Did UChicago's "Lyft Ride Smart" Program Significantly Increase Rideshare Usage?

I. Introduction

After a violent year including the loss of two University of Chicago students' lives to gun violence, the University responded with swift policy implementation in an effort to promote safety-conscious methods of transportation in Hyde Park. Beginning on October 1st, 2021, UChicago introduced the "Lyft Ride Smart" program which provides their students with 10 free Lyft rides per month (and up to $15 off per ride), as long as these rides start and end within the bounds of Hyde Park and are called between the hours of 5 pm to 4 am. We are ultimately interested in understanding whether the program meaningfully changed how UChicago students get around in Hyde Park. Our specific research question is testing if the implementation of the "Lyft Ride Smart" program impacted UChicago student decision-making by studying whether or not overall rideshare usage subsequently grew in the Hyde Park area.

We contribute to the existing literature in three distinct ways. First, whereas past studies have broadly examined the impacts of social influence in terms of university carshare programs' success (Li and Zhang, 2020), our study looks to focus on decision-making determined by a program's increasing popularity in conjunction with self-efficacy. Next, prior research has found that for universities unveiling new technologies to encourage ridesharing, "the resulting changes in commuter behavior have been dramatic" (Andrew, Lyons, and McCoy, 2016). We ultimately hope to uncover similar findings for UChicago, which is unique in terms of college campuses because of its modern culture juxtaposed with high crime rates. Finally, *Ride-Sharing, Fatal Crashes, and Crime* describes how "ride-sharing apps, such as Uber, contributed to decreases in crime" because "passengers spend less time standing on the street due to the electronic hailing" (Dillis and Mulholland, 2018). In our study, we have decided to take these findings to suggest that over time, people, as rational actors, tend to adapt to their surroundings and realize that they are less likely to become victims of crime if they choose rideshares as their means of transportation, especially at night in a particularly dangerous area. Thus, we hope to isolate local crime rates as a confounder in order to see whether or not the Lyft program actually increases ride-share usage, independent of crime levels in Hyde Park that are possibly increasing (which would, as suggested, make people more willing to utilize rideshare).

We believe that both students and faculty of the University would be interested in our findings in order to see whether the University's investment in reimbursing students' rideshare costs is worth the long-run costs versus its utility (calculated to provide over nine million dollars in value per year[1]), and is not simply a sunk cost. It is important to note that the University is not compensating students for each individual ride but rather buying expirable ride passes in bulk (UChicago, 2021), further illuminating the need for statistical testing. Additionally, our paper will help further establish decision-making social theory in terms of whether people gradually adopt the program in the long run (implying an increase in its popularity over time). Any findings we uncover will help complement research looking at how decision-making is formed and the general applications of behavioral theory. Finally, in terms of policy implementation, our study will be able to provide support as to whether a Lyft program of this kind is worthwhile (given a source with adequate funding), as well as whether it should be implemented at other universities or workplaces. While research has been conducted on ride-sharing programs in relation to public transportation usage, we can learn whether or not "Lyft Ride Smart" has the potential to increase rideshare usage over time across other university campuses.

[1] 9 months x 10 passes x ~7000 undergraduate students x \$15 off ≈ \$9,000,000 in value per year

In the rest of this research paper, we will be outlining our theoretical framework including the specific theory we hope to test, as well as the data we will be utilizing to analyze rideshare. Next, our empirical framework section will include our choices of variables, estimation strategy, along with the statistical tests we will be using to test our theory. Finally, in the results and discussion, we will be able to present our findings, limitations, and general takeaways regarding the "Lyft Ride Smart" program and its significance in Hyde Park in relation to the greater Chicago area.

II. Theoretical Framework

As outlined in *Intention of Chinese college students to use carsharing: An application of the theory of planned behavior*, the "Theory of Planned Behavior" states that decision-making boils down to "a favorable or unfavorable evaluation of the behavior, perceived social pressure to perform or not perform the behavior, and self-efficacy in relation to the behavior" (Li and Zhang, 2020). Specifically, Leiming Li and Yu Zhang found that social influence is a critical factor in the intention of Chinese students to utilize car-sharing programs. Additionally, they also found that younger generations in particular were keener to take part in car sharing services because they not only found it easier to "accept technological changes" but because they were influenced by the attitudes of their peers. These findings encapsulate the basis for our causal theory, as we try to uncover similar social phenomena in terms of students' adoption of the "Lyft Ride Smart" program.

In terms of our understanding of ride-share program's success in the past, we point to *Ridesharing, Technology, and TDM in University Campus Settings* which speaks about university initiatives in the ridesharing space and how they decrease rates of single-occupancy vehicle usage (Andrew, Lyons, and McCoy, 2016). They attribute these findings to universities utilizing recent technology innovations that make their programs more convenient, as well as to how these universities pair rideshare programs with supportive policies to improve their effectiveness. This finding is imperative in understanding the basis for our causal theory, and how we hope to extend this research into specifically understanding rideshare participation, as well as in grasping the various covariates that may play a role in the resulting ride-share participation in the same time frame as these university programs.

