Characterizing Ride-Hailing Driver Attrition and Supply in the City of Chicago Through the COVID-19 Pandemic

Sharika Hegde1, Hoseb Abkarian1, and Hani S. Mahmassani1

Abstract
The flexible nature of on-demand ride services provided by transportation network companies (TNC) has resulted in unique supply-side challenges as the industry deals with the COVID-19 pandemic. Early during the pandemic, there was a 70% decrease in the number of drivers accepting trips on TNC platforms, as individual drivers chose to reduce their risk of viral infection and abide by social distancing recommendations. Given the two-sided market nature of TNCs, the decrease was also the effect of reduced rider demand creating a less desirable driver experience. This paper characterizes and quantifies this change in supply as it relates to driver residency, tenure, attrition, and the number of trips provided. The distribution of drivers accepting trips shifted slightly toward the lower income and higher minority areas of Chicago. Using survival analysis methods, we find that retention among drivers who started in the early months of the pandemic was significantly lower than in reference years, after six months of driving. The results of the negative binomial regression show that drivers on a single TNC platform provided 20% less trips than drivers on multiple platforms. This difference increases to 30% during the pandemic. Additionally, new drivers joined multiple apps during COVID-19, likely to serve more trips and secure higher income. The results of this paper can be used to understand and target driver retention to accelerate the recovery of the TNC industry.

Keywords
on-demand mobility, transportation supply, mobility as a service, ride hailing, ridesharing, transportation network companies, driver attrition, COVID-19

Transportation network companies (TNCs), such as Uber, Lyft, and Via, have emerged as an important mode in the urban transportation landscape. These companies provide on-demand and scheduled point-to-point trips through apps on mobile devices. These services have grown increasingly popular with travelers using them for a variety of trip purposes such as first- and last-mile trips, commute trips, and recreational trips. TNCs, however, do not directly employ nor control drivers, operating instead as platforms that connect riders with drivers. Therefore, drivers are considered to be independent contractors who can choose when to work and for how long. As such, TNCs operate in a two-sided market where the supply, or the number of available drivers, varies over time depending on interactions between the number of riders (demand), the profit margins, operating costs, and—during the COVID-19 pandemic—the risk of viral transmission.

Although all urban transportation modes faced profound operational changes during the COVID-19 pandemic, TNCs have a unique supply-side challenge in a two-sided market. Given the flexible nature of TNC operation, the supply-side is particularly precarious in the face of disruption as drivers can easily choose not to provide trips. Thus, it is important to study the supply side of TNCs during the pandemic and characterize the behavior of the drivers to anticipate how they may react.
in future disruptions. Lessons can then be extracted from this pandemic to forecast future performance and quantify measures of uncertainty during similar disruptions in transportation supply and demand.

The number of TNC drivers offering rides within the City of Chicago dropped by over 70% between March and April 2020. In addition to the lockdown in the city that commenced on March 21, the drop in driver supply is almost certainly from reduced demand for rides and drivers’ choice not to accept rides to reduce their own risk of viral infection and to abide by the CDC’s recommendations to practice social distancing. Given the shift toward remote work and e-commerce, and the abundance of guidance to limit unnecessary travel, the reduction in supply also mirrored suppressed demand. The reduced demand likely worsened the probability that drivers would be matched with riders, increasing driver idle time, and diminishing the driver experience.

In addition to demand for rides, income prospects through fares and tips, driver expenses such as insurance and gas, and other factors that affect overall driver experience also affect the supply side of the rideshare market. However, limited by the scope and availability of data, this paper will characterize and quantify changes in the supply-side only as they relate to driver tenure, attrition, residency, and the number of trips provided. We identify differences in the entry and exit rates of drivers who initiated the drop in supply and explore the tenure distribution of drivers who quit between March and April of 2020 compared with those who quit between those same months in 2019 and 2021. Spatially, we track the change in driver residential location, which can be used to estimate changes in the socio-demographic characteristics of the driver populations before and after the start of the pandemic. An important aspect that this paper attempts to quantify is the impact of COVID-19 on attrition rates. Specifically, we study the probability of quitting the system in a survival-based analysis for two groups: those starting during COVID-19 and those who started before. This allows comparisons of these two groups to determine the “inertial” effect of the pandemic on driver retention. Finally, we run a negative binomial regression to understand differences in the distribution of the number of trips provided by drivers before and during the COVID-19 pandemic.

**Literature Review**

The rise of the “gig-economy,” by which people can sell their labor through apps managed by firms, has deeply transformed the transportation industry. As mobile computing infrastructure improved, TNCs emerged as platforms connecting drivers with their own vehicles to riders. Because the platforms facilitate interactions between two parties by appropriately charging and matching each side, TNC operation can be considered a two-sided market (1). In two-sided markets, the attractiveness of the platform for one group grows with the size of the other group, but the TNC market exhibits even greater dynamics. The rider’s utility is a function of their waiting time, which is determined by the number of drivers looking for passengers and the number of competing passengers (2). Similarly, the driver’s utility depends on their idle time, which is determined by the number of customers looking for rides but also the number of competing drivers (2). Both the riders’ and drivers’ utility depend heavily on the TNC platforms’ matching algorithms.

