THE COST OF CONVENIENCE:
RIDEHAILING AND TRAFFIC FATALITIES

John M. Barrios
Yael Hochberg
Hanyi Yi

The Cost of Convenience: Ridehailing and Traffic Fatalities
John M. Barrios, Yael Hochberg, and Hanyi Yi
NBER Working Paper No. 26783
February 2020
JEL No. I00,O3,R4

ABSTRACT

We examine the effect of the introduction of ridehailing in U.S. cities on fatal traffic accidents. The arrival of ridehailing is associated with an increase of approximately 3% in the number of fatalities and fatal accidents, for both vehicle occupants and pedestrians. The effects persist when controlling for proxies for smartphone adoption patterns. Consistent with ridehailing increasing congestion and road usage, we find that introduction is associated with an increase in arterial vehicle miles traveled, excess gas consumption, and annual hours of delay in traffic. On the extensive margin, ridehailing's arrival is also associated with an increase in new car registrations. These effects are higher in cities with prior higher use of public transportation and carpools, consistent with a substitution effect, and in larger cities. These effects persist over time. Back-of-the-envelope estimates of the annual cost in human lives range from $5.33B to $13.24B.

John M. Barrios
The University of Chicago
Booth School of Business
5807 S Woodlawn Ave
Chicago, IL 60637
john.barrios@chicagobooth.edu

Yael Hochberg
Jones Graduate School of Business
Rice University
6100 Main Street, MS-531
McNair Hall Room 331
Houston, TX 77005
and also NBER
hochberg@rice.edu

Hanyi Yi
Rice University
6100 Main Street
Houston, TX 77005
hy25@rice.edu

1. INTRODUCTION

The introduction of ridehailing services, such as Uber and Lyft, has fundamentally changed how many individuals are transported in cities and towns across the United States. While the ability to easily hail a ride through a smartphone app has undoubtedly increased convenience for people seeking transportation and flexible employment, critics increasingly argue that ridehailing creates offsetting negative effects, such as increases in traffic congestion and car-exhaust pollution. Are there, in fact, significant externalities that accompany ridehailing? In this paper, we present evidence suggesting that such costs exist, are not trivial, and can be measured in human lives—specifically, in increased rates of major traffic accidents and traffic fatalities. Using the staggered introduction of ridehailing across U.S. cities, we show that its introduction in a metropolitan area leads to an economically meaningful increase in overall motor vehicle fatalities. This increase is consistent with acknowledged macro trends in motor vehicle accidents and pedestrian fatalities, both of which had been falling steeply in the United States over the period 1985 to 2010, when ridehailing first launched, and have since reversed course (Figure 1).[1]

Whether ridehailing (RH from here on) should lead to higher accident rates is not apparent at first glance. A naïve view of the effects of RH sees it as removing drivers who would have driven themselves with their cars and replacing them with RH drivers for the same mileage. Under this view, RH substitutes self-drivers with RH drivers on a one-to-one basis. Moreover, one might also argue that many of the users who are substituting to being driven are often doing so because they are (or will be) inebriated or otherwise impaired. This substitution of impaired drivers with sober RH drivers thus potentially increases the quality of driving, while (in theory) holding car utilization fixed. Under this view, there will be no increase in the vehicle miles traveled and a possible increase in driver quality, and consequently, there should be no increase in accident rates—in fact, there might even be a reduction.

This naïve view, however, ignores some of the subtler effects of substituting driving oneself with being driven by a RH driver. First, the advent of RH platforms transforms vehicles into productive assets for individuals who now find it lucrative to provide services as drivers. Moreover, these RH drivers have riders in their car for only a fraction of the time that they are on the road: they must drive from fare to fare, and they drive from location to location in the city seeking better fare

---

[1] Figure 1 Panel A was created by Dennis Bratland and is reproduced under creative commons license. The figure uses NHTSA FARS and CrashStats data to depict total U.S. motor vehicle deaths, deaths per VMT, deaths per capita, VMT and population for the period 1920–2017. Figure 1 Panel B is excerpted from the Governors Highway Safety Association (2019) report on 2018 Pedestrian Traffic Fatalities by State and is constructed using data from FARS and State Highway Safety Organizations.

prospects, as there is not always a fare available. Furthermore, RH companies often subsidize drivers to stay on the road, even when utilization is low, to ensure that supply is quickly available.

The naïve view also assumes that only those who would have otherwise driven themselves are now using RH services, which is unlikely. The convenience and lower pricing of RH apps suggest that there may be a significant number of additional riders substituting away from other modes of transportation, such as subways, buses, biking or walking; persons who would have used these modes in the absence of the convenience and low cost of RH. Indeed, surveys report that fewer than half of RH rides in nine major metro areas actually substitute for a trip that someone would have made in a car (Schaller, 2018). Moreover, a survey conducted by the University of California at Davis of over 4,000 residents in seven major metros areas found that only 39% of respondents would drive themselves, carpool, or take a taxi if RH had not been available. The rest substitute from rail, biking, walking or not traveling at all (Clewlow and Mishra, 2017).[2] This survey evidence runs counter to the naïve view's notion that utilization remains fixed, though some evidence suggests that RH can also complement public transit (Hall et al., 2018).[3]

Here, we take a more nuanced view of the overall effect of RH on road safety. We begin by proposing a conceptual framework for considering how ridehailing's introduction may affect accident rates.[4] While the naïve view holds the utilization and supply of drivers constant, our view incorporates rational choice theory to drivers' and riders' decisions. Our framework models accidents as a function of vehicle miles traveled and average driver quality, both of which are in turn affected by the introduction of RH technology. The advent of RH makes car travel easier for riders, which, in turn, should decrease the marginal cost of making a trip for them, thus spurring more rides. In the case of potential drivers, the monetary value assigned to driving via the platform also increases the net benefit for individuals with vehicles of heading out to give rides. These two forces, combined with the addition of driving from fare to fare, should lead to an overall increase in vehicle miles traveled. Depending on the quality of the new RH drivers, as compared to the driving quality of former drivers who now become riders, this may also lead to a change in the average

---

[2] Similar numbers emerge from studies conducted by the Boston Metropolitan Area Planning Commission (MAPC, 2018), the New York Department of Transportation (NYDOT, 2018), and other researchers (Clewlow and Mishra, 2017; Henao, 2017; Circella et al., 2018).

[3] From a supply perspective, a local report that examines detailed ridehailing data in New York City suggests that ridehailing companies put 2.8 new vehicle miles on the road for each mile of personal driving they eliminated (a 180% overall increase). Moreover, the same report suggests that ridehailing has added 5.7 billion miles of annual driving in the Boston, Chicago, Los Angeles, Miami, New York, Philadelphia, San Francisco, Seattle, and Washington, D.C., metro areas (Schaller, 2018). While pooling services, such as UberPool and LyftLine, can reduce the overall increase in vehicle miles, these modes of ridehailing currently represent a relatively small (20%) share of overall rides.

[4] Our theoretical analysis of ridehailing's effect on safety can be thought of along the lines of the traditional offsetting behaviors literature (Peltzman, 1975).

quality of drivers on the road. We outline the potential effects of the introduction of RH through each of these two components in our conceptual framework.

We then turn to an empirical analysis of the effects of RH on accident rates. We define the entry of RH into cities using rollout dates obtained directly from Uber and Lyft. We use the launch date of the first service to arrive in each city to determine the first quarter of treatment. Our outcome measures are a variety of fatal traffic accident-related measures from the Fatal Accident Reporting System (FARS) maintained by the National Highway Traffic Safety Administration (NHTSA). This data does not distinguish accidents in which a RH driver-partner car was involved from those where one was not; rather, we examine overall motor vehicle fatalities in a city. From a policy perspective, this distinction is not critical, as we wish to explore how the introduction of RH as a phenomenon shapes *total* accident rates.

We begin our analysis by examining changes in the level of accidents in the treated cities around the introduction of RH. Figure 2 plots the raw quarterly average accidents over event time in RH cities. At the time of RH initiation—time zero—we see a distinct break in the trend of accident incidence in the RH cities: accident numbers begin to rise sharply, relative to the pre-event time trend. We investigate this increase formally using a difference-in-differences (DD) specification with fixed effects for location and time (quarter-year) and location-specific linear and quadratic trends. Our DD specification allows us to capture macroeconomic changes, such as the Great Recession, fuel costs, and improvements in vehicle technology, as well as city-specific conditions such as average weather patterns and city topology. The location-specific time trend captures location-specific pre-trends in accidents that existed prior to the arrival of ridehailing. To capture potential time-and-city varying confounders, such as population changes or increases in employment or income, we further control for population level and per capita income (which survey data also suggests is highly correlated with smartphone adoption and usage).[5]

Consistent with the raw data plotted in Figure 2, the DD specification documents a 2% to 4% increase in the number of fatal accidents and fatalities that persists throughout the week, on weekends, at night, and on weekend nights. The estimates are robust to the inclusion of a variety of additional control variables such as unemployment, population growth, gas prices and gas taxes, as well as to the inclusion of a location-specific quadratic trend, and are similar for a variety of different specifications of the left-hand-side accident measure.

---

[5] http://www.pewinternet.org/fact-sheet/mobile/ and http://www.pewresearch.org/fact-tank/2017/03/22/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/

Of course, a natural concern is that our models are merely picking up increases in smartphone usage while driving that is unrelated to RH, (though RH by necessity likely drives to an increase in smartphone usage) and that is not captured by the per capita income control variable. To address this concern directly, we show that our documented effects persist even after controlling for gross receipts, employment, and the number of establishments for the wireless telecommunications carrier industry at the local level. Furthermore, the effect is robust to controlling for the volume of google searches for smartphone keywords, as a further proxy for smartphone adoption.

Having established our primary finding, we proceed to examine differentials in outcomes for pooled versus nonpooled services. Despite allowing for more carpools and therefore, a potential reduction of total vehicle miles traveled, the introduction of UberPool and LyftLine do not reverse the documented increase in fatal accidents. Instead, the estimates suggest that either the share of pooled rides is insufficiently high relative to single rides, or that any positive effects of pooled services in reducing VMT—and accordingly, accidents—may be offset by an increase in overall ridership, due to the lower cost of the pool service.

Presumably, the intensity of RH use should be related to the documented increase in accident rates. We proxy for the intensity of RH driver adoption using the intensity of Google searches for terms such as "Uber" and "Lyft" in the treatment cities. When we substitute the indicator for city treatment with our Google intensity proxy for the adoption of RH services within a city, we obtain similar results to those in our main specifications: fatal accidents and fatalities increase with the intensity of adoption, as proxied for by the Google Trends measure.

Next, we separate traffic accidents and fatalities into those of car occupants and non-occupants (pedestrians, bicycle riders, etc.). Doing so allows us to examine externalities of RH more directly: pedestrians and bike riders represent a population that is neither using a RH car nor riding in or driving a private vehicle. Here, we find a similar magnitude increase in the number of fatal accidents, the number of pedestrians and bike riders involved in these accidents, and the number of fatalities in such accidents. These results suggest that the introduction of RH imposes a negative externality on pedestrians and bike riders, in addition to affecting vehicle occupants.

Of course, the effects of RH on fatal accident incidence may vary with city characteristics. We explore this next. We find that the accident increases are concentrated in large cities (high population), and cities with higher levels of income inequality (Gini coefficient). They are primarily concentrated in cities where the ex-ante use of public transportation was higher, consistent both with riders substituting away from an alternative mode of transportation or with RH serving to fill last-mile needs for the use of public transport (Hall et al., 2018). They are also concentrated in cities

with high ex-ante levels of vehicle ownership, consistent with RH increasing usage of existing vehicles.

We then turn to examine the quantity and quality mechanisms suggested by our conceptual framework. On the quantity side, we document that at the intensive margin, VMT, excess gas consumption, and annual hours of delay in traffic all rise following the entry of ridehailing. Furthermore, at the extensive margin, we find a 3% increase in new car registrations, consistent with RH services creating a new productive use for vehicle ownership. Consistent with our estimates for fatal accident rates, this increase in registrations is more substantial in cities with high ex-ante use of public transportation, again consistent with substitution from public transport or with use of RH for last-mile transport.

While it is challenging to explore driver quality directly, given the absence of data explicitly involving RH drivers (the RH companies do not provide such data), we provide several pieces of evidence that suggest the presence of the quality channel. First, we show that the effect persists even when controlling for VMT. Second, we demonstrate that the effect is concentrated in RH-eligible vehicles (relatively new, four-door vehicles), and is not present for accidents involving RH-ineligible vehicles (two-door vehicles). Third, we provide additional indirect evidence by looking at an outcome that many believe to be a likely quality *improvement* brought about by RH—whether and how drunk driving related accident changes after RH is launched. We separate out the fatal accidents that involve an impaired driver, as we would expect positive quality effects to primarily be present in the reduction of drunk accidents and fatalities. We find that accidents and fatalities related to drunk driving do not decrease post-RH: if anything, we find evidence of a small increase, albeit at a smaller magnitude and weaker statistical significance. There appears to be no decrease that would be consistent with a large quality improvement. Thus, while there may be some reduction in drunk driving associated with RH arrival, our results suggest that this is not enough to dominate any quantity (and/or unobserved nondrunk quality) effects.

We note that the documented effects may be short to medium-term, as pooling services such as LyftLine and UberPool increase ridership. Furthermore, as RH drivers become more experienced, both the VMT effects and the driver quality effects may be attenuated. In our sample through the end of 2016, however, we observe no such reversion; instead, the estimates appear to be increasing with time since RH launch, and the persistence is statistically significant. Still, many cities only saw the introduction of RH services in the last three years, and pooling services are not available in all cities. It may be too soon to tell whether the effect we document is a short-term adjustment or a

longer-term pattern; our initial evidence suggests that the effect is still present three years after the entrance of ridehailing.

We conclude our study with a back-of-the-envelope discussion of potential costs from an increase in fatalities of the magnitude estimated in this paper. Utilizing estimates of the value of a statistical life from the Department of Transportation, we estimate a potential cost of just under $10 billion. We end with a discussion of other potential societal costs suggested by some of our findings, which suggest the need for further research. Importantly, as we discuss, the current manner in which RH insurance is implemented (self-insurance by the RH companies) is unlikely to get the firms to internalize the accident externality fully. Moreover, some of this cost is borne by non-RH drivers and users, in the form of higher insurance premiums overall, and in time delays and lost productivity due to the congestion increases.

Our study offers several contributions to the existing literature. First and foremost, our work speaks to the importance of considering externalities—both positive and negative—associated with the introduction of new technologies. Often, discussion of externalities of new technologies focuses on positive externalities and benefits to society (e.g., Klenow and Rodriguez-Clare, 2005). Here, in contrast, we consider that some technologies may also impose negative externalities. When new technologies are introduced in markets that account for these externalities, they often induce competition with existing products and services that enhance welfare. If these negative externalities are not accounted for, even if the private costs are exceeded by the private benefits for the individual user, the social costs may not be. It is the *sum* of social and private costs as compared to the *sum* of social and private benefits that is key to welfare effects.

In the case of RH, whether there were associated negative externalities was not clear to economists ex-ante. In a 2014 University of Chicago Initiatives on Global Markets (IGM) survey of a panel of 43 top academic economists,[6] all the panelists either agreed or strongly agreed with the statement "Letting car services such as Uber or Lyft compete with taxi firms on equal footing regarding genuine safety and insurance requirements, but without restrictions on prices or routes, raises consumer welfare." Many commented on the contribution of competition to consumer welfare; none suggested any potential negative externalities (one Nobel Prize winner noted specifically that he did not see any externalities involved). The comments were consistent with the panelists considering private welfare, rather than social welfare. Here, we shed light on the potential social costs of ridehailing. Our results speak to a growing literature on the social and economic impacts of digitization (Brynjolfsson and McAfee, 2014). In this spirit, our paper joins

---

[6] http://www.igmchicago.org/surveys/taxi-competition

contemporaneous work by Hasan and Kumar (2018), who also explore social costs of technology adoption, but in the setting of online school ratings and their effects on educational inequality.

Second, our study contributes to a growing literature exploring the RH industry and its workers. Hall and Krueger (2018) and Chen et al. (2018) explore the importance of flexibility to RH drivers. Berger et al. (2018) examines the effect of Uber on taxi drivers. Cook et al. (2018) examine the gender earnings gap for Uber drivers and find that experience and preferences fully explain the gap. Ge et al. (2016) explore racial and gender discrimination in RH. Liu et al. (2018) examine the extent to which RH reduces driver moral hazard relative to taxis, while Cramer and Krueger (2016) explore the efficiency gains and lower transaction costs associated with Uber's matching algorithm. Other work in this category has focused on RH's effect on other modes of transportation, finding mixed evidence. Nie (2017) finds Uber has reduced taxi ridership, Hall et. al. (2018) finds RH complements public transit, while Cramer (2016) finds that the wages of taxi drivers and chauffeurs have not decreased. Finally, using Uber's individual-level data and its unique use of surge pricing, Cohen et al. (2016) estimate that UberX created $6.8 billion of consumer surplus in 2015.

Our paper is not the first to attempt to examine the effects of ridehailing's introduction on traffic accidents. A number of recent papers have explored this issue, primarily through the lens of the potential for reduction in drunk driving as a result of the availability of RH (see e.g., Brazil and Kirk, 2016; Martin-Buck, 2016; Greenwood and Wattal, 2016).[7] These studies are primarily focused on measures of alcohol-related fatal accidents, fatalities, and citations for driving while under the influence of alcohol (DUIs). They typically use the introduction of UberX as their measure of treatment and find evidence either for reduction or no significant change in drunk accidents/fatalities and DUIs. The findings of these studies, however, do not necessarily contradict our results. In the Online Appendix, we show that drunk driving fatality results are sensitive to the selected sample period and to accounting for pre-existing trends in accident rates, due to an important change in how FARS classified "drunk accidents" that took place in 2007 and was implemented starting in 2008. Importantly, in our research design, we do not place our focus solely on fatalities resulting from drunk driving or alcohol consumption, as we are interested in overall effects on fatal accidents. Rather, we focus on total fatal accidents, using a broad sample, and utilize the introduction of both Uber and Lyft, including the different types of Uber and Lyft service.[8] While RH may indeed

---

[7] Other studies include Dills and Mulholland (2018), who study crime in the wake of RH introduction.

[8] When we do not account for location-specific trends in our sample, we too observe a negative coefficient for alcohol-related accident measures. However, the inclusion of the location-specific trend aligns our results for these measures with those we obtain for all other accident measures: an increase in overall accidents and fatalities for vehicle occupants and pedestrians.

displace some drunk drivers, our estimates suggest that *overall* accident rates and fatalities increase in the wake of RH introduction, despite the possible benefits from limiting impaired driving.

An examination of ridehailing's effects on accident rates is particularly useful in providing insights into changes in motor vehicle fatality trends. Prior to 2011 and for the preceding 20+ years, motor vehicle accident fatalities, in total, per population, and per VMT, had been falling. The 2010s saw a reversal of these trends. If this reversal relates partly to increased vehicle miles on the road, due to the introduction of ridehailing, this may have implications for policy discussions around decreasing motor vehicle accident rates.

Accordingly, our findings also contribute to the current discussion regarding cities' response to the rapid growth of ridehailing. While much of the resistance to ridehailing has been presented as a case of entrenched incumbents (taxis) seeking rents, our findings suggest that other societal costs are also at play. In ridehailing's case, delays in the diffusion of this new technology may be optimal, if we consider offsetting costs such as increased accident rates or pollution or the need for learning-by-doing on the part of users. Introduction of new technology can have unintended effects: it may impose externalities not priced into the cost for the individual user. Overall, whether RH is welfare-enhancing or -decreasing depends on the value of the increase in convenience and other consumer surplus effects versus the offsetting costs in time, material, and human life. As a result, our findings do not on their own advocate for a specific policy recommendation. To do so requires equating all social benefits from the technology to total social costs, and then comparing the resulting changes derived from various policy recommendations. These related issues require careful consideration but fall outside the scope of the present work.

