TI 2020-080/VIII
Tinbergen Institute Discussion Paper

Accident Externality of Driving: Evidence from the London Congestion Charge

Cheng Keat Tang[1]
Jos van Ommeren[2]

[1] Nanyang Technological University
[2] Vrije Universiteit Amsterdam

Accident Externality of Driving: Evidence from the London Congestion Charge

Cheng Keat Tang
Nanyang Technological University*

Jos Van Ommeren
VU University Amsterdam†

November 18, 2020

Abstract

This paper estimates the marginal accident externality of driving in Central London by exploiting variation in traffic flow induced by the London Congestion Charge Zone using an instrumental variable approach. The charge attributed to a 9.4% reduction in traffic flow, which resulted in a less than proportional 6.0% and 7.6% decrease in accidents and slight injuries, and a 6.5% increase in serious injuries/fatalities. Our preferred estimates indicate that the accident, slight injuries, and serious injuries/fatalities rate elasticities with respect to traffic flow are -0.36, -0.19 and -1.65 respectively. These estimates imply that the marginal external benefit of road safety from an additional kilometre driven is approximately £0.16. The marginal accident externality is positive, as the marginal driver along congested roads decreases the risk and severity of traffic collisions for other road users by slowing others down and increasing awareness.

JEL: H23,I18,R41,R48

Keywords: Accidents, Injuries, Fatalities, Congestion Charge, Externalities

*Department of Economics, Email address: c.k.tang@ntu.edu.sg

†Department of Spatial Economics, VU Amsterdam, De Boelelaan 1105 1081 HV Amsterdam, email: jos.van.ommeren@vu.nl. Jos van Ommeren is a Fellow of the Tinbergen Institute, Amsterdam

1 Introduction

When drivers decide to take the road, they do not only undertake the risk of getting into a traffic accident themselves, but they also increase or decrease the accident risk of other road users, inducing negative or positive accident externality. The externality is negative when the accident elasticity with respect to flow is larger than one, i.e. an additional driver leads to more than proportional increase in traffic accidents (Vickrey, 1968), causing the social marginal cost of driving to exceed private marginal cost. While some studies have shown that the accident externality of driving is negative (Vickrey, 1968; Parry, 2004; Edlin & Karaca-Mandic, 2006; Romem & Shurtz, 2016), this relationship is less clear for congested cities.

Positive accident externalities from driving are more likely to be found in heavily congested cities such as New York, Jakarta, Bangkok, Rome and London.[1] Although heavier traffic may increase the risk of accidents, the marginal driver along congested roads slows down other motorists, reducing the severity of traffic collisions (Shefer & Rietveld, 1997; Verhoef & Rouwendal, 2004). More vehicles on roads could also heighten the awareness of drivers, reducing the probability of collisions. Under these circumstances, an additional driver is likely to cause a less than proportional increase, or even a decrease in the number and severity of traffic accidents, making roads safer. Knowledge about the sign and the size of the marginal accident externality of driving is particularly relevant for congested cities where a road tax is considered. Policymakers wish to be informed whether the optimal Pigouvian road tax that addresses the two main road externalities – congestion and accident externalities – is higher or lower than a road tax that considers only travel time delays.[2]

Quantifying the accident externalities of driving and the impacts of policies aiming to mitigate these externalities are important given the enormous cost accidents impose to society. In 2015, road accidents inflict more than 160,000 injuries and 1,700 fatalities across the United Kingdom (UK), imposing an estimated cost of £10.3 billion to society. While the average accident cost per driver appears high, these figures do not inform about the magnitude or sign of the external accident costs from an additional driver.[3] Consider a car collision, e.g. on a pedestrian, or between two cars,

[1] To clarify, accident externality is positive (negative) when the additional driver leads to a less (more) than proportional increase in the risk of accidents.

[2] Environmental externalities of car use are relatively small. For example, on highways, the marginal travel time external cost is of the same size as the marginal accident external cost, whereas the marginal environmental external cost is an order of magnitude smaller, see Small & Verhoef (2007).

[3] As highlighted by studies such as Vickrey (1968) and Edlin & Karaca-Mandic (2006), the tort law system, which recovers damages from the party liable for the accident, provides sufficient incentives for drivers to drive carefully, but does not provide adequate incentives on how much to drive, which makes the marginal accident externality of driving of key interest.

that would not have happened if one driver chooses to drive less. The driver only expects to incur the average accident cost for the additional kilometer driven - i.e. the accident cost of all drivers combined divided by the total distance driven - instead of the marginal accident cost to society. If the marginal external cost is positive (negative), i.e. the marginal cost exceeds (are less than) the driver's expected cost, then driving may be too cheap (expensive), causing people to drive beyond (below) socially-optimal levels (Jansson, 1994).

In this study, we quantify the accident externality associated with driving by estimating how traffic flow affects the likelihood and severity of traffic collisions using an instrumental variable (IV) approach. We provide causal estimates by exploiting the plausibly exogenous variation in traffic flow induced by the London Congestion Charge Zone (CCZ). The CCZ is a preferred instrument as it allow us to examine the accident externality from driving along congested roads, a margin that policy makers is particularly interested in. The CCZ was implemented in 2003 to reduce traffic along the most congested roads in Central London. The policy curtails road traffic by imposing a fee on drivers commuting into the zone during chargeable hours. Put differently, we are comparing changes in traffic flow and collision outcomes for areas inside and outside the charge zone before and after the charge is enforced. Because collision outcomes are discrete variables, we estimate count models with the CCZ as an IV using a control function approach.[4]

Our headline finding is that while the charge attributed to a 9.4% reduction in traffic flow, it resulted in a less than proportional 6.0% and 7.6% decrease in accidents and injuries, and a 6.5% increase in serious injuries/fatalities. From these estimates, one can derive the accident, slight injury, serious injury and fatality rate elasticity with respect to traffic flow by taking the estimated elasticity of various collision outcomes with respect to traffic flow and subtract by 1. The accident, slight injuries, and serious injuries/fatalities rate elasticities with respect to traffic flow are -0.36, -0.19 and -1.65 respectively. These findings show that the accident externality of driving is positive as the marginal driver along congested roads slows down the average driving speed, decreasing the risk and severity of traffic collisions for other road users. Further analyses show that the implied marginal external benefit of road safety from an additional kilometer driven around Central London is £0.16. These findings suggest that curtailing traffic flow is unlikely to be an efficient strategy to improve road safety for congested roads around city centre.

Our identification strategy hinges on the assumption that the CCZ is affecting traffic collisions between the charge and non-charged zone only through changes in traffic flow induced by

[4]For linear specifications, such an approach provides identical estimates as a two-staged least square (2SLS) approach. However, the latter cannot be applied to count and other non-linear models (Wooldridge, 2015). In the control function approach, one uses the traffic flow residuals from the first-stage regression as a control variable in the second-stage regression to partial out the endogenous variation in traffic flow. We provide additional details in Section 5.

the charge. This assumption could be violated if the charge is affecting traffic collisions through changing the composition of traffic, vehicles and drivers in the zone. This is plausible as some vehicle types like motorcycles and bicycles are omitted from paying the charge. Substitution to other non-charged transportation modes could attribute to a surge in 2-wheel vehicles or pedestrians (using public transit) along footpaths in the zone that could increase the probability and severity of traffic collisions.[5] We adopt the following strategies in response to these challenges. First, we progressively constrain our analysis to areas close the charge boundary (up to 2km). The idea is that, given that motorists driving into the charge zone must pass by areas right outside the zone, traffic composition is likely to be similar between areas around the charge boundary even if the charge has substantially affected the composition of traffic. Second, we directly test whether there is a change in composition of traffic inside the charge zone with a battery of balancing tests. Specifically, we examine whether there are significant changes in the (1) characteristics of drivers and vehicles involved in collisions, (2) the type of traffic flow and (3) the type of accidents in the charge zone after the charge is enforced. Finally, we control for driver and vehicle characteristics involved in collisions and limit our analyses to accidents that involve charged vehicles only in our robustness tests. Overall, all these analyses suggest that the charge is unlikely to have affected collision outcomes through means other than changes in traffic flow.

Despite the importance of estimating the accident externality associated with driving, empirical evidences are few and far between, with often contrasting results reported by studies that apply different empirical strategies for different countries. In his seminal paper, Vickrey (1968) focuses on highways in California and concludes that the accident rate elasticity is about 1.5. Shefer & Rietveld (1997) suggest that more traffic does not necessarily increase accident costs in Germany, Israel and US after observing fewer traffic fatalities during morning peak hours when vehicle miles travelled are the highest. They suggest that lower driving speeds could explain why collisions are less deadly. Several studies document findings suggesting positive accident externalities associated with driving. Fridstrøm & Ingebrigtsen (1991) report that the elasticity between traffic flow and accident rates is around 0.47 for Norway. Zhou & Sisiopiku (1997) further observe an inverse "U"-shaped relationship between accident rates and traffic flow in Detroit, US. Accident rates are the highest when traffic flow is the lowest and decrease dramatically as traffic flow increases. Accidents involving serious injuries/fatalities, conversely, decrease with traffic flow.

The major issue with these cross-sectional studies is that they do not account for the endogenous relationship between traffic flow and accidents. If drivers prefer safer routes that are less prone to accidents, roads with heavy traffic could end up being less accident prone, causing a

[5]In cities such as New York and London, roads are typically shared with cyclists and pedestrians, so a large share of accident injuries involving cars is incurred by non-car road users, whereas single-car accidents are rare. This is in sharp contrast to highways where single-car accidents are more common (Parry, 2004).

downward bias to the relationship between traffic flow and collision outcomes. In addition, roads with heavier usage typically have different designs and receive more funding for maintenance, inducing a spurious relationship between traffic flow and collision outcomes. Using temporal variation in traffic flow and collision outcomes is unlikely to address these issues as minor road works are often unobserved but have major influence on traffic flow and collision outcomes. This could explain why the estimated relationship between traffic density and accidents is often mixed and inconclusive between studies.

There are a few papers that address these empirical challenges. Edlin & Karaca-Mandic (2006) estimate traffic accident externalities by examining how annual traffic flow increases the cost of insurance across U.S states. After instrumenting traffic density with the number of registered vehicles and licensed drivers due to concerns of measurement error, their study shows that an additional driver increases insurance premiums for all other drivers by around \$1,725- \$2,150 per annum in California. These effects are the strongest for states with the heaviest traffic. One concern raised by Parry *et al.* (2007) is that insurance cost, which covers the property damage associated with traffic collisions but not the injury costs, constitutes a small proportion of cost associated with accidents. Romem & Shurtz (2016) extend the literature by exploiting the exogenous variation in traffic from the observance of Jewish Sabbath in Israel to measure how traffic flow affects the probability of collisions. Using a regression discontinuity design, they report that the accident rate elasticity is between -0.3 and 1. An additional driver only increases accident risk for others when the traffic is heavy after the Sabbath ended. The authors conclude that the relationship between traffic flow and accidents is likely non-monotonic.[6]

This study contributes to the existing literature in at least three ways. First, we improve the estimation between traffic flow and accidents from exploiting the plausibly exogenous variation in traffic flow from the implementation of the London Congestion Charge. We carefully estimate the causal relationship by eliminating other plausible factors that could bias our estimates through a battery of specifications. Second, we estimate the marginal effect of driving on accidents induced by road pricing, measuring how the marginal driver affects road safety along highly congested roads. The margin we are examining is paramount to policy makers who are most concerned about reducing driving externalities along congested roads. Given the non-linear (and even non-monotonic) relationship between traffic flow and accidents, previous estimates for other countries at a different margin are unlikely to be applicable to our context. Our findings also inform about the

[6] In these studies, differences in accident externalities due to differences between vehicles are ignored. When a driver chooses a heavier vehicle, he/she imposes an externality on other road users who are more likely to get more severely injured when involved in a collision. Van Ommeren *et al.* (2013) and Anderson & Auffhammer (2014) report almost identical estimates for the marginal external costs of car weight per accident for the Netherlands and the US, suggesting that the weight externality is proportional to the marginal external accident cost of flow estimated in the current paper.

impact of the congestion charge on road safety.[7] Third, we employ detailed micro information on accident severity (e.g. light and serious injuries/deaths) and road users involved in accidents (e.g. car drivers, pedestrians, cyclists) allow us to examine how additional traffic impacts the severity of collisions. This allow us to accurately quantify the external accident cost associated with driving that is not covered in insurance premiums.

Our study shows that the charge has effectively reduced traffic flow and accidents in the zone, but accidents have become more severe probably due to faster driving speeds after congestion is alleviated. These results imply that the optimal road toll could be less than the road toll based on travel time only. Furthermore, these findings highlight the importance of employing complementary instruments, such as speed bumps, speed limits and cameras, to regulate driving speed or the variance of driving speed between vehicles once traffic becomes more free-flowing. As in any other study, one may question the external validity of our results. We emphasise, therefore, that our results are applicable to other major congested cities around the world, but are not informative, for example, about the accident externalities of driving on non-congested highways.

The remainder of this paper is structured as follows. In section 2, we provide the theoretical framework. Section 3 provides an overview on the Congestion Charge in London. Section 4 outlines the data and Section 5 illustrates the identification strategy. Findings are then discussed in Section 6 and 7, and Section 8 concludes.

2 Theoretical framework

In this section, we explain the main theoretical concepts introduced by Vickrey (1968). We are interested to estimate the marginal external accident cost of driving, which is the external cost imposed on society from an additional driver through the change in the probability and severity of traffic collisions. We define traffic flow as F and an additional driver who decides to take the road will increase F. We focus on the marginal driver with a demand for a trip that depends negatively on the price paid for the journey, which includes the expected accident cost, $A(F)$. When deciding to drive, the driver increases traffic flow and changes the expected cost of accidents for other drivers. More specifically, we wish to know the consequences of a congestion charge (as in

[7] Our study differs from studies such as Li *et al.* (2012) and Green *et al.* (2016) that focus on the (reduced form) impact of a congestion charge on traffic accidents. In this paper, we rely on the policy to identify how the marginal driver affects road safety. One cannot derive the marginal external accident cost from these studies without considering the impact of the policy on traffic flow. Furthermore, their results could be driven by changes in traffic and driver composition due to the charge. We attempt to address these identification challenges using a variety of strategies that we will discuss in-depth in Section 5.