Unlike formal employees, TNC drivers are allowed to set their own work hours and can work for multiple platforms at the same time. Drivers must abide by platform rules such as keeping their vehicles clean and registered, passing certain background checks, and maintaining reasonable trip acceptance rates (3). The flexible nature of driving for on-demand ride sourcing introduces new transportation models that focus on the supply side. Using an ordinal logit model, Berliner and Tal estimated the willingness to drive for TNCs and found that those who generally report high vehicle-miles traveled and have more children are most likely to become TNC drivers (4). While some consider TNC driving a full-time job, most do not join out of unemployment. An Uber- commissioned survey analysis of 601 drivers in 2014 and 833 drivers in 2015 found that just 8% were unemployed before joining, 48% viewed their earned income from Uber as a supplement to their existing income, and approximately 60% of drivers worked full- or part-time on another job (5). Using administrative data provided by Uber, a weekly continuation analysis found that among 11,267 drivers who gave their first trips between January and June of 2013, 11% had become inactive by the next month and after half a year only 70% of drivers remained active (5).

As the COVID-19 pandemic began in cities around the world, and the U.S. Centers for Disease Control and Prevention (CDC) began to encourage social distancing and remote work, travelers became more skeptical of shared mobility transportation modes such as public transit, taxis, and ride-sourcing platforms, causing dramatic reductions in ridership (6, 7). Many studies focusing on travel demand have found significant changes before and after the initial lockdown orders. Ridership on the Chicago Transit Authority dropped by nearly 68% (8). In addition to transit ridership changes, studies on the short-term impact of the COVID-19 lockdowns on road transportation in the City of Chicago have found that, although the number of automobile collisions did reduce, the severity of those collisions increased during the pandemic (9).
The effect of COVID-19 on ride-sourcing and taxi platforms rests both in the changed travel demand but also in the supply of drivers. In a study of Shenzhen’s taxi industry during the pandemic, Zheng et al. (10) found that the demand was reflected in longer search times to find passengers and lower productivity for drivers, and that drivers adjusted their work hours to focus on serving peak-time trips, either in response to the demand change or to mitigate their own infection risk. The number of taxis in operation in Shenzhen fell by 60% from the first to the last week of January (10). Similar studies of the effect of COVID-19 on taxi supply found that the March 2020 stay-at-home order issued in the City of Chicago reduced ridership by 95% in taxis and the number of operating taxis by 85% (11).

This paper presents a comprehensive study to characterize the change in the supply-side of TNC operation in the City of Chicago before and during the COVID-19 pandemic. Specifically, this paper will use spatial analysis, survival analysis methods (12), and a negative binomial regression analysis to understand changes in driver residency, tenure, attrition, and the number of trips provided. The findings provide insight into the dynamics of TNCs as two-sided mobility markets, and the implications for service levels experienced by users in a changing post-pandemic mobility landscape; they can also help in targeting driver retention to accelerate the recovery of the TNC industry.

Data and Exploratory Analysis

The data used in this study was provided by the City of Chicago’s Open Data Portal (13, 14). The dataset includes all drivers that were reported by TNCs in Chicago as part of the City’s required licensing and reporting process. We believe the dataset provided by the City of Chicago is quite accurate as the data has been collected through a legal reporting process and made available for public use. For each report month (the first report month being in February 2015), the dataset provides individual driver-level data, including their start month, their city of residence, the state of their driver’s license, the ZIP code of their residence, the number of trips they provided in the report month, and whether the driver was reported by multiple TNCs that were operating. If a driver was included in a monthly report, then they were eligible for trips in the City of Chicago and were registered as a driver who could accept trips on any of the TNC apps, even if they did not actually provide any rides in that month. If a driver was eligible for trips in multiple months, they will be reflected in multiple entries in the dataset, one for each of the eligible months. The dataset, however, does not provide an identification number for the same driver across report months. Individual drivers, therefore, cannot be tracked over time. However, aggregate changes in the composition of drivers and the number of drivers at each tenure level in each report month may be inferred.

As the TNC market is still growing in the Chicago area, the time frame for this study will include the most recent data of drivers reported between August 2018 and May 2021 to understand changes in TNC driver supply caused by the COVID-19 pandemic. During this period, three TNC companies were in operation in the City of Chicago: Uber Technologies, Lyft, and Via Transportation.

Change in Supply Size

In the 2019 calendar year, the number of TNC drivers that accepted rides in Chicago hovered near 71,000 with a minimum of 67,221 drivers in February and a maximum of 73,386 drivers in August. Figure 1 tracks the supply of ride-accepting drivers since January 2019 overlaid with the rolling average COVID-19 case rate per 100,000 of population for the seven-day period ending on that date (15).

The number of drivers begins to fall at the start of 2020, coinciding with the World Health Organization’s declaration of a public health emergency of international importance (16). However, the largest drop occurred after Illinois’s stay-at-home order (17), when the number of drivers accepting rides dropped to 17,195 in April 2020. Compared with March 2020, this is a 71.6% decrease in the number of drivers completing trips in Chicago.

Throughout the pandemic, the number of drivers closely mirrored the number of COVID-19 cases in the City of Chicago. There was a noticeable decrease in drivers during the second wave of cases in November of 2020. Since the start of the Illinois stay-at-home order in March 2020, the number of drivers accepting trips had not recovered at the time of writing. However, the highest number of drivers in operation was 31,135 at the most recent data point in May 2021. While this hints at some level of recovery, there was still a considerable gap in driver supply from pre-pandemic levels as of May 2021.