The paper proceeds as follows. Section 2 provides an overview of RH services and outlines our conceptual framework. Section 3 describes our data and sample. Section 4 presents our main results. Section 5 explores the quantity mechanism described in our conceptual framework. Section 6 presents an estimate of costs and discusses welfare considerations. Section 7 concludes.

2. RIDEHAILING AND CONCEPTUAL FRAMEWORK

Before the advent of RH, the primary forms of private for-hire transportation were traditional taxis, limousines, and larger vehicles, such as bus and van services. Of these, only traditional taxis did not need to be reserved in advance, all came at fairly substantial costs, and the number of cars available varied widely from city to city. Most municipalities heavily regulate the traditional taxi industry, placing restrictions on the number of vehicles, prices charged, and the licensing and

insurance requirements for drivers and cars. Quantity restrictions, in particular, were thought to lead to shortages of taxis during periods of high demand and inconveniences for riders.

Uber was the first RH firm in the United States, launching in San Francisco in May 2010, and was followed two years later by Lyft and Sidecar. RH then expanded rapidly across the country. By the end of 2014, RH firms operated in 80% of U.S. cities with a population of 100,000 or more. Much of the spread in RH was driven by the convenience for users, stemming from new technology easing the matching of riders and drivers and enabling seamless payment through an app. RH firms' exemptions from (or willful disregard for) taxi and livery restrictions allowed them to expand supply during periods of high demand and adjust prices to encourage more riders and drivers to participate in the market.[9] This has in turn engendered backlash from advocacy groups and policymakers concerned with the effects of RH technology in their cities.[10]

2.1.   Conceptual Model

To better understand the expected effects of RH on accident rates, we develop a simple conceptual model in which accident rates are a function of two elements that are impacted by the introduction of RH technology: the number of vehicle miles traveled (VMT) on roads and the average quality of drivers. As noted by Vickrey (1968), Edlin (2003) and Edlin and Karaca-Mandic (2006) and others, with every mile driven by a driver, that driver exposes themselves and others to the risk of an accident.[11]

For notational purposes, we denote the accident rate for city i in period t as $A_{i,t}$ and the new technology (ridehailing) as θ. Accident rates can then be thought of as:

$$A_{i,t} = f(VMT_i(\theta); Q_{i,t}(\theta)),$$

where $VMT_i(\theta)$ is the number of vehicle miles traveled on the road in city i in period t (potentially a function of whether RH is available) and $Q_{i,t}(\theta)$ is the quality of the average driver on the road in city i in period t.

---

[9] Many major ridehailing companies adjust pricing in real time to better match supply and demand, charging higher "surge pricing" fares during periods with high demand.

[10] In many ways, ridehailing has become the modern poster child for the classic battle between what are argued to be outdated regulations, supported by rent-seeking incumbents, and the introduction of a welfare-enhancing technology. Many new technologies face frictions that slow diffusion (Grubler, 1991). Parente and Prescott (1994) argue that one such friction is resistance on the part of sectoral interests. Indeed, emphasizing barriers to technology adoption, economic historians, such as Rosenberg and Birdzell (1986), argue that the reason why the West grew rich before the rest of the world was that active resistance to technology adoption was weaker there. Most economic histories of technological adoptions provide cases in which adoption was met with fierce resistance (Mokyr 1990).

[11] Notably, these effects are compounded by the congestion and pollution effects of driving; we leave this topic to future research.

The number of VMTs can further be broken down into three sub-categories: (i) the number of VMTs generated by people driving themselves from origin to destination, denoted by VMT^{own}; (ii) the number of VMTs generated by RH drivers carrying passengers from origin to destination, denoted by VMT^{RH}; and (iii) the number of VMTs generated by RH drivers while driving in-between passengers, denoted by VMT^{btwnRH}. Thus,

$$VMT_i = VMT^{own} + VMT^{RH} + VMT^{btwnRH}.$$

Note that, even if VMT^{own} and VMT^{RH} simply offset as people move from driving themselves to being driven in a RH vehicle, there is still "between driving" (between fares, waiting for fares, going from fare location to fare location) that is introduced by the advent of RH in a city. While VMT^{own} is almost certainly decreased by the introduction of RH, the technology leads to the introduction of additional vehicle miles in the form of VMT^{RH} and VMT^{btwnRH}. Thus, the effect of the introduction of RH in a city on the number of VMTs on the road depends on whether the decrease in VMT^{own} is more than offset by VMT^{RH} and VMT^{btwnRH} that are introduced with the technology. Taking the model naïvely (and ignoring for the moment the UberPool and LyftLine services), each person who no longer chooses to drive himself or herself is now driven by a RH driver, thus precisely offsetting the effect on the overall vehicle miles traveled. But unless there are absolutely no between-fare miles driven by a ride-sharing driver, we would expect to see an increase in overall VMTs after RH arrives.

The limited evidence to date suggests that there is considerable between-fare travel by drivers. Henao (2016) reports statistics suggesting RH drivers only have passengers in the car 39% of the time and 59% of the miles they drive while active on the app. Schaller (2018), using detailed data from New York City, shows that RH drivers on average drive 2.8 miles while waiting for a fare, 0.7 miles to pick up the fare, and 5.1 miles with a passenger in the car, implying a 59% utilization rate. Furthermore, RH companies initially offered subsidies designed to induce drivers to spend more time out on the road active in the app, so as to decrease wait time for passengers. Finally, while not the focus of their study, the analysis of Chen et al. (2018) is consistent with a mismatch between rider demand and the supply of drivers, particularly given the flexibility afforded to the drivers.

More formally, we can write the first-order condition for the effects on accident rate A_i from the introduction of RH technology θ as:

$$\frac{\partial A_i}{\partial \theta} = \frac{\partial A_i}{\partial VMT_i} \frac{\partial VMT_i}{\partial \theta} + \frac{\partial A_i}{\partial Q_i} \frac{\partial Q_i}{\partial \theta}$$

where

$$\frac{\partial VMT_i}{\partial \theta} = \frac{\partial VMT^{own}_i}{\partial \theta} + \frac{\partial VMT^{RH}_i}{\partial \theta} + \frac{\partial VMT^{btwnRH}_i}{\partial \theta}.$$

Clearly, $\frac{\partial A_i}{\partial VMT_i}$ is positive, as every additional vehicle mile travelled will increase the likelihood of an accident and thus the overall accident rate. $\frac{\partial VMT^{own}_i}{\partial \theta}$ is negative. $\frac{\partial VMT^{RH}_i}{\partial \theta}$, however, will either equal or, more likely, due to substitution away from other forms of transport, be larger in absolute magnitude than $\frac{\partial VMT^{own}_i}{\partial \theta}$, and $\frac{\partial VMT^{btwnRH}_i}{\partial \theta}$ is positive. Thus the overall effect $\frac{\partial VMT_i}{\partial \theta}$ is positive: vehicle miles traveled increase with the introduction of RH.

Of course, in some cities, at later dates, the option to "carpool" via RH was introduced, in the form of Uber Pool and Lyft Line. With the introduction of these services, the reduction in own drive car hours may not be fully offset by RH drive hours, as multiple people may be substituting away from driving themselves into a single RH car. While Uber and Lyft have both heavily invested in promoting their shared services, Uber reports that UberPool accounts for only 20% of trips in cities where it is offered, and Lyft reports that 37% of users in cities with LyftLine request a Line trip, and many trips are not matched, thus leaving a single rider (Schaller, 2018). Pooled rides are also cheaper, potentially inducing more substitution from other modes of transport. It is not clear what fraction of rides must be pooled to counteract VMT^{btwnRH}, but Schaller (2018) suggests that, even if half of the rides were pooled, total VMT would still increase. Furthermore, stepping away from the naïve model, survey evidence suggests that $\frac{VMT^{RH}}{VMT^{own}} > 1$, as many riders are substituting away not from driving themselves but rather from other forms of transportation, including walking, biking, and, more importantly, public transportation (Clewlow and Mishra, 2017). Thus, it is likely that pooled ride adoption would need to be extremely high to offset such substitution effects.

Assessing the effects of the introduction of RH on the quality of the average driver on the road is less straightforward. On the one hand, the people substituting into a ridehailing vehicle, rather than driving themselves, may be low quality drivers (impaired or unskilled or may just prefer not to drive), but they may be high quality drivers who simply dislike driving. On the other hand, there is no guarantee that the driver who substitutes for them is of higher quality. Put another way, the introduction of RH makes it less costly to have someone else drive you, but also makes the gains from getting out on the road as a driver greater (as you can make money by doing so). Lower quality drivers, who in the absence of compensation may not have driven, now have an incentive to drive. Moreover, more affluent people are more likely to use RH (Pew Research Center, 2016), and the

less affluent are more likely to become RH drivers. To the extent that this substitution leads to more vehicle miles driven by lower quality drivers or in lower quality cars, this may positively affect accident rates. Finally, ridehailing VMT is also different from seld-driving VMT, as ridehailing drivers often stop in random locations mid-street to pick up or drop off a rider. This type of haphazard dropoff and pickup stoppage may also lead to additional accidents (even if the accidents do not involve the ridehailing vehicle).[12] Yet RH drivers, especially those with more experience from more hours driven, may in fact represent improved quality. To the extent that the substitution goes the other way and lower quality drivers are substituted by better drivers, this may reduce accident rates if the increase in quality offsets the increase in VMT.

Formally, $\frac{\partial A_i}{\partial Q_i}$ is negative: better drivers reduce accident rates, all else equal. The effect of RH on the quality of the average driver on the road, $\frac{\partial Q_i}{\partial \theta}$, however, is ambiguous. If the quality of the average driver increases, this could offset the quantity effect above. If it decreases or does not change, the quantity effect will prevail. Which effect dominates, of course, is an empirical question.

Many indicators suggest that both total VMT and driver quality may adjust over time. Cook et al. (2018) note that, even in the relatively simple production of a passenger's ride, experience is valuable for drivers. A driver with more than 2,500 lifetime trips completed earns 14% more per hour than a driver who has completed fewer than 100 lifetime trips, in part because he learns where and when to drive, which may decrease VMT^{btwnRH}. Similarly, Haggag et al. (2017) show that experience is important for taxi drivers. At the same time, not all learning-by-doing is necessarily good for accident rates. For example, learning by doing to maximize earnings could lead to behavior, on the part of certain driver populations, that directly or indirectly increases the probability of accidents, such as gaming time-and-distance pay algorithms by taking longer routes, speeding, etc.

3. DATA AND SAMPLE

Our sample consists of all incorporated "places"[13] in the continental United States with population greater than or equal to 10,000 in 2010.[14] Our sample covers the period 2001 to 2016;

---

[12] Examples of haphazard pickup and dropoff stoppage include blocking bike lanes and crosswalks, suddenly pulling over, not pulling over completely (blocking lanes), and similar. While taxis often engage in similar behavior, taxis are clearly labeled, such that other drivers and pedestrians may know to expect erratic driving.

[13] We use incorporated places, rather than Census Designated Places (CDPs), because CDP annual population estimates are not readily available, except by individual place download, whereas population data is available for incorporated places for mass download through the census.

[14] Some places in our sample had lower populations than 10,000 during the sample period, most notably during the period of 2001–2010. We impose the cutoff on population as measured in 2010. As an example, consider Hutto, Texas, a suburb of the Austin-RoundRock metro area. In 2001, Hutto had a population of 3,030, the lowest in our sample. By

all results are robust to employing shorter pre-RH sample windows. Our list of incorporated places is obtained from the Census Bureau and covers all self-governing cities, boroughs, towns, and villages in the United States.[15] (For ease of interpretation, we interchangeably refer to these as "cities" or "locations" throughout the text.) Our observations are measured at the quarterly level. The sample thus contains 190,080 quarterly observations on 2,970 "places" from 2001 to 2016, among which 1,199 adopt RH prior to 2017. Figure 3 shows the diffusion of RH across the United States, by cities and population. Diffusion of RH across U.S. cities began slowly, accelerating rapidly after 2013. Diffusion by population follows a standard S-curve, consistent with general historical patterns of new technology diffusion.[16]

3.1. Fatal Accidents

We obtain data on accidents involving at least one fatality ("fatal accidents") from the National Highway Traffic Safety Administration (NHTSA) Fatal Accident Reporting System (FARS). To qualify as a FARS case, a crash must involve a motor vehicle traveling on a traffic way customarily open to the public and must have resulted in the death of a motorist or a nonmotorist within 30 days of the crash. Importantly, the data identify whether any drivers involved are under the influence of alcohol. We aggregate the incident-level FARS data into quarterly totals for each place/city. When the data contain geographic coordinates, we use Google Map's Geocoding API service to determine the corresponding place/city. When the coordinates are not available, we use the city and state identifier codes to assign observations to the appropriate place. Geographic coordinates are present in 98% of FARS's observations, and we successfully match more than 99% to incorporated places in the U.S.

We construct various measures of accident volume from the FARS data. *Total Accidents* is the total number of fatal accidents according to the definition used by NHTSA. *Total Fatalities* is the total number of fatalities across all fatal accidents. We further classify our various categories of accidents based on their time of occurrence: (i) weekday: Monday through Thursday; (ii) weekend: Friday through Sunday; (iii) night: after 5 pm and before 2 am; (iv) Friday and Saturday night: after 5 pm and before 6 am on Friday and Saturday. Additionally, we further separate out accidents involving pedestrians and calculate three measures of pedestrian-involved accidents. *Pedestrian-Involved Accidents* is the number of fatal accidents involving at least one pedestrian. *Pedestrian-*

2010, it had grown to over 14,000, mimicking the growth of the Austin metro area. As it has population above 10,000 in 2010, it is included in our sample. Our results are robust to permutations to this cutoff.

[15] https://www.census.gov/content/dam/Census/data/developers/understandingplace.pdf

[16] In the Online Appendix, we further demonstrate the robustness of our results to using shorter pre-sample periods.

Involved Fatalities is the total number of fatalities in all accidents involving at least one pedestrian. Finally, *Pedestrians Involved in Fatal Accidents* is the total number of pedestrians involved in fatal accidents. When we refer to accident "rates," these are defined as the number of accidents per 100,000 people or the number of accidents per billion city VMT, as indicated.[17]

3.2. Ridehailing Launch and Driver Enrollment Intensity

Data on RH launch dates for each city are obtained directly from Uber and Lyft.[18] The companies provided dates of service launch for each type of service launched: (i) UberBlack/UberTaxi, which allows customers to hail a livery or taxi vehicle; (ii) UberX/Lyft, which allow customers to hail regular cars driven by driver-partners; and (iii) UberPool/Lyft Line, which allow customers to share a hailed vehicle with others. We merge these dates with Census Bureau's incorporated place directory in 2010.

While Uber and Lyft declined to provide data on driver enrollment and usage for this project, other researchers have shown a strong correlation between Google trends for searches for RH keywords and actual driver uptake (Cramer and Krueger, 2016). To measure the intensity of driver adoption, we thus follow the spirit of the work of Cramer and Krueger (2016) and Hall et al. (2018) and use Google searches for the terms "Uber," "Lyft," and "rideshare."[19] Rather than using the standard Google Trends index, which scales results from 0 to 100 based on the most popular term entered, and that does not easily allow comparisons across geographic areas and time periods, we use data from the Google Health Trends API, which describes how often a specific search term is entered relative to the total search volume on Google search engine within a geographic region and time range, and returns the probability of a search session that includes the corresponding term, which makes comparisons across locations and time feasible.[20] We track trends for searches for

---

[17] In analysis in the Online Appendix, we further separate out accidents by whether or not a drunk individual was involved. *Total Drunk Accidents* is the total number of fatal accidents involving any drunk drivers. *Total Drunk Fatalities* is the total number of fatalities in all drunk-driver accidents. *Total Non-Drunk Accidents* is the total number of fatal accidents not involving any drunk drivers. *Total Non-Drunk Fatalities* is the total number of fatalities in all nondrunk-driver accidents. We discuss these further in the Online Appendix.

[18] In this version, we use the exact cities indicated by Uber and Lyft, even if we suspect or believe that the launch covered adjacent cities as well (e.g., San Francisco launched in 2010, and there is no separate launch date for San Jose or Palo Alto). Since this means some places we include in our control may in fact be treated in later years in the sample as service expands slowly out beyond original boundaries, we are biasing against finding an effect of treatment.

[19] We use the freebase identifiers for term "Uber" (/m/0gx0wlr) and "Lyft" (/m/0wdpqnj). Freebase identifiers denote all searches that were classified to be about this topic.

[20] These probabilities are calculated on a uniformly distributed random sample of 10%-15% of Google web searches. Mathematically, the numbers returned from the Google Trend API can be officially written as:

$$Value_{[time, term\ restriction]} = P(term - restriction | time\ and\ geo - restriction) * 10M$$

This probability is multiplied by 10 million in order to be more human readable.

these terms using the Google Health Trends API for all Nielsen Designated Market Areas (DMAs) at monthly frequency from January 2004 to December 2016. We aggregate the data to the quarter level and match the DMAs to Census incorporated places using a crosswalk provided by Nielsen. Thus, in the accident specifications that use log search volume as a proxy for driver enrollment intensity, we interpret the coefficients in terms of percentage change in search volume.

3.3. Other Data

We use a number of measures to explore heterogeneity by city characteristics and as control variables in our models. We obtain annual city population estimates and population density from the U.S. Census and annual county income per capita from the Bureau of Economic Analysis. Household vehicle ownership and means of transportation to work at the city level are gathered from the 2010 American Communities Survey. Controlling for population and per capita income which vary by time and location are of first order importance as they provide a proxy for specific concerning confounders. For example, a reasonable concern might be that Uber and Lyft specifically chose cities to enter based on smartphone adoption trends, and any increase in accidents we document could be due to increased levels of smartphone adoption in those specific cities relative the ones not entered by RH companies, with smartphone usage leading to distracted driving, which in turn leads to increased accidents. While data on smartphone adoption by city is not publicly available, smartphone adoption is known be highly correlated with per capita income, which we thus include as a control variable in our models.[21]

To explore mechanisms that may drive any change in accident rates upon arrival of RH, we use a variety of data sources. We obtain data on new car registrations by zip code on a monthly level from Polk Automotive. We aggregate the data at city and quarter level using UDS Mapper's zip code-to-ZCTA crosswalk[22] and Census' ZCTA-to-place crosswalk. We obtain estimates of city and freeway vehicle miles traveled, total annual excess fuel consumption, and total annual hours of traffic delay for a sample of 101 urban areas from the Texas A&M Transportation Institute Urban Mobility Scorecard, covering the period of 1982–2014. Of the 101 urban areas covered by TAMU in their report, 99 fall into our sample of continental U.S. cities. For a set of tests regarding road use and driver quality, we use the census's urban area-to-place crosswalk to aggregate our main sample

---

[21] http://www.pewinternet.org/fact-sheet/mobile/ and http://www.pewresearch.org/fact-tank/2017/03/22/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/
[22] The crosswalk can be found at https://www.udsmapper.org/zcta-crosswalk.cfm. The crosswalk is recommended by Missouri Census Data Center, http://mcdc.missouri.edu/geography/zipcodes_2010supplement.shtml.

at urban area and annual level to merge the information with TAMU's dataset. We discuss the TAMU data construction methodology in further detail in the Online Appendix.

3.4.   Summary Statistics

Table 1, Panel A presents summary statistics for the places in our sample over the sample period. The places average 54,500 in population, have an income per capita of $39,720, and a population density of roughly 3,000 people per square mile. Prior to the arrival of RH, 2.96% of residents in our average city/place used public transportation to commute, 10.6% commuted by carpool, and 33% owned vehicles. The average city in our sample had 670 new car registrations per year. As can be seen from the distributional statistics in the table, there is wide variation across all these characteristics across the sample. The table further presents summary statistics on rate (per 100,000 population) of accidents for the cities in our sample over the sample period. We present statistics for total accidents and fatalities and total pedestrian-related accidents and fatalities. Pedestrian accidents and fatalities are approximately 20% of the total.