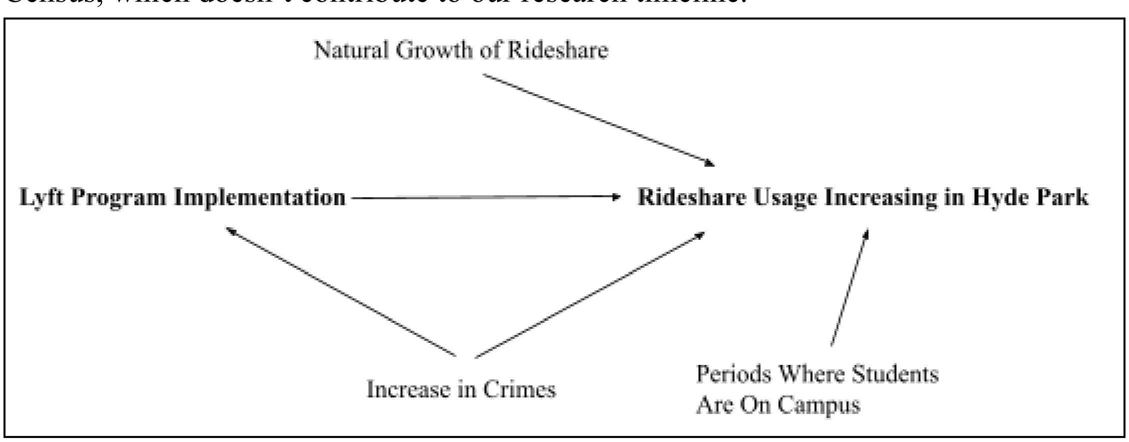
Regarding our causal theory, we will be testing if the implementation of the program impacted UChicago student decision-making, by studying whether or not overall rideshare usage increased in the Hyde Park area. Originally, we wanted to perform an observational study on Hyde Park rideshare data that resembled the "quasi-experimental" study in *The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis* (Campbell and Ross, 1968). The plan was to inspect whether the frequency of rideshare trips significantly differed on the two sides of the date threshold for the program's rollout. We wanted to compare rideshare data in the few months before the program to the months after its release. However, there are two main problems with this design. First, we run into the problem of seasonal confounding. If our "before" and "after" data lies on one continuous strip of the year, the two groups will inevitably lie in different seasons. This is a problem because people are more inclined to use rideshare during the colder winter months, even without the existence of any Lyft program. Since our "after" data would lie in these colder months (after October), we would struggle to attribute any positive ridership change we see to the program. Similarly, by looking only at data points near the program's start date, COVID becomes a serious confounder in the "before" stage because city-wide restrictions were still heavily up in the air at that time and there was no guarantee that the university would operate normally for the academic year. This is less of a problem for the "after" stage since most restrictions had already officially been lifted by then. Ideally, we would like the COVID status to stay constant throughout each side/group of dates that we look at rideshare data for.

Thus, an ideal solution groups the data into "before" and "after" stages that correspond with each other in regards to seasonality and independently maintain a more-or-less constant COVID risk level. We do this by employing a historical date comparison model. In such a design, the "after" group remains the same as described earlier, but the "before" group now becomes an exact mirror of the "after" dates we look at, except that they are exactly 2 years prior. Under such a framework, we declare our independent and dependent variables as follows:

The ***independent variable*** is the number of days since the implementation of the Lyft Rideshare program. Specifically, the first date post-implementation of the program on 10/01/2021 is labeled as 0, the second date 1, and so forth. Additionally, our ***dependent variable*** is the difference between the number of rides taken for every active program day and analogous non-program day (same month and day as the program date, but two years ago). For example, when the independent variable is 0, the corresponding dependent variable looks at the number of rides in Hyde Park taken on 10/01/2021 subtracted by the number of rides taken on 10/01/2019.

Our ***covariates*** relate to the fluctuation of monthly crime counts in Hyde Park, overall rideshare tendencies in Chicago (not just Hyde Park), and student break status (whether students are generally on campus or not for each active program day). First, the fluctuation of monthly crime rates will be accounted for by mapping each program date to the difference in monthly crime count in Hyde Park for that month and the corresponding month exactly 2 years ago (structured analogously to the dependent variable). We ultimately believe that higher crime rates may have an impact on increasing the number of Lyft rides taken, as outlined in the literature developed by Dillis and Mulholland. Next, we will control for the daily rideshare counts in Chicago by applying yet another equivalent “historical difference” process in which we compare city-wide rideshare data on each program date to an analogous date 2 years ago. This allows us to control for potential growth/decline in Lyft rideshare throughout the whole Chicago region that is not caused by the “Lyft Ride Smart" program, but still may manifest itself within the Hyde Park data. Finally, student break status will be a binary variable coded for each of our IV dates with a 0 aligning with students being on campus (no break), and 1 indicating that students are on seasonal break. We believe that this covariate will have an impact on ridership count as it directly correlates with the number of students within our studied population that are present on campus to potentially utilize the program.

While we initially also wanted to control for the growth of the student population at UChicago, we realized it wouldn’t be informative since we only have access to such data on a yearly basis. Similarly, we wanted to control for the overall Hyde Park population as a potentially confounding variable, but the only data available is collected every ten years by the U.S. Census, which doesn’t contribute to our research timeline.

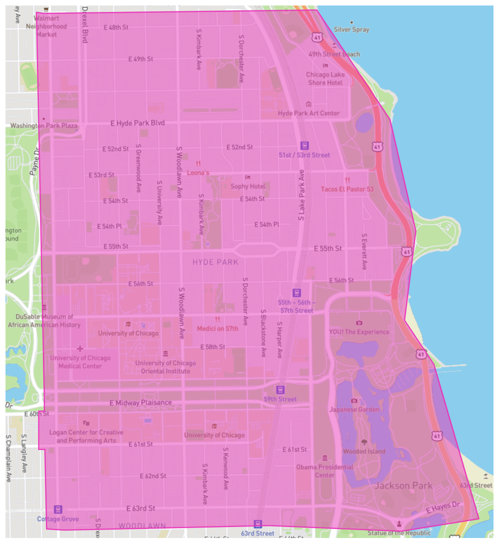


DAG (Directed Acyclic Graph) for our research study where causal relationships are represented by arrows between the variables, pointing from cause to effect.