In addition to active drivers, the dataset includes drivers in each report month who were eligible for trips in that month, regardless of whether the driver completed any trips. Because TNC drivers are considered independent contractors, their schedules are flexible, and drivers may choose not to accept trips in a month while remaining eligible on TNC platforms. Figure 2 documents the change in the number of eligible drivers in relation to the change in the number of drivers accepting trips.
The supply of eligible drivers dropped throughout the pandemic, but not as drastically as the supply of accepting drivers. While the number of drivers accepting trips began to increase in the months after the initial drop in April 2020, the number of eligible drivers continued to drop. This indicates that, among existing eligible drivers, there was less hesitancy to provide trips, but the overall supply of drivers has continued to decrease over the course of the pandemic.

To characterize the change in driver supply more thoroughly, we further consider the entry of new drivers and the exit of existing drivers. Figure 3 depicts the number of new drivers that started each month. The dataset contains underreporting of the number of drivers starting in a month (because of possible truncation), so Figure 3 assumes the same number of drivers were active in the month after they had begun driving. As reflected in the supply curve shown in Figure 3, the drop in new drivers joining the TNC apps began in February 2020. There was a 42.9% decrease in the number of new drivers joining TNCs, from 10,874 new drivers in February to 6,203 in May 2020. New drivers may have been more cautious, as little was known about the specifics of viral transmission. However, between April and June 2020 the number of drivers that quit each month briefly decreased. This could indicate that those existing drivers had a better grasp on the viral risk associated with driving in practice—or could not afford to forgo the income.

**Figure 1.** Drop in transportation network company (TNC) drivers accepting trips, overlayed with Chicago COVID-19 milestones.

**Figure 2.** Supply change between eligible drivers and eligible drivers that accepted trips.

The supply of eligible drivers dropped throughout the pandemic, but not as drastically as the supply of accepting drivers. While the number of drivers accepting trips began to increase in the months after the initial drop in April 2020, the number of eligible drivers continued to drop. This indicates that, among existing eligible drivers, there was less hesitancy to provide trips, but the overall supply of drivers has continued to decrease over the course of the pandemic.

To characterize the change in driver supply more thoroughly, we further consider the entry of new drivers and the exit of existing drivers. Figure 3 depicts the number of new drivers that started each month. The dataset contains underreporting of the number of drivers starting in a month (because of possible truncation), so Figure 3 assumes the same number of drivers were active in the month after they had begun driving. As reflected in the supply curve shown in Figure 3, the drop in new drivers joining the TNC apps began in February 2020. There was a 42.9% decrease in the number of new drivers joining TNCs, from 10,874 new drivers in February to 6,203 in May 2020. New drivers may have been more cautious, as little was known about the specifics of viral transmission. However, between April and June 2020 the number of drivers that quit each month briefly decreased. This could indicate that those existing drivers had a better grasp on the viral risk associated with driving in practice—or could not afford to forgo the income.
Although there was variation, the driver exit rate was higher than the entry rate on average after the initial COVID-19 shocks, constituting a drop in driver supply. Although the drop in driver supply was certainly the result of pandemic-related effects, as drivers tried to protect themselves from contracting COVID-19 through self-quarantine and social distancing, the driver supply must also be considered within the context of the market. In response to the CDC’s guidance to limit unnecessary travel and contact with people from outside of one’s home, the demand for TNC services decreased as well. Figure 4 shows that the drop in the supply constrained TNC demand in Chicago at the start of the pandemic.

The number of trips dropped by 80% from a five-day average of 262,989 trips per day on March 13 to an average of 52,161 on March 27. The decrease in passenger demand likely reduced the probability that drivers were quickly matched with riders by TNC platforms. In turn, the driver experience likely worsened as idle time increased. Among other factors, such as decreased demand and viral transmission worries, the decrease in the rate and reliability of income-generating trips could have encouraged drivers to quit from TNC platforms.

**Spatial Change in Supply**

In addition to the pandemic’s effect on the number of available drivers, the characteristics of the drivers that joined during or remained working through the pandemic differ from those of the pre-pandemic workforce. First, we consider the spatial distribution of drivers’ residences before and after the initial lockdown period in

---

**Figure 3.** Entry (green line) and exit (red line) of eligible drivers by month.

**Figure 4.** Daily trips with transportation network companies (TNC) in Chicago.
Chicago. Figure 5 spatially plots the fraction of Cook County drivers from each zip code in the county. Because there are many zip codes in Cook County, the magnitude of the fraction of drivers from each zip code itself was quite small. The 60606 zip code was overwhelmingly reported as a driver residence location, which was a gross over-representation compared with the reported population of that zip code and compared with similar zip codes. That zip code was therefore removed from the dataset as a non-representative datapoint. This zip code is not readily visible in Figure 5 but is in downtown Chicago in a business-heavy area.

The pre-pandemic figure from February of 2020 shows many drivers from outside of the downtown region. In April 2020, about 10% of active TNC drivers were from zip codes, 60617, 60619, 60620, and 60629, the four darker red zip codes to the south of the city, while they only housed 7% of drivers in February of
2020. Although many of the zip codes continued to have similar fractions of drivers in April 2020, the spatial distribution of drivers slightly shifted toward a larger fraction of drivers coming from zip codes in the south of Chicago than before the pandemic.

The comparison between before and after the start of the pandemic shows changes in the spatial distribution of drivers; however, seasonal effects might also contribute to variations in the distribution. Figure 6 tracks the change in the fraction of ride-accepting drivers in the month of April. The plot shows the change from 2019 to 2020, 2020 to 2021, and 2019 to 2021 to evaluate the overall change.