4.   EMPIRICAL ANALYSIS

To assess the impact of RH on fatal accident rates, we employ a standard generalized difference-in-differences approach. We index cities by c and time by t. We estimate models of the following form:

$$\log(1 + accidents_{t,c}) = \propto_c + \gamma_t + \beta' X_{t,c} + \theta_c t + \delta POST_t * TREATED_c + \varepsilon_{t,c},$$

where $accidents_{t,c}$ is our measure of accidents in city c in quarter t, $\propto_c$ is a city fixed effect, γ_t is quarter-year fixed effect, $X_{t,c}$ is a vector of time-varying, city specific control variables, and $\theta_c t$ is a city-specific linear time trend.

Of course, RH companies' choice of cities to launch in first is unlikely to be random. However, it is unlikely that Uber and Lyft were specifically selecting cities to roll out services based on trends in fatal accident rates. Rather, RH companies appear to have been selecting cities with higher populations and higher per capita income, and empirically, we observe no significant relationship between trends in accident rates and entry order.[23] We include these controls in all our models. For robustness, we also estimate all our models with the inclusion of a location-specific quadratic trend

---

[23] In the Online Appendix, Table A1, we run a multinomial logit to predict entry, where the observation level is a city. The dependent variable takes the value of 0 if RH launched in the city in 2010, 2011 or 2012 (early entry), a value of 1 if RH launched in the city in 2013 or 2014 (middle entry) and a value of 2 if RH launched in the city in 2015 or 2016 (late entry). While both population and per capita income load positively and significantly in predicting earlier entry, the change in accident rates over the 3, 5 or 10 years prior to entry does not load significantly.

as well, with qualitatively similar results. We use robust standard errors, clustered at the city level. Our observations are at the quarterly level, and cover the first quarter of 2001 through the fourth quarter of 2016.

5. MAIN RESULTS

In our estimations, we use a number of measures for $accidents_{t,c}$. In Panel A of Table 2, we employ our two main measures of total fatal accidents. Columns (1), (2) and (3) use total accidents, and columns (4), (5) and (6) use total fatalities. The first column of each set reports estimates without the inclusion of the city-specific linear time trend or quadratic trend, the second column of each pair adds the city-specific linear trend, and the third column further adds a city-specific quadratic trend. For brevity, we report only the coefficient on the variable of interest— $POST_t * TREATED_c$ in the table. We report OLS specifications, but our results remain robust to the use of count models instead.

For both total accidents and total fatalities, regardless of specification, we observe a consistently positive and significant coefficient on the $POST_t * TREATED_c$ variable. Before accounting for the location-specific time trend, the effect ranges in magnitude from an increase of 1.83% in total fatalities (column (4)) to 1.91% increase in total fatal accidents (column (1)). Once we include the location-specific time trend, the magnitudes of the increase are approximately 3.6% for both measures of accidents. The magnitude of the effect decreases slightly, to a little over 3.3%, once we include the quadratic trend. In the Online Appendix, Table A2, we further break out the effect of introducing each element of our main specification in turn. Figure A1 graphs the coefficient and associated confidence interval for the variable $POST_t * TREATED_c$, first itself, then adding year fixed effects, city fixed effects, and the city-specific linear trend, and city-specific quadratic trend, each in turn. For brevity, in much of the remainder of the reported analysis, we report only the DD models with city-specific linear trends, but all results remain robust to adding quadratic trends as well, and all model specification estimates are available upon request.

The top two graphs of Figure 4 graphically presents the difference-in-differences estimators (with each dot representing annual-coefficients) for the ten years preceding and four years following RH adoption for total accidents and total fatalities. In both panels, the counterfactual treatment

17

effects in the pre-RH periods are statistically indistinguishable from zero, providing support for our inferences (parallel trends in the pre-period).[24] Post-RH, we see a clear increasing treatment effect.[25]

In Panel B of Table 2 we demonstrate that our results are robust to alternative formulations of the analysis. Specifically, we demonstrate that similar estimates obtain when using population weighting, using accident rates as the LHS variable, or when employing the inverse hyperbolic sine instead of *log(1+accidents)*. In the Online Appendix Table A3, we further demonstrate robustness to controlling for the population growth rate, retail gas prices, the change in retail gas prices, and the unemployment rate. The coefficients are graphically presented in Figure A2. When we conduct a Bacon Goodman (2019) decomposition, we obtain similar estimates from variation within the ever-treated sample and within the treated versus never-treated.

In many cases, RH was rolled out at the MSA or CBSA level rather than to an individual city within the CBSA. In Panel C of Table 2, we demonstrate that our estimates remain robust and statistically significant when (a) clustering standard errors at the CBSA level, and (b) re-running our models at the CBSA level rather than the place-level, utilizing the earliest adoption date within the CBSA and clustering standard errors at the CBSA level. When we re-run at the CBSA level, we observe a somewhat larger magnitude of the effect, at roughly 5% (vs. 3.6% in main specification).

In Table 3, we break out weekend accidents, nighttime accidents, weekday accidents, and weekend night accidents for total fatal accidents (Panel A) and total fatalities (Panel B). We observe similar patterns to those exhibited in the models in Table 2. Accident and fatality increases are lowest on weekend nights (Friday and Saturday, after 5 pm, and before 6 am) at 2.50% and 2.69% respectively, consistent with weekend nights being the most likely period in which RH may be reducing the number of impaired drivers. For total weekend and nighttime accidents and fatalities, the magnitudes of the estimated increases are between 3% and 4%. We graph these estimates in Figure 5. Panel A presents the estimates and confidence intervals for total fatal accidents and total fatalities on weekends and nights. Panel B further splits the sample into large (highest quartile) and small (lowest quartile) cities by population, and graphs the estimates for accidents and drunk

---

[24] As an additional (closely related) way to assess the validity of the parallel trends assumption, we plot univariate trends separately for the treatment and control groups in the pre-ridehailing period (unreported, available upon request). A visual inspection provides no indication of differential trends between the groups for any of the four primary outcome variables, which provides further assurance that the parallel trends assumption is valid in our analyses.

[25] In the appendix, we conduct a placebo test using locations that did not adopt RH (to eliminate contamination from cities that adopted). We use various cutoffs for the city population. For each run, we simulate 100 cities adopting RH by assigning a random adoption date from within the set of actual launch dates in the cities that adopted. We run the simulation 100 times. We then plot the number of accidents per 100,000 population in simulated event time. We observe no discernable pattern.

accidents on weekends and nights for each. The panel hints at what we will see shortly in further estimations: that the effects of RH appear to be larger in larger cities.

We then examine the persistence of the documented RH effect by breaking the post-RH variables into quarters past. Doing this allows us to examine the dynamics of the effect up to two years after the introduction of RH in the cities. Table 4 reports the estimates of the dynamics of RH. The table shows that ridehailing's increase in accidents and fatalities persists and, in fact, appears to be increasing six quarters after introduction in the city, consistent with a time gap between launch and widespread adoption in a location.

5.1. Ride-Hailing or Smartphones?

Of course, a natural concern is that our models are somehow merely picking up increases in smartphone usage during driving, which leads to distracted driving and in turn to additional accidents. By necessity, RH involves smartphone usage, and thus, one possibility is that in fact RH leads to increased accidents through RH drivers' adoption and usage of smartphones and the associated distracted driving. This, however, would still mean the increase in accidents is as a result of RH entry. The alternative we would like to rule out, however, is that RH merely coincides perfectly with smartphone adoption patterns unrelated to RH itself, and thus that our results spuriously pick up this unrelated increase in smartphone usage. While we expect that the structure of the DD models and FE should absorb much of these types of trends, we attempt to more directly address this concern by controlling in our models for proxies for smartphone adoption. We utilize data from PowerAnalytics, which provides aggregate gross receipts, employment and number of establishments for NAICS Code 517312: Wireless Telecommunications Carriers (Except Satellite) at the MSA level on an annual basis. As an additional proxy for smartphone adoption, we also use Google Health Trends API search volume for smartphone related keywords ("iPhone," "Android," "Samsung Galaxy," "Smartphone," and "cellphone"). Table 5 Panels A and B demonstrate the robustness of our main finding to the inclusion of these four control variables. Accidents load positively on the cellphone sales and employment, but do not affect the significance or the magnitude of the main effect, which remains robust in all specifications.

Finally, in Table 6, we attempt to shed light on the notion that RH may drive smartphone adoption in places it was not prevalent before, and that smartphone adoption in turn (by RH drivers and passengers alike) may lead to higher accidents. We utilize the ex-ante stock of smartphones as proxied for by cumulative cellphone sales in the three years prior to RH entry in 2010. We interact this variable with $POST_t * TREATED_c$. In locations where cellphone adoption was higher pre-RH, the increase in accidents in lower, suggesting that in fact, the mechanism behind our findings may

be that RH induces people who did not previously have smartphone to purchase smartphones (so that they can drive for the service), and as a result, there is an increase in smartphone-distracted driving; this increase, however, remains attributable to the entry of RH as a service in the market.

5.2. Additional Services and Intensity of Driver Adoption

In Table 7 Panel A, we separate out the treatment effect of the different types of services: those that are single rides (UberBlack/taxi/X, Lyft) versus pooled rides (UberPool, LyftLine). We pool UberBlack/taxi with UberX, due to the very small number of cities that have (had) UberBlack/taxi service. We thus report the treatment effect for pooled versus nonpooled service. The estimates in the table suggest that the rollout of pooled ride services does not reverse the overall treatment effect of nonpool RH. The coefficients for pool launch are negative, and roughly half the magnitude of those for single ride (nonpool) RH launch, but are not statistically significant at conventional levels. This is consistent with reported low adoption rates for pooled rides in cities that offer the service.

In Panel B of Table 7, we explore the effect of the intensity of service adoption. In the main models presented in Table 3, we employ the first launch of a RH service, irrespective of the type of service, as our treatment date. Take up of these services, however, is likely to intensify over time. To explore this issue, we now interact our $TREATMENT$ indicator with the intensity of Google searches measure and re-estimate our models. Table 7 Panel B presents the results of this estimation where $accidents$ measure are measured as total accidents in columns (1), (2) and (3) and total fatalities in columns (4), (5) and (6). The estimates are consistent with an increase in accidents following an increase in our Google Trends intensity measure. For all six models, the coefficient estimate on $POST * INTENSITY$ is positive and statistically significant. Thus, as our proxy for adoption intensity (Google trends search intensity) increases, so do fatal accidents.

In unreported estimations, we perform a small falsification exercise, using only the sample of never-treated cities, and regress our accident measures on the Google trend search volume. We include our control variables, city and year-quarter fixed effects, and city-specific linear (and quadratic) trends. We observe no relationship between search volume for RH related terms and accident rates.[26]

---

[26] An ideal additional test would be to look at U.S. cities where RH was introduced and then withdrawn. Unfortunately, these cities are few, and the circumstances do not allow for the types of tests we would want. For example, Uber and Lyft both withdrew from the Austin market at one point in 2016 in a regulatory dispute, but at least five other RH services were still operating and took up the slack. Uber and Lyft then returned to the Austin market within a year, after Texas passed HB100, creating looser statewide rules that superseded Austin's (their return led to immediate massive drops in volume for the competitors that sprung up in their absence). In Las Vegas, the other city we are aware of, RH was introduced, then outlawed after only one month of service.

5.3.   Pedestrians versus Vehicle Occupants

An important question is whether the increase in accidents and fatalities suggested by the estimates in Table 2 are concentrated among vehicle occupants versus the alternative of potentially imposing an externality on pedestrians (nonvehicle occupants). The increase in accidents could primarily affect vehicle occupants, or it could additionally affect bystanders. The FARS data allow us to separate out accidents in which pedestrians were involved. We code an accident as pedestrian-involved if the FARS database indicates it involves persons that are not motor vehicle occupants or riders (motorcycle).[27] Thus, "pedestrian" in our context refers to both pedestrians in the usual sense, as well as bicycle, skateboard and scooter riders, etc.

In Table 8, we present the estimates from models similar to those in Table 2, substituting our measures of total fatal accidents with similar measures that solely count accidents in which a pedestrian was involved. Our $accidents$ measure in columns (1) and (2) is the total number of accidents in which a pedestrian was involved; in columns (3) and (4), it is the total number of fatalities in accidents that involved a pedestrian; and in columns (5) and (6), it is the number of pedestrians involved in fatal accidents. The estimates from these models follow the same pattern as the estimates of our main models, suggesting that the increase in accidents, following RH entry, imposes an externality on nonvehicle occupants. The magnitudes of these increases mirror those in our main models, ranging from a 2.45% increase in total fatal accidents involving a pedestrian and in fatalities in accidents involving a pedestrian, to an increase of 2.77% in the number of pedestrians who are involved in fatal accidents.

The bottom two graphs of Figure 4 graphically presents the difference-in-differences estimators (with each dot representing two quarter-coefficients) for the eight quarters preceding and following RH adoption for pedestrian accidents. As in our main models, the counterfactual treatment effects in the pre-RH periods are statistically indistinguishable from zero, again providing support for our inferences (parallel trends in the pre-period).

5.4.   Heterogeneity of Effects

In Table 9, we break out our results across a variety of city characteristics—population, income inequality, and population density—as well as by ex-ante vehicle ownership, public transport usage, and carpool usage, as reported by the American Community Survey. For each characteristic, we divide cities into quartiles and re-estimate our models, interacting $POST * TREATMENT$ with the four quartile indicators for the city characteristic. For each city characteristic, we estimate four

---

[27] FARS defines a pedestrian as "any person not in or upon a motor vehicle or other vehicle."

models, in which *accidents* measures total fatal accidents, total fatalities, total fatal accidents involving a pedestrian, and total fatalities in accidents involving a pedestrian. As before, all models include location and year-quarter fixed effects, a location-specific linear time trend, and control variables.

Panel A presents the estimates for the models using quartiles of city characteristics. Column (1) presents the estimates where the city characteristic of interest is city population. For both measures of total fatal accidents and fatalities and for measures of pedestrian accidents and fatalities, the estimates suggest that the increase in accidents observed in our main models is concentrated in large cities (fourth quartile). The estimates for $POST * TREATED * Q4$ are significant and range from 6.3% to 7.7%; in contrast, the estimates for the bottom three quartiles of city population are an order of magnitude smaller and insignificant at conventional levels. This is consistent with the larger magnitude estimates that arise in Table 2 in the population-weighted analyses. It is also consistent with findings in Edlin and Karaca-Mandic (2006) on accident externalities from driving more generally: high traffic locations have economically large externalities, while in contrast, the accident externality per driver in low-traffic locations appears to be quite small.

Column (2) repeats this exercise, breaking cities into quartiles by Gini coefficient. Here, we see stronger effects for cities in the top quartile of income inequality, and for pedestrian-involved accidents and fatalities, the effect appears to be fully concentrated in the top three quartiles of city income inequality. In column (3), we break cities into quartiles by population density. Here, we observe no clear pattern; the only outliers are the estimates for the coefficients for the least dense cities in the models for pedestrian accidents and fatalities, which, unlike the rest of the coefficients, are insignificant and much smaller in magnitude. These estimates and the associated confidence intervals are also graphically presented in Figure 6, Panel A.

Panel B of Table 9 turns to measures of ex-ante vehicle ownership, public transport usage, and car pool usage from the ACS. Some interesting patterns emerge. First, from column (1), we see that the increase in accidents following the launch of RH services appears to be concentrated in cities in the top quartile of ex-ante vehicle ownership. This is consistent with a lower cost of driving for those individuals who already had a car with which to drive for RH. This is also consistent with many of the RH firms' arguments that RH allows for better utilization of cars already present in the cities, inducing those cars to be on the road, instead of sitting idle.

Second, in column (2), we see that the increase in accidents is concentrated in cities with higher ex-ante usage of public transportation; the coefficients of interest are positive and significant for the top two quartiles of public transport use, are insignificant for the second quartile, and are even

negative and significant at the 5% level for cities in the lowest quartile of public transport use, when the dependent variable is calculated using pedestrian accidents or pedestrian accident fatalities. Finally, consistent with the estimates for the prior two columns, column (3) suggests that the increase in accidents, post-RH, is concentrated in cities that had above-median carpool usage. These estimates would be consistent with a substitution effect to RH and away from public transport and carpooling, but would also be consistent with RH serving to fill last-mile transport needs, complementing public transport use (Hall et al., 2018). We graph the estimates and associated confidence intervals in Figure 6, Panel B.

6. MECHANISMS

Having established a robust pattern of estimates consistent with an increase in fatal accidents and fatalities following the launch of RH services in a city, we now consider one of the two mechanisms discussed in our conceptual framework: increases in quantity (road utilization in the form of VMT) and changes in driver quality. We begin with an exploration of the effects of RH on measures of road congestion. Studies such as Mangrum and Molnar (2017) have demonstrated the relationship between provision of additional taxis and increased congestion; our conceptual model suggests there should be such as an effect for introduction of RH vehicles.

6.1. Quantity

Road-utilization and congestion data for city roads are not readily available for most cities (in contrast to highway VMT, which are readily available from the Department of Transportation). To examine this channel, first, on the intensive margin, we use annual estimates of arterial vehicle miles traveled, excess gas consumption,[28] and hours delay in traffic for 99 urban areas reported by the TAMU Transportation Institute for the years 2000–2014.

In Table 10, we estimate similar models to our main specification, replacing the *accidents* variable as our dependent variable with arterial street daily VMT (column (1)), annual excess fuel consumption (column (2)), and annual hours of delay (column (3)). Due to the limited availability of data relative to the full sample, the models in Table 10 aggregates locations up to the urban area.[29]

---

[28] Excess fuel consumption and excess hours of travel are calculated as the difference between the observed fuel consumption or hours of travel and the free-flow fuel consumption or hours of travel. The free-flow speed iss estimated using the speed at low volume conditions (for example, 10 p.m. to 5 a.m.) for each roadway section and hour of the week.

[29] TAMU uses the Department of Transportation (DOT) urban area boundaries. DOT urban areas were adopted from Census   urban   areas   but   have   slight   adjustments   for   transportation   purposes.   See

Moreover, we can estimate only for the years up to 2014, for these 99 urban areas, leaving us with 1,386 observations (as compared to 190,080 in our other models). Still, for all three models, we obtain a positive and significant estimate for the coefficient on our variable of interest, $POST * TREATMENT$, though with lower statistical significance levels. The economic magnitudes for both measures are roughly on the order of a 1.6% increase: approximately one half the magnitude of the effect of RS on total fatal accidents, and suggesting that quantity effects alone are unlikely to fully explain the increase in accidents estimated in the main models.[30]

Next, in Table 11, we examine the extensive margin in usage by estimating similar models where the dependent variable is the logarithm of new car registrations as reported by Polk Automotive. Both Lyft and Uber often report numbers from surveys of users, suggesting some of their riders forgo owning their own cars, and thus argue that they are removing vehicles from the road. These surveys, however, do not account for the possibility that, while some of the *rider* population is forgoing owning a vehicle, others may be purchasing vehicles precisely in order to work as RH *drivers*. While the advent of RH may reduce personal car usage for some, it also transforms cars into a productive asset, as RH now makes it lucrative to drive. Both Uber and Lyft offer programs subsidizing the purchase or leasing of vehicles for those willing to become driver-partners on their platform. Consistent with this notion, Buchak (2018), in contemporaneous analysis, documents that RH entry coincides with sharp increases in auto loans, auto sales, employment and vehicle utilization among low-income individuals. Thus, while RH may enable reductions in vehicles purchased on the rider (demand) side, it also provides strong incentives for the purchase of more vehicles on the supply (driver) side. Which effect dominates is an empirical question.

Panel A of Table 11 reports the estimates from models with and without the location-specific trend. The estimates suggest that the initiation of RH leads to an increase in new car registrations, rather than an overall decrease. This increase is in the range of 3-5% when including the location-specific time trend. A caveat to the magnitudes is that we cannot separate out regular car purchases from RH-service intended vehicle registrations.