London, Singapore) that reduces travel time losses by reducing traffic flow, and therefore, by design, affecting the external accident cost of driving. For simplicity, we do not distinguish between different types of accidents and assume that the monetary cost of an accident is standardised to 1.[8]

The number of accidents, Y, is assumed to be an increasing function of traffic density, D, as well as of speed, S, in line with previous literature (Verhoef & Rouwendal, 2004; Edlin & Karaca-Mandic, 2006).[9] Traffic density increases with traffic flow ($\partial D/\partial F > 0$), while speed is decreases with traffic flow ($\partial S/\partial F < 0$). It follows that the number of accidents is a function of flow ($Y = Y(D(F), S(F)) = Y(F)$), but the sign of the flow effect is ambiguous. On roads with low or moderate levels of congestion, it is plausible that the density effect dominates such that the number of accidents is an increasing function of traffic flow. However, along highly congested roads, speed effect could dominate such that the number of accidents could reduce with an increase in traffic flow.

We observe the total number of accidents (per area) in our data. The driver's expected accident cost, i.e. the accident rate $A(F)$, can be written as Y/F. We assume that the effect of flow on the number of accidents, Y, has the following functional form: $Y = exp^{\lambda + \beta logF}$, where β is the elasticity of the number of accidents with respect to traffic flow. This is the main parameter of interest from our estimations when we regress collision counts against traffic flow holding all other factors constant. It then follows that the marginal number of accidents $\partial Y/\partial F$ is equal to $\beta Y/F$, hence the marginal accident externality equals $\beta Y/F - Y/F$ = $[\beta - 1]Y/F$. This is also known as the accident rate elasticity with respect to traffic flow. Consequently, there is no externality when $\beta = 1$. There is a marginal external cost when $\beta > 1$, and a marginal external benefit when $\beta < 1$. In other words, the marginal driver only increases the collision risk for other road users when $\beta > 1$. Otherwise, the presence of an additional driver actually reduces the risk of other road users getting into accidents, conferring a benefit to other road users.

3 Background

The Congestion Charge Zone (CCZ) was implemented on the 17th February 2003. It covered a total of 21 square kilometres and encompassed the financial centre (Bank), parliament and government offices (Palace of Westminster), major shopping belts (Oxford Circus) and tourist attractions

[8]We relax this assumption in subsequent analyses as we examine how traffic flow affects accident probability and severity, before monetizing these estimates to compute the marginal external cost/benefit associated with driving.

[9]We assume away drivers' decisions regarding vehicle weight, and self-chosen speed, both of which may also imply an externality. Hence, we assume away that drivers optimally trade off between choosing their own speed and accident risk (Verhoef & Rouwendal, 2004). One justification for this assumption is that tort law provides sufficient incentives to drive carefully.

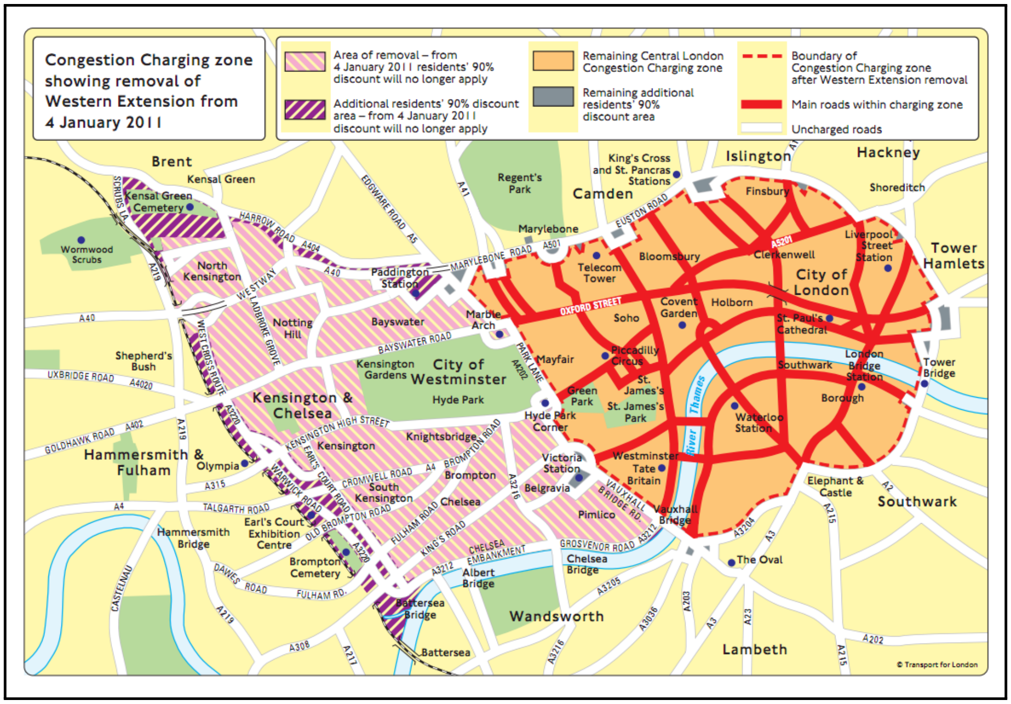

Figure 1: Map of the Original Congestion Charge Zone (CCZ) & Western Extension Zone (WEZ) Source: Transport for London (TfL)

(Trafalgar Square, Westminster Abbey, Big Ben, St Paul Cathedral etc). To reduce confusion and for convenience, the initial Congestion Charge Zone will be abbreviated as the CCZ and the Western Extension Zone, which was later on introduced but quickly abolished, will be abbreviated as WEZ. As the WEZ was short-lived and did not materially improve traffic conditions, we will see that the use of the WEZ cannot convincingly be used as an instrument. Hence, our paper will focus on the CCZ.

Figure 1 shows the CCZ, the shaded area enclosed by a dash line. The boundary was drawn to isolate the most congested areas in Central London. It is bordered by major inner ring roads such as Edgeware, Vauxhall Bridge, Pentonville, Park Lane, Marylebone, Tower Bridge and Victoria to divert traffic displaced by the charge. Drivers travelling on these roads are not required to pay unless they turn into the zone.

Commuters driving into the CCZ between 7:00am and 6:30pm from Monday to Friday excluding public holidays are required to pay a one-time daily flat fee of £5.00.[10] Residents living in the zone and within discount zones outside are entitled to a 90% waiver to the charge for their first

[10]The rationale for levying a flat fee, other than the difficulty in imposing time-varying fees to reduce congestion during peak hours, is that vehicular flow on roads seems fairly uniform across the day.

registered vehicle.[11] Other groups excluded from paying the charge include public transport (taxis and buses), motorcycles, bicycles, environmentally-friendly vehicles (battery powered or hybrid cars), vehicles driven by disabled individuals (blue badge holders), vehicles with 9 seaters or more and emergency service vehicles.

The tax levied increased to £8.00 on the 4th July 2005 to further reduce traffic and raise revenues. On the 19th of February 2007, charging was extended to Central West London (known as the Western Extension Zone) because of congestion in that area. Operating hours of the charge were reduced by half an hour from 7:00am to 6:00pm. The westward extension is circumvented by Harrow Road, Scrubs Lane, West Cross Route, the Earls Court One-Way system, Chelsea Embankment and the River Thames to the South. This area is marked with pink stripes in Figure 1. However, under tremendous pressure from residents and businesses in West London, the WEZ was scrapped on the 24th of December 2010.[12]

Impact assessment by Transport for London (TfL) showed significant improvement in traffic conditions after the charge was enforced in 2003. These results are very consistent to those reported in this study. Travel speeds were almost 20% higher (from 14.3km to 16.7km per hour) and minutes of delay fell by 30% (TfL, 2003a). This was due to a 27% overall drop in the number of private automobiles in Central London.[13] No evident displacement of traffic into neighbouring uncharged roads and weekends were recorded as traffic conditions were fairly similar compared to those during pre-charged periods. A change in composition of inbound traffic into the zone was also observed: the flow of bicycles, buses and taxis went up by 28%, 21% and 22% respectively. Surveys conducted echoed similar findings with the majority of the drivers switching to public transport and others travelling during off-charging hours (TfL, 2005). Though the number of commuters using rail did not increase, the number of bus passengers during morning peak periods were 38% higher (TfL, 2004).

These results are expected as these vehicle types are omitted from the charge. We acknowledge that increases in bicycles and buses suggest that the charge could have affected collision outcomes via changing the composition of traffic in the charge zone, violating the exclusion restriction. We adopt a variety of strategies to alleviate this concern, including one that limits the analysis to areas very close to the charge boundary as these areas are more likely to share similar traffic composition. We provide more details in Section 5. Overall, evidences provided by TfL suggest that the charge

[11] These discount zones are shaded in grey for the CCZ and in purple-striped for the WEZ as shown in Figure 1. Residents living in these areas are entitled to the discount due to parking and severance issues.

[12] We observe that the WEZ did not materially improve traffic conditions after its introduction. For more information, refer to Table A2 in Data Appendix. We provide detail interpretation of these findings in subsequent sections.

[13] We present similar results in this paper. Depending on specifications, we document traffic reductions of between 9.4% and 13.0%. For more information, refer to Tables 3 and 4.

significantly reduced traffic flow in the charge zone and we exploit this substantial improvement in traffic conditions to measure the accident externality of driving.

4 Data

To estimate the accident externality of driving, we rely on two main data sets together with a variety of auxiliary data. First, we rely on STATS 19 Road Accident Database that provides detailed information for each reported accident that involved an injury or fatality to the Police Force.[14] For each accident, we observe location, time, date, road conditions, age and gender of driver, vehicle type, number of injuries, seriousness of injuries and fatalities (pedestrians and inside the vehicle). We calculate *annual* collision outcomes per Lower Super Output Area (LSOA) from the year 2000 to 2014, distinguishing between number of accidents, slight injuries, serious injuries/fatalities.[15] Fatal accidents rarely occur (there are slightly more than hundred fatal traffic accidents in London annually) and almost all LSOAs (about 94%) do not have any fatal accident recorded in a year. Hence, we aggregate serious injuries and fatalities into one category when measuring severity of traffic collisions.[16]

Second, we obtain the Average Annual Daily Traffic Flow (AADF) collected at each count point (CP) from the Department of Transport (DfT). There are a total of 2,523 CPs in London and most of them clustered around Central London as shown in Figure 2. Each site is counted manually by a trained enumerator on a *neutral day* for a twelve hour period to provide junction-to-junction traffic flow. A *neutral day* is a weekday between March and October, excluding public holidays and school holidays. Traffic on these days is reflective of an ”average” day across the year. Traffic flow is reported for different vehicles, distinguishing between buses and coaches, cars and taxis, light and heavy good vehicles, pedal cycles and motorcycles. We compute the annual average daily traffic flow for each LSOA by taking the average of traffic flow reported by count points within the LSOA boundary.

[14]It is possible that accidents involving light injuries are under-reported to the Police but this should be less of an issue for accidents involving serious injuries as well as fatalities. As long as the probability of reporting of accidents is random across time and is not correlated with the enforcement of the CCZ, which seems plausible, then this under-reporting should not affect our estimates.

[15]Lower Layer Super Output Areas (LSOAs) are boundaries delineated for census data collection and reporting. There are a total of 4,835 LSOAs across London with an average size of 0.325 square kilometers, housing around 1,700 residents from 675 households. These areas align to form 33 Local Area Districts (LAD) across London.

[16]We have also estimated models for serious injuries and fatalities separately. These results can be found in Table A3 in Data Appendix. We find that our effects are largely driven by serious injuries. Hence, when computing the monetary value of accident externalities associated with the charge in Table 7, we rely only on the monetized values of serious injuries to more conservatively measure the cost of accidents.

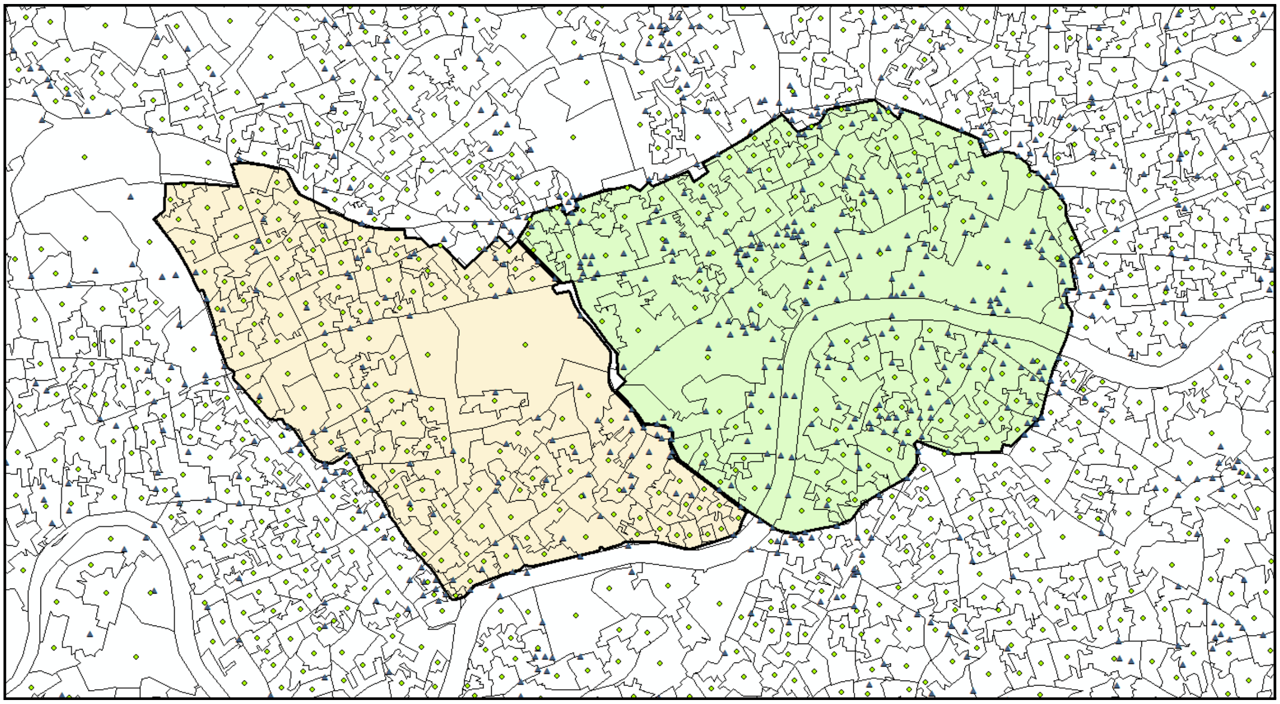

Figure 2: The London Congestion Charge Zone (CCZ & WEZ) and location of count points (denoted in triangle) where average annual daily traffic flow is recorded

Information on the boundaries of the CCZ and WEZ are collected from the shapefiles provided by Transport of London (TfL). Using Geographic Information Systems (GIS) mapping, we categorize whether LSOAs are inside or outside the charge zone based on whether the centroid of the LSOA is within the zone. For 17 LSOAs, the charge boundary cuts through the LSOA. [17] This shows that the charge boundary is not determined by census administrative zones. To avoid erroneously classifying LSOAs outside the charge zone as treated areas and vice versa, we omit these 17 areas, which are at both sides of the charge boundary, from our analysis. [18]

We supplement additional information on the average earnings, number of hours worked, number of jobs and job density from the Annual Labour Force Survey. We also collected data on population size and % of population from 18 to 25 years old from Nomis Population Estimates. All these data are reported annually at Local Authority District levels. We further collected information on daily weather conditions, which include precipitation, temperature, relative humidity and wind speed, from 212 monitoring stations across London maintained by London Air Quality Network (LAQN). We match each LSOA to the nearest monitoring station based on proximity from the centroid and annual measures are computed by averaging daily measures across the year. For

[17] To visualize, refer to the shaded LSOAs that overlapped with the CCZ in Figure A1 in Data Appendix.