Our theory is inherently unique in execution in that it distinctly combines the sub-findings of a number of previously developed research papers and then applies these findings to a particular case in the real world. Not only does our theory combine elements from the “Theory of Planned Behavior” by understanding how student decision-making is altered, but it takes a very specific business model in the “Lyft Ride Smart” program and looks at its application in an area as unique as Hyde Park. Additionally, *Ridesharing, Technology, and TDM in University Campus Settings* focused primarily on contrasting rideshare programs with public transportation, but did not dive into the reasoning as to why university programs had the success they did. Because of this and the fact that our theory looks to not only control for the natural growth of Lyft users but also unique covariates such as whether increases in local crime had an impact on Lyft rideshare, our theory can be considered novel.

III. Data and Descriptive Statistics

To source data about rideshare trips, the city of Chicago released a public CSV of individual rideshare rides taken from November 2018 to the present. This is a trustworthy source because rideshare companies cannot refuse a city’s decision to publicize their rideshare data after a Supreme Court case determined that Uber and Lyft trip data could not be considered a trade secret when they tried to withhold it from the city of Seattle (Gutman, 2018). This dataset is cross-sectional where every row in the dataset corresponds to a single rideshare ride taken in Chicago and contains pickup and dropoff coordinates, which we will use in conjunction with a Python script to only filter for data that start and end in Hyde Park. Looking at Google maps, the boundary we chose for Hyde Park is a rectangle outlined by the coordinates (46.781, -87.606) and (41.803, -87.581). To formulate a time series design, we will accumulate the dataset so that each row represents a full day of rideshare trips (tracking the aggregated total number of rides per day, instead of just a single trip). We will also use the original rideshare dataset (pertaining to the entire city of Chicago) to add a column in our final dataset that details the total trip count throughout all of Chicago for each day, in addition to tracking the number of rides in Hyde Park each day.



Hyde Park Map where the pink area represents eligible pick-up and drop-off locations for "Lyft Ride Smart" program usage (UChicago, 2021)

Next, the data we used to track crime rates in Hyde Park is derived from incident reports published by the UCPD Daily Incident Report Log. Every weekday, the UChicago Police Department posts daily crime incidents and fire incidents that were reported to the UCPD over the past 24 hours. Their patrol area includes the area between 37th and 64th streets and Cottage Grove Avenue to Lake Shore Drive. We believe the data to be trustworthy as UChicago Safety & Security is required to provide such data to the public, and a public database is available at the local police office (which we were not permitted to obtain a local copy of). We were able to manually web-scrape the data from a web-based HTML format on the UCPD website, so this was not an issue.

The structure of the crime data is cross-sectional in nature, in that it looks at every incident report, in time series, ordered by date. Within this scraped dataset, each row represents a single reported crime, along with its corresponding date and incident description. We will accumulate the number of total crimes reported for each month, since there is only ever a low single-digit number of reported crimes per day. Now, each row corresponds to a month/year combination and its respective total crime count. From here, we take the differences in the monthly crime counts in a similar fashion to our dependent variable for each active program month (monthly crime count in Hyde Park for that month subtracted by the corresponding crime count for that month 2 years ago). To merge this with our rideshare data (where each row represents a single day instead of a month), we will append to each row the corresponding monthly crime count difference for that day.

For the last few covariates, we will append the following information to each row of our global dataset. First, a boolean variable (0/1) which we will manually append to each row to describe whether or not students are on official holiday break for that program date. Next, in a similar fashion to our Lyft rideshare dataset for Hyde Park, we take the total number of rides in Chicago as a whole for each date post-program implementation and subtract it by the number of rides on that date two years ago. This allows us to control for overall growth in Chicago while maintaining a dataset that is consistent with the dependent variable's estimation strategy. Collectively, this yields a global dataset that technically only includes rows for dates after the program's rollout but still sources information from further back in time and maps it to these dates using the historical difference technique. Our overall data sample is inherently unique in that it mainly represents more educated and technologically adept college students.

Summary Statistics (R Data Output):

IV (Numerical)

- **Days since the program started on 10/01/2021 (0, 1, …)**
    - *Min: 0, Mean/Median: 90.50, Max: 181*
- The maximum value suggests that we have 182 individual data points/rows in our study, which implies that the Lyft program has been running for 182 days (City of Chicago rideshare data only exists up to March 31, 2022).

DV (Numerical):

- **Historical daily ride count difference in Hyde Park**
    - *Median: -2057.0, Mean: -2015.6*
      Min: -8251.0, 1st Q: -3706.0, 3rd Q: -457.2, Max: 6535.0
- Every summary statistic (except for the maximum) is negative. This implies that the ride counts corresponding to active program days were generally lower than what they were on each corresponding date 2 years ago. On average, there were about 2000 fewer rides in Hyde Park during each program day (when compared to the corresponding historic dates).

Covariates:

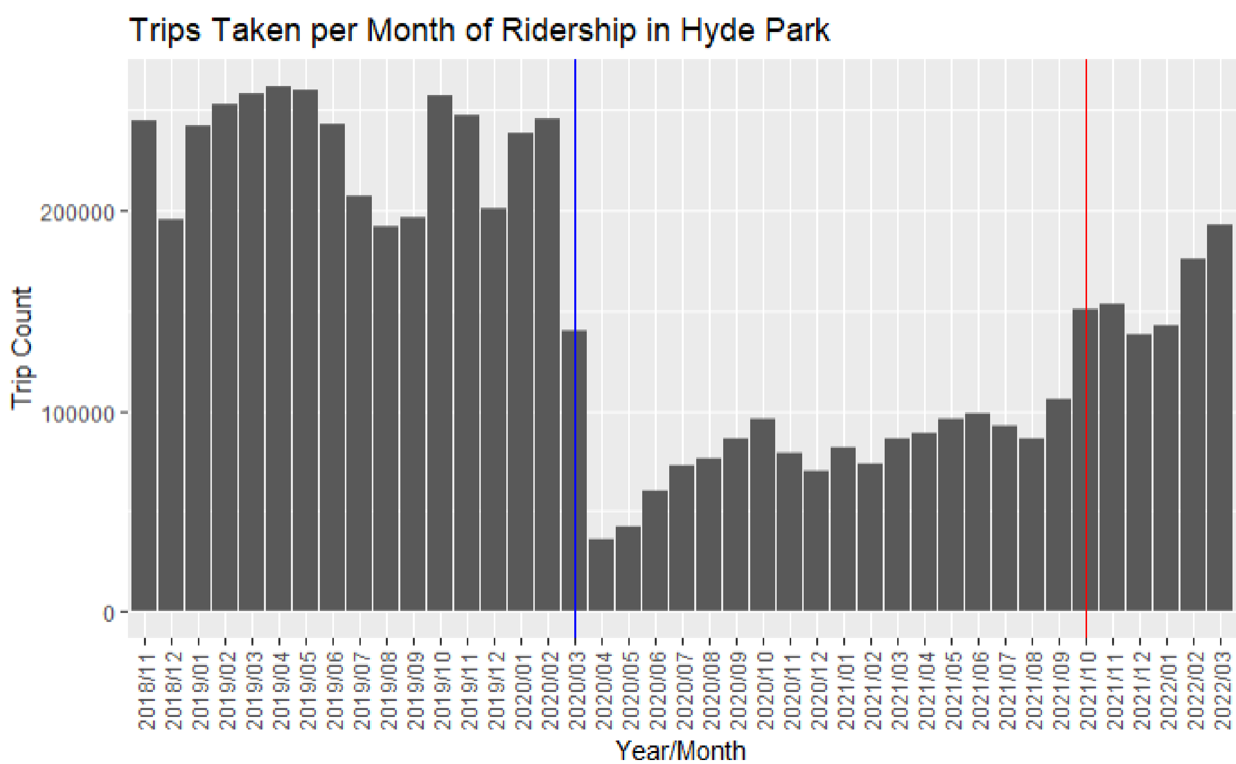
- **Historical daily ride count difference in Chicago as a whole (Numerical)**
    - *Median: -112207, Mean: -111916*
      Min: -327670, 1st Q: -176841, 3rd Q: -59647, Max: 208917
    - A similar picture is painted here as was with the Hyde Park ridership data. The majority of program days saw a decrease in ride count when compared to their historic dates. On average, the ride count for a program day was about 112,000 less than the ride count for that same date 2 years ago.

- **Historical monthly crime count difference in Hyde Park (Numerical)**
    - *Min: -6, 1st Q: 15.75, Median: 30.5, 3rd Q: 52.5, Max: 69*
    - Every summary statistic (except for the minimum) is positive, implying that crime has generally gone up in Hyde Park over the past 2 years (it may also be the case that crimes are being reported/logged more frequently). On average, a program day within the past year corresponded to a month with about 30 more crimes than were reported during that month 2 years ago. However, there was still a single recent month that had 6 fewer crimes reported than were reported 2 years ago.

- **Students on break during Program Days? (Categorical 0/1)**
    - *Mean: 0.2637, Mode: 0*
      Frequencies: 134 days off break, 48 days on break
    - The mean suggests that out of all days that the Lyft program has been active, students have been on break (and probably away from campus) for around ~26% of the days. The mode suggests that students are not on break (0) more often than they are on break (1). The 182 data points further indicate that our data spans 182 days of the Lyft program's operations.

Descriptive Figures:

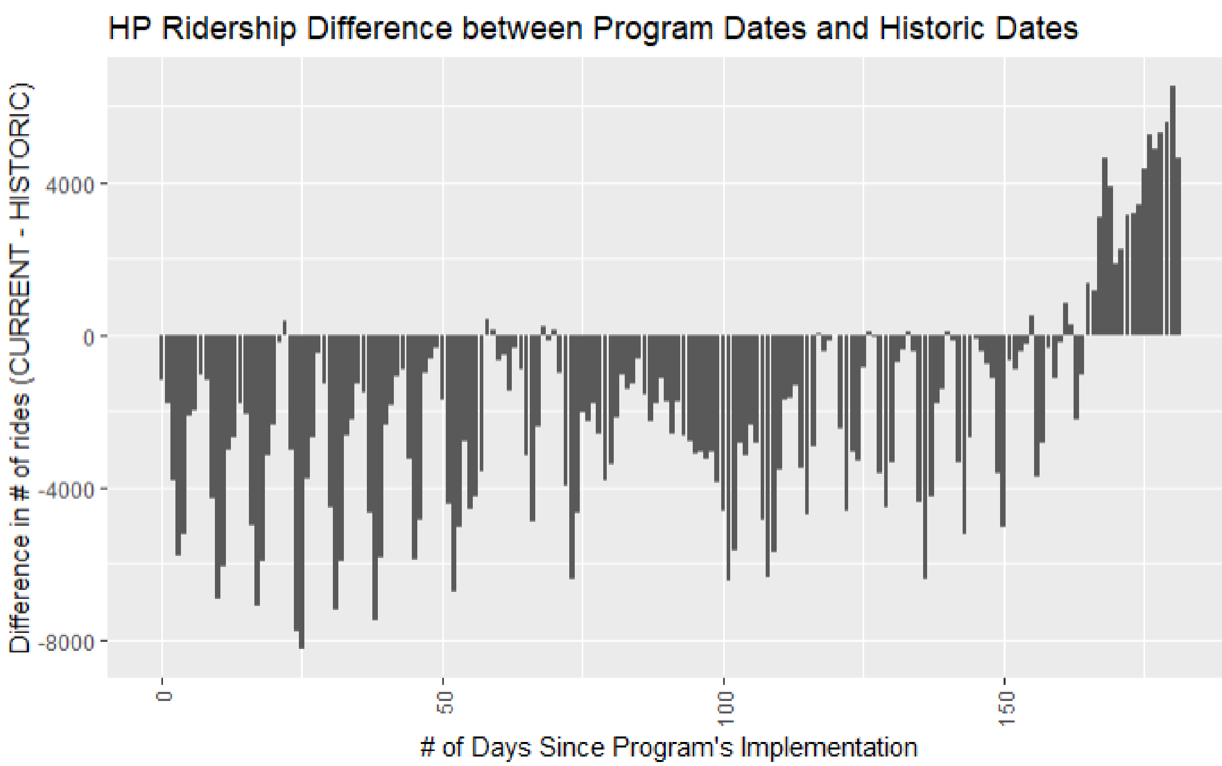
The first bar plot represents a more intuitive look at the kind of ridership data that we have access to. This plot summarizes monthly trip counts in Hyde Park across the entire time period that we have official city data for (as a function of time) and resembles the naive design that we described in the theoretical framework but ended up stepping away from. Thus, these are not the actual variables we inspect for our study, but they are good at conveying the overall ridership trends at a surface level. It is apparent that the start of COVID (3/2020) results in a stark decline in the number of trips taken per month as represented by the blue line below. However, we can see that at the start of the Lyft Ride Smart program (10/2021), the number of trips appears to increase towards pre-COVID normalcy at a more rapid rate.



