensitivity R group.

1

2 $H_{Null\ (NR)}$ = The $\bar{P}(Transit|CTA)$ of the observed NR group is not statistically
3 different from the $\bar{P}(Transit|CTA)$ of the sensitivity NR group.

4   If the t-statistic is greater than the critical value (1.365), then we reject the null hypothesis
5 and refer to the alternative hypothesis. The alternative hypothesis opposes the null by concluding
6 that there is a statistically significant difference between the observed and the sensitivity
7 condition data. Meaning, the influence of the decision variable on the $P(Transit|CTA)$ is
8 expected to have an effect on the whole population, similar to the effect of the sensitivity
9 condition.
10   The following equation was used to compute the t-statistic per variable and group for
11 each sensitivity condition (Equation 8):

$$t_i = \left| \frac{\bar{P}_i - \bar{P}_{observed}}{\sqrt{\left(\frac{s_i^2}{n_i}\right) + \left(\frac{s_{observed}^2}{n_{observed}}\right)}} \right| \tag{7}$$

12 Where,
13 i = sample trips of group g (R or NR), decision variable *var*, and percent-change condition *%Δ*.
14 P_i = mean $P(Transit|CTA)$
15 $P_{observed}$ = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-
16 change condition *%Δ*.
17 s_i = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-change
18 condition *%Δ*.
19 $s_{observed}$ = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-
20 change condition *%Δ*.
21 n_i = sample size for group g (R/NR), decision variable *var*, and percentage-change condition *%Δ*.
22 $n_{observed}$ = mean $P(Transit|CTA)$ for group g (R/NR), decision variable *var*, and percentage-
23 change condition *%Δ*.

24 **RESULTS AND DISCUSSION**

25 The following section introduces the results from the replaceability and sensitivity analyses for
26 the four sensitivity variables: TTT, WT, HHI, and SiT.

27 **Replaceability Analysis (Probability of Selecting CTA)**
28 The P(Transit|CTA) estimation is a function of the abovementioned procedures and their
29 respective outputs. To reduce the paper's length, we will summarize each group in terms of size
30 and P(Transit|CTA).. For ease of recall, the group for a trip, *T*, is categorized by the following
31 conditional:

32 $$replaceability\ group_T = \begin{cases} Not\ Replaced\ (NR), & 0 \leq P(Transit|CTA) \leq 0.45 \\ Replaced\ (R), & 0.55 \leq P(Transit|CTA) \leq 1.0 \\ Buffer\ Zone, & 0.45 < P(Transit|CTA) < 0.55 \end{cases}$$

33   Of the 7,949,902 trips output from the route analysis, approximately 8% (646,808 trips)
34 had a probability lying within the buffer range of 0.45<P<0.55. Moving forward, we will be
35 summarizing findings in terms of the R and NR groups only.

1 **Sensitivity Analysis**
2 Results from our sensitivity analyses were extensive and will be summarized for the submission
3 guidelines of this paper.
4   For the aforementioned sensitivity parameters, t-tests were conducted at the 95%
5 confidence level. Under each sensitivity condition, all variables exhibited t-values greater than
6 the critical value. This implies that with 5% error, we can assume the adjustment of each of those
7 four parameters will have a statistically meaningful impact on the P(Transit|CTA).
8   For each trip, the variable of interest was adjusted, the P(Transit|CTA) was recalculated,
9 and the trip was recategorized based on its magnitude. Thus, it is important to consider that for
10 each percent-change condition, the sample for R and NR trip groups will vary in size. Moreover,
11 means of both groups will fluctuate with the sensitivity condition.
12   We will introduce stacked bar charts per sensitivity variable that depicts the overall
13 weighted mean P and trip count per sensitivity-condition. Equation 10 below is used to compute
14 the total weighted mean P(Transit|CTA):

$$P_{weighted} = P_R \left(\frac{n_R}{n_T} \right) + P_{NR} \left(\frac{n_{NR}}{n_T} \right) \tag{8}$$