In Panel B, we advance the intuition of this extensive margin effect by examining how new car registrations respond to the interaction of RH intensity, as proxied by the Google search intensity variable used in Section IV.B. The estimates suggest that new car registrations increase with the

---

https://www.fhwa.dot.gov/planning/census_issues/archives/metropolitan_planning/faqa2cdt.cfm#q24 and https://www.fhwa.dot.gov/legsregs/directives/fapg/g406300.htm.

[30] In early August 2019, over a year after the initial draft of this paper, Uber and Lyft released a joint study in which they admitted that RH cars contribute to increased overall congestion in six major U.S. cities (Boston, San Francisco, Washington DC, Chicago, and Los Angeles). https://drive.google.com/file/d/1FIUskVkj9lsAnWJQ6kLhAhNoVLjfFdx3/view

intensity of Google searches for Uber/Lyft/ridesharing. This relationship intensifies when RH begins in a treated city. These results suggest that new vehicle purchases increase as RH services become more intensely used.

Turning to Panel C of the table, the heterogeneity in this increase along city characteristics lines up with the heterogeneity in the increase in accidents documented in Section IV.D: the new car registrations are concentrated in cities with above median population and in cities with above median ex-ante vehicle ownership. Moreover, the increase in new car registrations is larger in cities with high ex-ante public transport usage and car pool usage. They are decreasing only in the cities with the *lowest* quartile of ex-ante carpool usage. We graph the estimates and confidence intervals in Figure 7.

Interestingly, the estimates in Panel C of Table 11 suggest that the increase in new car registrations is higher in cities with high population density: the estimates imply a 9% increase in new registrations in the cities in the highest quartile, a 6% increase in cities in the second quartile, a 2% increase for cities in the third quartile, and a statistically insignificant 3% *decrease* in cities in the lowest quartile. Overall, this fact pattern suggests increases in congestion prompted by RH. The increase in new car registration also appears to be concentrated in cities with higher population levels in general, consistent with our findings regarding VMT (the intensive margin). In contrast, the increase in new car registrations is stronger for cities with lower income inequality. This is perhaps unsurprising; while the more affluent are more likely to use RH, the less affluent lower tiers of society who are the likely RH drivers, the less likely they are to be able to purchase or lease new cars in order to become drivers. Overall, this may then lead to RH driving being done in existing older or lower quality cars, leading to a decrease in the quality channel, consistent with our finding that accident rates increase more in cities with higher income inequality.

6.2. *Driver Quality*

Examining driver quality is challenging given the nature of the available data. As seen in the previous analyses, the magnitude of the association between RH arrival and VMT is only about half the magnitude of the association with fatal accident measures. This difference would be consistent with the presence of decreased quality overall. Moreover, in additional analysis in the Online Appendix, we utilize the subsample of 99 urban areas for which the TAMU VMT data is available, and re-estimate our difference-in-difference models, this time additionally controlling for VMT, to see whether the increased VMT (quantity channel) absorbs the entire main effect. When estimating at the aggregated Urban Area level, the coefficient of interest remains positive, even controlling for

the increase in VMT. When we estimate the models at the place level, we observe positive and statistically significant coefficients, controlling for VMT. Thus, the evidence suggests that part of the effect we document comes from outside the quantity channel, consistent with a reduction in overall driver quality.

Furthermore, if the effect is primarily driven by RH drivers, as opposed to just a VMT increase overall, the effect should be more prominent for vehicles more likely to be RH vehicles. RH companies restrict the types and ages of vehicles that can be used by driver-partners. While these restrictions vary somewhat from company to company, generally speaking, all RH companies require vehicles to be four door, and relatively new. We take advantage this to provide further evidence in support of a quality channel.

In Table 12, we utilize the vehicle level dataset provided by NHTSA to explore this notion. Panel A columns (1) and (2) presents estimates from models at the crash-vehicle level. We use linear probability models where the dependent variables are indicators for whether the crash vehicle has RH-eligible characteristics. We define a RH-eligible vehicle as a four door sedan, SUV or minivan that is less than 5 or 10 years old. We include *City x Year*, *City x Day of Week* and *City x Hour of Day* fixed effects in each model to account for differences in passenger flows due to work days and rush hours. The estimates demonstrate that the effect holds at similar magnitudes for accidents involving at least one RH-eligible vehicle. Column (3) presents the estimates for a placebo test where we estimate the same model, but where the dependent variable is the log of one plus the number of accidents involving a two-door vehicle (which are not eligible for RH). Here, we observe no significant effect post-RH (either economically or statistically).

Furthermore, if accidents post-RH entry are primarily RH vehicle driven (which would support the lower quality argument), we should observe more passengers per accident, as each RH vehicle involves a driver and at least one passenger, whereas before RH, there could be a single driver only. We demonstrate that this is the case in Table 12 Panel B, where we estimate our previous DD model, but where the dependent variable is log of one plus the number of passengers involved in accidents. As in our main models, we estimate three version of each model, once with no trends, once with a location-specific linear trend, and once with both linear and quadratic location-specific trends. In all three models (columns (1) to (3)) we a significant increase of between 2.5% and 6.6% in the number of passengers involved in accidents post RH entry. In columns (4) through (6), we estimate models where the dependent variable is log number of passengers per vehicle involved in accidents. In order

to calculate number of passengers per vehicle, we must restrict the sample solely to observations where there is at least one accident, resulting in a loss of some statistical power. Here too the estimates indicate an increase in the number of passengers per vehicle involved in accidents.

All this said, many of the arguments supporting the notion of improved quality concentrate primarily on reduction of drunk driving specifically. To the extent that the substitution of inebriated drivers with sober RH drivers leads to improved quality along this dimension, we would expect to see a reduction in the rate of drunk fatal accidents and fatalities following the introduction of RH. However, any such reduction may be limited or somewhat offset by incidence of drunk driving by ridehail drivers themselves (see e.g. https://fortune.com/2017/04/13/uber-drunk-drivers/), or increases in alcohol consumption more generally due to behavioral responses stemming from risk compensation (Burgdorf et al., 2019).

We proceed to break the sample of accidents into those involving a drunk driver and those that do not involve a drunk driver. We define outcome measures as follows: *Total Drunk Accidents* is the total number of fatal accidents involving any drunk drivers. *Total Drunk Fatalities* is the total number of fatalities in all drunk-driver accidents. Importantly, given that the NHTSA changed the manner in which accidents were recorded as alcohol-related in 2008, leading to a sharp *mechanical* drop in the incidence of alcohol-related accidents from 2008 onward as graphed in Figure A4.[31] We thus restrict our models to the period of 2008 to 2016 for the estimation of our models. (In the Online Appendix, we show the sensitivity of the results when including the years prior to the change in the estimation: the models document a negative coefficient when linear and quadratic trends are not included, similar to the results that have been found in past studies that did not account for trends and used the prior years; the coefficients flips to a positive once when we remove the years prior to the definition change, or when we add linear (and/or quadratic) trends, as would be expected.)

Table 13 Panel A presents the estimates from our difference-in-differences models for these measures for the post-definition change period. Column (1) presents estimates for the model without city-specific trends: we observe a positive, but statistically insignificant coefficient. Once we include linear or linear and quadratic trends, we observe a positive and weakly statistically significant coefficient, on the order of roughly 1.5% increase in fatal accidents post RH introduction.

---

[31] From 1999 through 2007, alcohol-related accidents were recorded as any fatal accident involving at least one vehicle occupant (driver or non-driver) or pedestrian being impaired (in the legal sense). From 2008 onwards, alcohol-related accidents are recorded as such only if the *driver* was impaired.

Thus, the estimates suggest that any quality improvements from RH in the form of removal of drunk drivers from the road do not seem to swamp the quantity effect or other quality effects which we cannot directly observe.[32]

While we cannot offer evidence on whether some of the effects we document are driven directly by a reduction in overall driver quality, the results above suggest that there is not a significant quality improvement post-RH, at least in terms of removal of drunk drivers, to reduce the number of accidents overall.

7. DISCUSSION AND COST ESTIMATION

Up until this point, our study has documented a societal cost associated with the introduction of RH. To make a welfare calculation, we must also consider its benefits. Benefits come from, for example, the consumer surplus provided by convenience. Cohen et al. (2018) use Uber's "surge" pricing algorithm and the richness of its individual-level data to estimate demand elasticities at several points along the demand curve and then use these elasticity estimates to estimate consumer surplus. They estimate that, in 2015, the UberX service generated about $2.9 billion in consumer surplus in the four U.S. cities they examine. Moreover, their calculations suggest that the overall consumer surplus generated by UberX in the United States in 2015 was $6.8 billion.

Our estimates allow us to attempt to quantify the cost of the RH's increase in fatal accidents, using estimates of the value of a statistical life. Assuming RH services are eventually made available across the entire United States, we can do a back-of-the-envelope calculation of the costs of the increase in accidents we document. In 2010, the year before RH began, there were 32,885 motor vehicle fatalities in the U.S.[33] The 3% annual increase associated with the introduction of RH in fatalities represents an additional 987 lives lost each year.[34] The U.S. Department of Transportation estimates the value of a statistical life (VSL) at $9.6 million for 2015; the DOT recommends analysts use a test range of $5.4 million (low) to $13.4 million (high) in 2015 dollars. Applying the VSL and assuming an annual increase of 987 lives lost per year, the annual cost of the increase in fatalities

---

[32] The fact that there is some increase is not surprising or contradictory to our findings. There is a robust legal industry dedicated to lawsuits involving drunk RH drivers. For the period 8/2014 to 8/2015 in CA alone, Uber faced fines for 151 separate violations for drivers driving under the influence (and that is for drivers who get caught; the number that manage to avoid detection is likely higher). RH companies in some states are exempted from drug and alcohol screening programs imposed in other commercial driving operations. Media reports (e.g. https://www.sfgate.com/business/article/California-tells-Uber-it-s-sloppy-about-11069749.php) suggest that impaired driving by RH driver-partners is widespread. Moreover, we note that given the overall increase in VMT, there should also be an increased chance that a sober driver might hit or be hit by any remaining impaired drivers on the road.

[33] https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811552

[34] We round the estimated number of fatalities to the nearest whole number.

associated with RH can be estimated as roughly $9.48 billion per year, with a range of $5.33 billion (low) to $13.24 billion (high).

A comparison of our cost estimate with Cohen et al.'s (2018) estimates of consumer surplus generated by RH services suggests that the costs from the increase in fatal accidents match or surpass the benefits of convenience to direct consumers of RH. Our estimates, moreover, do not include the costs imposed by nonfatal accidents, for which data is not readily available. We can assume that the pattern for fatal accidents is repeated for nonfatal accidents, leading to costs in material and healthcare that may dwarf these VSL estimates. The incremental cost derives from the externalities associated with driving and traffic congestion, where riders of RH do not bear the full cost of being on the road—some of this cost is borne by pedestrians, as we document above.

Of course, consumer surplus is not the only benefit generated by RH services, and thus conducting overall welfare calculations would require more research on the overall impact of RH in the economy. Our study documents only one particular social cost associated with RH, much as Cohen et al. (2018) documents a particular type of surplus. Our findings, however, suggest significant additional costs beyond the loss of life associated with increased traffic fatalities. Nationally, the number of traffic accidents in which individuals are injured is an order of magnitude higher than the number of those in which there is a fatality. Detailed data on such accidents is generally unavailable, but the economic and societal cost of all accidents in the United States in 2010 totaled $836 billion (https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812013). An increase in non fatal accidents is also likely to be present here, with large associated societal costs.

One potential suggested solution to the allocation problem imposed by the increase in accidents associated with RH is insurance. Insurance alters the economic incentives of agents by internalizing some of the costs. Here, however, the externalities we discuss will only be internalized in the presence of a compulsory insurance monopoly. To date, however, RH companies such as Uber self-insure, and as a result, may not fully internalize this cost in the price of their service. Moreover, IPO incentives in the near future for RH companies may cause RH firms to under-insure, as provision for this cost may deteriorate financial performance. Furthermore, some of the costs involved under the current regime will be imposed on non-RH users/drivers, through increases in insurance premiums that result from the overall increase in accident probability.

Additionally, our findings suggest an increase in road utilization and congestion that imposes additional costs on society. While an increase in congestion may impose incremental costs on individuals driving to work or to a social event, it can impose much greater costs on first responders

and those being assisted by them (Beland and Brent, 2018). For illustration, suppose there are 100 heart attack victims transported to the emergency room each day by car or ambulance. These individuals face much higher costs from congestion and road delays. As congestion increases, a higher proportion of these 100 cases may encounter a delay in receiving life-saving medical attention. The disutility of the externality imposed by congestion is heterogeneous, however, unlike, say, the case of congestion in broadband telecommunication services, it is not straightforward to solve this with differential pricing. Other costs, such as pollution, are also not solved by insurance.

In sum, our findings suggest more research will be needed to better quantify both the societal cost and benefit of RH. More generally, our work points to the need for better consideration of societal costs and externalities associated with the introduction of new technologies.

8. CONCLUSION

Beginning in the mid-1980s the United States experienced a dramatic decrease in fatal accidents per capita and per vehicle mile driven. In 2010, 32,885 people died in motor vehicle traffic crashes in the United States—the lowest number of fatalities since 1949 (NHTSA, 2012). This decline halted and then reversed shortly after the introduction of RH into U.S. cities. This increase has not been restricted to occupants of motor vehicles; the Governors Highway Safety Association recently noted that the 2018 pedestrian fatality pedestrian fatality figure was at its highest since 1990 and 35 percent higher than it was 10 years ago, reversing a longstanding trend of decline in pedestrian deaths from motor vehicle crashes. In 2017, the NHTSA noted:

> There were 37,461 people killed in crashes on U.S. roadways during 2016, an increase from 35,485 in 2015. … Fatalities increased from 2015 to 2016 in almost all segments of the population—passenger vehicle occupants, occupants of large trucks, pedestrians, pedal cyclists, motorcyclists, alcohol-impaired driving, male/female, and daytime/nighttime … [W]ith the large increases in fatalities in 2015 and 2016, [the] decade-long downward trend of 21 percent has been reduced by more than one-third.

In this paper, we provide evidence consistent with RH imposing an increase in fatal accidents and fatalities on the motor vehicle occupants and pedestrians in the cities it serves. We document a roughly 2% to 4% increase in the number of fatal accidents, which persists throughout the week, on weekends, at night, and on weekend nights. We develop a conceptual framework for analyzing how the introduction of RH may affect accident rates, which models accidents as a function of vehicle

miles traveled and average driver quality. We document increases in the intensive margin of quantity. For example, VMT, measures of excess gas consumption, and annual hours spent in traffic rise following the entry of RH. Furthermore, at the extensive margin, we find a 3% increase in new car registrations. Consistent with our estimates for fatal accident rates, this increase in registrations is more substantial in cities with high ex-ante use of public transportation, consistent with the findings of Hall et al. (2018), who find that RH serves as a complement to public transportation by serving to fill last-mile transportation needs.

While our documented effects alone are unlikely to fully explain the reversal of accident rate trends in recent years, they are worth further investigation and discussion. Moreover, while RH appears associated with more motor vehicle deaths, it does also bring many benefits. These include improved mobility for the disabled and minorities, flexible job opportunities that are especially valuable to those otherwise at high risk of unemployment, and customer convenience and resulting consumer surplus.

Still, the annual cost in human lives is nontrivial, and it is higher than estimates for annual consumer surplus generated. And on top of this, our estimates do not include the costs imposed by nonfatal accidents, for which data is not readily available. We can assume that the pattern for fatal accidents is also repeated for nonfatal accidents, leading to costs in material and healthcare that may dwarf the costs in human lives. An essential contribution of our study is to point to the need for further research and debate about the overall cost-benefit tradeoff of RH and ways to increase the benefits or reduce the costs. Further research on this issue will likely necessitate unrestricted access to private data generated by RH companies.

Finally, given the relatively short period in which RH has existed, our results are necessarily short term. The long-term consequences of RH may differ, as individuals may change behavior as time passes. For example, some drivers may exit the market, and those who remain may gain knowledge and improve their driving with the platforms. Additionally, as competition increases in the market, the massive subsidies provided by RH companies for drivers and riders may decline, reducing the number of riders. If usage of pooled ride services increases, car utilization may rise, lowering the number of vehicle miles traveled overall. Thus, any regulatory actions should proceed cautiously, considering the short-term effects of RH documented here, the real and potential benefits and the necessity for further research on the outcomes.

9. ACKNOWLEDGMENTS

10. FUNDING

11. CONFLICTS OF INTEREST

The authors have no known conflicts of interest associated with this paper.

REFERENCES

Beland, Louis-Philippe, and Daniel Brent. "Traffic Congestion and the Performance of First Responders: Evidence from California Fire." (2018).

Berger, Thor, Chinchih Chen, and Carl Benedikt Frey. "Drivers of disruption? Estimating the Uber effect." European Economic Review 110 (2018), 197-210.

Brazil, Noli, and David S. Kirk. "Uber And Metropolitan Traffic Fatalities in the United States," American Journal of Epidemiology, 184 (2016), 192–198.

Buchak, Greg. "Financing the Gig Economy." University of Chicago Booth Working Paper, 2018.

Burgdorf, Jacob, Conor Lennon, and Keith Teltser. "Do Ridesharing Services Increase Alcohol Consumption?." Working Paper, 2019.

Chen, M, Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen, "The Value of Flexible Work: Evidence from Uber Drivers," NBER Working Paper No. w23296, 2017.

Clewlow, Regina R., and Gouri Shankar Mishra, "Disruptive Transportation: The Adoption, Utilization and Impacts of Ride-Hailing in the United States," Institute of Transportation Studies, University of California, Davis, Working Paper, 2017.

Circella, Giovanni, Farzad Alemi, Kate Tiedeman, Susan Handy, and Patricia Mokhtarian, "The Adoption of Shared Mobility in California and Its Relationship with Other Components of Travel Behavior," Institute of Transportation Studies, University of California, Davis Working Paper, 2018.

Cohen, Peter, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe, "Using Big Data to Estimate Consumer Surplus: The Case of Uber," NBER Working Paper No. w22627, 2016.

Cook, Cody, Rebecca Diamond, Jonathan Hall, John A. List, and Paul Oyer, "The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers," Stanford Graduate School of Business Working Paper, 2018.

Cramer, Judd, and Alan B. Krueger, "Disruptive Change in the Taxi Business: The Case of Uber," American Economic Review, 106 (2016), 177–182.

Dills, Angela K., and Sean E. Mulholland, "Ride‐Sharing, Fatal Crashes, and Crime," Southern Economic Journal, 84 (2018), 965–991.

Edlin, Aaron S., "Per-Mile Premiums for Auto Insurance." In Economics for an Imperfect World: Essays in Honor of Joseph E. Stiglitz, edited by Richard Arnott, Bruce Greenwald, Ravi Kanbur, and Barry Nalebuff. Cambridge, MA: MIT Press, 2003.

Edlin, Aaron S., and Pinar Karaca-Mandic, "The Accident Externality from Driving." Journal of Political Economy, 114 (2006), 931-955.

Ge, Yanbo, Christopher R. Knittel, Don MacKenzie, and Stephen Zoepf. "Racial and gender discrimination in transportation network companies." NBER Working Paper No. w22776, 2016.

Goodman-Bacon, Andrew. "Difference-in-differences with variation in treatment timing." NBER Working Paper No. w25018, 2018.

Greenwood, Brad N., and Sunil Wattal, "Show Me the Way to Go Home, An Empirical Investigation of Ride-Sharing and Alcohol Related Motor Vehicle Fatalities," MIS Quarterly, 41 (2017), 163–187.

Grübler, Arnulf, "Diffusion: Long-Term Patterns and Discontinuities," Technological Forecasting and Social Change 39 (1991), 159–180.

Haggag, Kareem, Brian McManus, and Giovanni Paci, "Learning by Driving: Productivity Improvements by New York City Taxi Drivers," American Economic Journal: Applied Economics, 9 (2017), 70–95.

Hall, Jonathan D., Craig Palsson, and Joseph Price. "Is Uber a Substitute or Complement for Public Transit?," Journal of Urban Economics, 108 (2018), 36-50.