[18] In unreported analyses, we incorporate these areas in our regressions and document largely similar findings. These results are available upon request.

further details on the definition of these variables, refer to Table A1 in Data Appendix.

5 Identification Strategy

To estimate the effects of traffic on the probability and severity of traffic collisions, a typical specification takes the following form:

$$Y_{jt}^{n} = exp(\alpha_j^n + \beta^{\mathbf{n}}\mathbf{logF_{jt}} + X'_{jt}\phi^n + \tau_t^n + \varepsilon_{jt}^n), \qquad \varepsilon_{jt}^n = \nu_{\mathbf{jt}}^{\mathbf{n}} + \epsilon_{jt}^n \tag{1}$$

where Y_{jt}^n is the counts for n collision outcome, which includes Accidents, Injuries, Serious Injuries & Fatalities, in area j at year t. Because collision outcomes follow an implicit count process and only take non-negative integer values, using Ordinary Least Squares (OLS), which specifies a conditional mean function that takes negative values, could yield inconsistent estimates (Cameron & Trivedi, 2013). Hence, we implement Poisson count regression models estimated using Pseudo Maximum Likelihood. $\mathbf{F_{jt}}$ is the average annual daily traffic flow (AADT) in neighbourhood j at year t. The key variable of interest, $\beta^{\mathbf{n}}$, measures the percentage change for different collision outcomes from a 1% change in traffic flow, i.e. the elasticity of the expected number of accidents/injuries with respect to traffic flow.[19] To minimise salient differences between neighbourhoods, we further control for time-varying neighbourhood specific characteristics X'_{jt} (e.g population size, % of young drivers etc) that could be correlated with $\mathbf{F_{jt}}$ and affect Y_{jt}^n. α_j^n denotes area fixed effects that partials out time-invariant area-specific unobservables that could correlate with traffic flow and affect collision outcomes. τ_t^n represents time fixed effects that captures the general trends in for different collision outcomes over time. ε_{jt}^n denotes the error term.

For consistent estimation, the least square estimator of $\beta^{\mathbf{n}}$ requires $E[\varepsilon_{jt}^n|F_{jt}] = 0$. This assumption, however, is likely to be violated because of the endogenous relationship between traffic flow and collision outcomes, which could be driven by omitted variable bias, measurement error and reverse causality. The endogenous variation in traffic flow is denoted by $\nu_{\mathbf{jt}}^{\mathbf{n}}$. For instance, roads that are frequently used could receive more funding for maintenance, which could result in fewer accidents. Failure to adequately control for these factors, which could be unobserved to researchers, could underestimate $\beta^{\mathbf{n}}$. Traffic collisions can also affect traffic flow if drivers are attracted to safer roads that are less congested. This reverse causality, which cannot be easily

[19] For an additional driver to impose a negative accident externality to the rest of the road users, $\beta^{\mathbf{n}}$ should exceed 1 as a 1% increase in traffic corresponds to a more than 1% increase in collisions.

addressed by controlling for observable differences, could again attribute to a downward bias to $\beta^{\mathbf{n}}$.[20] ϵ_{jt}^{n} is assumed to be uncorrelated with all the right-hand side variables.

To overcome these challenges, we adopt a control function approach relying on the exogenous variation in traffic induced by the London Congestion Charge. This method, which has been adopted widely by Rivers & Vuong (1988), Blundell & Powell (2003), Blundell & Powell (2004) and Wooldridge (2015), allows for the estimation of non-linear models with continuous endogenous variables. This is particularly relevant to our context as we are modelling discrete non-negative collision outcomes with Poisson count regressions with continuous traffic flow as the endogenous variable. Put differently, we are exploiting the local changes in charged traffic flow induced by the Congestion Charge to understand how an additional driver can affect the risk and severity of traffic collisions. The control approach is essentially a two-step approach, which resembles closely to an instrumental variable approach, includes:

$$logF_{jt} = \lambda_j + \gamma \mathbf{CCZ_{jt}} + X'_{jt}\rho + \psi_t + e_{jt}, \qquad e_{jt} = \nu_{\mathbf{jt}} + \mu_{jt} \tag{2}$$

$$Y_{jt}^{n} = exp(\alpha_j^n + \beta^{\mathbf{n}}\mathbf{logF_{jt}} + X'_{jt}\phi^n + \tau_t^n + \delta^{\mathbf{n}}\mathbf{\hat{e}_{jt}} + \epsilon_{jt}^{n}) \tag{3}$$

where F_{jt} is the annual average daily traffic flow for LSOA j at year t. CCZ_{jt} is an indicator variable that takes the value of 1 if LSOA j is located in the Congestion Charge Zone after the charge is implemented. Equation 2 is analogous to a first-stage regression and γ measures the effectiveness of the charge on reducing traffic flow in the charge zone. We also present reduced form estimates that measure the impact of the charge on road safety by replacing F_{jt} with various collision outcomes. λ_j and α_j^n in equation 2 and 3 represent LSOA fixed effects that partial out area-specific time-invariant unobservables, while ψ_t and τ_t^n denote year fixed effects that control for general trends in traffic and collisions over time. We further control for a vector of time-variant area specific characteristics (population size, job density, proportion of population from 18 to 25 years old, annual pay and hours worked) and weather conditions (temperature, wind speed, rainfall and humidity). We also partial out the effects of the implementation of the Western Extension Zone (WEZ) on traffic flow and collisions. As emphasised above, we do not focus on the WEZ as we report a negligible impact of the policy on traffic flow in our preferred specification, implying that the extension of the zone is not a suitable instrument. [21] The ineffectiveness could explain

[20] This is reflected in our analysis as our estimates from naive Poisson regressions are 10-20 times smaller than our estimates from control function approach in which we exploit the exogenous variation in traffic flow from the Congestion Charge.

[21] These results are summarized in Table A2 in Data Appendix. We run several specifications to examine the impact of the WEZ on traffic flow. First, we aggregate both areas (CCZ and WEZ) in column 1, before

why the WEZ was prematurely terminated by the end of 2010 three years after enforcement.[22] All these controls are subsumed under vector X'_{jt}.

The first stage error term, e_{jt}, can be decomposed into two components: (*i*) endogenous variation in traffic flow that is captured by ν_{jt}, and (*ii*) random variation in traffic flow that is denoted by μ_{jt}. While we do not directly observe ν_{jt}, we are able to estimate e_{jt} that captures the endogenous variation in traffic flow ($\nu_{\mathbf{jt}}$) from equation (2). We plug in the first-stage residuals, $\hat{e}_{jt}$, into Equation (3) as a control variable to partial out the endogenous variation in traffic flow. Equation (3) is a non-linear control function that measures the elasticity between traffic flow and various collision outcomes. This is essentially equation (1) other than the inclusion of $\hat{e}_{jt}$. For the control function to consistently estimate β^n, the assumption is that after controlling for $\hat{e}_{jt}$, $E[\epsilon^n_{jt}|F_{jt}] = 0$ as the error term, ϵ^n_{jt}, is uncorrelated with traffic flow. This assumption only holds if the Congestion Charge substantially reduces traffic flow such that $\gamma < 0$ (instrument relevance), and is only affecting collision outcomes through changes in traffic flow (exclusion restriction).

There are several reasons why the exclusion restriction could be violated. For instance, the charge could affect traffic collisions through changing the composition of traffic. As mentioned earlier, two-wheelers, such as motorbikes and bicycles, are not required to pay the charge and the CCZ could encourage commuters to switch to these modes.[23] Transport for London (TfL) also increase the number of bus routes and the frequency of buses to encourage drivers to switch to public transit after the charge is enforced. Furthermore, the charge could also increase the number of bigger more luxurious cars in the zone as wealthier drivers are more able to afford the charge. These changes in traffic composition that correlate with the policy could increase the variance in the weight of the vehicle fleet in the zone, affecting both the seriousness and probability of traffic accidents (Van Ommeren *et al.*, 2013; Anderson & Auffhammer, 2014).

We adopt the following approaches to mitigate these concerns. First, we minimize differences in traffic composition by constraining the analysis to neighbourhoods around the charge boundary. The idea is that commuters are likely to bypass these areas when commuting into the charge zone. Hence, we progressively constrain our analysis up to areas within 2km from the charge boundary.

splitting these areas up to measure independently their effects on traffic in column 2. Next, in column 3, we further examine the removal of the WEZ before focusing the analysis on WEZ only, removing areas in the CCZ from the analysis in column 4. We further relax the parallel trend assumption by including local authority specific trends, before limiting the analysis to neighbourhoods within 2km from the charge boundary in columns 5 and 6. None of these specifications indicate that the WEZ has a significant impact on traffic flow.

[22] The mayor of London then Boris Johnson was quoted saying that there was no significant downside for the removal of the WEZ as it did not substantially increase congestion (BBC, 2011). These findings seem fairly consistent with our results indicating that the policy is ineffectual to begin with.

[23] It is unlikely that residents living inside the charge zone will switch to un-charged vehicle modes as they are entitled to a 90% discount to the charge as mentioned earlier.

Second, we conduct a battery of balancing tests on driver and vehicle characteristics involved in collisions. This is possible because the administrative database provided detailed information on the age and gender of drivers, the engine size, the number and type of vehicles for each accident. Therefore, we can directly test whether the driver and vehicle characteristics involved in collisions in the zone changed after the charge is implemented. These regressions are analogous to equation (2) but we replace the dependent variable with the average driver and vehicle characteristics involved in collisions. Finally, we estimate models that control for driver and vehicle characteristics and models that exclude accidents related to uncharged vehicles only. This again ensures that our estimates are not driven by changes in the composition of drivers or vehicle types due to the congestion charge. Further details will be provided in the robustness section.

6 Results

6.1 Summary Statistics

Table 1 reports summary statistics for traffic flow, various collision outcomes, observable characteristics associated with LSOAs, and average driver and vehicle characteristics involved in collisions inside and outside of the charge zone. We then present similar statistics for areas outside but within 2.5km from the CCZ boundary. [24] In total, there are 20,858 observations from 1661 LSOAs (of which a minority, 68 LSOAs are inside the zone). When we limit the analysis to areas 2.5km from the CCZ boundary, the number of observations drop by more than 3158 for 226 LSOAs, of which almost one third is inside the zone.

Overall, we observe that roads within the CCZ are more dangerous. The number of collisions, injuries and fatalities are higher compared to areas outside. On average, we are observing 28.2 accidents every year for LSOAs within the CCZ, attributing to 27.4 slight injuries, 4.1 serious injuries and 0.15 fatalities. These collision statistics are substantially lower for areas just outside the charge zone, although traffic flow differences are not that large (average daily traffic flow is only about 20 % lower inside the charge zone (19,008) compared to areas right outside (24,475)). There are on average 11.8 collisions that resulted in 11.7 slight injuries, 1.62 serious injuries and 0.07 fatalities. This disparity in collision outcomes is likely driven by differences in the road network, traffic and pedestrian composition. For instance, there could be more pedestrians in the CCZ because of the higher concentration of economic activities (e.g Central Business District, Shopping Belts along Oxford Street), which could elevate the severity of collisions. This is supported by our data as the number of jobs and job density are considerably higher for neighbourhoods within the charge

[24] We choose 2.5km as the cutoff because this distance incorporates all the LSOAs within the CCZ.

Table 1: Summary statistics for LSOAs inside and outside of the Congestion Charge Zone

	Outside CCZ	Inside CCZ	2.5km Outside	Within 2.5km
Collision & Traffic Outcomes				
Accidents	8.55	28.21	11.75	16.86
	(0.06)	(1.24)	(0.23)	(0.44)
Slight Injuries	9.21	27.40	11.70	16.57
	(0.07)	(1.21)	(0.24)	(0.43)
Serious Injuries	1.15	4.10	1.62	2.39
	(0.01)	(0.19)	(0.04)	(0.07)
Deaths	0.07	0.15	0.07	0.10
	(0.00)	(0.01)	(0.01)	(0.01)
Traffic Flow (all)	24316.19	19007.53	24474.87	22778.23
	(142.20)	(289.59)	(330.70)	(249.21)
Accident/Flow (per thousand)	0.82	1.78	0.88	1.16
	(0.02)	(0.09)	(0.06)	(0.05)
Slight Injuries/Flow (per thousand)	0.89	1.73	0.86	1.13
	(0.03)	(0.09)	(0.06)	(0.05)
Serious Injuries/Flow (per thousand)	0.11	0.26	0.12	0.16
	(0.00)	(0.02)	(0.01)	(0.01)
Deaths/Flow (per thousand)	0.01	0.01	0.00	0.01
	(0.00)	(0.00)	(0.00)	(0.00)
Local Authority Characteristics				
Gross Annual Pay	28109.12	33733.15	32619.46	32965.07
	(33.42)	(182.97)	(107.13)	(93.62)
Average Number of Hours Worked	37.53	37.15	37.22	37.20
	(0.00)	(0.03)	(0.02)	(0.02)
Average Number of Jobs	135596.44	353700.00	281513.31	303914.50
	(654.00)	(6076.61)	(3992.70)	(3389.44)
Job Density	0.80	7.04	1.74	3.39
	(0.00)	(0.55)	(0.02)	(0.18)
% of Population from 18 to 25	0.10	0.13	0.11	0.12
	(0.00)	(0.00)	(0.00)	(0.00)
Population Size	253179.69	218617.96	226277.20	223900.36
	(385.41)	(2241.46)	(849.76)	(911.53)
Driver & Vehicle Characteristics (involved in collisions)				
% female	0.23	0.12	0.16	0.15
	(0.00)	(0.00)	(0.00)	(0.00)
Average age band	5.63	5.87	5.66	5.72
	(0.01)	(0.02)	(0.02)	(0.02)
Average number of vehicles	1.40	1.33	1.37	1.36
	(0.00)	(0.00)	(0.00)	(0.00)
Average engine size	1879.17	2224.53	1988.90	2062.88
	(5.31)	(25.31)	(17.51)	(14.54)
Average age of vehicle	5.57	4.52	5.02	4.86
	(0.02)	(0.05)	(0.05)	(0.04)
Sample Size	19878	980	2178	3158
No of LSOAs	1593	68	158	226

Table 1 records the means and standard error of means (in parenthesis) for (1) LSOAs outside the CCZ, (2) inside the CCZ, (3) outside the CCZ but within 2.5km from the boundary, and (4) within 2.5km from the boundary inside and outside the CCZ.

zone. Higher concentration of commercial activities could also explain why residential population is lower in the zone compared to areas outside. We also observe that households residing within the charge zone are, on average, earning higher incomes than those living outside. These differences in observable characteristics are smaller once we reduce our sample to areas within 2.5km from the CCZ boundary.