15   The data labels (percentages) within each bar corresponds to the percentage of total trips
16 (n_T) that each group contains. Per the legend, the blue portion of the stacked bar corresponds to
17 the replaced (R) trips, whereas the orange portion corresponds to the not replaced (NR) trips. It
18 should be clarified that these percentages are independent from the portion heights of the stacked
19 bars.
20   In summary, the first three variables (*TTT*, *WT*, and *HHI*) exhibit an inverse correlation
21 with P(Transit|CTA) whereas the remaining variable, *SiT*, has a positive relationship with
22 P(Transit|CTA).
23   The first two variables are associated to travel time: *WT* and *TTT*. The P(Transit|CTA) is
24 more sensitive to adjustments in *TTT* when compared to WT. The range of the weighted
25 probability for *TTT* is approximately 0.20, as opposed to 0.10 for WT. The relatively heightened
26 sensitivity for *TTT* can be explained by *TTT*'s formula and the nesting of *WT* in its value. Recall
27 that the *TTT* is the sum of the *IVTT*, wait time, and walk time. Meaning, a 50% decrease in *TTT*
28 includes a 50% decrease in *IVTT* and wait time in addition to a decrease in walk time. Whereas a
29 50% decrease in *WT* does not include the reduction in *IVTT* and walk time. When the new *TTT* is
30 calculated, it uses the observed wait time and *IVTT*, but changes the *WT*. While these differences
31 exist, they are relatively small in comparison to their distribution patterns and weighted
32 probability values. In Figure 6 and Figure 7, the volumetric share between *R* and *NR* trips is
33 indistinguishable.
34
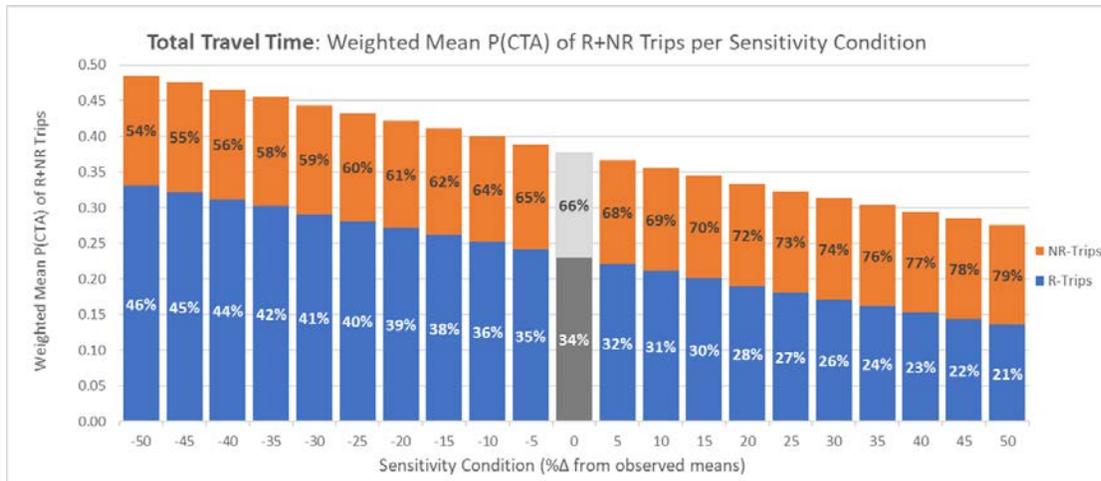
Figure 6: **Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Total Travel Time (TTT)**