Hall, Jonathan V., and Alan B. Krueger, "An Analysis of The Labor Market for Uber's Driver-Partners in the United States," ILR Review, 71 (2017), 705–732.

Henao, Alejandro, "Impacts of Ridesourcing – Lyft and Uber – on Transportation including VMT, Mode Replacement, Parking, and Travel Behavior," Working Paper, 2017.

Hasan, Sharique and Kumar, Anuj, "Digitization and Divergence: Online School Ratings and Segregation in America", SSRN Working Paper, 2018.

Klenow, Peter, and Andres Rodriguez-Clare, "Externalities and Growth," Handbook of Economic Growth, 1 (2005), 817-861

Landier, Augustin, Daniel Szomoru, and David Thesmar, "Working in the On Demand Economy: An Analysis of Uber-Driver Partners in France," MIT Sloan School of Management Working Paper, 2016.

Liu, Meng, Erik Brynjolfsson, and Jason Dowlatabadi, "Technology, Incentives, and Service Quality, the Case of Taxis and Uber," Working Paper, 2018.

Mangrum, Daniel, and Alejandro Molnar. "The marginal congestion of a taxi in New York City." Vanderbilt University Working Paper, 2017.

Martin-Buck, Frank, "Driving Safety, An Empirical Analysis of Ridesharing's Impact on Drunk Driving and Alcohol-Related Crime," University of Texas at Austin Working Paper, 2017.

Metropolitan Area Planning Council, "Fare Choices: A Survey of Ride-Hailing Passengers in Metro Boston," February 2018.

Mokyr, Joel, "Punctuated Equilibria and Technological Progress," The American Economic Review, 80 (1990), 350–354.

New York City Department of Transportation, "NYC Mobility Report," June 2018.

Nie, Yu Marco, "How Can the Taxi Industry Survive the Tide of Ridesourcing? Evidence from Shenzhen, China," Transportation Research Part C: Emerging Technologies, 79 (2017), 242–256.

Parente, Stephen L., and Edward C, Prescott, "Barriers to Technology Adoption and Development," Journal of Political Economy, 102 (1994), 298–321.

Peltzman, Sam, "The Effects of Automobile Safety Regulation," Journal of Political Economy, 83 (1975), 677–725.

Peltzman, Sam, "A Reply," Journal of Economic Issues, 11 (1977), 672–678.

Pew Research Center, May, 2016, "Shared, Collaborative and On Demand: The New Digital Economy."

Rosenberg, Nathan, Luther E. Birdzell, and Glenn W. Mitchell, How the West Grew Rich, Mumbai: Popular Prakashan, 1986.

Retting, Richard. "Pedestrian Traffic Fatalities by State." Governors Highway Safety Association , 2019

Schaller, Bruce, "The New Automobility: Lyft, Uber and the Future of American Cities," Schaller Consulting Report, 2018.

Vickrey, William, "Automobile Accidents, Tort Law, Externalities, and Insurance: An Economist's Critique." Law and Contemporary Problems 33 (1968), 464–87.

Table 1
Summary Statistics

	Mean	Median	Std. Dev.
Population (thousands)	54.50	23.51	199.99
Income per capita (thousands $)	39.72	37.47	12.17
Population density	2,999.28	2,169.70	3,159.57
Carpool usage	10.62	10.05	3.98
Public transportation usage	2.96	1.19	4.96
Household vehicle ownership (thousands)	32.72	15.45	80.62
New car registration	670	264	2,340
Accident rate	3.49	0.96	5.66
Fatality rate	3.84	0.98	6.51
Pedestrian-involved accident rate	0.58	0.00	1.80
Pedestrian-involved fatality rate	0.59	0.00	1.86

Notes: The sample contains 190,080 quarterly observations on 2,970 census incorporated places from 2001 to 2016. Population density measures population per square mile. Carpool usage measures the percentage of population commuting to work using a carpool. Public transportation usage measures the percentage of population commuting to work using public transportation. Household vehicle ownership measures the total number of available vehicles in households. New car registration measures the total number of new vehicle registrations. All rates are measured per 100,000 population. Accident is the number of fatal accidents, according to the definition used by NHTSA. Fatality is the total number of fatalities across all fatal accidents. Pedestrian-involved accident is the number of fatal accidents involving at least one pedestrian. Pedestrian-involved fatalities is the total number of fatalities in all accidents involving at least one pedestrian.

Table 2
Effect of Ridehailing on Traffic Safety

Panel A: Overall Effect

	Log (1+Total Accidents)			Log (1+Total Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Post_t * Treated_c$	0.0191***	0.0360***	0.0332***	0.0183***	0.0360***	0.0335***
	(0.0062)	(0.0074)	(0.0091)	(0.0065)	(0.0077)	(0.0095)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.61	0.62	0.63	0.60	0.60	0.61

Panel B: Variation in Empirical Specifications

	Variations in Accident Specifications			Variation in Fatality Specifications		
	(1)	(2)	(3)	(4)	(5)	(6)
	Weighted by Population	Accident Rate	$\sinh^{-1}$(Accidents)	Weighted by Population	Fatality Rate	$\sinh^{-1}$(Fatalities)
$Post_t * Treated_c$	0.0610***	0.2219***	0.0443***	0.0624***	0.2222***	0.0439***
	(0.0128)	(0.0669)	(0.0095)	(0.0125)	(0.0749)	(0.0099)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.92	0.40	0.61	0.91	0.38	0.59

Panel C: CBSA Aggregation

	Cluster Standard Errors at CBSA Level		Diff-in-Diff at CBSA Level	
	(1)	(2)	(3)	(4)
	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Total Accidents)	Log (1+Total Fatalities)
$Post_t * Treated_c$	0.0360***	0.0360***	0.0524***	0.0579***
	(0.0095)	(0.0097)	(0.0148)	(0.0153)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Observations	190,080	190,080	52,928	52,928
R2	0.62	0.60	0.81	0.79

Notes: This table presents results from generalized difference-in-difference regressions. The dependent variables are listed at the top of each column. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. Panel A presents the overall effect of ridehailing on two traffic safety measures. Total accidents is the number of fatal accidents according to the definition used by NHTSA. Total fatalities is the total number of fatalities across all fatal accidents. Panel B presents population-weighted regression results, and regression results using accident and fatalities rates (per 100,000) and the inverse hyperbolic sine of accidents and fatalities. In Panel C column (1) and (2), we run the same specification as in Panel A, but cluster standard errors at the CBSA level instead. In Panel C column (3) and (4), the difference-in-difference specifications are run at CBSA level, utilizing the earliest adoption date within the CBSA. City-specific linear trends are included in all regressions in Panel B and C. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors—clustered at city level in Panel A and B, and clustered at CBSA level in Panel C—are reported in parentheses. ∗∗∗, ∗∗, and ∗ represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 3
Effect of Ridehailing on Traffic Safety by Day and Time

Panel A: Total Accidents

	(1)	(2)	(3)	(4)
	Log (1+Weekday Accidents)	Log (1+Weekend Accidents)	Log (1+Accidents at Night)	Log (1+Accidents at Fri. and Sat. Night)
$Post_t * Treated_c$	0.0272***	0.0350***	0.0396***	0.0250***
	(0.0069)	(0.0065)	(0.0065)	(0.0054)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080
R2	0.52	0.54	0.55	0.45

Panel B: Total Fatalities

	(1)	(2)	(3)	(4)
	Log (1+Weekday Fatalities)	Log (1+Weekend Fatalities)	Log (1+Fatalities at Night)	Log (1+Fatalities at Fri. and Sat. Night)
$Post_t * Treated_c$	0.0276***	0.0356***	0.0407***	0.0269***
	(0.0072)	(0.0069)	(0.0069)	(0.0058)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080
R2	0.54	0.52	0.54	0.44

Notes: This table presents the effect of ridehailing on accidents and fatalities, respectively in Panel A and B, by day and time. The dependent variables are listed at the top of the columns. Total accidents is the number of fatal accidents according to the definition used by NHTSA. Total fatalities is the total number of fatalities across all fatal accidents. Weekday is defined as Monday through Thursday. Weekend is defined as Friday through Sunday. Night is defined as 5 pm through 2 am. Friday and Saturday Night is defined as 5 pm Friday through 6 am Saturday and 5 pm Saturday through 6 am Sunday. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 4
Dynamic Effect of Ridehailing on Traffic Safety

Ridehailing Tenure	Log (1+Total Accidents)			Log (1+Total Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
Ridehailing Tenure						
1 - 2 Quarters	0.0246***	0.0346***	0.0323***	0.0242**	0.0348***	0.0325***
	(0.0094)	(0.0100)	(0.0109)	(0.0098)	(0.0104)	(0.0114)
3 - 4 Quarters	0.0255**	0.0371***	0.0350***	0.0237**	0.0356***	0.0337**
	(0.0105)	(0.0112)	(0.0125)	(0.0110)	(0.0117)	(0.0131)
5 - 6 Quarters	0.0215*	0.0343***	0.0279*	0.0241**	0.0372***	0.0305*
	(0.0114)	(0.0125)	(0.0149)	(0.0120)	(0.0132)	(0.0157)
7 - 8 Quarters	0.0189	0.0394***	0.0295*	0.0187	0.0402***	0.0293*
	(0.0121)	(0.0130)	(0.0164)	(0.0127)	(0.0137)	(0.0173)
9 - 10 Quarters	0.0060	0.0350**	0.0213	0.0007	0.0309*	0.0164
	(0.0141)	(0.0154)	(0.0190)	(0.0147)	(0.0161)	(0.0200)
11 - 12 Quarters	-0.0004	0.0452**	0.0243	-0.0014	0.0460*	0.0253
	(0.0216)	(0.0228)	(0.0272)	(0.0226)	(0.0239)	(0.0284)
> 12 Quarters	-0.0198	0.0839**	0.0109	-0.0227	0.0826**	0.0074
	(0.0280)	(0.0350)	(0.0424)	(0.0294)	(0.0359)	(0.0444)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.61	0.62	0.63	0.60	0.60	0.61

Notes: This table presents the dynamic effects of ridehailing on traffic safety. The dependent variables are listed at the top of the columns. Total accidents is the number of fatal accidents, according to the definition used by NHTSA. Total fatalities is the total number of fatalities across all fatal accidents. Ridehailing tenure variables are dummy variables that take the value of one if ridehailing has been in effect for the specified periods of time. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ∗∗∗, ∗∗, and ∗ represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 5
Ridehailing Effects and Cellphone Usage Controls

Panel A: Total Accidents

	(1)	(2)	(3)	(5)
$Post_t * Treated_c$	0.0304***	0.0307***	0.0301***	0.0400***
	(0.0091)	(0.0091)	(0.0091)	(0.0078)
Log (Cellphone Store Gross Receipts)	0.0061			
	(0.0082)			
Log (Cellphone Store Employment)		0.0340**		
		(0.0143)		
Log (Number of Cellphone Store Establishments)			-0.0096	
			(0.0107)	
Log (Smartphone-Related Google Search Vol)				-0.0004
				(0.0023)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Observations	92,316	92,316	92,316	154,440
R2	0.67	0.67	0.67	0.63

Panel B: Total Fatalities

	(1)	(2)	(3)	(5)
$Post_t * Treated_c$	0.0309***	0.0312***	0.0307***	0.0406***
	(0.0096)	(0.0096)	(0.0096)	(0.0082)
Log (Cellphone Store Gross Receipts)	0.0073			
	(0.0087)			
Log (Cellphone Store Employment)		0.0352**		
		(0.0152)		
Log (Number of Cellphone Store Establishments)			-0.0064	
			(0.0117)	
Log (Smartphone-Related Google Search Vol)				-0.0001
				(0.0025)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Observations	92,316	92,316	92,316	154,400
R2	0.66	0.66	0.66	0.61

Notes: This table presents results from generalized difference-in-difference regressions when controlling for various additional control variables related to cellphone adoption. Gross receipts, employment, and the number of establishments are measured for industry with NAICS 517312: Wireless Telecommunication Carriers (Except Satellite) by MSA and year from 2007-2016. Log Smartphone-Related Google Search Vol is the natural logarithm of Google search volume for the terms "iPhone", "Android", "Samsung galaxy", "smartphone", and "cellphone". Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 6
Ridehailing and Smartphone Adoption

	Log (1+Total Accidents)			Log (1+Total Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Post_t * Treated_c$	0.0215**	0.0387***	0.0478***	0.0185**	0.0381***	0.0479***
	(0.0084)	(0.0097)	(0.0118)	(0.0088)	(0.0103)	(0.0125)
$Post_t * Treated_c * Cumulative\ Cellphone\ Sales$	-0.0063*	-0.0058	-0.0119**	-0.0053	-0.0055	-0.0113**
	(0.0033)	(0.0039)	(0.0050)	(0.0035)	(0.0042)	(0.0053)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	152,576	152,576	152,576	152,576	152,576	152,576
R2	0.65	0.66	0.66	0.63	0.64	0.65

Notes: This table shows how the effects of ridehailing on traffic safety varies with ex-ante cellphone adoption. Cumulative cellphone sales are the MSA-level cumulative gross receipts for industry with NAICS 517312: Wireless Telecommunication Carriers (Except Satellite) from 2007-2010. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ∗∗∗, ∗∗, and ∗ represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 7
Variation of Ridehailing Service and Adoption Intensity

Panel A: Single Ride Services vs. Pooled Ride Services

	Log (1+Total Accidents)			Log (1+Total Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
Single Ride Service (UberBlack/Taxi/X, Lyft)	0.0215***	0.0372***	0.0337***	0.0204***	0.0371***	0.0340***
	(0.0065)	(0.0075)	(0.0091)	(0.0067)	(0.0079)	(0.0096)
Pooled Ride Service (Uber Pool, Lyft Line)	-0.0233	-0.0141	-0.0191	-0.0199	-0.0130	-0.0191
	(0.0145)	(0.0150)	(0.0163)	(0.0153)	(0.0158)	(0.0172)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.61	0.62	0.63	0.60	0.60	0.61

Panel B: Google Trends Intensity

	Log (1+Total Accidents)			Log (1+Total Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Post_t * Treated_c * Log RideHail oogle \ Search Vol_{ct}$	0.0027***	0.0049***	0.0028**	0.0026***	0.0049***	0.0028**
	(0.0007)	(0.0009)	(0.0012)	(0.0008)	(0.0010)	(0.0013)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	154,440	154,440	154,440	154,440	154,440	154,440
R2	0.61	0.62	0.63	0.60	0.61	0.62

Notes: This table shows how the effect of ridehailing on traffic safety varies with the intensity of service. In all panels, the dependent variables are listed at the top of the columns. Total accidents is the number of fatal accidents, according to the definition used by NHTSA. Total fatalities is the total number of fatalities across all fatal accidents. In Panel A, Single (Pooled) Ride Service is a dummy variable that takes the value of one if any single (pooled) ride service is adopted. In Panel B, Log RideHail Google Search Volume is the natural logarithm of Google search volume for the terms "Uber," "Lyft," and "rideshare." Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 8
Externality of Ridehailing on Non-Vehicle Occupants

	(1) Log (1+Pedestrian-Involved Accidents)	(2) Log (1+Pedestrian-Involved Fatalities)	(3) Log (1+Pedestrians Involved in Fatal Accidents)
$Post_t * Treated_c$	0.0245***	0.0245***	0.0277***
	(0.0058)	(0.0058)	(0.0062)
City and Quarter Fixed Effects	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes
Observations	190,080	190,080	190,080
R2	0.54	0.54	0.55

Notes: This table presents results from generalized difference-in-difference regressions. The dependent variables are listed at the top of each column. Pedestrian is defined as any person not in or upon a motor vehicle or other vehicle. Pedestrian-involved accident measures the number of fatal accidents involving at least one non-vehicle occupants. Pedestrian-involved fatalities measures the total number of fatalities in all accidents involving at least one non-vehicle occupants. Pedestrians involved in fatal accidents measures the total number of non-vehicle occupants involved in fatal accidents. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. City-specific linear trends are included in all columns. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ∗∗∗, ∗∗, and ∗ represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 9
Heterogeneous Effects of Ridehailing on Traffic Fatalities

Panel A: City Characteristics

	Population				Income Inequality (Gini Index)				Population Density			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Pedestrian-Involved Accidents)	Log (1+Pedestrian-Involved Fatalities)	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Pedestrian-Involved Accidents)	Log (1+Pedestrian-Involved Fatalities)	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Pedestrian-Involved Accidents)	Log (1+Pedestrian-Involved Fatalities)
$Post_t * Treated_c * Q4$	0.0746***	0.0772***	0.0632***	0.0640***	0.0510***	0.0521***	0.0476***	0.0478***	0.0354***	0.0347***	0.0173*	0.0176*
	(0.0113)	(0.0117)	(0.0100)	(0.0102)	(0.0123)	(0.0128)	(0.0117)	(0.0119)	(0.0109)	(0.0113)	(0.0089)	(0.0091)
$Post_t * Treated_c * Q3$	0.0049	0.0038	-0.0046	-0.0047	0.0342**	0.0351**	0.0255**	0.0265**	0.0323**	0.0327**	0.0346***	0.0342***
	(0.0136)	(0.0144)	(0.0095)	(0.0096)	(0.0144)	(0.0149)	(0.0112)	(0.0114)	(0.0136)	(0.0143)	(0.0108)	(0.0110)
$Post_t * Treated_c * Q2$	0.0041	0.0011	-0.0156*	-0.0159*	0.0283**	0.0251*	0.0207**	0.0212**	0.0504***	0.0481***	0.0527***	0.0541***
	(0.0154)	(0.0161)	(0.0090)	(0.0091)	(0.0139)	(0.0146)	(0.0101)	(0.0102)	(0.0163)	(0.0169)	(0.0131)	(0.0134)
$Post_t * Treated_c * Q1$	0.0035	0.0007	-0.0025	-0.0018	0.0281**	0.0293**	-0.0013	-0.0014	0.0265	0.0303*	-0.0077	-0.0068
	(0.0128)	(0.0132)	(0.0075)	(0.0077)	(0.0141)	(0.0147)	(0.0098)	(0.0098)	(0.0165)	(0.0176)	(0.0118)	(0.0120)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.62	0.60	0.54	0.54	0.62	0.60	0.54	0.54	0.62	0.60	0.54	0.54

Panel B: Ex-ante Behavior

	Ex-ante Vehicle Ownership				Ex-ante Public Transportation Usage (% of Population)				Ex-ante Car Pool Usage (% of Population)			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Pedestrian-Involved Accidents)	Log (1+Pedestrian-Involved Fatalities)	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Pedestrian-Involved Accidents)	Log (1+Pedestrian-Involved Fatalities)	Log (1+Total Accidents)	Log (1+Total Fatalities)	Log (1+Pedestrian-Involved Accidents)	Log (1+Pedestrian-Involved Fatalities)
$Post_t * Treated_c * Q4$	0.0769***	0.0793***	0.0656***	0.0664***	0.0381***	0.0380***	0.0434***	0.0431***	0.0448***	0.0446***	0.0310**	0.0317**
	(0.0111)	(0.0115)	(0.0099)	(0.0101)	(0.0112)	(0.0117)	(0.0096)	(0.0097)	(0.0151)	(0.0162)	(0.0126)	(0.0128)
$Post_t * Treated_c * Q3$	0.0024	0.0006	-0.0061	-0.0063	0.0537***	0.0566***	0.0273***	0.0289***	0.0633***	0.0622***	0.0469***	0.0468***
	(0.0145)	(0.0153)	(0.0108)	(0.0109)	(0.0123)	(0.0128)	(0.0102)	(0.0103)	(0.0131)	(0.0134)	(0.0113)	(0.0114)
$Post_t * Treated_c * Q2$	-0.0054	-0.0068	-0.0099	-0.0097	0.0175	0.0130	0.0190*	0.0176	0.0174	0.0201	0.0124	0.0130
	(0.0148)	(0.0158)	(0.0097)	(0.0098)	(0.0155)	(0.0163)	(0.0115)	(0.0117)	(0.0129)	(0.0134)	(0.0094)	(0.0095)
$Post_t * Treated_c * Q1$	0.0136	0.0110	-0.0097	-0.0092	0.0181	0.0192	-0.0290**	-0.0264**	0.0120	0.0106	0.0021	0.0025
	(0.0136)	(0.0141)	(0.0071)	(0.0071)	(0.0190)	(0.0199)	(0.0116)	(0.0119)	(0.0128)	(0.0135)	(0.0094)	(0.0097)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.62	0.60	0.54	0.54	0.62	0.60	0.54	0.54	0.62	0.60	0.54	0.54