While we do not observe stark differences in the average age of the drivers and the average number of vehicles involved in collisions per accident between areas inside and outside the charge zone, we find that cars involved in accidents have larger engine size and are, on average, newer in the charge zone. These results are suggesting that drivers involved in collisions in the charge zone are more affluent driving newer and bigger cars. Drivers involved in accidents in the charge zone are also less likely to be females. Noticeably, these differences in driver and vehicle characteristics are less discernible as soon as we limit our analysis to areas around the charge boundary.

6.2 Baseline Estimates

Table 2 reports the impact of the London Congestion Charge on traffic flow (Columns 1) and various collision outcomes (Columns 2-4). These estimates are interesting as they measure the efficacy of the CCZ in curtailing traffic and improving road safety. Our estimates suggest that traffic flow is around 13.5% lower after the charge is implemented, which is around 2566 fewer vehicles every day.[25] Strong first-stage F-statistics (> 10) further verify the strength of the CCZ as an instrumental variable. Corresponding to this reduction in traffic flow in the charge zone, we are bserving a 6.9% and 9.7% reduction in accidents and minor injuries. This is around 1.9 and 2.6 fewer accidents and slight injuries per area every year. Our results are consistent with the findings of Green *et al.* (2016) although our estimates are considerably smaller.[26]

An important departure in our findings is that the congestion charge increases severity of traffic collisions. After the charge is enforced, the number of serious injuries/fatalities increase by

[25] This is computed by taking the $exp(-0.1450)$ before subtracting by 1. Absolute effects are computed by multiplying percentage changes with the mean dependent variable.

[26] Green *et al.* (2016) document a substantial 35% reduction in accidents, 25% in serious injuries and 35% in fatalities. Differences in our estimates could stem from the disparity in the empirical strategy as we are exploiting variation in collision outcomes between areas around London close to the charge boundary. In contrast, they compare collision outcomes in the charge zone with 20 major cities across United Kingdom using a difference-in-difference and synthetic control strategy. Another possible explanation is that there are spillover reductions in traffic from the charge for areas outside the charge zone. This is supported by Green *et al.* (2016) who document a 10 to 12% decrease in accidents for areas within 4km from the charge zone. Nevertheless, subtracting these spillover effects from their headline estimates still yield a net reduction of around 25% (35% - 10%), which is still considerably larger than our estimates. Hence, their results could be capturing possible changes in traffic composition driven by the charge that could affect collision outcomes.

Table 2: Baseline Estimates: Poisson First Stage & Reduced Form regressions

	(1) Traffic	(2) Accidents	(3) Slight	(4) Serious Inj & Fatalities
CCZ	-0.1450***	-0.0716***	-0.1017***	0.1100***
	(0.0270)	(0.0248)	(0.0288)	(0.0387)
Obs	20863	20858	20852	20354
1st Stage F-stats	40.39			
Mean Dep Variable	19007.53	28.21	27.40	4.31
No.of LSOAs	1663	1661	1660	1581
% Δ (CCZ)	-13.50	-6.91	-9.67	11.63

Dependent variable is the counts of various collision outcomes or the natural log of traffic flow as denoted in the column headers. **CCZ** is a binary variable denoting LSOAs in the charge zone after the charge is implemented. All regressions are estimated with LSOA and year fixed effects. Other control variables include job density, % of population from 18 to 25 years old, total population size, gross annual income,hours worked collected at Local Authority Level at year t and mean wind speed, temperature, precipitation and relative humidity collected at LSOA level at year t. We estimate the first stage and reduced form effects of the CCZ on traffic flow (Columns 1) and various collision outcomes (Columns 2-4) using Ordinary Least Squares and Pseudo Poisson Maximum Likelihood regressions respectively. Mean Dep Variable is the average daily traffic flow or traffic outcomes in the CCZ before the charge is implemented. %Δ is the percentage change in traffic flow and various collision outcomes and they are computed by taking exponential of the estimated effects before subtracting by 1. 1st Stage F-stats reported is the Kleibergen-Paap rk Wald F statistic from first stage regressions. Robust standard errors (in parenthesis) are clustered at LSOA. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

11.6%, which is around 0.50 more serious injuries/fatalities per area every year. This finding is consistent with the idea that removing traffic along busy roads reduces bottlenecks and increases travelling speed, resulting in more serious injuries. Indeed, initial estimates by TfL has shown that driving speed in the CCZ are almost 20% higher, with time delays dropping by 30% (TfL, 2003a).[27] Overall, this result suggests that an unintended consequence of alleviating congestion along busy roads using the Congestion Charge is that accidents can become more severe.

Next, in Panel A of Table 3, we present control function estimates from the estimation of equation (3) to estimate the causal effect of traffic flow on various collision outcomes. As mentioned earlier, we plug the residuals from first stage regressions (reported in Column 1 Table 2) in equation 2 to partial out the endogenous variation in traffic flow. This is akin to the second stage of an instrumental variable (IV) regression. We document that higher traffic flow is associated with

[27]Ideally, we will like to produce similar estimates between driving speed and various collision outcomes. Micro-data for driving speed, however, is not readily available for us to examine this relationship.

Table 3: Baseline Estimates: Control Functions & Naive Poisson Regressions

	(1) Accidents	(2) Slight	(3) Serious Inj & Fatalities
Panel A: Control Function			
Ln(Traffic) (β^n)	0.48***	0.68***	-0.74***
	(0.16)	(0.19)	(0.26)
Collision Rate Elasticity(β^n-1)	-0.52***	-0.32***	-1.74***
	(0.16)	(0.19)	(0.26)
Panel B: Naive Poisson Regressions			
Ln(Traffic) (β^n)	0.02	0.02	-0.05
	(0.02)	(0.02)	(0.04)
Collision Rate Elasticity(β^n-1)	-0.98***	-0.98***	-1.05***
	(0.02)	(0.02)	(0.04)
Obs	20858	20852	20354
Mean Dep Variable	28.21	27.40	4.31
No.of LSOAs	1661	1660	1581

Dependent variable is the counts of various collision outcomes denoted in the column headers. **Ln(Traffic)** is the natural logarithm of annual average daily traffic flow. The sample size, mean dependent variable and number of LSOAs are consistent across panel A and B. All regressions are estimated with LSOA and year fixed effects. Other control variables include job density, % of population from 18 to 25 years old, total population size, gross annual income,hours worked collected at Local Authority Level at year t and mean wind speed, temperature, precipitation and relative humidity collected at LSOA level at year t. In Panel A, we estimate the elasticity between traffic flow and various collision outcomes using non-linear control functions. In particular, we plug in residuals from first stage traffic flow regressions to address the endogeneity concerns between traffic flow and collision outcomes. In Panel B, we report the elasticity between traffic flow and various collision outcomes from naive Poisson regressions. Collision Rate Elasticity is the elasticity of various collision rates with respect to traffic flow and is computed by subtracting the estimated β^n with 1. Mean Dep Variable is the average daily traffic flow or traffic outcomes in the CCZ before the charge is implemented. Robust standard errors (in parenthesis) are clustered at LSOA. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

more accidents and slight injuries, congruent with the reduced form estimates reported in Table 2.[28] Specifically, a 1% increase in traffic flow corresponds to a less than proportional 0.48% and 0.68% increase in the absolute number of accidents and slight injuries respectively. These results imply that the marginal driver reduces the accident risk for other road users as the accidents and

[28]We also estimate similar regressions using a traditional IV approach using OLS. These results are summarized in Table A5 in Data Appendix. To avoid omitting observations with 0 collision outcomes, we transform our dependent variable by taking the natural logarithm of 1 + various collision outcomes. These results are very similar in terms of direction and magnitude for accidents and slight injuries, but the estimates for serious injuries/fatalities are quite different. The disparity in the estimated effects could be driven by the substantial number of zeros for serious injuries/fatalities as severe accidents rarely happen. Logarithm transformation for these outcomes could materially affect the estimated effects.

slight injuries rate elasticity with respect to traffic flow are -0.52 and -0.32 respectively. This is computed by subtracting β^n with 1. Based on the standard errors reported in parenthesis below, these collision rate elasticities across various outcomes are quite precisely estimated and are statistically smaller than one, confirming that the additional driver do not heighten the collision risk for other road users.

While these estimates are noticeably smaller than previously reported by Romem & Shurtz (2016) for Israel, a plausible explanation is that the accident-traffic flow relationship could be non-monotonic and that we are estimating this relationship at a different margin. Roads are considerably more congested around Central London and an additional driver along a busy road can impose congestion externalities to other drivers. By slowing down driving speeds for other motorists, the marginal driver could reduce the severity of traffic collisions for other road users. This is supported by our results as in column 3 we document that a 1% increase in traffic flow leads to a 0.74% reduction in the serious injuries/fatalities, which means that the serious injury and fatality rate elasticity with respect to traffic flow is -1.74.

In Panel B, we further present elasticity estimates from naive Poisson regressions as a benchmark. These estimates are remarkably smaller. A 1% increase in traffic flow corresponds to a 0.021% and 0.020% increase in accidents and slight injuries, and a 0.047% decrease in serious injuries/fatalities. None of these estimates is precise enough to be statistically significant from zero. These results corroborate with the idea that, without accounting for the endogenous relationship between traffic density and collision outcomes, naive poisson estimates could underestimate the accident-traffic relationship. For instance, planners could invest more in maintaining heavily used roads to make them less accident prone. Failure to account for these road infrastructure budgets (or omitted variables) could severely underestimate the relationship between traffic density and accidents. Another explanation is that there could be due to attenuation bias driven by classical measurement error. This could arise when traffic flow is imprecisely estimated and interpolated, and could be further exacerbated when we exploit changes in traffic flow within LSOA over time. The differences between naive Poisson estimates and our control function estimates suggest that these empirical challenges are adequately addressed with our strategy.

6.3 Robustness Tests

Table 4 summarizes a battery of robustness and placebo tests to ensure that earlier estimates are not spuriously driven by factors unrelated to traffic flow. Here, we report the **reduced form effects (Red)** and **control function elasticity estimates (Ctr)** of traffic flow on Accidents, Slight

Injuries, Serious injuries/Fatalities in Panel A to C. Panel D further reports first-stage regression estimates that measures the impact of the Congestion Charge on traffic flow.[29]

Limit to Proximate areas: Our analyses so far incorporate areas across London. As mentioned, the Congestion Charge could influence collision outcomes by changing the composition of traffic in the charge zone. Some vehicle types, such as motorcycles, buses and bicycles, are omitted from paying the charge. Incorporating areas further from the charge boundary runs the risk that composition of traffic could be very different inside and outside the charge zone. Areas nearer to the charge boundary, however, should share the same traffic composition as vehicles entering the charge zone will need to bypass these areas.

Hence, in columns 1 and 2, we limit our analysis to areas within 2.5km from the charge boundary. This is measured based on the euclidean distance of the centroid of the LSOA from the charge zone boundary. This strategy constrain the analysis to around 15% of full sample, to 226 LSOAs around the charge boundary. However, our results remain fairly comparable to earlier estimates. Like before, we find that an additional driver do not impose a negative accident externality to other road users as the estimated elasticity of the expected number of accidents/injuries with respect to traffic flow is less than 1. Specifically, a 1% increase in traffic corresponds to a 0.64% and 0.81% increase in accidents and minor injuries that is statistically different from zero and statistically smaller than one. Although we find that a 1% increase in traffic reduces serious injuries/fatalities by 0.65%, this relationship is no longer statistically different from zero at any conventional levels. Nevertheless, we can conclude that it is improbable for the marginal driver to increase the severity of traffic collisions around Central London.

Notably, compared to our baseline estimates from the full sample, these elasticity estimates are larger because the impact of the CCZ on traffic flow around the charge boundary is less pronounced. This is consistent with the findings from Green *et al.* (2016) who show that the congestion charge has spillover reduction of traffic for areas circumventing the zone. We prefer estimates from these analyses as they are less susceptible to possible changes in vehicle composition driven by the congestion charge.

To visualize how these elasticity estimates change with distance from the CCZ boundary, we further replicate our analysis for areas from 2km to 10km from the charge boundary at 1km intervals and plot these estimates for various collision outcomes in Figure 3. In other words, we are repeating the analysis by constraining the analysis to areas proximate to the charge boundary. For instance, elasticity estimates at 5km incorporates areas within 5km from the charge boundary. As observed, these elasticity estimates are quite stable in size across the distance bandwidths.

[29]Take note that all the control functions, estimated for the different collision outcomes, share a similar first stage as the only endogenous variable here is traffic flow.