When analyzing the *WT*, it is important to consider that a trip's replaceability is contingent on the persons' capabilities of physical exertion. 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. Additionally, the replaceability is influenced by the user's safety, which is dependent upon the perception of the route's surrounding physical environment(s). These conditions produce a bias to act more conservatively such that hazardous events are mitigated. All of these conditions, concerns, and exceptions cannot be explicitly accounted for in our model.
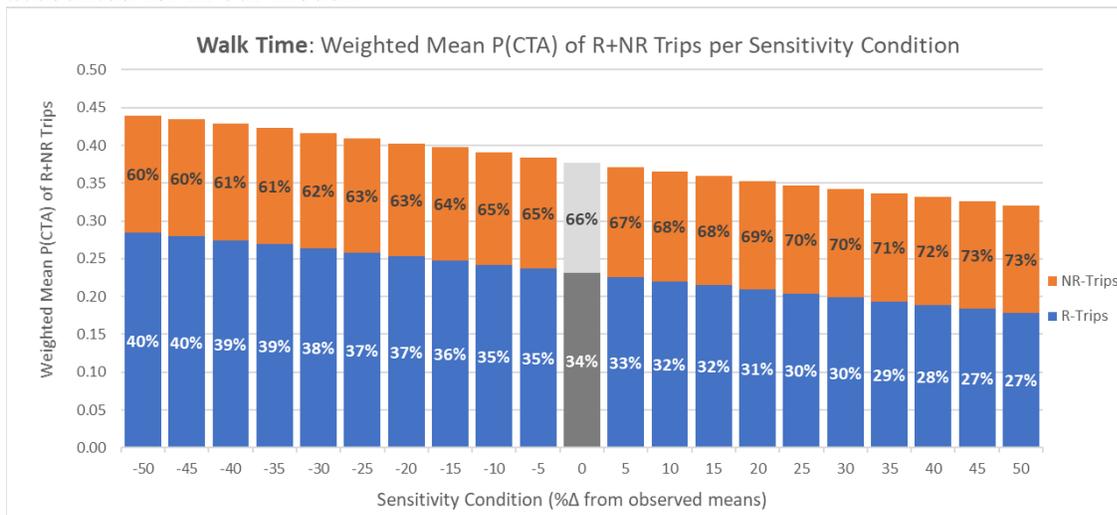
Figure 7: **Weighted Mean Probability and Trip Distribution between Groups per Condition for Sensitivity Variable: Walk Time (WT)**

The next variable, average *HHI*, has less impact on P(Transit|CTA) per Figure 8. In existing literature, it was determined that ride-hailers exhibit demographic characteristics that are at variance with the average American. Ride-hailers were found to be more educated and be of a