Comparing April 2019 with April 2020, the same trend emerges where the initial COVID-19 shock resulted in a larger fraction of drivers coming from areas toward the south side of the city. The change from April 2020 to April 2021, however, may indicate a form of recovery as the fraction of drivers began to increase in the northern areas of the city and shifted the distribution back toward that of April 2019. The third plot shows the net effect of the initial COVID-19 lockdown and the slow recovery in 2021. However, the spatial distribution will need to be compared once again when there is a full recovery, and the number of drivers is restored to pre-pandemic levels, to fully understand the pandemic shift in driver residential locations.

Although the change in fraction was small, it does hint toward differences in the socio-demographic makeup of the driver population. As presented in Figure 7, the areas to the south of Chicago tend to be lower income and higher minority populations (where “minority” refers to all persons except white non-Hispanics) The per capita income and minority population percentages displayed in the figure were taken from the 2014 to 2018 American Community Survey and compiled in the CDC/ATSDR Social Vulnerability Index dataset for the State of Illinois (18). It should be noted that the color scale for the per capita income figure is centered at $32,000, which is the median individual income in Cook County. Census tracts in red, therefore, fall below this median income.

We see that in the initial stages of the COVID-19 pandemic, the fractions of drivers from lower income and high minority areas increased slightly. This indicates a small change in driver demographics, either indicating that drivers with lower incomes were more likely to drive during the pandemic, perhaps to secure additional income and offset any potential losses incurred because of the pandemic, or that those drivers with higher incomes were more likely stop accepting trips out of caution and to reduce their risk of viral exposure. Overall, the correlation between the spatial distribution of drivers, income, and minority status, indicates that the supply of drivers during the pandemic shifted, although slightly, toward drivers with lower incomes and those of minority status.

Methodology

Survival Analysis

Survival analysis is commonly used in epidemiological studies to analyze data where the variable of interest is
time until an event occurs. During the COVID-19 pandemic, the supply of TNC drivers changed, with fewer drivers joining the platform and existing drivers quitting at higher rates. Survival analysis can be used to understand changes in the time until drivers quit before and after the COVID-19 pandemic. A driver’s tenure before they quit will be referred to as the driver’s survival time.

When considering disaggregate data, each individual in the dataset will have a survival time given by:

\[ T = \text{survival time (where } T \geq 0) \]  

From each individual’s survival time, an aggregate curve, the survival function, is then developed. The survival function represents the probability that a driver has survived beyond a specified tenure level, \( t \).

\[ S(t) = P(T > t) \]  

The survival function, which determines the survival probabilities for specific tenure levels, is central to survival analysis because it provides a summary of the dataset \((I2)\). Theoretically, at \( t = 0 \), \( S(t) = 1 \), as no individual should be able to quit at a tenure of 0 and at \( t = \infty \), \( S(t) = 0 \) as all drivers should eventually quit as their tenure increases.

To determine a survival curve from aggregate data, the aggregate number of drivers that quit after each tenure level in the dataset must be constructed. From the original dataset, the tenure of each driver is found as the difference between the driver’s start date and the report month. Thus, the following is obtained:

\[ n_t^m = \text{number of eligible drivers in report month } t \text{ with tenure } m, (m, t) \]

\[ y_t^m = \text{number of eligible drivers at } (m, t) \text{ that quit before } (m + 1, t + 1) \]

such that:

\[ n_{t+1}^{m+1} = n_t^m - y_t^m \]  

Thus, the total number of drivers that were tracked over time is given by:

\[ N = \sum_{t=0}^{T} n_t^0 \]  

From the aggregate data, the number of drivers that quit at each tenure is given by:

\[ q_m = \text{number of drivers at tenure } m \text{ that quit before tenure } m + 1 \]

\[ q(m) = \sum_{t=0}^{T} y_t^m \]  

The probability of quitting at each tenure, \( P(m) \), and the corresponding cumulative probability of quitting, \( C(m) \), is found as follows:

Figure 7. Income and demographic distribution among census tracts in Cook County, IL.
Under this model form, the vector of coefficients can be given by:

\[ P(m) = \frac{d_m}{N}; \quad C(m) = \sum_{m = 0}^{M} P_m \]  

An aggregate survival curve can then be determined by \[ \hat{S}(m) = 1 - C(m). \]

This formulation is equivalent to the non-parametric Kaplan-Meier estimator or product-limit estimator that is given by:

\[ \hat{S}(t) = \prod_{t_i \leq t} \left( 1 - \frac{d_i}{n_i} \right) \]  

where \( d_i \) is the number of drivers that quit at time \( t_i \) and \( n_i \) is the number of drivers known to have driven until time \( t_i \). An advantage of using the Kaplan-Meier estimator is that the formulation accounts for right-censored data which occurs if a driver has not yet quit over the analysis period.

Finally, to understand differences in the tenure distribution of drivers that quit before and after the pandemic, an empirical cumulative distribution curve is built conditioned only on drivers that do quit. If drivers do not quit at a certain tenure, then the same number of eligible drivers is expected at the next report month with a tenure increased by one month. The number of drivers quitting is therefore calculated as the difference between the number of drivers in month \( m \) with tenure \( t \), and those in month \( m + 1 \) with tenure \( t + 1 \).