Table 4: Robustness Tests

	<=2.5km		Alt Matching		LA Linear Trends		Driver/Veh Char		Charged Veh		4-Wheels		Placebo Time		Placebo Area	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	Red	Ctr	Red	Ctr	Red	Ctr	Red	Ctr	Red	Ctr	Red	Ctr	Red	Ctr	Red	Ctr
Panel A: Accidents																
CCZ	-0.0625**		-0.0517*		-0.0581**		-0.0674**		-0.0579*		-0.0570*		0.0355		0.0477	
	(0.0298)		(0.0264)		(0.0285)		(0.0285)		(0.0298)		(0.0297)		(0.0267)		(0.0446)	
Ln(Traffic) (β^n)		0.6419**		0.3716**		0.5094**		0.6424**		0.4962**		0.5848*		3.4857		0.8174
		(0.3027)		(0.1888)		(0.2452)		(0.2690)		(0.2520)		(0.3006)		(2.6046)		(0.7685)
Collision Rate Elasticity (β^n-1)		-0.36***		-0.63***		-0.49***		-0.36***		-0.50***		-0.42***		2.49*		-0.18***
Mean Dep Variable		28.21		21.21		28.21		29.02		28.21		24.81		36.66		14.43
Panel B: Slight Injuries																
CCZ	-0.0793**		-0.0873***		-0.0738**		-0.0865***		-0.0764**		-0.0699**		0.0022		0.0232	
	(0.0334)		(0.0298)		(0.0316)		(0.0321)		(0.0334)		(0.0345)		(0.0311)		(0.0485)	
Ln(Traffic) (β^n)		0.8098**		0.6260***		0.6412**		0.8190***		0.6470**		0.7129**		0.2615		0.3769
		(0.3402)		(0.2146)		(0.2728)		(0.3036)		(0.2823)		(0.3507)		(3.0326)		(0.8352)
Collision Rate Elasticity (β^n-1)		-0.19***		-0.37***		-0.36***		-0.18***		-0.35***		-0.29***		-0.74**		-0.62**
Mean Dep Variable		27.40		20.56		27.40		28.19		27.40		23.89		36.02		14.19
Panel C: Serious Inj & Fatalities																
CCZ	0.0634		0.0881*		0.0476		0.0489		0.0704		0.0582		0.0594		0.0365	
	(0.0511)		(0.0523)		(0.0526)		(0.0504)		(0.0511)		(0.0565)		(0.0724)		(0.0801)	
Ln(Traffic) (β^n)		-0.6459		-0.6230*		-0.4095		-0.4672		-0.6031		-0.5934		5.7260		0.6232
		(0.5197)		(0.3719)		(0.4541)		(0.4754)		(0.4300)		(0.5750)		(7.0957)		(1.3981)
Collision Rate Elasticity (β^n-1)		-1.65***		-1.62***		-1.41***		-1.47***		-1.60***		-1.59***		4.73		-0.38
Mean Dep Variable		4.31		3.28		4.31		4.41		4.31		3.68		5.60		2.12
Panel D: First Stage Regressions (Log Traffic Flow)																
CCZ	-0.0983***		-0.1395***		-0.1159***		-0.1059***		-0.1190***		-0.0983***		0.0102		0.0568	
	(0.0292)		(0.0141)		(0.0283)		(0.0278)		(0.0310)		(0.0292)		(0.0071)		(0.0345)	
Obs	3158		6011		3158		3032		3156		3158		3924		4367	
Mean Dep Variable	19007.53		18676.85		19007.53		19229.70		19007.53		19007.53		23863.95		30763.32	
No.of LSOAs	226		414		226		226		226		226		1315		332	
1st Stage F-stats	15.93		75.43		21.04		15.93		14.66		15.93		2.72		1.85	

For Panel A to C, we report reduced form estimates (Red) of the CCZ on various collision outcomes in columns 1,3,5,7,9,11, 13 & 15 and elasticity between traffic flow and various collision outcomes from non-linear control functions (Ctr) in columns 2,4,6,8,10,12,14 & 16. These estimates are from Pseudo Poisson Maximum likelihood regressions. For Panel D, we report first stage regression estimates of the CCZ on traffic flow. The sample size and number of LSOAs are similar across reduced form and control function regressions and across Panel A to D. Collision Rate Elasticity is the elasticity of various collision rates with respect to traffic flow and is computed by subtracting the estimated β^n with 1. Dependent variable is the counts of various collision outcomes as denoted in the Panel . **CCZ** is a binary variable denoting LSOAs in the charge zone after the charge is implemented. **Ln(Traffic)** is the natural logarithm of annual average daily traffic flow. Specification adopted is similar to Table 3 unless otherwise stated. 1st Stage F-stats reported is the Kleibergen-Paap rk Wald F statistic from first stage regressions. Robust standard errors (in parenthesis) are clustered at LSOA. * p<0.10, ** p<0.05, *** p<0.01. In Columns 1-2, we constrain the analysis to LSOAs within 2.5km from the CCZ boundary measured based on the centroid of the LSOA from the charge boundary. In Columns 3-4, we computed weighted average traffic flow for LSOAs using countpoints (CPs) within 1km from the centroid of LSOAs using inverse distance weighting method. In Columns 5-6, we control for local authority districts specific linear trends. In Columns 7-8, we control for average driver and vehicle characteristics involved in traffic collisions. In Columns 9-10, we examine the relationship between charged flow and various collisions, holding constant the flow of uncharged vehicle flow, which includes pedal cycles, buses and motorcycles. In Columns 11-12, we adopt a specification similar to columns 1-2 but our dependent variable involves collision outcomes with at least 1 four-wheel vehicle. In Columns 13-14, we conduct the analysis using a placebo window 2 years before the actual implementation and constrain the analysis to observations before 2003. In Columns 15-16, we consider LSOAs outside the charge zone but within 1km from the CCZ boundary as treated areas and we exclude areas within the CCZ and WEZ from this analysis.

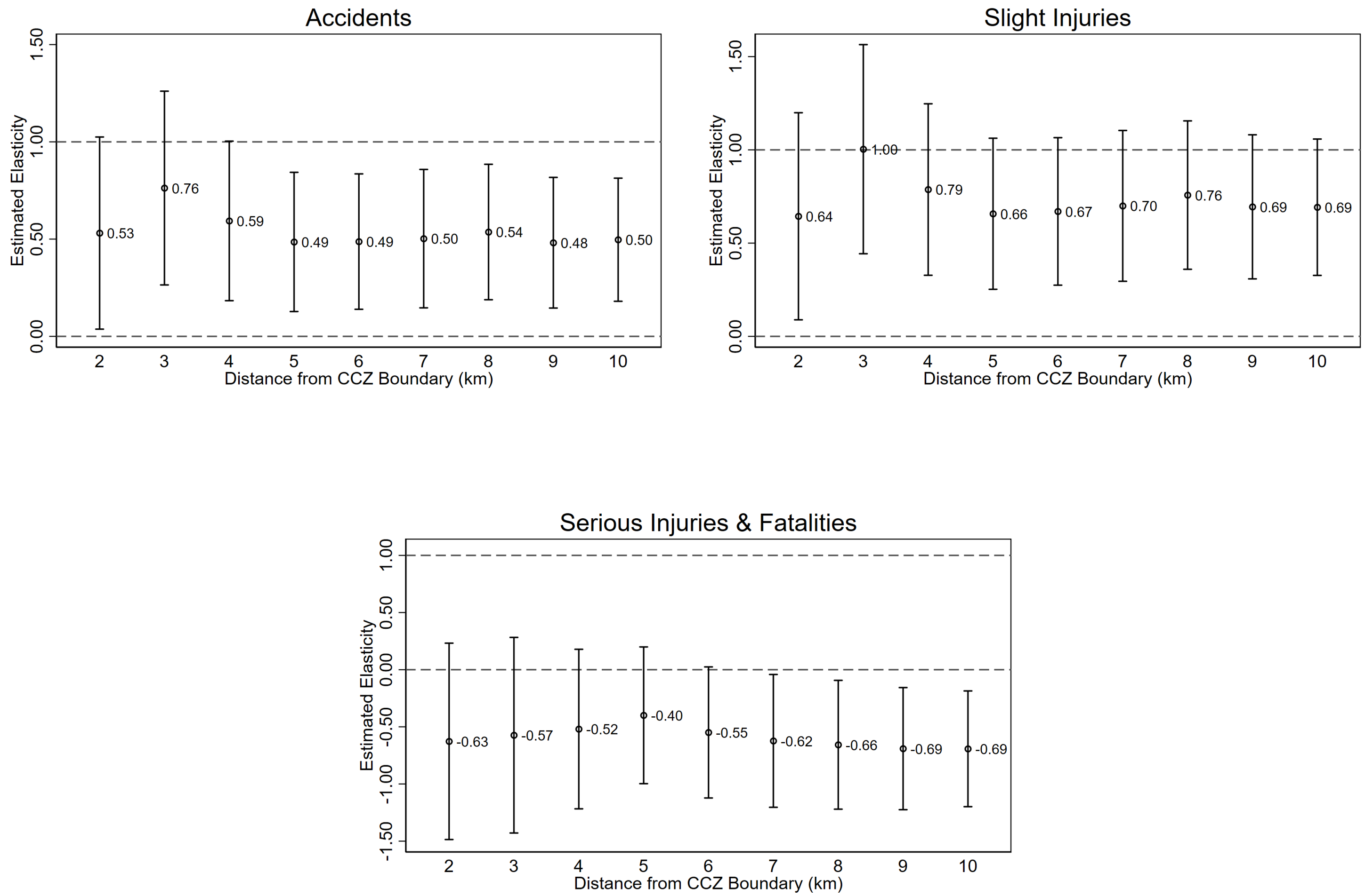

Figure 3: Plotted coefficients are the elasticity estimates between charged traffic flow and various collision outcomes for LSOAs across different distance bandwidths from the CCZ boundary. Tails denote 90% confidence intervals constructed from robust standard errors clustered at LSOA. Every point estimate is from a different regression that limits the analysis to LSOAs within the stipulated distance bandwidth from (e.g 5km limits the analysis to LSOAs within 5km from CCZ boundary). Specification adopted is similar to columns 1 and 2 of Table 4 in which we limit the analysis to areas <=2.5km from the charge boundary.

Specifically, a 1% increase in traffic is associated with a 0.48-0.76% increase in accidents, a 0.64-1.00% increase in slight injuries and 0.40-0.69% reduction in serious injuries/fatalities. These estimates are noticeably less precise for serious injuries/fatalities and we can no longer conclude they are statistically different from zero for areas within 6 km from the charge boundary. Even so, these results show that estimated elasticity of the expected number of accidents with respect to traffic flow is less than one across most of our estimates (8 out of 9 estimates), indicating that there is no negative accident externality associated with the marginal driver along congested roads in Central London.

Alternate Matching: Due to the way that we are matching traffic data to LSOAs, our analysis only incorporates LSOAs with at least 1 count points (CPs) within the boundaries. This reduces our sample considerably to 1663 LSOAs across London, which is around 35% of the 4833 LSOAs across London (See Table 3). Here, we adopt an alternate matching approach that allows us to match traffic to almost all the LSOAs. This is as follow: first, we identify all the CPs that are within 1km from the centroid of the LSOAs. We then remove any CPs that are outside the charge zone but that are matched to LSOAs within the charge zone. Third, we compute the annual daily traffic flow for each LSOA by taking the weighted average of traffic flow based on inverse distance weights. That is, traffic flow in CPs that are furthest or closest from the LSOA are given the least or most weight to more accurately measure local traffic conditions. This matching procedure increases our sample size from 227 to 413 LSOAs within 2.5km from the charge boundary.

Figure 4 illustrates how we match the CPs to LSOAs. All the CPs within the 1km buffer will be matched to the LSOA (denoted in dots). Results, summarized in columns 3 and 4, are quite similar to earlier findings although these elasticity estimates are smaller and more precise. This is largely due to the larger reductions of traffic flow driven by the charge (see Panel D) as we now incorporate areas further from the charge boundary due to the matching process. As mentioned earlier, these areas are less susceptible to spillover effects from the CCZ. Reduced form estimates on collision outcomes, on the other hand, are quite comparable with earlier estimates in terms of magnitude and direction, explaining why elasticity estimates are more conservative under this matching procedure.

Local Authority Linear Trends: In columns 5 and 6, we repeat the analysis in columns 1 and 2 but allow traffic and collision trends to vary linearly across 33 local authorities. This is estimated with the inclusion of 33 local authority linear year trends, which relaxes the assumption that areas inside and outside the CCZ must follow parallel trends in traffic flow and collisions in the absence of the charge enforcement. Doing so produces slightly smaller elasticity estimates due to larger first stage effects on traffic flow, but these results are largely in line with earlier findings. These findings suggest that our estimates are unlikely to be driven by differential trends between charged and uncharged areas.

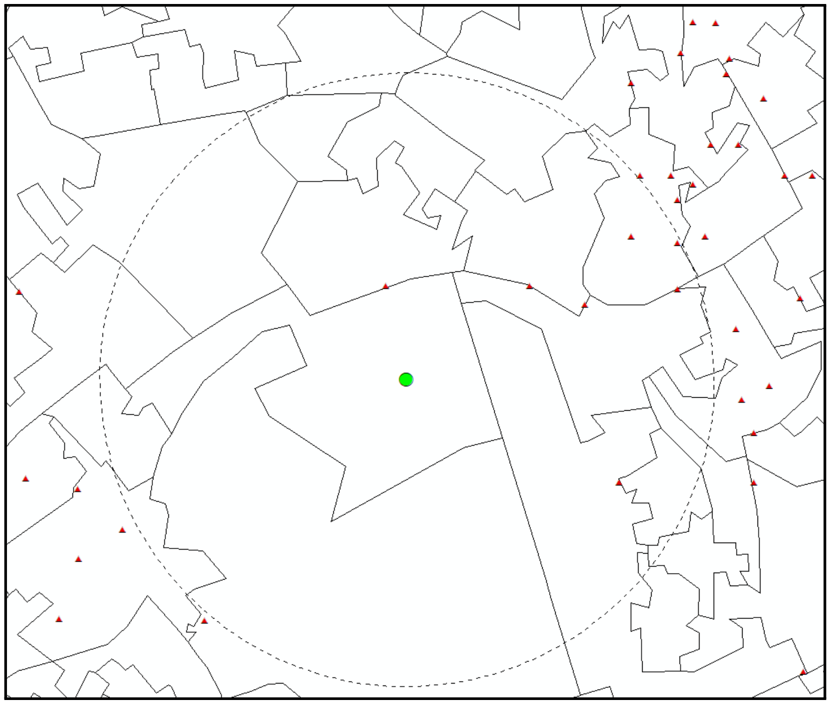

Figure 4: An example illustrating how Count Points (in triangle) within 1km buffer (in dash line) are matched to LSOAs (centroid denoted in dot).

Driver and Vehicle Characteristics: In columns 7 and 8, we include a vector of controls that measure the average characteristics of drivers and vehicles involved in traffic collisions. We interpret these estimates with caution as these variables are likely to be proxy "bad" controls - control variables that might partially control for omitted variables but are likely to be impacted by the enforcement of the charge. For instance, the charge could have affected the engine capacity of vehicles in the charge zone as affluent drivers are more likely to be able to afford the charge. Nevertheless, these estimates allow us to understand whether earlier results hold after controlling for possible changes in vehicle or driver composition. Estimates are slightly larger after controlling for these differences but they remain comparable with earlier results.

Uncharged Vehicle Flow: In columns 9 and 10, we replace traffic flow with charged vehicle traffic flow and control for uncharged vehicle flow, which includes buses and motorcycles, in our analysis. The objective is to provide estimates for charged vehicles only, which constitutes bulk of the traffic flow). The concern is whether the charge increases uncharged vehicle flow as drivers substitute from charged to uncharged modes to avoid the charge. This could directly affect collision outcomes and invalidate out instrument. Although subsequent balancing tests in Table 5 indicate this is not an issue, with no significant changes in the flow for 2-wheelers and buses documented after the charge is implemented, we control for changes in uncharged traffic flow as a precaution. We find almost identical results when focusing on charged vehicles only.