1   higher income class. Therefore, the use of the average *HHI* may undervalue that of the average
2   ride-hailer and the accuracy of these results could be challenged. Nonetheless, the trend and
3   behavior HHI has on P(Transit|CTA) is transposable.
4
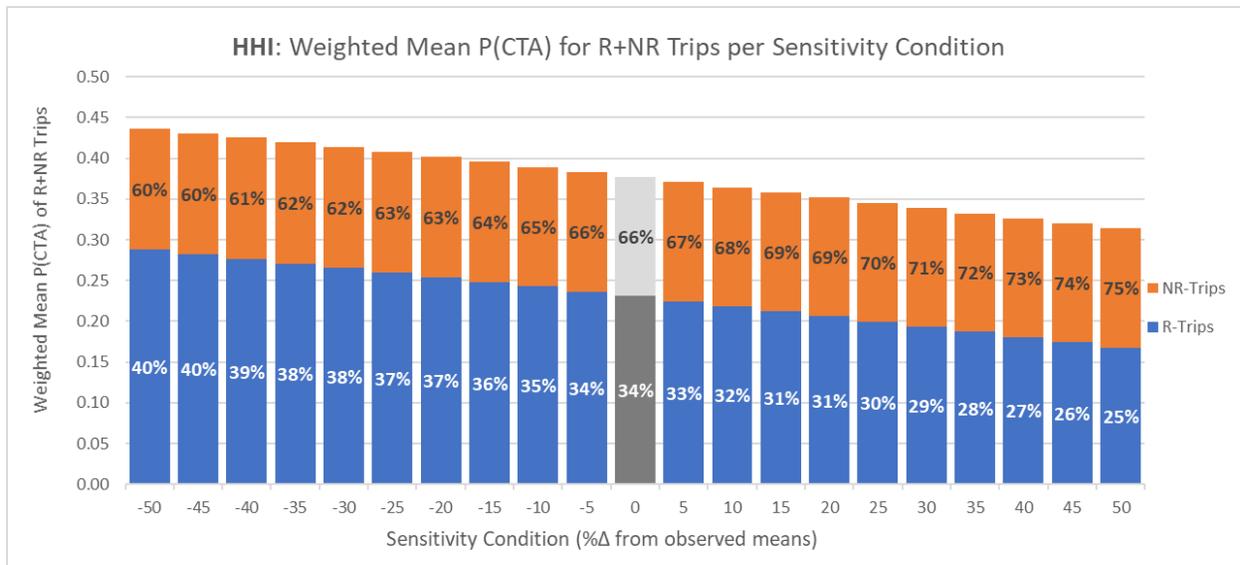
5
6   **Figure 8: Weighted Mean Probability and Trip Distribution between Groups per Condition for**
7   **Sensitivity Variable: Average Household Income (HHI)**
8
9   In opposition to the prior variables, the number of transit stops per census tract (*SiT*) was
10   exhibited a positive correlation with the P(Transit|CTA), as illustrated in Figure 9. The number
11   of transit stops in a network has many implications on operations and ridership. An increase in
12   transit stops implies an increase in route LOS. As the distance between consecutive stops is
13   decreased, the average access and egress distance decreases. With this increase in accessibility,
14   the volume of serviceable patrons increases.
15
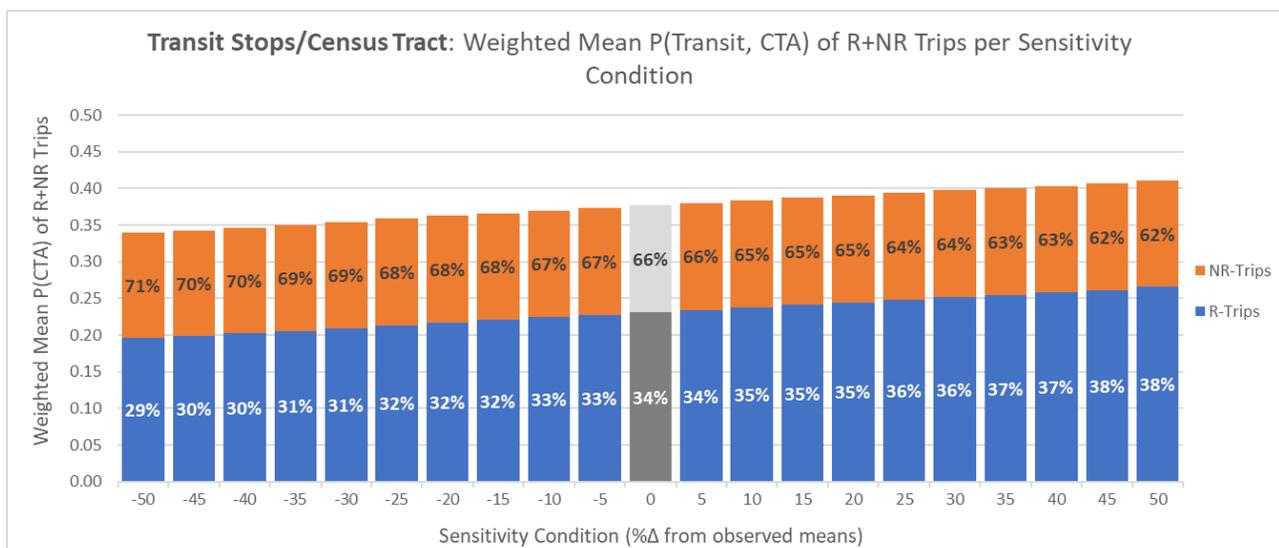
16
17   **Figure 9: Weighted Mean Probability and Trip Distribution between Groups per Condition for**
18   **Sensitivity Variable: Transit Stops per Census Tract**