**Modeling Count of Provided Trips**

The count of trips provided by drivers in a given report month can be modeled as a function of explanatory variables, for the eligible drivers who have provided at least one trip. First, we separate our analysis into pre-pandemic data (report month before March 2020 starting January 2019) and during pandemic data (report month after or on March 2020 until April 2021). Since the dependent variable is discrete and significant heterogeneity is expected across individuals, leading to the conditional variance of the dependent variable being greater than the conditional mean, we use a negative binomial regression, which is formulated as follows (19, 20). Let \( y_{it} \) be the number of trips provided by driver \( i \) during time \( t \), which under a Poisson model has a probability of:

\[ P(y_{it}) = \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!} \]  

where the log of the expected trip frequency \( \ln(\lambda_{it}) = \ln(E[y_{it}]) = \beta X_{it} \), \( X_{it} \) is a vector of exogeneous characteristics of driver \( i \) and time \( t \) capturing systematic variation and \( \beta \) is a vector of the associated coefficients. Under this model form, the vector of coefficients can be estimated using maximum likelihood where the likelihood of the sample is:

\[ L = \prod_{i,t} \frac{e^{-\lambda_{it}} \lambda_{it}^{y_{it}}}{y_{it}!} \]  

To remedy the problem of overdispersion, the assumption of equal mean and variance is relaxed in negative binomial regression, which adds an error term as below:

\[ \ln(\lambda_{it}) = \beta X_{it} + \epsilon_{it} \]  

where \( \exp(\epsilon_{it}) \) is a Gamma distributed error term (with a mean of 1 and variance \( \alpha \)). Under this formulation, the probability of driver \( i \) providing \( y_{it} \) trips during time \( t \) follows a negative binomial distribution:

\[ P(y_{it}) = \frac{\Gamma(\theta + y_{it})}{\Gamma(\theta) y_{it}!} \mu_{it}^{\theta} (1 - \mu_{it})^{y_{it}} \]  

where \( \mu_{it} = \frac{\theta}{\theta + \lambda_{it}} \) and \( \theta = \frac{1}{\alpha} \). This model allows the mean to be different than the variance, such that:

\[ \text{var}(y_{it}) = \lambda_{it} + \alpha \lambda_{it}^2 \]  

Under the negative binomial distribution, the likelihood of the distribution is:

\[ L = \prod_{i,t} \frac{\Gamma(\theta + y_{it})}{\Gamma(\theta) y_{it}!} \left[ \frac{1}{\theta + \lambda_{it}} \right]^{y_{it}} \left[ \frac{\lambda_{it}}{\theta + \lambda_{it}} \right]^\theta \]  

This likelihood is then maximized to estimate values of the coefficients \( \beta \) and \( \alpha \). If the estimated value of \( \alpha \) is significantly different than zero, then the negative binomial model outperforms the Poisson model. Otherwise, a Poisson model is preferred.

**Discussion and Results**

**Empirical Survival Function Results**

Using the survival analysis methodology, four groups of drivers were tracked over a period of 12 months to determine their survival functions (see Figure 8). As mentioned previously, survival analysis considers all eligible drivers, regardless of the number of trips they provide. Group 1 tracks 42,008 drivers that joined as TNC drivers in July, August, and September 2018. Group 2 tracks 39,007 drivers that joined in October, November, and December 2018. Group 3 tracks 36,259 drivers that joined in January, February, and March 2019. Tracking the retention of these driver-groups for 12 months includes data points before the observed change in driver supply caused by COVID-19. Group 4, however, includes 21,438 drivers that started in March, April, and May 2020. This represents the drivers who started during the
These four groups are not mutually exclusive, as drivers could start driving within Group 1, then stop driving and could restart and be counted in another group.

The curves for groups 1, 2, and 3 are similar, indicating a fairly stationary process in the continuation rate of drivers pre-pandemic, despite the potential for varying seasonal effects. These curves show that, after six months of driving, only 30% of drivers were still on TNC apps and only 16% were still eligible after 11 months. For drivers that started in COVID-19, however, only 20% of drivers were still eligible for rides after six months, and 4% after 11 months. Both curves show a shift in the dropout rate around the six-month mark, where drivers that have attained six months of experience were less likely to drop out. Therefore, experience in the first few months could be quite important for driver retention.

The above curves can be compared using a log-rank test, with the null hypothesis being that there are no differences between the probability of quitting between the groups at any time. This test makes no assumptions about the distribution of survival times and is non-parametric as the log-rank test statistic compares the observed survival rates with the expected survival rate given the null hypothesis.

The log-rank statistic is computed as follows:

$$\chi^2 = \sum \left( \frac{O_{jt} - E_{jt}}{E_{jt}} \right)^2$$

which is distributed approximately as chi-squared with one degree of freedom (2).

Given the similarity of the three groups of drivers that started before the pandemic, the log-rank test is conducted between the combined survival curve of all three groups pre-pandemic and the curve representing drivers that started during the pandemic. The results of the log-rank test are displayed in Table 1, below.

The log-rank statistic has degrees of freedom equal to the number of comparison groups minus one. Thus, as

**Table 1. Survival Analysis Log-Rank Test Results**

| Tenure (months) | Quitters (COVID-19) | Quitters (2018–2019) | Expected quitters (COVID-19) | Expected quitters (2018–2019) |
|-----------------|--------------------|---------------------|-----------------------------|-----------------------------|
| 1               | 3,418              | 19,633              | 3,563                       | 19,488                      |
| 2               | 3,093              | 15,863              | 2,953                       | 16,003                      |
| 3               | 2,568              | 13,323              | 2,453                       | 13,438                      |
| 4               | 2,719              | 10,985              | 2,096                       | 11,608                      |
| 5               | 2,446              | 10,309              | 1,832                       | 10,923                      |
| 6               | 2,809              | 10,162              | 1,717                       | 11,254                      |
| 7               | 1,086              | 4,110               | 551                         | 4,645                       |
| 8               | 1,005              | 3,081               | 372                         | 3,714                       |
| 9               | 652                | 3,580               | 302                         | 3,930                       |
| 10              | 416                | 3,416               | 226                         | 3,606                       |
| 11              | 336                | 3,954               | 219                         | 4,071                       |
| Total           | 20,548             | 98,416              | 16,284                      | 102,680                     |

$$\chi^2 = \frac{(20548 - 16284)^2}{16284} + \frac{(98416 - 102680)^2}{102680} = 1293.6$$
the statistic is larger than the critical value of 3.84, there is significant evidence of a statistical difference between the two survival curves.