Four wheel accidents: In columns 11 and 12, we limit our analysis to accidents that involve at least 1 four-wheel vehicle. We are concerned whether earlier estimates could be spuriously driven by accidents caused by vehicles unaffected by the congestion charge. Here we are assuming that accidents that do not involve any charged vehicles (4 wheels or more) are unrelated to the reduction in charged vehicle flow driven by the congestion charge. Therefore, this analysis is likely to provide us with a lower bound elasticity estimate between traffic flow and accidents. As expected, we observe that the estimated effects, although slightly smaller, are fairly comparable with earlier results. A plausible explanation is that only a small proportion of accidents (around 5.5%) in our sample does not involve any four-wheel vehicles.

Placebo Time & Area: In columns 13 to 14, we summarize findings from our placebo test that shifts the treatment year to 2002, one year before the CCZ is enforced. The sample size is remarkably smaller because we remove any observations from 2003 onwards to avoid capturing any effects from the CCZ. First stage regression estimates in Panel D show that we are unlikely to document any spurious effects on traffic flow and collisions prior to the enforcement of the charge. We revisit this point more formally in latter event study regressions. Finally, in columns 15 and 16, we further create placebo treatment areas by shifting the charge boundary 1km outwards. Put differently, LSOAs outside of the CCZ and WEZ, but are within 1km from the charge boundary, are now defined as treated areas. LSOAs within the CCZ and WEZ are omitted from the analysis to avoid capturing any effects from the charge. For an illustration of the placebo treatment area, refer to Figure 5. Similar to earlier findings, we do not document any significant changes in traffic flow and collisions in these placebo treatment areas, suggesting that earlier results are unlikely to be spuriously driven.

Event Study Regressions: Next, we examine whether there are any pre-trends in traffic flow and collision outcomes prior to the enforcement of the CCZ by estimating the following Poisson **event study** regression represented by equation 4. Pre-existing trends in traffic flow or collisions could bias our first stage or reduced form estimates, affecting the exogeneity of the instrument and the consistency of the elasticity estimates.

$$Y_{jt}^{n} = exp(\vartheta_j^n + \sum_{\mathbf{g=<-3}}^{\mathbf{11}} \varphi^{\mathbf{ng}}\mathbf{CCZ_{jt}^{g}} + X_{jt}'\zeta^n + \pi_t^n + \eta_{jt}^n), \tag{4}$$

where g represents the **number of years from the year the CCZ is enforced** (e.g g = 0 represents the year the charge is enforced (in 2003), g = -3 represents 3 years before the charge is enforced). CCZ_{jt}^{g} takes the value of 1 if LSOA j is within the CCZ g years from the year the CCZ is enforced. These estimates are plotted in Figure 6. Plotted estimates (φ^{ng}) are the effects of the CCZ on traffic flow and different collision outcomes across the different years relative to the LSOAs **one year**

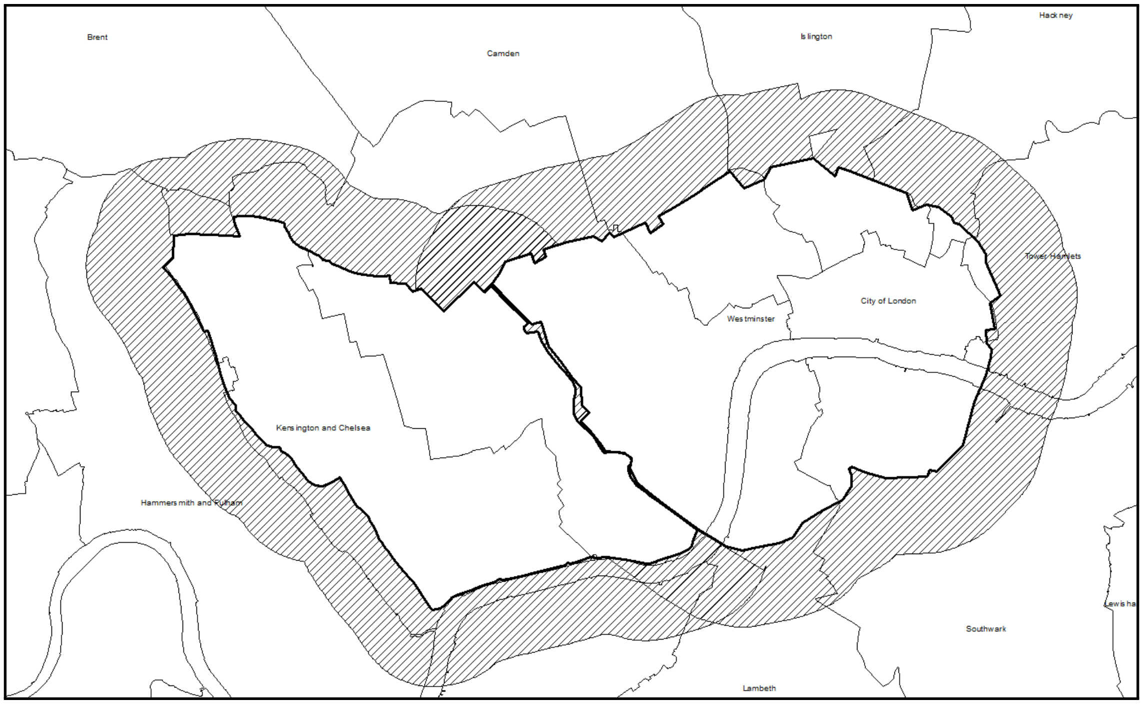

Figure 5: Placebo CCZ (shaded) 1km outside the CCZ and the WEZ.

before the CCZ is enforced (denoted by dash line at year -1). Tails denote the 90% confidence intervals. If there are no spurious effects before the charge is implemented, we expect φ^{ng} to be close to 0 when $g < 0$ (before the CCZ is implemented). Similar to equation 3, these regressions are estimated using Pseudo Maximum Likelihood to account for the count nature of the dependent variable and we constraint our analysis to LSOAs within 2.5km from the CCZ boundary.

As observed from Figure 6, we do not report any significant effects on traffic flow and various collision outcomes prior to the enforcement of the CCZ. These results suggest that our reduced form and first stage estimates are unlikely to be spuriously driven by pre-trends in traffic flow or collision outcomes. After the charge is implemented, we observe a significant reduction in traffic flow that is stable across post-enforcement years. Although the estimated effects on collision outcomes are considerably noisier once we break down the effects into different years, we document a general downward (upward) trend in accidents and slight injuries (serious injuries/fatalities) that is quite stable across the years. Overall, these results are consistent with the results when we include local authority district linear trends in Table 4, suggesting our estimates are unlikely to be driven by differential trends in collisions and traffic between areas inside and outside the charge zone.

Balancing Tests: A major threat to our identification strategy is that the CCZ could have affected collision outcomes through means other than traffic flow, violating the exclusion restriction assumption. To mitigate these concerns, we conduct a battery of balancing tests to examine whether the composition of drivers and vehicles involved in collisions, and the composition of uncharged

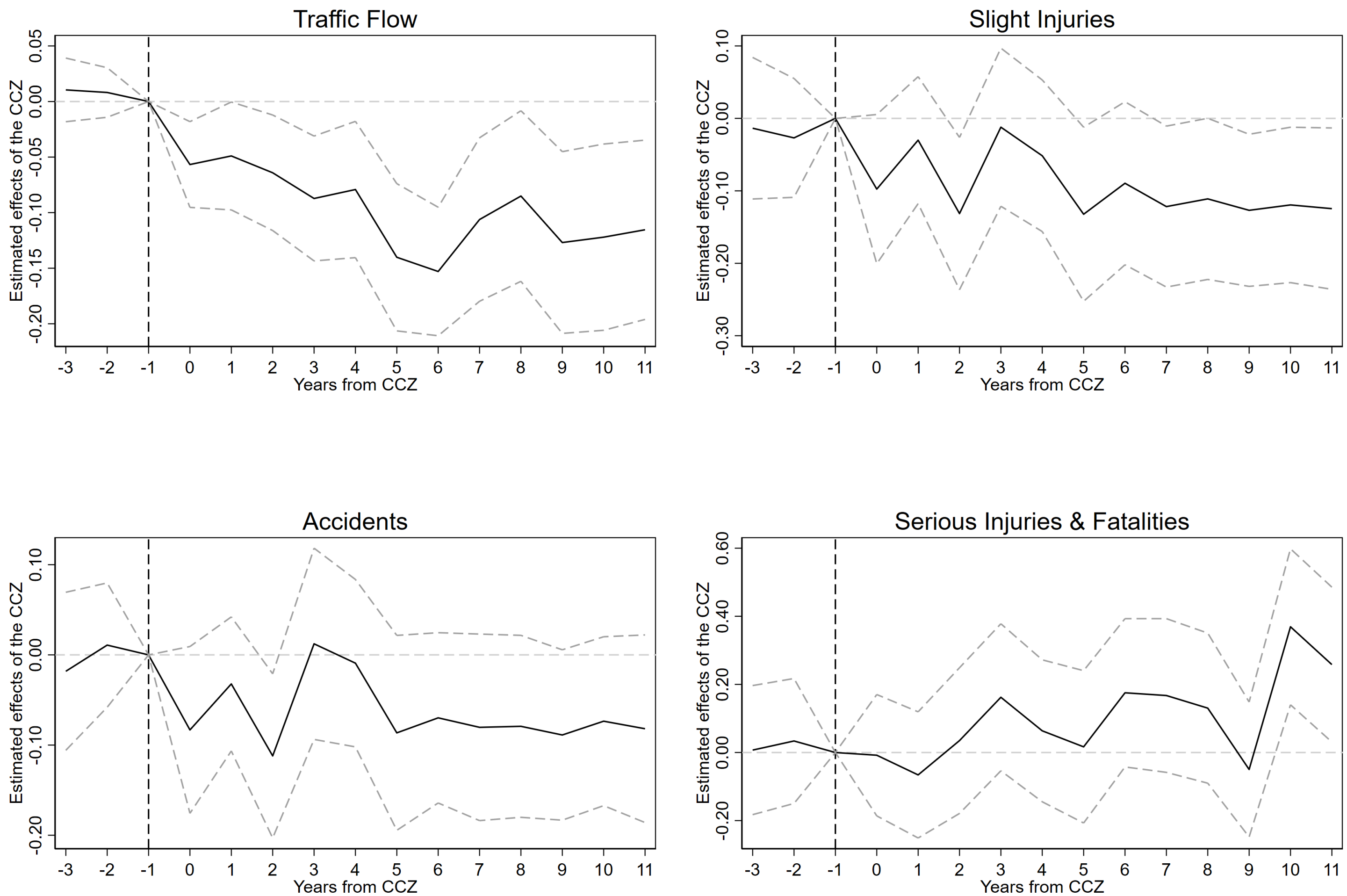

Figure 6: Event study first stage and reduced form impacts of the CCZ on traffic flow and various collision outcomes. Specification adopted is similar to columns 1 and 2 of Table 4 in which we limit the analysis to areas <=2.5km from the charge boundary. Omitted reference group consists of observations 1 year before the CCZ is implemented in 2002. 90% confidence intervals constructed from standard errors clustered at LSOA are represented by dashed lines.

Table 5: Balancing Tests for observed driver and vehicle characteristics involved in traffic accidents within 2.5km from charge boundary

	(1) Share Female	(2) Age Band	(3) Vehicle Count	(4) ln(Engine Capacity)	(5) Vehicle Age
CCZ	-0.0023	0.0765	0.0035	0.0803***	0.0137
	(0.0097)	(0.0693)	(0.0081)	(0.0299)	(0.2076)
Obs	3085	3073	3085	3050	3040
Mean Dep Variable	0.12	5.87	1.33	7.65	4.53
	(6) 2-Wheel Traffic	(7) Bus Traffic	(8) 2-Wheel Accidents	(9) Bus Accidents	(10) Pedestrians
CCZ	-0.0336	-0.0007	-0.0615	-0.0060	-0.0930**
	(0.0444)	(0.0361)	(0.0427)	(0.0732)	(0.0451)
Obs	3158	3158	3158	2983	3158
Mean Dep Variable	1566.70	1204.30	15.54	5.27	9.38

Dependent variable is the characteristics of drivers involved in collisions or the composition as labelled in the headers. In Column 1, the dependent variable is the ratio of the accidents with at least 1 female driver involved. In Column 2, the dependent variable is the average age band of drivers involved. In Column 3, the dependent variable is the average number of vehicles involved in the collisions. In Columns 4 and 5, the dependent variable is the average engine capacity and average vehicle age of vehicles involved in the collision respectively. Regressions for Columns 1 to 5 are estimated using OLS. In Columns 6 and 7, the dependent variable is the average annual daily traffic flow of two-wheel and buses respectively. In Columns 8, 9 & 10, the dependent variable is the counts of two-wheelers, buses and pedestrians involved in accidents. Regressions for Columns 6 to 10 are estimated using Pseudo Poisson Maximum Likelihood regressions due to the count nature of the dependent variable. Each coefficient is from a different regression and measures whether the implementation of the CCZ affects the composition of drivers and vehicles involved in collisions. CCZ is a binary variable denoting LSOAs in the charge zone after the charge is implemented. All regressions are estimated with LSOA and year fixed effects. Other control variables include job density, % of population from 18 to 25 years old, total population size, gross annual income,hours worked collected at Local Authority Level at year t and mean wind speed, temperature, precipitation and relative humidity collected at LSOA level at year t. Sample is constrained to 226 LSOAs within 2.5km from the CCZ boundary. Robust standard errors (in parenthesis) are clustered at LSOA. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

traffic flow and uncharged collision outcomes change after the charge is enforced.[30] These results are summarized in Table 5. Columns 1 to 5 are estimated using OLS, while columns 6 to 9 are estimated using poisson regressions capturing the count nature of the dependent variable. Similar to before, we limit our analysis to areas within 2.5km from the CCZ boundary.

In columns 1 to 3, we examine whether the share of female drivers, the average age of drivers, and the average number of vehicles involved in collisions change after the CCZ is enforced. In columns 4 and 5, we investigate whether vehicles involved in accidents after the charge is implemented have larger engine capacity and whether they are older. Across the board, we do not find any significant changes in driver or vehicle characteristics involved in traffic collision inside the charge zone, suggesting that these factors are unlikely to drive changes in collision outcomes in the charge zone. Next, in Columns 6 and 7, we test whether the enforcement of the charge leads to a substantial spike in the two-wheel and bus traffic flow in the zone. Traffic composition remains

[30]The specification adopted is similar to equation 2 but we replace traffic flow with various driver and vehicle characteristics involved in collisions, and uncharged traffic flow and collision outcomes.

fairly similar with no discernible changes reported for these vehicle types. These findings corroborate with the idea that traffic entering the charge zone are likely to bypass areas around the charge zone, explaining why we are not observing stark differences in traffic composition.