1
2   Although there exist caveats with the more extreme positive sensitivity conditions that
3   are not captured in the utility model used. The addition of transit stops to existing routes must be
4   optimized to account for added lost time. Delay has the opportunity to incur at every transit stop
5   during the approach, boarding and dwelling, and exit. Boarding and dwell times can quickly
6   accumulate during peak period hours when there are large platoons of entering riders, and there
7   is discontinuity in payment forms. Additionally, this is a consequence of increased ridership. The
8   second source is called the 're-entry' delay; this is the time required for the driver to merge into
9   oncoming traffic. For every additional transit stop, one re-entry delay is incurred per cycle. The
10   summation of these delays per stop and per cycle can adversely affect the travel time between
11   stops, and the TTT of each rider. In summary, the addition of transit stops increases the
12   accessibility and consequently, utility. Although designing addition to increase ridership must
13   strategically consider implications it has on the existing travel times and LOS attributes.

14   **CONCLUSIONS**

15   The impact of ride-hailing services on the recent decline in public transit ridership has not been
16   widely explored. The current body of research is constrained to empirical studies that vary in
17   methodologies used and analyze relatively small samples. To our knowledge, there are no studies
18   that explore the research question using a massive dataset that contains individual trips over an
19   extended period of time. Furthermore, our approach to exploring the research question is
20   resourceful and novel. We define the replaceability of a RH trip by a series of spatiotemporal and
21   mathematical analyses. First, the real-time transit equivalent trip is computed using the GTFS-
22   integrated ArcGIS Route Analysis. Then, the probability of choosing transit over all other
23   alternatives identified any potential viable transit-equivalent trips.
24   Our findings indicate that 31% of ride-hailing trips were executed where the transit
25   alternative exhibited a competitive utility, with respect to travel times, fare/expenses, and
26   workload. Consequently, over the month of June, the total revenue lost from trips replaced by
27   ride-hailing is estimated to be \$6,114,450[4]. If we assume the percentage of replaced trips and trip
28   counts for each month can be represented by June 2019, then the total loss in fare revenue in
29   Chicago over one year would be approximately 73 million dollars. Further, the ramifications of
30   the demand transfer to ride-hailing services is not fully represented by the loss in revenue. As
31   such, public transit agencies should employ strategies to increase transit utility such that a
32   significant portion of this estimate can be recovered.
33   Publicly available ride-hailing trip data will likely maintain its anonymity by recording
34   origins and destinations as their census tract centroids. Given it is unlikely for the precision to
35   increase, studies that are macroscopic and encompass all attribute types (temporal, spatial,
36   monetary) should be executed. However, the use of our methodologies and approach is only
37   doable for regions that mandate the submission of all ride-hailing trips and their corresponding
38   spatial and temporal parameters of the origin and destination. Recording and releasing this data
39   will enable institutions to publish research that will provide a greater understanding of how ride-
40   hailing impacts the transportation network and economy in varying geographic settings.
41   Moving forward, future research should focus on mode-choice behavior to thoroughly
42   understand the factors favoring and disfavoring the use of transit. Replaced trips have

---

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

1 comparable transit alternatives, hence there exists preferences and attitudes exclusive from LOS
2 attributes that favor ride-hailing. Regarding NR trips, transit agencies should turn inwards and
3 evaluate services, or the lack thereof, in the corresponding origin and destination zones.

4 AUTHER CONTRIBUTIONS

5 The work described in this article is the collaborative development of all authors,
6 conceptualization, J.D.and H.A.R.; methodology, L.B., J.D., and H.A.R.; software, J.D. and
7 L.B.; validation, L.B., J.D., and H.A.R.; formal analysis, L.B., J.D., and H.A.R.; investigation,
8 L.B., J.D., and H.A.R.; writing—review and editing, L.B., J.D., and H.A.R.