Blue Line: Start of COVID | Red Line: Start of "Lyft Ride Smart" Program

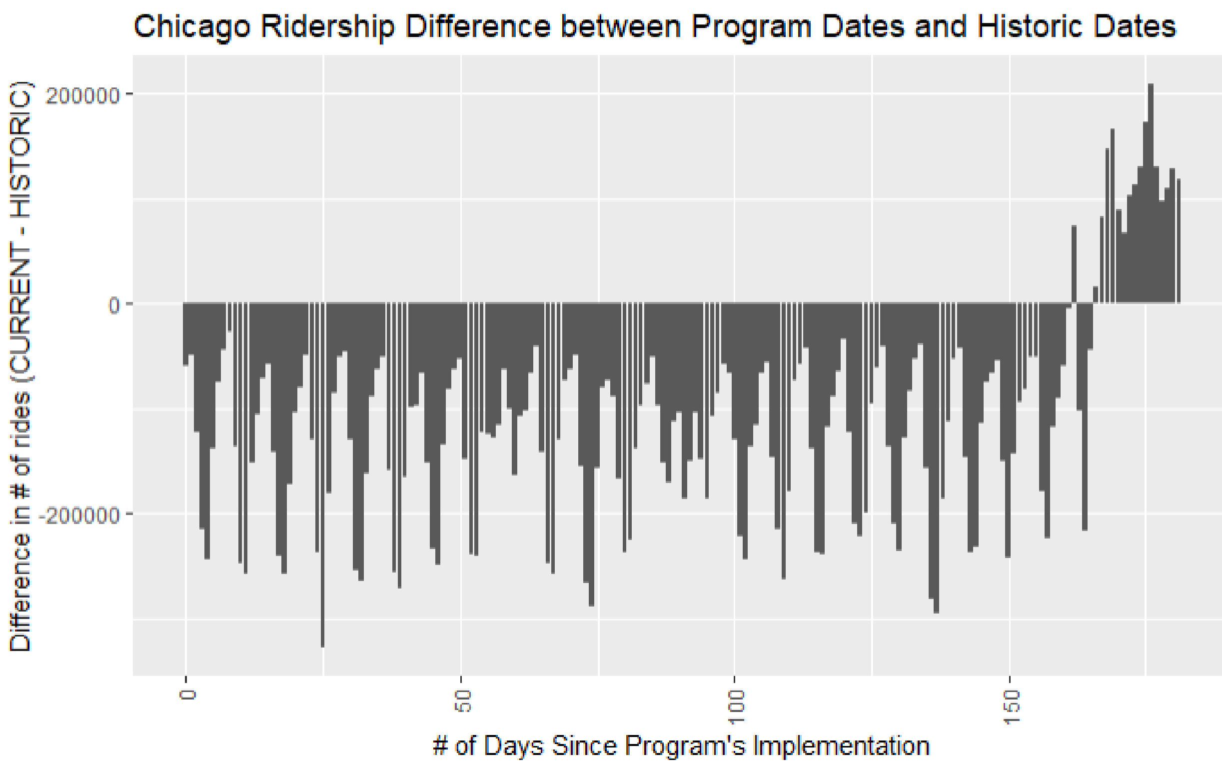
Trips Taken per Month of Ridership in Hyde Park

Our second bar plot represents how the historic difference in daily ride count (DV) changes as the Lyft program's implementation runs for additional days (IV). Specifically, the value at x = 0 represents the number of rides in Hyde Park on the first day of implementation subtracted by the number of rides on the corresponding date two years ago. The number of days since the program's implementation maps directly to the date range 10/1/2021 to 3/31/2022. For the first ~150 days of the program, we can see that the number of pre-implementation rides generally exceeds the number of post-implementation rides.



(Second bar plot: relation of IV and DV)

However, on the far right side of our graph, we can see that the number of post-implementation rides finally eclipses the number of rides on that date two years ago and maintains a positive difference up until our data stops. This is because we were not entirely able to avoid COVID in the historic data group, but it is the best we can do given our current rideshare data source. If we were to use dates 1 year ago for the historical counts, the COVID status would become even more of a confounding factor since it fluctuated much more during that time period. If we were to use dates 3 years ago, we would end up losing a month of data since the City of Chicago rideshare data starts only from November 2018 and we would want to track rideshare differences starting from October 2018 (3 years ago from the start of the program). It is also important to note that we do account for this COVID-19 spike through our addition of overall ridership difference in Chicago per day as a covariate (which was drastically lowered during COVID-19, just as it was in Hyde Park). The third bar plot shows that the trends of the overall Chicago ridership difference covariate mimic the Hyde Park data in this regard.