Notes: This table presents heterogeneous effects of ridehailing on traffic safety. The dependent variables are listed at the top of each column. Panels A and B break out results across a variety of city characteristics and ex-ante behaviors, respectively. The variable used for sample cut is listed at the top of each panel. Population measures annual city population. Income inequality is measured using city Gini index. Population density measures population per square mile. Vehicle ownership measures the total number of available vehicles in households. Public transportation usage measures the percentage of population commuting to work using public transportation. Carpool usage measures the percentage of population commuting to work using carpool. Total accidents is the number of fatal accidents according to the definition used by NHTSA. Total fatalities is the total number of fatalities across all fatal accidents. Pedestrian-involved accident measures the number of fatal accidents involving at least one pedestrian. Pedestrian-involved fatalities measures the total number of fatalities in all accidents involving at least one pedestrian. The independent variables of interest are the interaction of $Post_t * Treated_c$, a dummy variable that equals one if city c adopted at least one ridehailing service at time t, and an indicator for the quartile the observation falls in. In addition to the natural logarithm of population and the natural logarithm of income per capita, all interacted variables are included separately as control variables. All columns include city-specific linear trends. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 10
Effect of Ridehailing on Road Utilization and Congestion

	(1) Log Arterial Street VMT	(2) Log Excess Fuel Consumption	(3) Log Hours of Delay in Traffic
$Post_t * Treated_u$	0.01799**	0.01562**	0.01563**
	(0.00834)	(0.00715)	(0.00715)
Urban Area and Year Fixed Effects	Yes	Yes	Yes
Urban Area Linear Trend	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes
Observations	1,386	1,386	1,386
R2	0.998	0.999	0.999

Notes: The sample contains 1,386 annual observations on 99 urban areas from 2001 to 2014. The dependent variables are the natural logarithm of congestion-related measures listed at the top of each column. Arterial Street VMT measures the total number of vehicle miles traveled on arterial streets in an urban area. Excess fuel consumption measures the extra fuel consumed, due to inefficient operation in slower stop-and-go traffic. Hours of delay measures the amount of extra time spent traveling, due to congestion. $Post_t * Treated_u$ is a dummy variable that equals one if urban area u adopted at least one ridehailing service at year t. Urban area-specific linear trends are included in all regressions. Control variables include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the urban area level, are reported in parentheses. For more detailed information on the dependent variables, please refer to https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-scorecard-2015-wappx.pdf. ∗∗∗, ∗∗, and ∗ represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 11
The Effect of Ridehailing on New Car Registrations

Panel A: Overall Effect

	Log (1+New Car Registrations)		
	(1)	(2)	(3)
$Post_t * Treated_c$	0.0291***	0.0490***	0.0296***
	(0.0079)	(0.0069)	(0.0065)
City and Quarter Fixed Effects	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes
City Quadratic Trend	No	No	Yes
Control Variables	Yes	Yes	Yes
Observations	190,080	190,080	190,080
R2	0.94	0.97	0.98

Panel B: Intensity of Effect

	Log (1+New Car Registrations)	
	(1)	(2)
Google Search Volume		
$Post_t * Treated_c * Log\ RideHail\ Google\ Search\ Vol_{ct}$	0.0075***	
	(0.0009)	
Ridehailing Service Type		
Single Ride Service (UberBlack/Taxi/X, Lyft)		0.0463***
		(0.0067)
Pooled Ride Service (UberPool, Lyft Line)		0.0300***
		(0.0107)
City and Quarter Fixed Effects	Yes	Yes
City Linear Trend	Yes	Yes
Control Variables	Yes	Yes
Observations	154,440	190,080
R2	0.97	0.97

Panel C: Heterogeneous Effects

	(1)	(2)	(3)	(4)	(5)	(6)
Dep: Log (1+ New Car Registration)	Population	Income Inequality (Gini Index)	Pop Density	Public Transport	Carpool	Vehicle Ownership
$Post_t * Treated_c * Q4$	0.0842***	0.0415***	0.0926***	0.0620***	0.1572***	0.0809***
	(0.0096)	(0.0115)	(0.0118)	(0.0110)	(0.0144)	(0.0093)
$Post_t * Treated_c * Q3$	0.0347**	0.0470***	0.0554***	0.0673***	0.0674***	0.0427***
	(0.0142)	(0.0095)	(0.0110)	(0.0123)	(0.0108)	(0.0132)
$Post_t * Treated_c * Q2$	0.0149	0.0456***	0.0168	0.0299**	0.0195*	-0.0001
	(0.0168)	(0.0123)	(0.0118)	(0.0142)	(0.0109)	(0.0195)
$Post_t * Treated_c * Q1$	0.0041	0.0704***	-0.0274	-0.0020	-0.0619***	0.0219
	(0.0147)	(0.0164)	(0.0186)	(0.0194)	(0.0127)	(0.0155)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.97	0.97	0.97	0.97	0.97	0.97

Notes: This table presents the effect of ridehailing on new car registrations. In all panels, the dependent variables are the natural logarithm of one plus new car registrations. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. Panel A presents results from generalized difference-in-difference regressions without city time trend, with linear trend, and with quadratic trend. Panel B shows how the effect varies with the intensity of ridehailing service. Log RideHail Google Search Volume is the natural logarithm of Google search volume for the terms "Uber," "Lyft," and "rideshare". Single (Pooled) Ride Service is a dummy variable that takes the value of one if any single (pooled) ride service is adopted. Panel C breaks out results across a variety of city characteristics and ex-ante behaviors. The variable used for sample cut is listed at the top of each column. Population measures annual city population. Income inequality is measured using city Gini index. Population density measures population per square mile. Pop Density measures the population per square mile. Public Transport measures the percentage of the population commuting to work using public transportation. Carpool measures the percentage of the population commuting to work using carpools. Vehicle Ownership measures the total number of available vehicles in households. In addition to the natural logarithm of population and the natural logarithm of income per capita, all interacted variables are included separately as control variables. In Panel B and C, all regressions include city linear time trend. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 12
Ridehailing Eligible and Non-Eligible Vehicles

Panel A: Uber-Eligible Vehicles

	Characteristics of RH-Eligible Vehicles		Placebo (Non-Eligible)
	(1)	(2)	(3)
	5-Year or Newer (4 door)	10-Year or Newer (4 door)	2-Door
$Post_t * Treated_c$	0.0293***	0.0246**	0.0021
	(0.0099)	(0.0116)	(0.0057)
City X Year FE	Yes	Yes	Yes
City X Day of Week FE	Yes	Yes	Yes
City X Hour of Day FE	Yes	Yes	Yes
Observations	397,839	397,839	397,839
R2	0.23	0.24	0.22

Panel B: Passengers Involved in Accidents

	Log (1+Passengers)			Log (Passengers Per Vehicle)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Post_t * Treated_c$	0.0253**	0.0657***	0.0499***	0.0178*	0.0419***	0.0185
	(0.0106)	(0.0125)	(0.0151)	(0.0093)	(0.0110)	(0.0132)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	190,080	190,080	190,080	97,599	97,599	97,599
R2	0.52	0.54	0.54	0.08	0.12	0.16

Notes: This table presents results from generalized difference-in-difference regressions where the dependent variables for each regression are listed at the top of the columns. In Panel A, the unit of observation is crash-vehicle, and the outcomes are indicators for whether the vehicle involved in the accident is newer than 5 years old (4 door vehicle), newer than 10 years old (4 door vehicle), and 2-door respectively in columns (1), (2), and (3). In Panel B, Passengers is the total number of persons in motor vehicles involved in fatal accidents. Passenger Per Vehicle is the average number of persons per motor vehicle involved in fatal accidents. Control variables include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 13

The Effect of Ridehailing on Alcohol-Involved Accidents and Fatalities

|	Log (1+Drunk Accidents)			Log (1+Drunk Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Post_t * Treated_c$	0.0059	0.0146**	0.0166*	0.0052	0.0148**	0.0173*
	(0.0050)	(0.0070)	(0.0087)	(0.0054)	(0.0074)	(0.0093)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	No	Yes	Yes	No	Yes	Yes
City Quadratic Trend	No	No	Yes	No	No	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	106,920	106,920	106,920	106,920	106,920	106,920
R2	0.46	0.48	0.50	0.45	0.47	0.49

Notes: This table presents coefficient estimates from generalized difference-in-difference regressions. The dependent variables are listed at the top of each column. Drunk accidents is the number of fatal accidents involving any drunk drivers. Drunk fatalities is the total number of fatalities in all drunk accidents. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: U.S. Motor Vehicle Death per VMT, Death per Capita, Total Death, VMT and Population
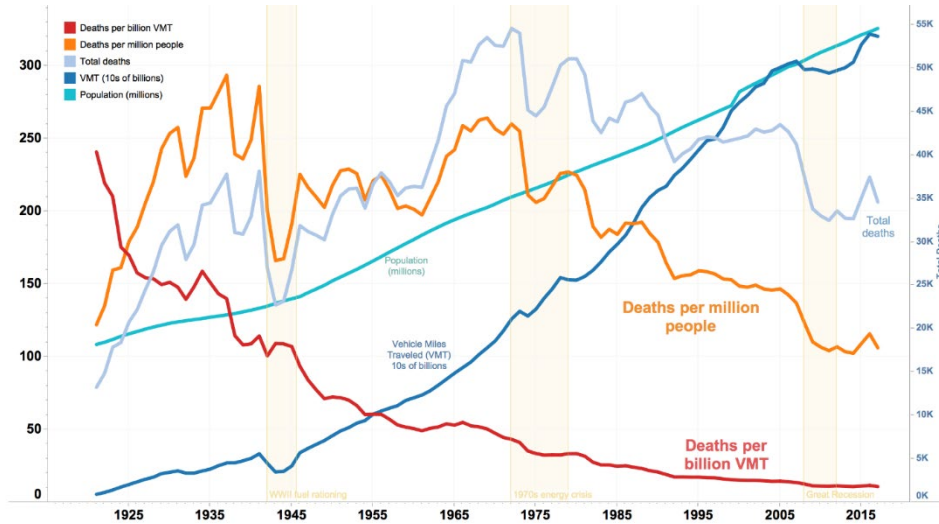

Panel B: U.S. Pedestrian Fatalities
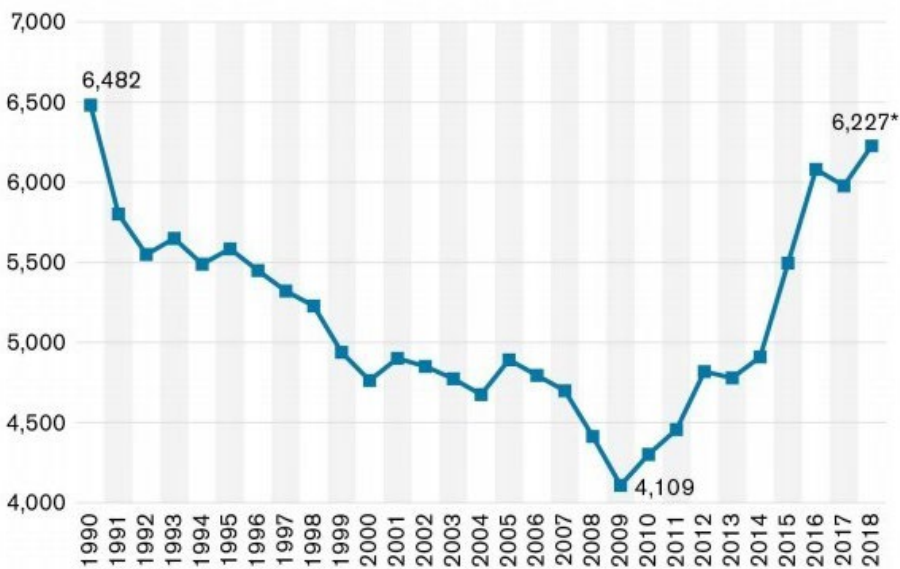
* 2018 estimate based on preliminary data and historical trends
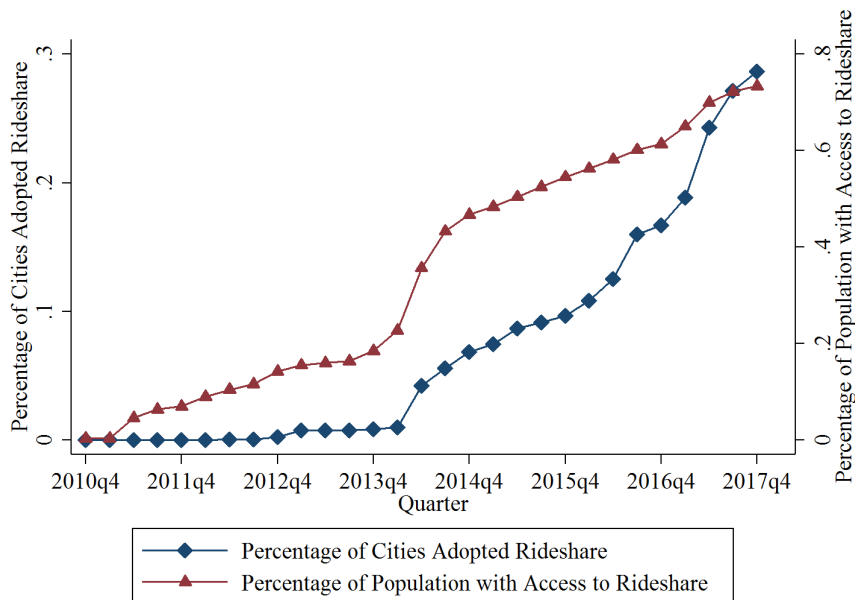
Figure 1
U.S. Traffic Safety Trends

Panel A of this figure was produced by Dennis Bratland and is reproduced here under creative commons license. It uses NHTSA FARS and CrashStats data to depict total U.S. motor vehicle deaths, deaths per VMT, deaths per capita, VMT and population for the period of 1920–2017. Panel B is excerpted from the Governors Highway Safety Association (2019) report on 2018 Pedestrian Traffic Fatalities by State. It is constructed using data from FARS and State Highway Safety Organizations for the period of 1990–2018.
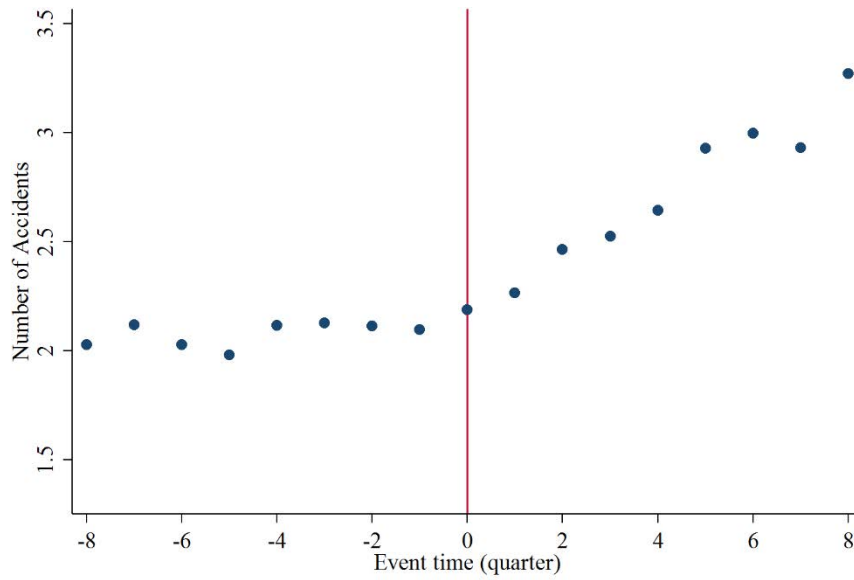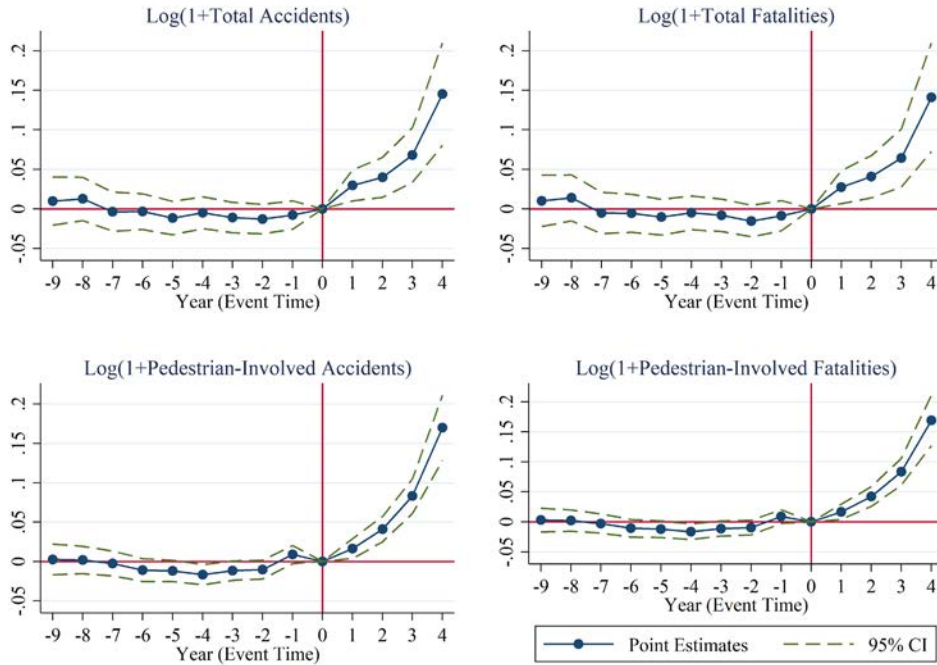
Figure 2
Average Accidents for Treated Cities in Event Time

This figure graphs the average level of accidents for treated cities in event time. The quarter of ridehailing entry is indicated by the red vertical line at event time zero.
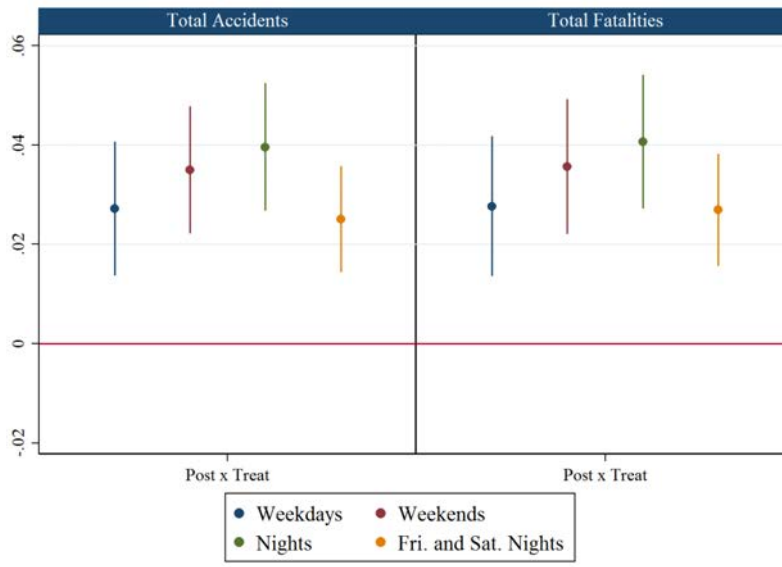

Figure 3
Ridehailing Diffusion

This figure shows the diffusion of ridehailing across the U.S. by cities and population. The sample consists of all census incorporated places in the United States. The navy (red) line graphs the percentage of cities (population) that adopted ridehailing in each quarter between the fourth quarter of 2010 and the fourth quarter of 2017.
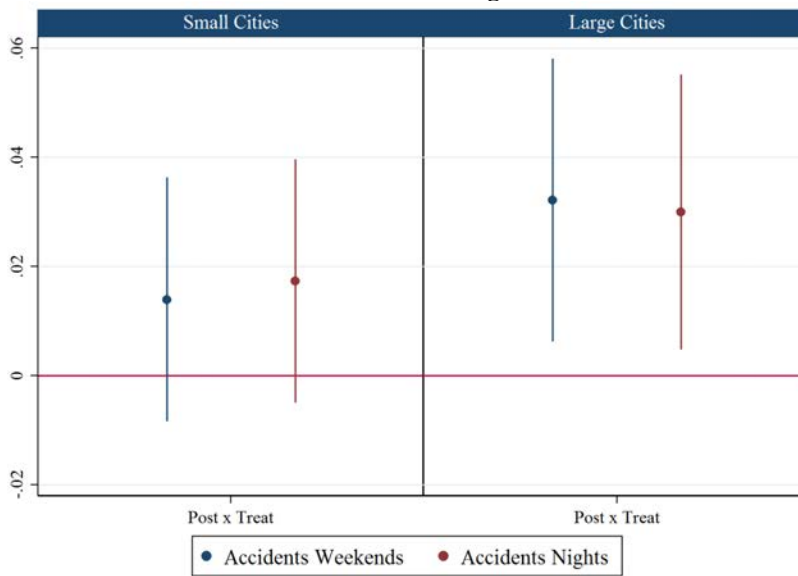
49

Figure 4
Difference-in-Differences Estimators

This figure displays the regression coefficient estimates and two-tailed 95% confidence intervals based on standard errors clustered at the city level. To map out the pattern in the counterfactual treatment effects, we regress the various outcome measures on lag and lead indicators (bunched by four quarters) for the entry of ridehailing. The vertical red line indicates the quarter of entry. We provide a description of the variables in section 2.