9 ACKNOWLEDGMENTS

11 REFERENCES

12 1.   *The history of Uber*. 2020 June 6, 2020]; Available from:
13   https://www.uber.com/newsroom/history/.
14 2.   Greiner, A., et al. *A History of Lyft, From Fuzzy Pink Mustaches to Global Ride Share*
15   *Giant*. 2019 April 2, 2019 [cited 2020 June 11, 2020]; Available from:
16   https://www.cnn.com/interactive/2019/03/business/lyft-history/index.html.
17 3.   Melendez, S., *How Uber Conquered the World in 2013*. 2014, Fast Company.
18 4.   Jiang, J., *More Americans are using ride-hailing apps*, in *FactTank*. 2019, Pew Research
19   Center: Online.
20 5.   *Ride-hailing Phenomenon Under Hood*. [cited 2020; Available from:
21   https://mobility.here.com/learn/smart-mobility/ride-hailing-phenomenon-under-hood.
22 6.   *APTA Ridership by Mode and Quarter 1990-Present*, APTA, Editor. 2019: Online.
23 7.   *Population by Age*. 2018, Kaiser Family Foundation: Online.
24 8.   Lo, J. and S. Morseman, *The Perfect uberPOOL: A Case Study on Trade-Offs*.
25   Ethnographic Praxis in Industry Conference Proceedings, 2018. **2018**(1): p. 195-223.
26 9.   Amirkiaee, S.Y. and N. Evangelopoulos, *Why do people rideshare? An experimental*
27   *study*. Transportation Research Part F: Traffic Psychology and Behaviour, 2018. **55**: p. 9-
28   24.
29 10.   Clewlow, R.R. and G.S. Mishra, *Disruptive Transportation: The Adoption, Utilization,*
30   *and Impacts of Ride-Hailing in the United States*. 2017, UC Davis Institute for
31   Transportation Studies. p. 38.
32 11.   Javanmardi, M., et al., *Mode Choice Modeling Using Personalized Travel Time and Cost*
33   *Data*. 2015, University of South Florida: Tampa, Flordia.
34 12.   Rayle, L., et al., *Just a better taxi? A survey-based comparison of taxis, transit, and*
35   *ridesourcing services in San Francisco*. Transport Policy, 2016. **45**: p. 168-178.
36 13.   Henao, A. and W.E. Marshall, *The impact of ride-hailing on vehicle miles traveled*.
37   Transportation, 2018.
38 14.   Contreras, S.D. and A. Paz, *The effects of ride-hailing companies on the taxicab industry*
39   *in Las Vegas, Nevada*. Transportation Research Part A: Policy and Practice, 2018. **115**: p.
40   63-70.
41 15.   Stiglic, M., et al., *Enhancing urban mobility: Integrating ride-sharing and public transit*.
42   Computers & Operations Research, 2018. **90**: p. 12-21.

1   16.   Graehler, M., R. Mucci, and G. Erhardt, *Understanding the Recent Transit Ridership*
2   *Decline in Major US Cities: Service Cuts or Emerging Modes?*, in *TRB 2019 Annual*
3   *Meeting*. 2018, Transportation Research Board: Washington, D.C.
4   17.   Sadowsky, N. and E. Nelson, *The Impact of Ride-Hailing Services on Public*
5   *Transportation Use: A Discontinuity Regression Analysis*. 2017, Bowdoin College.
6   18.   Murphy, C., *Shared Mobility and the Transformation of Public Transit*. 2016, Shared-
7   Use Mobility Center (SUMC).
8   19.   N.A. *DATA USA: Chicago, IL*. 2020 [cited 2020 May 25, 2020]; Available from:
9   https://datausa.io/profile/geo/chicago-il.
10 20.   *QuickFacts Chicago city, Illinois*, U.S.C. Bureau, Editor. 2020, United States Census
11   Bureau: Online.
12 21.   Hughes-Cromwick, M., *2018 Public Transportation Fact Book*. 2018, American Public
13   Transportation Association.
14 22.   *Facts at a glance*. 2017 [cited 2020 May 25, 2020].
15 23.   *Transportation Network Providers (Ride-Hail)*. 2020 [cited 2020 June 11, 2020];
16   Available from: https://www.chicago.gov/city/en/depts/bacp/supp_info/transportation-
17   network-providers.html.
18 24.   *Transportation Network Providers - Trips*, J. Levy, Editor. 2020, City of Chicago:
19   Online.
20 25.   *CTA GTFS*, GTFS, Editor. 2020, OpenMobilityData.
21