The survival curve for drivers who started during COVID-19 closely follows the pre-pandemic curves for shorter tenures. The curve represents drivers who started after the pandemic began, therefore they were likely aware of the risks involved in driving passengers. However, the curve strays further from the pre-pandemic curves at longer tenures. This is likely a reflection of the pandemic demand on driver experience. As mentioned previously, lower matching probabilities and longer wait times for matches likely affected the driver supply as the higher ratio of idle time to revenue-generating time on the platforms may have caused drivers who started in the pandemic to drop out at higher rates. The survival analysis has an important implication for managing driver supply. If the driver experience on the application is driving the difference in retention, TNCs could employ tactics such as alerting drivers about which areas to travel to for statistically higher matching probabilities. However, equity and the rider-side fairness of such a strategy must be considered.

Finally, we compare the distribution of drivers that quit during the initial COVID-19 shock to the expected distribution before the pandemic. The supply of drivers accepting trips fell between March and April 2020, so we consider the distribution of tenures for drivers who quit in March 2020 to characterize the pandemic-induced quitting behavior. Figure 9 plots the cumulative probability of quitting at each tenure level given the total population of drivers quitting in the report month. This is an empirical cumulative distribution function that is conditioned on those drivers who quit. Because of the flexible nature of TNC driving, where eligible drivers may choose not to provide rides in certain months, quitting analysis can only be conducted on the number of eligible drivers, regardless of whether they accepted a trip or not.

Compared with March of 2019 and 2021, the curve representing the conditional probability of quitting in March 2020 is shifted. The conditional cumulative probability of a driver quitting at six months or less was around 83% in March of 2019 and 2021, but the same cumulative conditional probability corresponds to a tenure of less than 12 months in March 2020. This indicates that, of the drivers who quit in March 2020, fewer had very short tenures. Specifically, the difference in the probabilities of short tenure is likely the effect of international media coverage of the development of the coronavirus in late 2019 and early 2020 and the World Health Organization’s January declaration that the virus was an international public health emergency. Thus, drivers with short tenures in March 2020 likely had already weighed the risks involved with driving passengers amid the pandemic before deciding to join a TNC platform. Additionally, the effect of the pandemic may have caused a larger percentage of drivers of mid-range tenure to quit than would typically be expected, causing the curve to shift downwards. Finally, the 2021 curve matches well with the 2019 curve, indicating that the driver exit process was recovering after a full year of the COVID-19 pandemic.

**Negative Binomial Regression Results**

**Model Specification.** In this paper, the exogeneous variables available to us are:
Table 2. Sample Statistics for Negative Binomial Regression Datasets

| Variable                     | Type     | Statistic | Pre-COVID | During COVID |
|------------------------------|----------|-----------|-----------|--------------|
| Driver start month           | Date     | Max       | Feb 2020  | May 2021     |
|                              |          | Min       | Aug 2012  | Mar 2013     |
| Report month                 | Date     | Max       | Feb 2020  | May 2021     |
|                              |          | Min       | Jan 2019  | Mar 2020     |
| Lives in Chicago             | Binary   | Min       | 0.00      | 0.00         |
|                              |          | Avg       | 0.48      | 0.50         |
|                              |          | Max       | 1.00      | 1.00         |
|                              |          | Std dev   | 0.50      | 0.50         |
| Driver tenure (years)        | Integer  | Min       | 0.000     | 0.000        |
|                              |          | Avg       | 1.074     | 1.530        |
|                              |          | Max       | 7.417     | 8.000        |
|                              |          | Std dev   | 1.168     | 1.588        |
| On multiple apps             | Binary   | Min       | 0.00      | 0.00         |
|                              |          | Avg       | 0.31      | 0.29         |
|                              |          | Max       | 1.00      | 1.00         |
|                              |          | Std dev   | 0.46      | 0.45         |
| Number of trips provided     | Integer  | Min       | 1.00      | 1.00         |
|                              |          | Avg       | 129.74    | 116.89       |
|                              |          | Max       | 996.00    | 998.00       |
|                              |          | Std dev   | 133.33    | 128.71       |

- Single App: 1 if the driver was on one app, 0 otherwise.
- Tenure: A piecewise linear specification with two breakpoints was specified (shown below).
- Live in Chicago: 1 if the zip code of the driver’s residence is between the zip codes 60600 and 60700, 0 otherwise.
- State: The state in which the driver’s license is registered, a categorical variable. This variable is a proxy for driver ID and it is necessary to include it to control for serial correlation, as the ID of the driver is unknown.
- Quarter and Year: The quarter and year of the report month, a categorical variable. This variable controls for seasonality.

Table 2 outlines sample statistics for the two datasets used in this regression analysis.