(Third bar plot: relation of IV and Chicago Ride Difference covariate)

IV. Empirical Framework

Our research design consists of running a multivariate regression in which we regress our DV over the IV and three defined covariates. The model will have the following slope coefficients (sample regression equation) and hypotheses:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \varepsilon$$

Y = Historical ride count difference in Hyde Park
x_1, x_2, x_3, x_4 = Days since program release, Ride count difference in Chicago, Crime count difference in Hyde Park, student break status (0/1) per program date

H_0: *Having more days of program implementation does not impact the historic ridership difference of current daily count - "historic count" ($b_1 = 0$)*
H_a: *Having more days of program implementation significantly impacts the historic ridership difference of current daily count - "historic count" ($b_1 \neq 0$)*

Our regression predicts [*current year ride count - 2 years ago count*] in Hyde Park as a function of our chosen variables. The b_1 coefficient signifies the expected change in the HP ride count difference for an additional unit of time (one extra program day) and is the main focus of our study.

If our proposed theory is correct, we should see a statistically significant positive b_1 coefficient which means that as the days go on, the "historic differences" tend to get larger and larger. The testable implication of this theory would then lead us to believe that the Lyft program is seeing progressively more usage with the passage of time, to the extent that it makes a noticeable impact on rideshare counts throughout the past year that trumps the natural city-wide growth of rideshare. These results would ideally still hold even when accounting for growth in the Hyde Park population, rate of violence in Hyde Park, and the overall growth of rideshare in

Chicago. Yet, we still test for a non-zero b_1 in our t-test because it is also plausible that the program's spread slows down over time as the initial excitement dies down.

Causality is not guaranteed by this strategy. Specifically, our strategy relies on as-if random variation conditional on observational data and on a set of controls, where we effectively believe that any variation that does not correspond to the program can be controlled by our chosen covariates. This form of controlled regression to isolate the effects of the "Lyft Ride Smart" program is not perfect. Because of the wide-reaching scope of our research, usage of observational data, and internal/external validity concerns that we outline in our conclusion, we cannot necessarily interpret our point estimate as a causal effect of the independent variable on the dependent variable via as-if random variation.

However, the results of our analysis are valuable in that they point us in the direction of what a causal study should look like. Our study also allows us to gain a baseline understanding of the effects of the "Lyft Ride Smart" program from an estimation perspective. Our coefficients will hold significance in the confines of the data we have utilized. To that extent, our test gets us closer to an accurate estimate of a causal effect than we were before, as we are ultimately testing a combination of theories in an entirely new setting. Our research is the first, to our knowledge, being conducted on the effects of UChicago's Lyft Program, and ultimately combines literature looking at its usage in conjunction with the "Theory of Planned Behavior", crime rates, and previous literature on University-funded rideshare initiatives.

V. Results

Our 4 regression models, and their corresponding coefficients and significance levels

	Model 1	Model 2	Model 3	Model 4
(Intercept)	−4474.440	−460.391	−1088.752	−921.053
	[−5182.841, −3766.039]	[−883.715, −37.067]	[−1781.684, −395.820]	[−1640.209, −201.898]
	p = 0.000***	p = 0.033*	p = 0.002**	p = 0.012*
	s.e. = 359.006	s.e. = 214.525	s.e. = 351.139	s.e. = 364.414
Days_Since_Start	27.170	9.905	11.969	11.737
	[19.530, 34.810]	[6.701, 13.109]	[8.292, 15.647]	[8.105, 15.370]
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
	s.e. = 3.872	s.e. = 1.624	s.e. = 1.864	s.e. = 1.841
‘Chicago Ride Difference‘		0.022	0.021	0.021
		[0.020, 0.024]	[0.019, 0.023]	[0.019, 0.023]
		p = 0.000***	p = 0.000***	p = 0.000***
		s.e. = 0.001	s.e. = 0.001	s.e. = 0.001
‘HP Crime Difference‘			10.767	9.609
			[2.108, 19.426]	[0.906, 18.312]
			p = 0.015*	p = 0.031*
			s.e. = 4.388	s.e. = 4.410
Student_Break				−288.983
				[−683.401, 105.435]
				p = 0.150
				s.e. = 199.861
Num.Obs.	182	182	182	182
R2	0.270	0.818	0.825	0.827
R2 Adj.	0.266	0.816	0.822	0.823
Std.Errors	HC2	HC2	HC2	HC2

The independent variable (elapsed days since the program’s start) proves to be significant across all stages of our model. In each case, the corresponding p-value is less than 0.0005, which is statistically significant at the most rigorous threshold of 0.001 (***). In Model 1, we can see that without any covariates, it is estimated that each additional day of the Lyft program correlates with an increase in the historical ridership difference by 27.170 rides. In Model 4, with all covariates included, this per-unit increase (in the difference) falls to 11.737 rides. The other models simply serve as “in-between” points that illustrate the behavior of our independent variable’s coefficient as the covariates are added one by one.

These results imply that Hyde Park’s ridership numbers stray further and further from their historical/expected values throughout the selected time period, even when controlling for local crime tendencies, broader rideshare trends in Chicago, and student break status.