Finally, in columns 8, 9 and 10, we examine whether the charge affected the number of 2-wheel accidents, bus accidents and pedestrians injuries in the zone. If the charge causes a substantial spike in 2-wheelers and buses in the charge zone, we could observe more 2-wheel and bus accidents in the zone after the charge is enforced. Consistent with earlier results, none of these estimates are sizable enough or precisely estimated for us to conclude that the charge causes more 2-wheel and bus accidents. Another concern is whether the charge induces drivers to switch to public transit that could increase the number of pedestrians within the charge zone. While we do not observe the number of pedestrians along walkways, we have information on the number of pedestrians involved in accidents before and after the charge is implemented. If the charge increases pedestrian traffic, we should expect more pedestrians to be involved in traffic collisions in the charge zone. Our estimates suggest that this is unlikely the case as we document a 8.9% decrease in pedestrians injured from traffic collisions. If anything, this effect is larger than our estimates in column 1 in Table 4 on slight injuries, suggesting that accidents in the charge zone, on average, involve fewer pedestrians after the charge is enforced.

6.4 Non-monotonic elasticity estimates

We have so far assumed that the relationship between traffic flow and various collision outcomes is monotonic. However, adding an additional car to a congested road could have very different marginal effects on collision outcomes compared to a less busy road (Shefer & Rietveld, 1997; Dickerson *et al.*, 2000). For instance, putting another vehicle on a bottleneck road could further slow down speed for the rest of the drivers, which could reduce the severity of traffic collisions. Hence, we allow elasticity estimates to vary across areas with different levels of traffic flow. Specifically, we divide our sample into quartiles based on the average traffic flow over sample period, before interacting these quartiles with traffic flows. Q1 (Q4) is an indicator variable denoting areas with the lowest (highest) quartile of traffic flow. Similar to earlier specifications, we limit our analysis to areas within 2.5km from the CCZ boundary. Estimated elasticities for various collision outcomes are plotted in Figure 7.[31]

Our estimates suggest that adding an additional vehicle in areas with different traffic conditions could have very different impacts on road safety. In particular, a marginal driver along moderately busy road (at Q2) has the largest impact on accidents and slight injuries. A 1% increase in traffic

[31]Corresponding estimates and standard errors can be found in Table A4 Data Appendix.

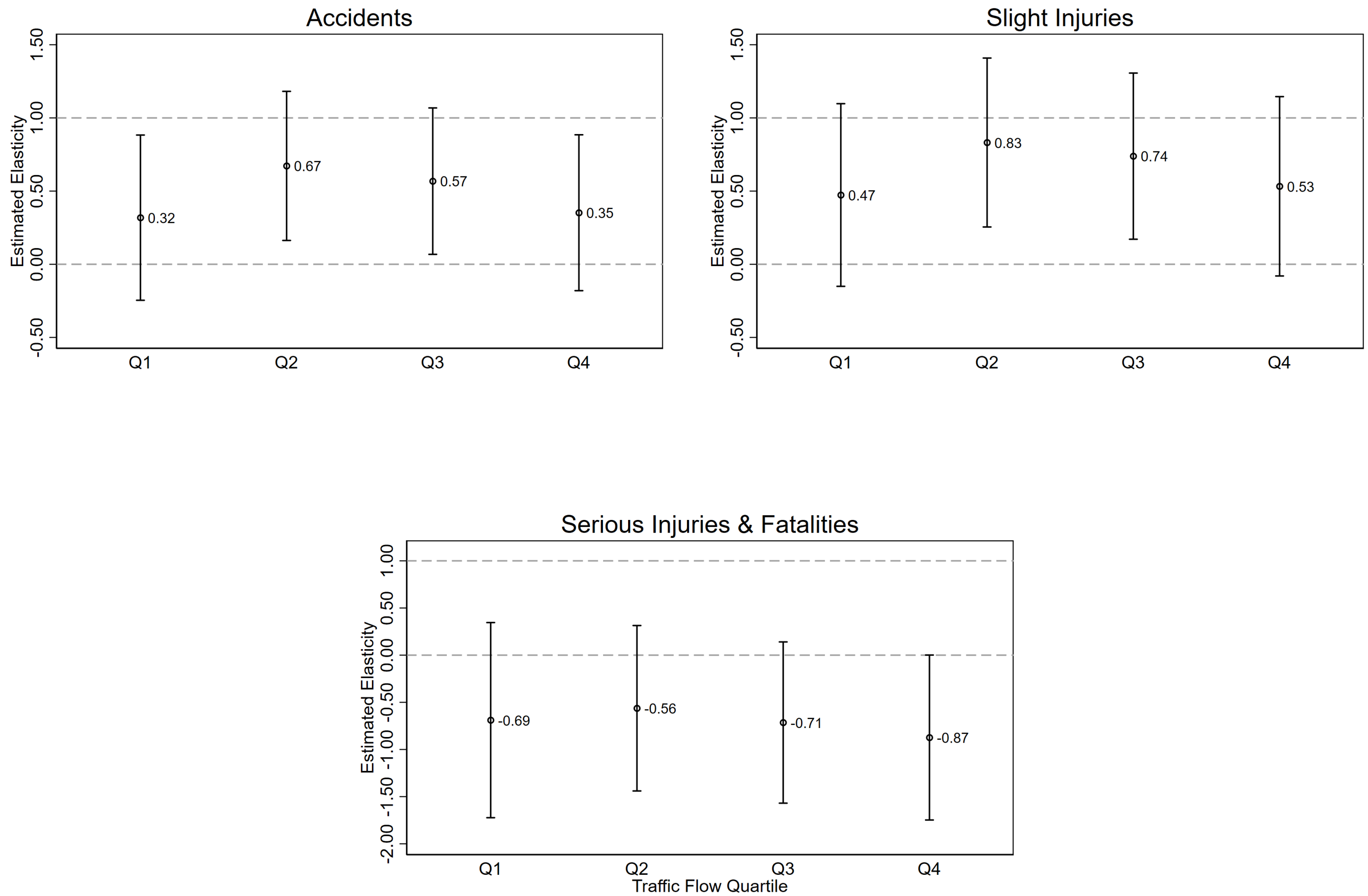

Figure 7: Plotted coefficients are the elasticity estimates between traffic flow and various collision outcomes across different traffic flow quartiles. Tails denote 90% confidence intervals constructed from robust standard errors clustered at LSOA. We classify LSOAs into different quartiles based on the average traffic flow over sample period and interact ln(Traffic), the natural logarithm of traffic flow with traffic density quartiles. Q1 and Q4 corresponds to the lowest and highest quartile of traffic flow respectively. Specification adopted is similar to columns 1 and 2 of Table 4 in which we limit the analysis to areas <=2.5km from the charge boundary.

corresponds to a 0.67% and 0.83% increase in accidents and slight injuries respectively. These marginal effects become smaller as we move up the traffic quartiles. Along the most congested roads (at Q4), an additional car is associated with a much smaller 0.35% and 0.53% increase in accidents and slight injuries that is no longer statistically different from zero at any conventional levels. These findings, again, suggest that the congestion externalities imposed by the additional driver on other road users along busy roads could reduce the probability of traffic collisions.

Next, we examine how the marginal driver affects the severity of accidents. Our estimates suggest that an additional driver in areas with light and moderate traffic (Q1, Q2 & Q3) do not significantly reduce the number of serious injuries and deaths. The elasticities for more serious injuries/fatalities are more precisely estimated and much larger in areas with heavier traffic. Specifically, we observe that an additional 1% increase in traffic flow attributes to a 0.87% reduction in serious injuries and deaths at Q4. We draw two main conclusions from these findings: first, similar to before, we do not find any evidence that the marginal driver imposes a negative accident externality to other road users. This is improbable for less congested areas and even more so for areas with heavy traffic. Second, we observe that the marginal driver only reduces the severity of collisions if existing traffic conditions are sufficiently heavy such that the additional driver imposes congestion externalities.

7 Discussion

We now compute the marginal external accident cost (MEC) associated with driving relying on earlier estimates between traffic flow and accidents. MEC is the monetized accident cost imposed to other road users by an additional kilometre driven, formally represented by the following equation:

$$MEC = \frac{\sum_{n=1}^{3} C^n(\beta^n - 1)A^n}{D}, \tag{5}$$

where n indexes the severity of accidents, which include property damage only accidents, slight injuries and serious injuries. C^n is the estimated monetary cost of type n accident reported by DfT and is summarized in column 5 of Table 6. Based on the estimates by the Department of Transport, monetary savings from avoiding an accident, slight injury and serious injury is £2,142, £15,450 and £200,422 respectively (2015 values).[32] The accident rate elasticity with respect to traffic flow for type n accident, calculated as β^n - 1, is summarized in column 3 of Table 6. We obtain these

[32] For more information, refer to https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254720/rrcgb-valuation-methodology.pdf

estimates from Columns 1-2 of Table 4, which is our preferred specification, that provides more conservative point estimates than our baseline estimates. Importantly, the accident rate elasticity is negative across all collision outcomes.

A^n represents the the average number of type n accidents per traffic flow. This is computed by dividing the annual collision outcomes by the annual total traffic flow (computed by multiplying annual daily traffic flow by 365 days). D denotes the average number of kilometers per traffic flow in London, which is about 1.6 km.[33] Plugging these parameters onto equation 5, we compute the MEC for type n accident in column 6.

The estimated total MEC per km driven is around -£0.16, suggesting that every additional km driven along congested roads reduces the cost of accident for other road users. In other words, an additional driver along roads in Central London actually improves road safety. Why is this so? Here, we are exploiting variation in traffic along highly congested roads around Central London to understand how the marginal driver can impact road safety. When traffic is sufficiently heavy, the marginal driver could worsen traffic congestion, slowing down the rest of the road users. Lower driving speed could reduce both the probability and severity of traffic collisions. Also, the presence of an additional car could heighten the awareness of other motorists, causing a less than proportional increase, or even lowering the risk of accidents. We further benchmark our estimated MEC of accidents against the estimated optimal congestion charge reported by Prud'Homme & Bocarejo (2005). The estimated accident benefit per km of £0.16 is about 20% of the optimal charge per km of £0.80 computed purely based on time delays.[34] Evidently, the accident benefit from the marginal driver pales in comparison with the marginal social cost of congestion delays, indicating that the London congestion charge is, still, likely to enhance social welfare.

Table 6: Marginal External Cost (MEC) of Accident from an additional km driven.

(1) **Outcomes** n	(2) **Estimated Elasticity** β^n	(3) **Collision Rate Elasticity** $\beta^n - 1$	(4) **Mean Daily Outcome/Flow** A^n	(5) **£/Outcome** C^n	(6) **MEC/km** MEC^n
Accident	0.6419	-0.3581	4.88×10^{-6}	£2,142	£-0.0023
Slight Injuries	0.8089	-0.1902	4.74×10^{-6}	£15,450	£-0.0087
Serious Injuries	-0.6459	-1.6459	7.12×10^{-7}	£200,422	£-0.1459
				MEC	**£-0.1569**

Estimated elasticity (in Column 2) between traffic flow and collision outcomes from Column 1-2 of Table 4. Collision rate elasticity with respect to traffic flow (in Column 3) computed by subtracting estimated elasticity (in Column 2) by 1. Mean Daily Outcome/flow (in Column 4) is calculated by dividing annual collision outcomes with the total traffic flow and 365 days. MEC/km (in Column 6) is computed by multiplying estimated collision rate elasticity with mean collision outcome per flow and monetary value per outcome, before dividing by distance travelled per traffic flow (approximately 1.6km) based on estimates by DfT.

[33] To estimate the kilometers driven per flow, we divide the average annual total kilometers driven by vehicles in London between 2000 and 2015 (approximately 30.5 billion km), provided by DfT, with the total annual traffic flow in our data (18.9 billion).

[34] This is adjusted from the £0.56 (or €0.81) reported by Prud'Homme & Bocarejo (2005) in 2003 values to 2015 values.

In terms of policy implications, we have also shown that the congestion charge has successfully curtailed traffic into the charge zone. This reduction in traffic attributed to fewer traffic accidents and minor injuries although collisions appear to be more severe. While a comprehensive cost-benefit analysis of the London Congestion Charge is out of the purview of this paper, it is interesting to measure whether the benefits from fewer accidents and minor injuries outweigh the cost of having more severe collisions. To provide an accurate assessment on the impact of the Congestion Charge across London, we rely on our baseline estimates from Table 2 that incorporate LSOAs across London.[35] We summarize our calculations in Table 7.

From Table 2, we observe that after the charge is implemented, the number of accidents and slight injuries are 6.91% and 9.65% lower and the number of serious injuries/fatalities are 11.62% higher. In absolute terms, we are looking at 187 fewer accidents, 257 fewer slight injuries and 46 more counts of serious injuries. We monetize these effects by multiplying the absolute effects with the monetary value.[36] Because the economic cost of serious injuries is very sizable, we observe that the CCZ imposes a net annual cost of around £4.8 million to the society. Assuming a discount rate of 3.0%, and that the CCZ remains operational for the next 30 years, the present value of accident costs impose by the CCZ is around £94 million. We further compute the upper and lower bound of the estimated effects of the CCZ on various collision outcomes based on the 95% confidence interval. Because the standard errors of our estimates are quite large, the net annual effects could ranged from a benefit of £3.9 million to a cost of £14.1 million. Nevertheless, the potential additional accident cost associated with the Congestion Charge indicates that reducing driving may not be the most effective policy in minimizing accident externalities.

8 Conclusion

This paper estimates the marginal accident externality of driving by exploiting the plausibly exogenous variation in traffic flow induced by the London Congestion Charge. Using the charge as an instrumental variable to negate endogeneity concerns, we estimate whether an additional driver affects the risk of accidents and injuries for other road users. Concern that the charge could

[35]As mentioned before, spillover effects from the Congestion Charge could possibly bias the reduced form impacts of the charge on traffic and accidents around the charge boundary. Traffic conditions for areas right outside the charge zone could have improved because there are fewer drivers driving into Central London after the charge is implemented. Hence, constraining our analysis to roads around Central London could underestimate the reduced form effects. This, however, is not an issue for our study as what we care about is whether the charge is impacting collision outcomes through changes in traffic flow.

[36]We do not include the cost of fatalities because estimates for fatalities are too imprecisely estimated to be statistically significant. Hence, if anything our estimates are overestimates (i.e. the estimates are even more negative than reported). See Table A3 for more information.