Panel A: Accidents and Fatalities on Weekend and at Nights

Panel B: Small vs. Large Cities

Figure 5
Nights and Weekends

This figure displays the regression coefficient estimates and two-tailed 95% confidence intervals based on standard errors clustered at the city level, broken down by accidents at night and on the weekend. Panel A presents estimates for accidents and fatalities, while Panel B presents the coefficients for accidents separately for small and large cities. We provide a description of the variables in Section 2.

51

Panel A: City Characteristics

Population | Income Inequality | Population Density

● Total Accidents ● Pedestrian Accidents

Panel B: Ex-Ante Behaviors

Vehicle Ownership | Ex Ante Public Transport | Ex Ante Car Pooling

● Total Accidents ● Pedestrian Accidents
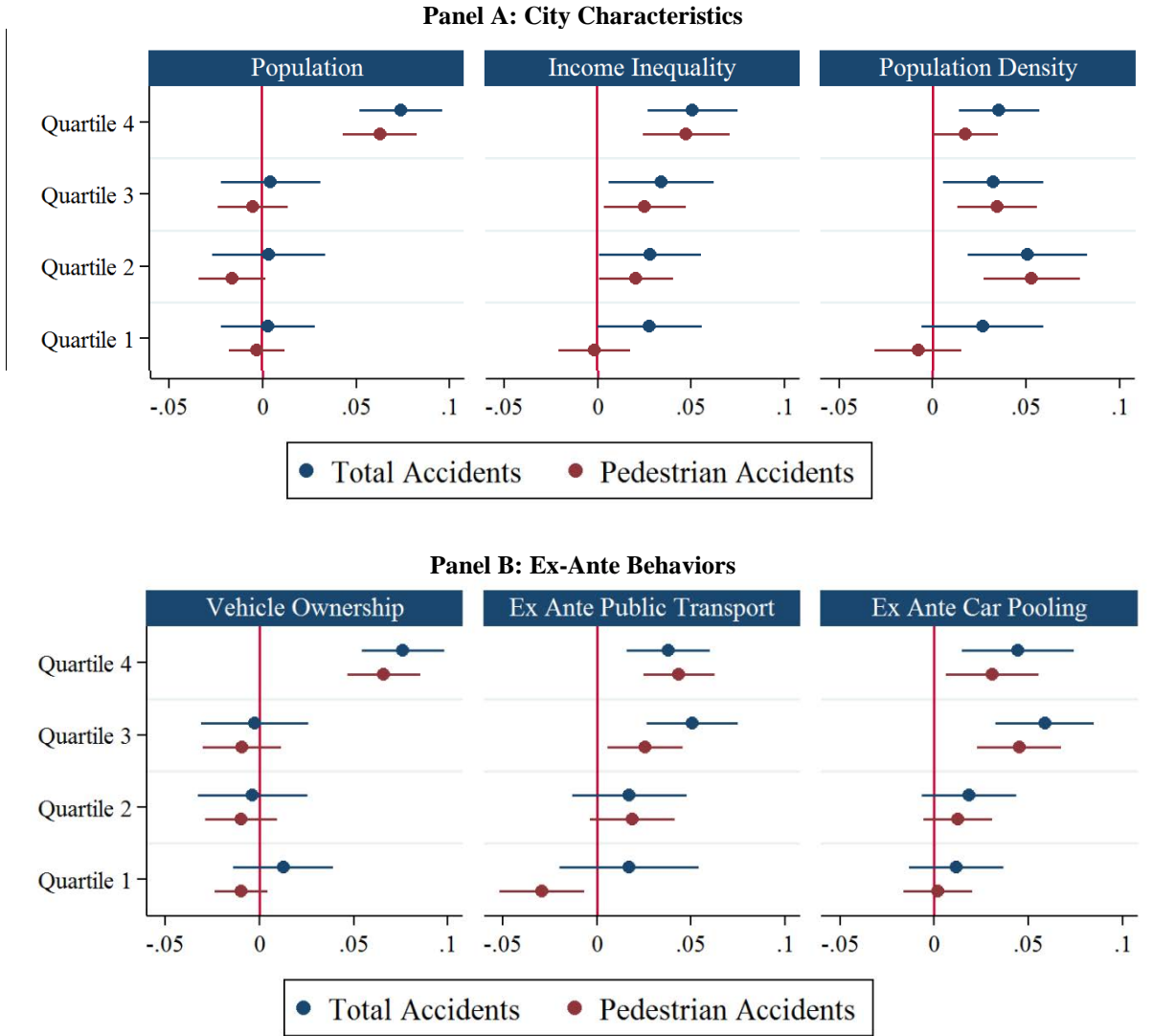
Figure 6
Heterogeneity by City Characteristics
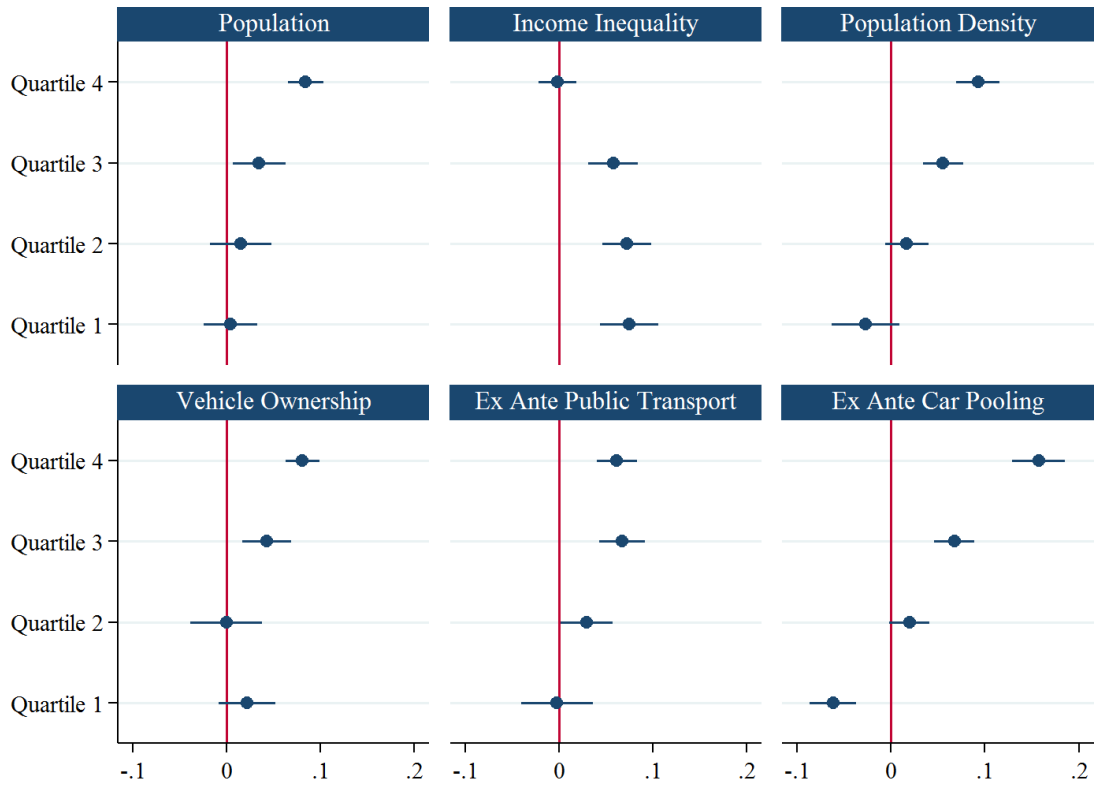
This figure displays the regression coefficient estimates and two-tailed 95% confidence intervals based on standard errors clustered at the city level, broken down by quartiles for six city characteristics: population, income inequality (measured by Gini Index), population density, vehicle ownership, ex-ante usage of public transportation, and ex-ante use of carpooling. The outcome variables for the regressions are listed at the bottom of each figure. We provide a description of the variables in Section 2

Figure 7
Heterogeneous Effect on Vehicle Registration by City Characteristics
This figure displays the regression coefficient estimates and two-tailed 95% confidence intervals based on standard errors clustered at the city level, broken down by quartiles for six city characteristics: population, income inequality (measured by Gini Index), population density, vehicle ownership, ex-ante usage of public transportation, and ex-ante use of carpooling. The outcome variable in all regressions is the natural logarithm of one plus new vehicle registrations. We provide a description of the variables in Section 2

ONLINE APPENDIX

BARRIOS, HOCHBERG AND YI (2019) – THE COST OF CONVENIENCE

BARRIOS, HOCHBERG AND YI (2019) – THE COST OF CONVENIENCE

This Online Appendix provides supplementary analysis to Barrios, Hochberg and Yi (2019).

1. MODELLING RIDEHAILING ENTRY

In the table below, we run multinomial logit models to predict city entry of ridehailing services. The dependent variable takes the value of 0 if RH launched in the city in 2010, 2011 or 2012 (early entry), a value of 1 if RH launched in the city in 2013 or 2014 (middle entry) and a value of 2 if RH launched in the city in 2015 or 2016 (late entry). While both population and per capita income load positively and significantly in predicting earlier entry, the change in accident rates over the 3, 5 or 10 years prior to entry does not load significantly.

Table A1 Multinomial Logit Estimation of Ridehailing Entry Decisions

	(1)	(2)	(3)
Early Entry			
Log (Population)	0.8410***	0.8396***	0.8363***
	(0.1329)	(0.1328)	(0.1320)
Log (Income)	2.8368***	2.8486***	2.8177***
	(0.6290)	(0.6291)	(0.6317)
3-Year Accidents Trend	0.3146		
	(0.7033)		
5-Year Accidents Trend		0.5028	
		(1.1834)	
10-Year Accidents Trend			0.9669
			(2.0542)
Late Entry			
Log (Population)	-0.3277***	-0.3284***	-0.3279***
	(0.0623)	(0.0624)	(0.0624)
Log (Income)	-0.8070***	-0.8114***	-0.8366***
	(0.2642)	(0.2646)	(0.2656)
3-Year Accidents Trend	0.0305		
	(0.1603)		
5-Year Accidents Trend		0.3406	
		(0.2595)	
10-Year Accidents Trend			0.6513
			(0.4620)
Observations	1199	1199	1199

Notes: This table presents multinomial logit estimates of ridehailing entry decisions. The outcome variable is defined as 0 if the city adopted ridesharing in 2010 through 2012, 1 if the city adopted ridesharing in 2013 and 2014 (the base outcome), and 2 if the city adopted ridesharing in 2015 or 2016. 3,5, or 10-Year Accidents Trends are the 3,5, or 10-year average quarterly change in accidents before ridesharing entry. Standard errors are reported in parentheses. ∗∗∗, ∗∗, and ∗ represent statistical significance at the 1%, 5%, and 10% levels, respectively.

2. ROBUSTNESS TESTS

In the table below, we demonstrate the effects of adding our control variable, fixed effects, and location-specific trends one-by-one to the model. We graph the coefficients of interest (*POST x TREATED*) in the corresponding Figure.

Table A2 Effect of Ridehailing on Traffic Safety in Different Model Specifications

**Panel A: Dependent Variable: Log (1+Total Accidents)**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Post_t * Treated_c$	0.2875***	0.1431***	0.0405***	0.0965***	0.0191***	0.0360***	0.0332***
	(0.0242)	(0.0113)	(0.0059)	(0.0133)	(0.0062)	(0.0074)	(0.0091)
Control Variables	No	Yes	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	Yes	Yes	Yes
Year-Quarter FE	No	No	No	Yes	Yes	Yes	Yes
City Linear Trend	No	No	No	No	No	Yes	Yes
City Quadratic Trend	No	No	No	No	No	No	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.009	0.346	0.605	0.351	0.610	0.619	0.626

Panel B: Dependent Variable: Log (1+Total Fatalities)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Post_t * Treated_c$	0.2878***	0.1454***	0.0411***	0.0964***	0.0183***	0.0360***	0.0335***
	(0.0247)	(0.0116)	(0.0061)	(0.0137)	(0.0065)	(0.0077)	(0.0095)
Control Variables	No	Yes	Yes	Yes	Yes	Yes	Yes
City FE	No	No	Yes	No	Yes	Yes	Yes
Year-Quarter FE	No	No	No	Yes	Yes	Yes	Yes
City Linear Trend	No	No	No	No	No	Yes	Yes
City Quadratic Trend	No	No	No	No	No	No	Yes
Observations	190,080	190,080	190,080	190,080	190,080	190,080	190,080
R2	0.008	0.335	0.590	0.340	0.595	0.604	0.612

Notes: This Table 1llustrates the changes in the generalized difference-in-difference regression coefficient estimates when using different model specifications. The dependent variables are in Panel A are the natural logarithm of one plus total fatal accidents, and in Panel B are the natural logarithm of one plus total fatalities. We provide a detailed description of these variables in section 2. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. Control variables include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Figure A1 Effect of Ridehailing on Traffic Safety in Different Model Specifications
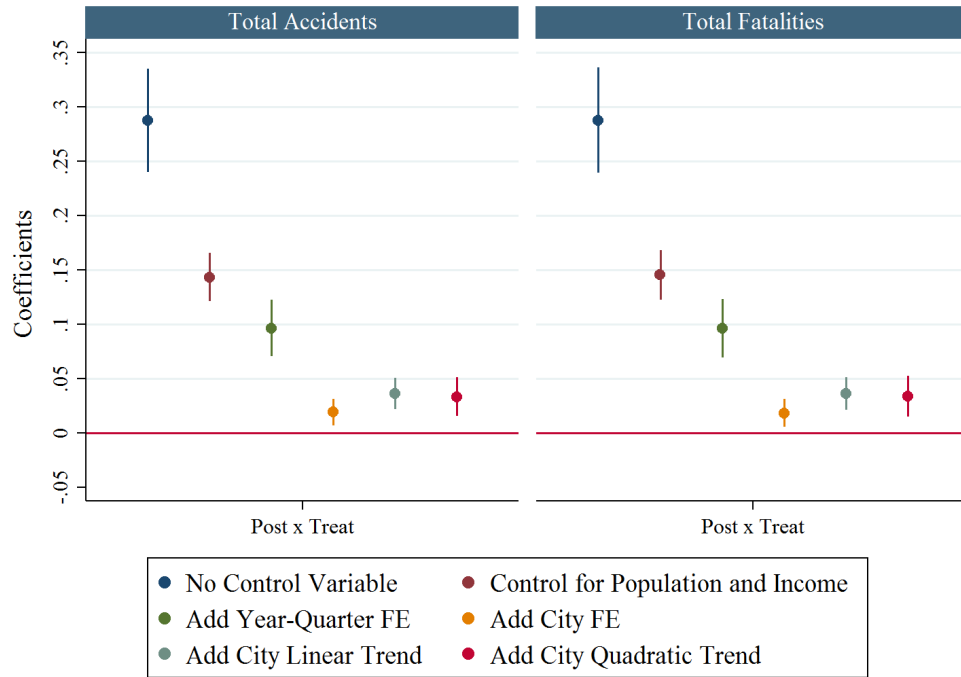
Notes: This figure shows the regression coefficient estimates and two-tailed 95% confidence intervals based on standard errors clustered at the city level. To demonstrate how the point estimates and standard errors change across different model specifications, we incrementally add control variables, fixed effects, and city-specific linear and quadratic time trends into the difference-in-difference regression specification. The outcome variables are displayed at the top of each graph. We provide a description of the variables in Section 2.

We next demonstrate the robustness of the results to controlling for additional variables, such as population growth (instead of population level), unemployment, gas prices and gas taxes. Coefficients are graphed in the corresponding Figure.

Table A3 Robustness to Additional Controls

	Log (1+Total Accidents)			
	(1)	(2)	(3)	(4)
$Post_t * Treated_c$	0.0361***	0.0292***	0.0335***	0.0361***
	(0.0074)	(0.0074)	(0.0074)	(0.0074)
Population Growth Rate	0.1007			
	(0.1012)			
Unemployment Rate		-0.0130***		
		(0.0013)		
Retail Gas Price			0.1163***	
			(0.0201)	
Retail Gas Price Change (%)				0.0012***
				(0.0004)
City and Quarter Fixed Effects	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes
Observations	187,110	190,080	190,080	190,080
R2	0.62	0.62	0.62	0.62

Notes: This table presents results from generalized difference-in-difference regressions when controlling for various additional control variables. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. Population growth rate is the annual percentage growth in city population. Unemployment rate is the quarterly average county unemployment rate. Retail Gas Price is average quarterly retail gasoline price (dollars per gallon). Retail Gas Price Change is the quarterly percentage change in retail gasoline price. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

Figure A2 Robustness to Additional Control Variables
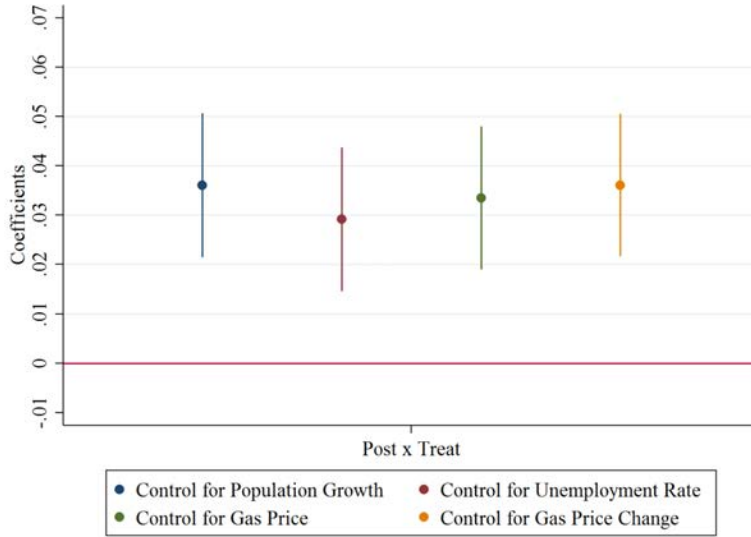
Notes: This figure illustrates how the regression coefficients vary when we include additional control variables. The outcome measures for all regressions are Log Total Accidents. Two-tailed 95% confidence intervals based on standard errors clustered at the city level are displayed in the figure. We provide a description of the variables in Section 2.