Equivalent regression to Model 4, but without March data

(Intercept)	−935.598
	[−1766.193, −105.003]
	p = 0.028*
	s.e. = 420.268
Days_Since_Start	12.458
	[7.245, 17.671]
	p = 0.000***
	s.e. = 2.638
‘HP Crime Difference‘	11.229
	[0.719, 21.739]
	p = 0.036 *
	s.e. = 5.318
Student_Break	−137.965
	[−581.933, 306.003]
	p = 0.540
	s.e. = 224.641
‘Chicago Ride Difference‘	0.022
	[0.020, 0.025]
	p = 0.000***
	s.e. = 0.001
Num.Obs.	151
R2	0.682
R2 Adj.	0.673
Std.Errors	HC2

Next, we attempt to entirely remove COVID as a possible confounder. Although this was our original goal in applying the historical difference model, we were unable to assemble a group of historical dates which maintained a constant COVID status throughout. In the 2-year model that we ran, our previous barplot[2] makes it clear that the differences suddenly become positive after being almost entirely negative for the first 151 days. The reason for this is that March 2020 (2 years ago) was the onset of COVID, so the historical ride counts for this month are way lower than they should have been, leading to a suddenly larger historical difference for that time period.

Therefore, as a compromise, we run another regression (with 2-year historical differences) that does not include the month of March. Our sample size falls from 182 to 151 days, but it is clear that the independent variable’s coefficient is still statistically significant at the most rigorous threshold of 0.001 (***) with a p-value < 0.0005. It is estimated that each additional day of the Lyft program correlates with an increase in the historical ridership difference by 12.458 extra rides.

[2] Referencing ‘Trips Taken per Month of Ridership in Hyde Park’ bar plot from ‘Data and Descriptive Statistics’

Furthermore, when comparing the two bar plots (that correlate our IV and DV) for Hyde Park and Chicago, it appears that the Hyde Park ride count difference data moves approximately in line with the Chicago data for the first few months of the program. To formally test this, we run another regression that only includes the first 2 months of the program's rollout. Note that the 'Crime Difference' covariate becomes highly insignificant in this regression simply because it is a monthly count, so we now effectively only observe 2 different values for it throughout the entire sample size of 61 days.

Equivalent regression to Model 4, but only for the first 2 program months

(Intercept)	−542.908
	[−2623.926, 1538.111]
	p = 0.603
	s.e. = 1038.826
Days_Since_Start	38.081
	[1.236, 74.926]
	p = 0.043*
	s.e. = 18.393
'HP Crime Difference'	2.819
	[−21.755, 27.393]
	p = 0.819
	s.e. = 12.267
Student_Break	−1549.242
	[−2492.506, −605.979]
	p = 0.002**
	s.e. = 470.869
'Chicago Ride Difference'	0.027
	[0.024, 0.030]
	p = 0.000***
	s.e. = 0.001
Num.Obs.	61
R2	0.833
R2 Adj.	0.821
Std.Errors	HC2

This is the first time that we see our independent variable's coefficient significant at only the 0.05 threshold (*). For all other regressions, it was significant with a p-value less than 0.001 (***). This implies that our original intuition is correct, and it is the later months of the program that lead to the stronger statistical significance that we identify in the historic ride count difference over the course of 182 days.

It is important to observe that we have essentially removed all the winter months from our regression in this stage. Thus, we interpret this particular result as implying that some combination of the colder winter months and word about the program being spread over time finally incentivized students to start using their free rides in a more significant way. This conclusion very much falls in line with the “self-efficacy in relation to the behavior” outlined by the “Theory of Planned Behavior”, where students’ decision-making is altered by a combination of social pressure and stronger actualization of their own needs.

We attempt to attribute the correlation between our independent and dependent variables to the rollout of UChicago’s “Lyft Ride Smart” Program by selectively looking at different periods of the observed 182 days and controlling for covariates all throughout. In regards to our research question, it would appear that the “Lyft Ride Smart” had a significant impact on altering student behavior over time, particularly in the later months. However, there still exist threats to the internal validity of drawing such a conclusion and ultimately determining causality.

VI. Discussion and Conclusion

After our tests, we have effectively seen that HP ridership growth outpaces Chicago growth throughout the duration of the program. While we cannot explicitly prove causality, we find significance in our estimation in terms of Lyft Ride Smart’s implementation’s effect on ridership growth. Specifically, we can see that when compared to the number of rides on the same date two years ago, there is a general increase in the ridership differential in Hyde Park following the implementation of UChicago’s Lyft program. These results are significant even when controlling for student breaks, crime rates in the area, and the overall growth of rideshare in Chicago as a whole. It appears that the Lyft program had a significant impact on altering student behavior, and with its increasing popularity, more students decided to opt in and utilize the program to its full potential.

Our findings ultimately correlate with the findings of the aforementioned literature. Specifically, based on our controls, the program appears to have significance in impacting the resulting ride share in Hyde Park. While the Dills and Mulholland literature implied that increased crime leads to increased rideshare usage, we were able to find that rideshare in Hyde Park still grew regardless of fluctuating crime levels. Additionally, our conclusions are able to add to those in Andrew, Lyons, and McCoy's literature that analyzed six universities and their success with rideshare programs. We are able to find similar success in popularity with the Lyft program, generalized to a campus that is unique in its urban density. Finally, our largest takeaway is derived from how the study is partially attributed to the "Theory of Planned Behavior" by Zhang and Li. We were ultimately able to find that the resulting social influence of the program is caused by its increase in popularity and usage by other college students over time.

In terms of the real-world significance of our research, our study ultimately points to the implications of social choice and how when presented with a program that improves self-efficacy and is increasing in popularity among one's peers, an individual is likely to have a shift in behavior and adapt to said program. In addition, our study is able to speak to the benefits of UChicago's program overall and attest to its reward vs. cost (in terms of the University budget). Finally, because of the success of the program that we found in our estimation, we ultimately recommend this program to other campuses or companies that are looking to add similar initiatives to promote safety and accessible transportation.