The final model specification is:

\[
\begin{align*}
\ln(\lambda_{it}) &= \beta_0 + \beta_1 (\text{Single App}) + \beta_2 (\text{Single App}) \\
&\times (\text{Tenure}_{p_{2i}}) + \beta_3 (\text{Single App}) \times (\text{Tenure}_{p_{3i}}) \\
&+ \beta_4 (1 - \text{Single App}) \times (\text{Tenure}_{p_{2i}}) \\
&+ \beta_5 (1 - \text{Single App}) \times (\text{Tenure}_{p_{3i}}) \\
&+ \beta_6 (\text{Live in Chicago}) + \beta_7 (\text{State}) \\
&+ \beta_8 (\text{Quarter and Year}) + \epsilon_{it}
\end{align*}
\]

where

\[
\begin{align*}
\text{Tenure}_{p_{2i}} &= \begin{cases} 
1, & \text{if Tenure} \leq 4 \\
0, & \text{otherwise}
\end{cases} \\
\text{Tenure}_{p_{3i}} &= \begin{cases} 
1, & \text{if Tenure} > 4 \\
0, & \text{otherwise}
\end{cases}
\end{align*}
\]

\(\beta_0\) to \(\beta_8\) are coefficients to be estimated (\(\beta_7\) and \(\beta_8\) are a vector of coefficients).

With this specification, we study the effect on tenure for drivers registered on a single app compared with multiple app drivers separately. Note that the two models were estimated separately to find differences between those who completed more trips during the pandemic and before it.

Model Estimation Results. The model was estimated in STATA. Table 3 provides the coefficients of the covariates as well as their incidence rate ratios. The incidence rate ratio is the exponent of the coefficient and shows the effect of the explanatory variable in percentage increase/decrease compared with a baseline, except the constant which provides an average predicted number of trips provided when all other explanatory variables are zero. Note that we omit the State coefficients and Quarter and Year coefficients; however, these coefficients were significant at a 5% level, indicating that the model captured serial correlation and seasonality effects.

First, the dispersion parameter in both models was found to be significantly different than zero (p-val < 0.001).
those living in the suburbs did not necessarily provide more trips. This indicates that models have similar findings, showing that Chicago residents provide nearly 52% more trips. This indicates that the newer drivers that joined multiple apps during COVID-19 provided 5.1% fewer trips. This indicates a change in the distribution of drivers. Specifically, it indicates that the newer drivers that joined multiple apps during COVID-19 provided more trips, likely to secure greater earnings. Finally, with respect to living in Chicago, both models have similar findings, showing that Chicago residents provide nearly 52% more trips. This indicates that those living in the suburbs did not necessarily provide more trips.

## Conclusion

This paper characterizes and quantifies the change in TNC driver supply in the Chicago area before and during the COVID-19 pandemic. There was a 71.6% decrease in the number of drivers accepting trips between March and April 2020, and the supply throughout the pandemic closely mirrors the COVID-19 case rate in the city. Spatially, the distribution of drivers accepting trips shifted slightly toward the lower income and higher minority area of the Chicago region. This could be because higher-income drivers could afford to lose the additional income TNC driving provided, while lower-income drivers may have needed to secure additional income to offset any hardships induced by the pandemic.

Overall, in addition to reduced demand for rides affecting the supply of drivers, it is probable that new drivers were reluctant to join TNC apps at the beginning of the pandemic as they tried to adhere to the stay-at-home orders and CDC guidelines. Drivers that did join, however, likely weighed these risks and were therefore less likely to quit at high rates in March 2020 than were drivers of longer tenure. The survival analysis conducted in this paper, however, showed that driver retention after six months was much lower than in reference months. While this could be explained by the increasing number of COVID-19 cases worldwide causing drivers to practice more social distancing, it could also be the effect of the reduced demand for trips creating a less desirable supply-side experience and income throughout the pandemic. The results of the negative binomial regression show that drivers on a single TNC platform generally provided 20% fewer trips than drivers on multiple platforms. This difference increased to 30% post the COVID-19 pandemic. Additionally, the newer drivers that joined multiple apps during COVID-19 provided more trips than did longer-tenure drivers on multiple apps, likely to secure better earnings.

As recovery begins in many cities across the United States, the increase in demand will need to be matched with an increase in driver supply. The characterization of

### Table 3. Negative Binomial Regression Estimation Results

| Coefficient       | During COVID | Pre-COVID |
|-------------------|--------------|-----------|
|                   | Estimate     | Exp (coefficient)  | Estimate     | Exp (coefficient)  |
| Intercept         | 4.495*** (0.010) | 89.595*** (0.939) | 4.757*** (0.004) | 116.431*** (0.506) |
| Single_App        | -0.356*** (0.009) | 0.700*** (0.006) | -0.229*** (0.004) | 0.795*** (0.004) |
| Single_Tenure_p2  | 0.0396*** (0.005) | 1.040*** (0.005) | 0.159*** (0.004) | 1.172*** (0.004) |
| Single_Tenure_p3  | 0.064*** (0.010) | 1.066*** (0.011) | 0.093*** (0.013) | 1.010*** (0.014) |
| Multiple_Tenure_p2| -0.090*** (0.009) | 0.914*** (0.008) | 0.177*** (0.005) | 1.194*** (0.006) |
| Multiple_Tenure_p3| -0.053*** (0.011) | 0.949*** (0.011) | 0.2319*** (0.010) | 1.261*** (0.012) |
| Live_in_Chicago   | 0.424*** (0.004) | 1.529*** (0.006) | 0.4177*** (0.003) | 1.519*** (0.004) |
| Dispersion parameter (alpha) | 1.239*** (0.003) | — | 1.207*** (0.002) | — |

Note: State reference = IL; Year/Quarter reference (during COVID) = 2020/1; Year/Quarter reference (pre-COVID) = 2019/1; Tenure reference = Tenure_p1 (Tenure = 1).