Table A4 Difference-in-Difference Estimation Using a Short Sample 2007-2016

	(1) Log (1+Total Accidents)	(2) Log (1+Total Fatalities)
$Post_t * Treated_c$	0.0309***	0.0315***
	(0.0084)	(0.0088)
City and Quarter Fixed Effects	Yes	Yes
City Linear Trend	Yes	Yes
Control Variables	Yes	Yes
Observations	118,800	118,800
R2	0.63	0.61

Notes: This table presents results from generalized difference-in-difference regressions using a shorter sample than our main specification. The sample includes all quarters from the first quarter of 2007 to the last quarter of 2016. $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, adjusted for clustering at the city level, are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

3. FURTHER DESCRIPTION OF TEXAS A&M MOBILITY SCORECARD DATA

In this section we describe the underlying data for the TAMU data for VMT, excess fuel consumption and hours of delay in traffic used in the paper. Our VMT measure comes from the Urban Mobility Scorecard which is produced annually from 1982-2014 by TAMU's Transportation Institute, one of the largest university-affiliated transportation research agencies in the U.S.

The base data for the TAMU's Urban Mobility Scorecard comes from two sources: INRIX and the U.S. Department of Transportation. INRIX is a private company that collects traffic speed data from a variety of sources including commercial vehicles, smart phones and connected cars with location devices. It uses a proprietary algorithm to filter inappropriate data (e.g., pedestrians walking next to a street) and provides TAMU with a dataset of average speeds for each road segment every 15 minutes. This dataset combined with the daily volume and roadway inventory data from Federal Highway Administration (FHWA)'s Highway Performance Monitoring System (HPMS) files are used to calculate travel mileage and delay statistics.

The key steps are the following. First, the daily traffic volume data are divided into the same time interval as the traffic speed data (15-minute intervals). The traffic volume is measured using average daily traffic (ADT) count. Second, because the geographic referencing systems are different for the speed and volume datasets, a geographic matching process utilizing Geographic Information Systems (GIS) tools are performed to assign traffic speed data to each HPMS road section. Third, daily vehicle-miles of travel are calculated as the ADT of a section of roadway multiplied by the length (in miles) of that section of roadway, and is aggregated at the annual level. Last, in order to calculate congestion measures such as excess fuel consumption and hours of delay, a free-flow speed was estimated using the speed at low volume conditions (for example, 10 p.m. to 5 a.m.) for each roadway section and hour of the week. Excess fuel consumption (hours of travel) are calculated as the difference between the observed fuel consumption (hours of travel) and the free-flow fuel consumption (hours of travel). Any observed speed faster than the free-flow speed is changed to the free-flow speed so that congestion measures are capped at zero, rather than providing a negative value.

While much work is put into generating these estimates, there is likely a good deal of imputation present: for most roads in most states, real measurements are infrequent, every three years according to guidelines (and state compliance with the FHWA guidelines are not

documented as far as we know). In principle, these potential VMT data issues would attenuate the coefficient towards not finding an effect, so the real effect on VMT might be larger than that which we document.

4.   ADDITIONAL QUALITY EVIDENCE

One way to determine whether the main effect we document is driven primarily by the quantity channel is to re-estimate our models, adding various VMT (quantity) measures as additional RHS variable, and seeing whether the VMT control absorbs the effect. We do this in Online Appendix Table A5 below. In Panel A, we run the specification on 99 urban areas. In Panel B, we disaggregate to the place (city) level within the urban area. In Panel A, controlling for VMT, we observe a positive coefficient on Post * Treated; the magnitude of the estimate is about one-quarter of the total accident increase we estimate in our main specifications. In Panel B, we have more power but are assuming VMT increases are spread evenly across incorporate places within the larger urban area. Here we also observe a positive coefficient of interest, controlling for VMT, this time statistically significant. Both panels are consistent with the documented increase in accidents being attributable to more than merely a quantity effect.

That said, there are caveats to this analysis. The sample size is significantly reduced, the VMT estimates are recorded only once a year, and considerable imputation is involved, as described in Online Appendix Section II above. Thus we expect to have measurement error which should attenuate the effect and issues of power given that we are trying to identify a relatively small and precise effect.

Table A5 Difference-in-Difference Estimation Controlling for Quantity Channel

Panel A: Urban Area Level

	Log (1+Total Accidents)			Log (1+Total Fatalities)		
	(1)	(2)	(3)	(4)	(5)	(6)
$Post_t * Treated_u$	0.0044	0.0034	0.0042	0.0034	0.0027	0.0033
	(0.0322)	(0.0320)	(0.0321)	(0.0323)	(0.0321)	(0.0322)
Log Arterial Street Daily VMT	-0.0021			0.0134		
	(0.1503)			(0.1487)		
Log Freeway Daily VMT		0.0522			0.0454	
		(0.1459)			(0.1262)	
Log Arterial Street and Freeway Daily VMT			0.0110			0.0168
			(0.1826)			(0.1652)
Urban Area and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Urban Area Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	990	990	990	990	990	990
R2	0.98	0.98	0.98	0.98	0.98	0.98

Panel B: City Level

	Log (1+Total Accidents)		Log (1+Total Fatalities)		Log (1+Fatal Accidents Involving Pedestrians)		Log (1+Fatalities in Accidents Involving Pedestrians)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Post_t * Treated_c$	0.0431***	0.0428***	0.0441***	0.0437***	0.0274**	0.0275**	0.0281**	0.0282**
	(0.0140)	(0.0140)	(0.0145)	(0.0145)	(0.0126)	(0.0126)	(0.0128)	(0.0128)
Log Arterial Street Daily VMT	0.0379		0.0391		-0.0295		-0.0282	
	(0.0464)		(0.0486)		(0.0424)		(0.0427)	
Log Arterial Street and Freeway Daily VMT		0.0827		0.0908		-0.0341		-0.0355
		(0.0554)		(0.0583)		(0.0616)		(0.0616)
City and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City Linear Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	87,584	87,584	87,584	87,584	87,584	87,584	87,584	87,584
R2	0.72	0.72	0.71	0.71	0.65	0.65	0.64	0.64

Notes: This table presents results from generalized difference-in-difference regressions when controlling for various proxies for vehicle miles traveled. In all panels, the dependent variables are listed at the top of the columns. Total accidents is the number of fatal accidents, according to the definition used by NHTSA. Total fatalities is the total number of fatalities across all fatal accidents. Pedestrian is defined as any person not in or upon a motor vehicle or other vehicle. Pedestrian-involved accident measures the number of fatal accidents involving at least one non-vehicle occupants. Pedestrian-involved fatalities measures the total number of fatalities in all accidents involving at least one non-vehicle occupants. Pedestrians involved in fatal accidents measures the total number of non-vehicle occupants involved in fatal accidents. In Panel A, we run the specification on 99 urban areas. $Post_t * Treated_u$ is a dummy variable that equals one if urban area u adopted at least one ridehailing service at year t. Urban area-specific linear trends are included in all regressions. In Panel B, the specifications are estimated at the city year level and $Post_t * Treated_c$ is a dummy variable that equals one if city c adopted at least one ridehailing service at time t. City-specific linear trends are included in all regressions in Panel B. Control variables in all regressions include the natural logarithm of population and the natural logarithm of income per capita. Standard errors, clustered at the urban level in Panel A and clustered at the city level in Panel B, they are reported in parentheses. ***, **, and * represent statistical significance at the 1%, 5%, and 10% levels, respectively.

5. DRUNK ACCIDENTS

While our main empirical question centers on the patterns for total fatal accidents, much has been made of the possible effects that RH has specifically on drunk fatal accidents, which comprise a quarter to a third of all fatal accidents. A reduction in alcohol-involved fatal accidents would not be inconsistent with our overall findings; drunk accidents could go down, but the decrease could still be swamped by increases in nondrunk accidents, leading to an overall increase.

In this section, we provide further insight into the patterns for accidents involving an impaired driver. We further discuss how our findings on drunk accidents and fatalities relate to prior work exploring the relationship between ridehailing and drunk accidents. The costs of drunk driving have been extensively documented. Drunk driving costs the U.S. tens of thousands of lives and billions of dollars in law enforcement, property damage, and lost productivity each year. The NHTSA estimates the direct economic cost of these accidents to be $44 billion and estimates the total societal costs at $201 billion.

As a result of these costs, extensive amounts of resources have been utilized to develop public policy to discourage intoxicated driving. Most policies focus on deterrence through fear of punishment and increasing the expected cost of a conviction (i.e. increasing the penalties for drunk driving convictions through higher fines, more jail time, and/or driver's license confiscation). Previous work on ridehailing services has proposed that ridehailing, by increasing the convenience of and decreasing the cost of alternative transportation, may in turn lead to fewer drunk drivers and thus reduce the level of fatalities due to impaired driving. Several previous studies have made attempts to examine the effect of RH on fatal accidents in U.S. cities (see e.g., Brazil and Kirk, 2016; Martin-Buck, 2016; Greenwood and Wattal, 2016, Dills and Mulholland, 2018). Most of these studies rely on data on fatal alcohol-related auto accidents from the NHTSA over the period of 2000-2014, and a number conclude that the advent of RH is associated with a significant reduction in alcohol related accidents. These studies employ a variety of specifications, including difference-in-differences specifications.

There are a number of reasons, however, why our secondary results looking at drunk accident and fatalities do not align with the conclusions of these prior studies. We discuss these in turn: (1) a definitional change in how alcohol-involved accidents are classified half way through our sample period; (2) smaller samples; (3) the specific coefficients that are interpreted by the studies.

5.1. *Definition Changes for Alcohol Involvement*

First and foremost, these studies do not appear to account for an important change in how accidents were classified as involving alcohol impairment, which took place in the beginning of 2008. Specifically, in the years prior to 2008, alcohol-related accidents were recorded by the NHTSA as any fatal accident involving at least one person—vehicle occupant (driver or non-driver) or pedestrian—having blood alcohol levels above the legal threshold for impaired driving. In other words, prior to 2008, an accident in which a sober taxi driver kills a drunk pedestrian would be classified as a drunk accident, and a sober driver driving a passenger who was impaired who gets into an accident that results in a death would also be classified as a drunk accident. From 2008 onwards, alcohol-related accidents are recorded as "drunk accidents" only if the *driver* himself was impaired.

This definitional change lead to a massive mechanical decrease in accidents classified as "alcohol-related" and to a corresponding increase in non-drunk accidents, as graphed below in Appendix Figure A3.

Figure A3: Accidents Classified as Alcohol-Related, Annual and Quarterly

| Panel A: Annual | Panel B: Quarterly |
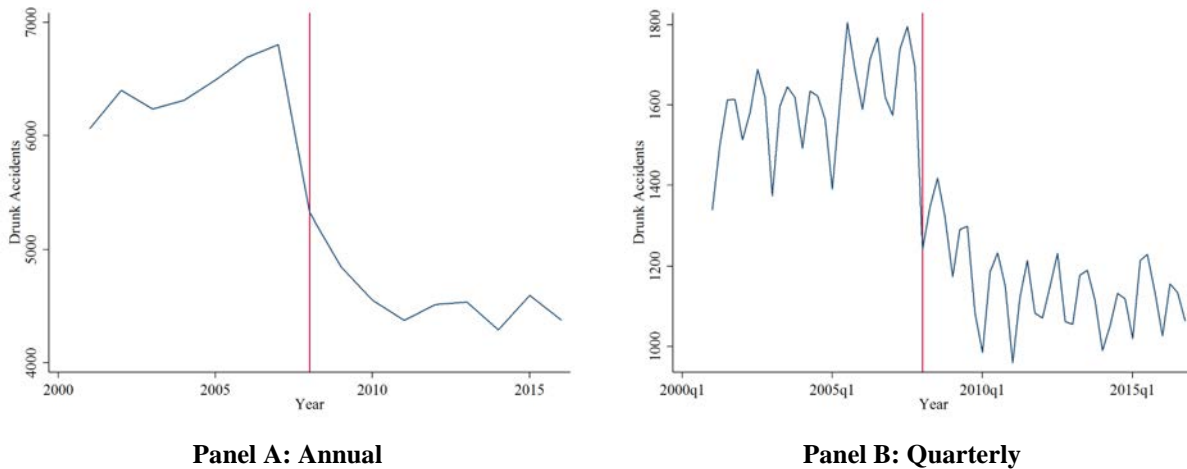
Notes: This figure displays the total number of drunk accidents through time. The vertical line indicates the year or quarter of the NHTSA drunk accidents categorization change. Specifically, prior to 2008, alcohol-related accidents were recorded as any fatal accident involving at least one vehicle occupant (driver or non-driver) or pedestrian being impaired (in the legal sense). From 2008 onwards, alcohol-related accidents are recorded as such only if the *driver* was impaired.

This mechanical drop has clear effects when estimating models that include the pre-change years. Figure A4 shows the coefficient of interest for the estimated "effect" of RH for drunk accidents when using different sample periods. The first four specifications include some years of

the period prior to the definitional change. The second two include only years after the change, when a drunk accident clearly involved an impaired *driver*. These specifications are estimated in similar fashion to those in prior studies, such as Martin-Buck (2016), which do not control for location-specific linear and quadratic trends (Martin-Buck (2016) uses a sample period of 2000-2014). In the four specifications contaminated by the old definition of drunk accident, we observe an estimated negative coefficient. In the two uncontaminated specifications, we observe a positive coefficient, consistent with our main results.

For contrast, in Figure A5 we provide the same coefficients for total fatal accidents, which are not affected in any way by the definition change for drunk accidents, as they are the sum of drunk and nondrunk. As expected, these coefficients are consistent across all the sample periods, and do not vary in sign based on sample years, unlike the drunk accident estimates, which are affected mechanically by the definition change leading to a drop in accidents close to RH entry.

Figure A4 Drunk Accident and Fatalities Coefficient Estimates in Different Sample Periods
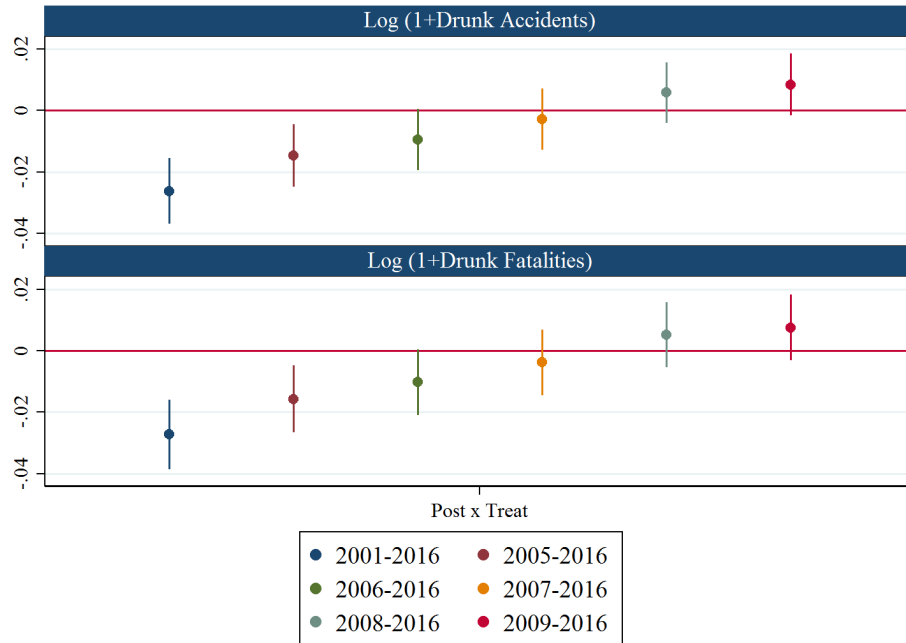
Notes: This Figure illustrates how the regression coefficient estimates of the effect of ridehailing on drunk accidents and fatalities vary with different sample time periods. The outcome measure for all each panel are displayed at the top of the panel. Two-tailed 95% confidence intervals based on standard errors clustered at the city level are displayed in the figure.

Figure A5 Placebo Test – Total Accident and Fatalities Coefficient Estimates in Different Sample Periods
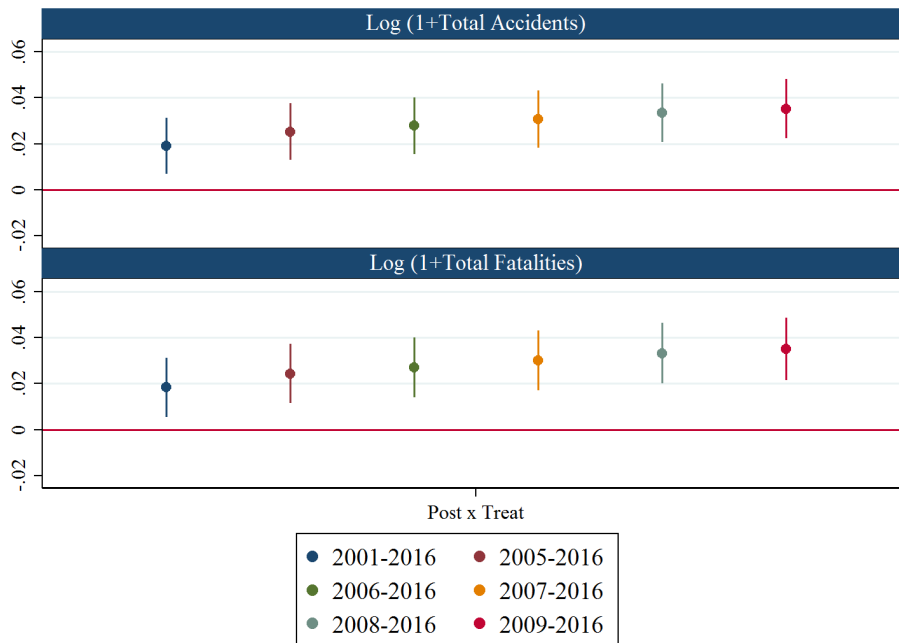
Notes: This Figure illustrates how the regression coefficient estimates of the effect of ridehailing on total accidents and fatalities vary with different sample time periods. The outcome measure for all each panel are displayed at the top of the panel. Two-tailed 95% confidence intervals based on standard errors clustered at the city level are displayed in the figure.

13

5.2.   Sample Coverage

Other papers in this vein have significant differences in geographic sample coverage and size relative to this study. For example, Greenwood and Wattal (2017) use a sample period and data in which the definition of alcohol-involved accident is uniform throughout the sample. They look at fatalities in alcohol-related accidents, using the period 2009-2013, but with a much smaller sample than ours—540 townships in California, only 7% of which were treated with UberX service during the sample period. Perhaps unsurprisingly as a result, their models (DD models that do not account for location-specific linear or quadratic trends) demonstrate relatively low R-sq (0.035-0.041). Notably, in our analysis, the positive relationship between RH and accidents are clearly concentrated in larger cities and cities with high ex-ante use of public transport, which many CA cities do not have.

5.3.   Model Specification

Finally, a number of other studies look at this relationship using the lens of alternative specifications which may be less appropriate for answering the question at hand. Brazil and Kirk (2016) use a negative binomial model rather than a difference-in-differences setup. They find no effect of RH on either accidents or drunk accidents using this specification. Dills and Mulholland (2018) use a difference-in-differences approach, looking at both total and drunk accidents in relationship to the launch of UberX at the county level. Their model estimates find an economically large positive coefficient on TREAT x POST, similar to our findings. However, they instead choose to draw conclusions from the coefficient on a different term included in the model: an interaction of TREAT x POST with a linear time trend (non-location-specific). Here they find a negative coefficient roughly $1/6^{th}$ to $1/10^{th}$ the magnitude of the main effect, which is statistically significant only in some specifications. They then interpret this coefficient as evidence that accidents are reduced post UberX entry, rather than as a reduction in the slope of the increase, which is what the specification would actually imply. Their models exhibit R-sq of 0.03-0.13, in contrast to our R-sq which is typically in the range of 0.4 to 0.6 or higher. Notably, their drunk accident sample also includes at least one year of data pre-definition change.