In terms of limitations, we outline a number of threats to both internal and external validity. In terms of external validity, the levels of violence within Hyde Park may systematically alter the usage of rideshares within the area, making it difficult to generalize these findings to other universities or cities. In addition, educated students are potentially more tech-savvy and may use the rideshare program at a significantly higher rate than other populations. Finally, Hyde

Park inhabitants across the board may utilize ridesharing programs more frequently because of Chicago's frigid temperatures, making the applicability of our findings even more difficult.

Regarding limitations based on measurement or data collection, the data we collected on rideshare usage in Chicago does not distinguish between rideshare services. Because of this, we have no way of knowing whether or not the rides we are plotting are Uber vs. Lyft (or even other ride-share apps like Curb). While we assumed that Lyft dominated the market share for this research study, there is a lack of confidence in this domain of the data collection process. It is possible that Lyft does not occupy enough market share in Hyde Park to significantly impact overall rideshare data. In a similar vein, students might also not make up enough of the rideshare usage in Hyde Park to make a significant impact.

One primary threat to internal validity is the way in which we compare historic dates. In our graphical visualizations of historic differences, it is clear that there is some degree of periodicity in the difference values (a dip followed by a rise every week or so). This is because the historic dates do not align with program dates on their days of the week. For example, 10/1/2021 falls on a Friday, while 10/1/2019 falls on a Tuesday. Intuitively, there will be more rides taken on any weekend (or Friday) than any nearby weekday. Thus, the individual differences are majorly affected by a discrepancy in the day of the week, rather than the overall change in ridership over the course of 2 years. To strengthen our study, we should compare historic days of the week instead of dates (i.e. first Friday of October in 2021 is compared with the first Friday of October in 2019). Yet, there is still an argument to be made that the overall trends of the differences over the course of 182 days should be unaffected by this discrepancy in the long run. The regression should still hold value because it can capture the trend in the differences between each of the 7-day periodicities instead of over the individual dates. The comparisons on a week-by-week basis should still be valid because the dates in each individual week end up being misaligned in exactly the same way. However, this requires a lot of assumptions and the results would certainly hold more weight if we implemented the described change.

Next, there are several other more concrete steps we could take to strengthen the internal validity of our results. For one, we could pinpoint a more accurate ridership comparison group within the Chicago area. Currently, we compare Hyde Park ridership to overall Chicago ridership to control for overall rideshare trends within the city. However, Hyde Park is a very unique region given the gentrification that has ensued in the nearby area due to the university, contrasting heavily with more dangerous surrounding South Side neighborhoods. There are a number of factors that differentiate Hyde Park from the average Chicago suburb, and thus it proves ever-so difficult to attribute the difference in ridership trends that we see in our results to just the program itself.

Ideally, instead of comparing to Chicago ridership as a whole, we find another college-centric Chicago neighborhood that is demographically similar to Hyde Park (but has no such Lyft program) to run our rider count control on. There are several universities near the heart of the city, but these urban neighborhoods are vastly different from the suburban setting of Hyde Park. A prime neighborhood choice would be Evanston, which is where Northwestern University is located. Yet, there are still some problems with this control. For one, Northwestern also has a partnership with Lyft, although it is nowhere near as extensive as UChicago's program. In their scenario, Northwestern can choose to sponsor transportation for certain events, visits, research, or extracurriculars on a case-by-case basis. Additionally, Evanston may serve as a poor control since the region has significantly less crime than the South Side. Since crime plays a big factor in our decision to analyze the usage of the UChicago program in the first place, we ultimately elected to run our rider count control on the entirety of Chicago, as we believe that it has more comparable overall crime demographics (and no Lyft partnership of any kind).

We believe that the outcomes of this project generate new research ideas that social scientists should pursue. Specifically, in the future, rideshare programs can serve as a proxy to understand student decision-making, as well as the application of the “Theory of Planned Behavior” in the real world. Future studies can look to build on our research and the results we found by using additional controls as they see fit, and more nuanced data, in an effort to reach causality. In addition, other causal approaches can be studied, such as the example with Northwestern, or even other universities, comparing UChicago’s unique Lyft program to other schools with or without similar functionalities. Finally, in our study, we ran into difficulty with trying to account for COVID-19 which drastically reduced rideshare. Social scientists in the future can pursue an advancement of our research as more transportation data will be available that is not as actively hindered by COVID-19.

VII. References

Dillis, Angela K, and Sean Eric Mulholland. "Ride-Sharing, Fatal Crashes, and Crime - Research Gate," January 2018. https://www.researchgate.net/publication/322664017_Ride-Sharing_Fatal_Crashes_and_Crime.

"Getting Around." Lyft Ride Smart at UChicago | Department of Safety & Security | The University of Chicago, 2021. https://safety-security.uchicago.edu/services/lyft_ride_smart_at_uchicago/.

Gutman, David. "Uber and Lyft May Have to Disclose Seattle Data They Claim Secret, Supreme Court Rules." The Seattle Times. The Seattle Times Company, February 20, 2019. https://www.seattletimes.com/seattle-news/transportation/uber-and-lyft-may-have-to-disclose-data-they-claim-secret-supreme-court-rules/.

McCoy, Kevin, James Andrew, William Lyons, and John A. Volpe National Transportation Systems Center (U.S.). "Ridesharing, Technology, and TDM in University Campus Settings : Lessons for State, Regional, and Local Agencies." Rosa P. United States. Federal Highway Administration. Office of Planning, May 1, 2016. https://rosap.ntl.bts.gov/view/dot/12297.

"The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis" - JSTOR." Accessed June 1, 2022. https://www.jstor.org/stable/3052794.

Zhang, Yu, and Leiming Li. "Intention of Chinese College Students to Use Carsharing: An Application of the Theory of Planned Behavior." Transportation Research Part F: Traffic Psychology and Behavior. Pergamon, November 2, 2020. https://www.sciencedirect.com/science/article/abs/pii/S1369847820305441.