***p-value < 0.001.

Standard error in parentheses.
the supply of drivers in this paper can be used to develop strategies to increase driver retention and experience and ultimately return TNC operation to pre-pandemic levels in the City of Chicago. A key question going forward is whether the attrition curves will return to their pre-pandemic levels, or whether they will remain closer to those reflecting the current situation. Many factors are likely to influence this process, though one could reasonably envision an adjustment process whereby current characteristics provide a lower bound on the survival probability while pre-pandemic conditions constitute the upper bound. Future work exploring the supply-side changes in the rideshare market could include survey methods to qualitatively and quantitatively understand the importance of pandemic, tenure, policy, demand-side, and income-related variables and more completely capture any causal relationships.

Author Contributions
The authors confirm contribution to the paper as follows: study conception and design: S. Hegde, H. Abkarian, H. S. Mahmassani; data collection: S. Hegde, H. Abkarian, H. S. Mahmassani; analysis and interpretation of results: S. Hegde, H. Abkarian, H. S. Mahmassani; draft manuscript preparation: S. Hegde, H. Abkarian, H. S. Mahmassani. All authors reviewed the results and approved the final version of the manuscript.

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

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is based on internal funds provided by the Northwestern Transportation Center.

ORCID iDs
Sharika Hegde https://orcid.org/0000-0002-8822-3844
Hoseb Abkarian https://orcid.org/0000-0003-2681-3002
Hani S. Mahmassani https://orcid.org/0000-0002-8443-8928

References
1. Rochet, J.-C., and J. Tirole. Two-Sided Markets: An Overview. Working Paper. Institut d’conomie industrielle, Toulouse, France, 2004.
2. Wang, X., F. He, H. Yang, and H. O. Gao. Pricing Strategies for a Taxi-Hailing Platform. Transportation Research Part E: Logistics and Transportation Review, Vol. 93, 2016, pp. 212–231.
3. De Stefano, V. The Rise of the Just-in-Time Workforce: On-Demand Work, Crowdwork, and Labor Protection in the Gig-Economy. Comparative Labor Law & Policy Journal, Vol. 37, No. 3, 2016, pp. 471–504.
4. Berliner, R. M., and G. Tal. What Drives Your Drivers: An In-Depth Look at Lyft and Uber Drivers. Institute of Transportation Studies, University of California, Davis, 2018.
5. Hall, J. V., and A. B. Krueger. An Analysis of the Labor Market for Uber’s Driver-Partners in the United States. ILR Review, Vol. 71, No. 3, 2018, pp. 705–732.
6. Ridership Data Portal. WMATA. https://www.wmata.com/initiatives/ridership-portal/. Accessed July 28, 2021.
7. Day-by-Day Ridership Numbers. MTA. https://new.mta.info/coronavirus/ridership. Accessed July 28, 2021.
8. Tyler, M. Funding in Illinois One Year Later. Illinois Economic Policy Institute, La Grange, 2021.
9. Tahliyan, D., and H. S. Mahmassani. Chicago Mobility Under COVID-19. Northwestern Transportation Center, Evanston, IL, 2020.
10. Zheng, H., K. Zhang, and Y. M. Nie. Plunge and Rebound of a Taxi Market Through COVID-19 Lockdown: Lessons Learned From Shenzhen, China. Transportation Research Part A: Policy and Practice, Vol. 150, 2021, pp. 349–366.
11. Ale-Ahmad, H., and H. Mahmassani. Impact of COVID-19 on Taxi Operation in Chicago. Northwestern Transportation Center, Evanston, IL, 2020.
12. Kleinbaum, D. G., and M. Klein. Survival Analysis. Springer, Berlin, Heidelberg, 2010.
13. City of Chicago. Transportation Network Providers - Drivers. https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Driver/jf6w-834c. Accessed July 28, 2021.
14. City of Chicago. Transportation Network Providers - Trips. https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p. Accessed July 28, 2021.
15. City of Chicago. COVID-19 Daily Rolling Average Case, Death, and Hospitalization Rates, City of Chicago, Data Portal. https://data.cityofchicago.org/Health-Human-Services/COVID-19-Daily-Rolling-Average-Case-Death-and-Hosp/e68t-c7fv. Accessed July 28, 2021.
16. World Health Organization. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV). https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov). Accessed July 28, 2021.
17. Pritzker, J. Executive Order 2020-10. Illinois.gov. https://www2.illinois.gov/pages/executive-orders/executiveorder2020-10.aspx. Accessed July 28, 2021.
18. CDC/ATSDR’s Social Vulnerability Index (SVI). https://www.atsdr.cdc.gov/placeandhealth/svi/index.html. Accessed July 28, 2021.
19. Chang, L. Y. Analysis of Freeway Accident Frequencies: Negative Binomial Regression Versus Artificial Neural Network. Safety Science, Vol. 43, No. 8, 2005, pp. 541–557.
20. Venables, W. N., and B. D. Ripley. Modern Applied Statistics With S-PLUS. Springer Science & Business Media, New York, NY, 2013.
21. Bland, J. M., and D. G. Altman. The Logrank Test. BMJ, Vol. 328, No. 7447, 2004, p. 1073.