Table 7: Annual monetary value of accident externalities associated with the London Congestion Charge

(1) Outcomes	(2) Estimated Effects (Mean)	(3) Pre-treatment Mean	(4) Absolute Effects	(5) £/Outcome	(6) Monetized Effects
Accident	-6.91%	2,713	-187	£2,142	£401,557
Slight Injuries	-9.65%	2,666	-257	£15,450	£3,983,044
Serious Injuries	11.63%	394	46	£200,422	-£9,183,777
				Net Effects	**-£4,799,176**
Outcomes	**Estimated Effects (Upper Bound)**		**Absolute Effects**		**Monetized Effects**
Accident	-11.33%		-307		£658,205
Slight Injuries	-14.61%		-390		£6,025,088
Serious Injuries	3.47%		14		-£2,743,109
				Net Effects	**£3,940,184**
Outcomes	**Estimated Effects (Lower Bound)**		**Absolute Effects**		**Monetized Effects**
Accident	-2.27%		-62		£132,087
Slight Injuries	-4.42%		-117		£1,822,372
Serious Injuries	20.38%		80		-£16,128,388
				Net Effects	**-£14,173,928**

Monetized effects of the CCZ (Column 6) computed by multiplying estimated effects (Column 2) with absolute pre-treatment mean in collision outcomes in the CCZ (Column 3) and the monetized value per collision outcome (Column 5). Upper and Lower bound estimates computed base on the 95% confidence interval (by adding or subtracting estimated effects (mean) with 1.96 multiplied by standard errors from Table 2).

have affected driver and vehicle composition in the charge zone, violating the exclusion restriction assumption, we test the robustness of our estimates through different specifications (i.e., we limit our analysis to areas proximate to the charge boundary, we focus on a subsample which exclude bicycle and pedestrian accidents). We further conduct a battery of balancing tests to show that the charge is unlikely to have affected traffic collisions through means other than the change in traffic flow.

From these analyses, we report that the charge attributed to a 9.4% reduction in traffic flow, and caused a less than proportional 6.0% and 7.6% decrease in accidents and injuries, and a 6.5% increase in serious injuries/fatalities. In other words, our results show that an additional driver actually decreases the risk of collisions for other road users. Specifically, the accident, slight injuries, and serious injuries/fatalities rate elasticities with respect to traffic flow are -0.36, -0.19 and -1.65 respectively. These findings are in line with the idea that the marginal driver along congested roads slows down the overall driving speed for others, reducing the probability and severity of traffic collisions. Putting a monetary value to these collision estimates, we find the implied marginal external accident benefit from an additional km driven around Central London is £0.16.

These results imply that the optimal road toll in congested city centres could be less than the road toll based on travel time losses only, a conclusion which sharply contrasts with conventional wisdom for highways. Our estimates are informative at a margin where traffic conditions are quite heavy and therefore relevant to policy makers of major cities who are concerned about traffic congestion and road safety around busy roads facing the ubiquitous problem of traffic congestion. Because of the high cost associated with serious accidents, a simple benefit-cost analysis indicates that the additional cost from increased collision severity from the CCZ likely outweighs the benefit from the reduction in accidents and slight injuries. More importantly, policy makers must be aware of the accident externalities that might arise after ameliorating traffic congestion along heavily used roads.

Our study is limited to examining the impact of the accident externality of driving. A more in-depth analysis on other driving externalities (e.g time delays, pollution externalities etc) is required to quantify whether the London Congestion Charge is too low or too high from a welfare perspective (Leape, 2006). Nonetheless, our study reveals the possible repercussions on road safety that policy makers must be aware off when implementing Pigouvian taxes to reduce congestion externalities. More broadly, our findings suggest that the conventional perception that driving imposes negative accident externalities does not hold for congested roads around the city center, and reducing traffic flow along these roads is unlikely to be an effective strategy to improve road safety.

References

Anderson, Michael L, & Auffhammer, Maximilian. 2014. Pounds that kill: The external costs of vehicle weight. *Review of Economic Studies*, **81**(2), 535–571.

BBC. 2011. 'Fewer cars' as London congestion charge zone cut back. *Retrieved from* `https://www.bbc.com/news/uk-england-london-13640707` *on 4th of August 2020*.

Blundell, Richard, & Powell, James L. 2003. Endogeneity in nonparametric and semiparametric regression models. *Econometric Society Monographs*, **36**, 312–357.

Blundell, Richard W, & Powell, James L. 2004. Endogeneity in semiparametric binary response models. *The Review of Economic Studies*, **71**(3), 655–679.

Cameron, A Colin, & Trivedi, Pravin K. 2013. *Regression analysis of count data*. Vol. 53. Cambridge University Press.

Dickerson, Andrew, Peirson, John, & Vickerman, Roger. 2000. Road accidents and traffic flows: an econometric investigation. *Economica*, **67**(265), 101–121.

Edlin, Aaron S, & Karaca-Mandic, Pinar. 2006. The accident externality from driving. *Journal of Political Economy*, **114**(5), 931–955.

Fridstrøm, Lasse, & Ingebrigtsen, Siv. 1991. An aggregate accident model based on pooled, regional time-series data-. *Accident Analysis & Prevention*, **23**(5), 363–378.

Green, Colin P, Heywood, John S, & Navarro, Maria. 2016. Traffic accidents and the London congestion charge. *Journal of Public Economics*, **133**, 11–22.

Jansson, Jan Owen. 1994. Accident externality charges. *Journal of Transport Economics and Policy*, 31–43.

Leape, Jonathan. 2006. The London congestion charge. *The Journal of Economic Perspectives*, **20**(4), 157–176.

Li, Haojie, Graham, Daniel J, & Majumdar, Arnab. 2012. The effects of congestion charging on road traffic casualties: A causal analysis using difference-in-difference estimation. *Accident Analysis & Prevention*, **49**, 366–377.

Parry, Ian WH. 2004. Comparing alternative policies to reduce traffic accidents. *Journal of Urban Economics*, **56**(2), 346–368.

Parry, Ian WH, Walls, Margaret, & Harrington, Winston. 2007. Automobile externalities and policies. *Journal of Economic Literature*, **45**(2), 373–399.

Prud'Homme, Remy, & Bocarejo, Juan Pablo. 2005. The London congestion charge: a tentative economic appraisal. *Transport Policy*, **12**(3), 279–287.

Rivers, Douglas, & Vuong, Quang H. 1988. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, **39**(3), 347–366.

Romem, Issi, & Shurtz, Ity. 2016. The accident externality of driving: Evidence from observance of the Jewish Sabbath in Israel. *Journal of Urban Economics*, **96**, 36–54.

Shefer, Daniel, & Rietveld, Piet. 1997. Congestion and safety on highways: towards an analytical model. *Urban Studies*, **34**(4), 679–692.

Small, Kenneth A, & Verhoef, Erik T. 2007. *The Economics of Urban Transportation*. Routledge.

TfL. 2003a. Congestion Charging: Six months on.

TfL. 2004. Impacts Monitoring Programme: Second Annual Report.

TfL. 2005. Impacts Monitoring Programme: Third Annual Report.

Van Ommeren, Jos, Rietveld, Piet, Hop, Jack Zagha, & Sabir, Muhammad. 2013. Killing kilos in car accidents: Are external costs of car weight internalised? *Economics of Transportation*, **2**(2-3), 86–93.

Verhoef, Erik T, & Rouwendal, Jan. 2004. A behavioural model of traffic congestion: endogenizing speed choice, traffic safety and time losses. *Journal of Urban Economics*, **56**(3), 408–434.

Vickrey, William. 1968. Automobile accidents, tort law, externalities, and insurance: An economist's critique. *Law and Contemporary Problems*, **33**(3), 464–487.

Wooldridge, Jeffrey M. 2015. Control function methods in applied econometrics. *Journal of Human Resources*, **50**(2), 420–445.

Zhou, Min, & Sisiopiku, Virginia P. 1997. Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record*, **1581**(1), 47–52.

9 Data Appendix

I. Description of Data

STATS 19 Accident Data

We constructed our measures for various collision outcomes at LSOA level from STATs 19 Database, which document each and every accident reported to the Police from 1979 on wards. Overall, there are a total of 217,497 accidents from 2000 - 2014 in our regression analysis. Take note that we are constraining our analysis to collisions from 2000 on wards because traffic flow data is reported from 2000. Here, we briefly describe the nature of collisions in the database.

0.72% of the collisions (or 1573) have at least one fatality and 12.73% (or 27,691) of the accidents have at least one party serious injured. 88.80% (193,126) of the accidents have at least one suffering from slight injuries. 93.81% of these accidents involve at least one four-wheel vehicle while only 31.79% of the accidents are single-vehicle accidents. Overall, every reported accident has at one party suffering from injuries or death. It is evident that these accidents are more severe in nature as they are reported to the police force.

To construct our dataset for analysis, we aggregate the total number of collisions, slight, serious and fatalities that occur in every LSOA annually. Take note that the measures for slight, serious injuries/fatalities are mutually exclusive. Put differently, a fatality from a collision will not be recorded as serious or slight injury and vice versa. Given that traffic collisions, especially with serious injuries or fatalities, rarely occur, we are concerned with the number of observations with 0 collision outcomes. This is not an issue of accidents and slight injuries. From 2000 to 2014, only 4% of the observations have 0 accident and injury outcomes. This proportion, however, increases considerably to 44.56% and 93.62% for serious injuries and deaths. Hence, to account for the count nature and the sizable proportion of zeros of our dependent variables, we adopt a Poisson count regression model.

Traffic Data

We constructed our measures of traffic flow in an LSOA per year based on Average Annual Daily Traffic Flow (AADT) collected at a count point (CP) reported by the Department of Transport (DfT). There are a total of 2563 CPs across London, with most of these points concentrated in Central London. Traffic flow is reported for different types of vehicles including cars, motorcycles, light and heavy good vehicles, pedal cycles, buses and coaches etc. As mentioned earlier, traffic flow is manually counted on a normal day that is representative of the average traffic flow across the entire year. We match traffic flow data to LSOAs based on the location of CPs. However, not all the LSOAs have CPs. After removing those without traffic data, we end up with 1675 LSOAs (out of 4833 LSOAs in total) with at least one CP. Understanding that this matching approach could remove a considerable number of LSOAs, we adopt an alternate approach of matching CPs to LSOAs based on distance. Specifically, we compute the average traffic flow for each LSOA based on CPs within 500 to 1000m from the centroid of the LSOA that is weighted inversely based on distance. Doing so allow us to match more than 95% of the LSOAs although our measures of traffic flow are less accurate now.

List of Variables employed in the analysis

Table A1: List of Variables

Variable	Source	Description
Dependent Variable		
Traffic Flow	DfT	Average daily traffic flow (collected at count point) in LSOA j in year t.
Accident	STATS19	Number of Accidents at LSOA j in year t
Slight Injuries	STATS19	Number of Slight Injuries at LSOA j in year t
Serious Injuries	STATS19	Number of Serious Injuries at LSOA j in year t
Deaths	STATS19	Number of Deaths at LSOA j in year t
Local Authority Characteristics		
Gross Annual Salary	Annual Labour Force Survey	Average Gross annual salary at LA j in year t
Hours worked	Annual Labour Force Survey	Average number of hours worked in LA j in year t
Job Density	Annual Labour Force Survey	Number of Jobs per unit area of LA j (hectare) in year t
Population Size	Nomis Population Estimates	Total population size living in LA j in year t
% of 18 to 25	Nomis Population Estimates	Percentage of population aged 18 to 25 in LA j in year t
Weather Controls		
Temperature	LAQN	Annual average temperature in LSOA j in year t
Precipitation	LAQN	Annual average precipitation in LSOA j in year t
Humidity	LAQN	Annual average humidity in LSOA j in year t
Wind Speed	LAQN	Annual average wind speed in LSOA j in year t
Driver & Vehicle Characteristics		
% female	STATS19	% of female drivers involved in collisions at LSOA j in year t
Average age band	STATS19	Average age band drivers involved in collisions at LSOA j in year t
Average number of vehicles	STATS19	Average number of vehicles in collisions at LSOA j in year t
Average engine size	STATS19	Average engine size of vehicles involved in collisions at LSOA j in year t
Average age of vehicle	STATS19	Average age of vehicles involved in collisions at LSOA j in year t

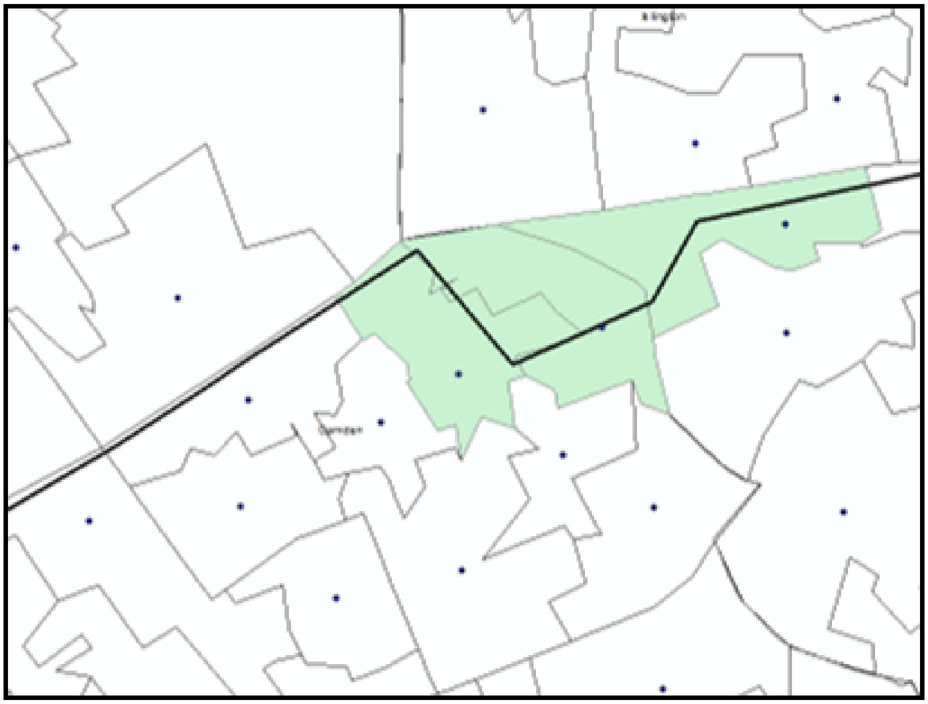

Figure A1: An illustration of some LSOAs (shaded) that overlapped with the CCZ boundary

II. Effects of the Western Extension Zone on Traffic Flow

Table A2: Effects of the CCZ & WEZ on Traffic Flow

	(1)	(2)	(3)	(4)	(5)	(6)
CCZ & WEZ	-0.1502***					
	(0.0237)					
CCZ		-0.1505***	-0.1505***		-0.1048***	-0.1243***
		(0.0237)	(0.0237)		(0.0275)	(0.0280)
WEZ		-0.0132	-0.0134	-0.0121	-0.0433	-0.0332
		(0.0135)	(0.0242)	(0.0244)	(0.0345)	(0.0297)
Remove WEZ			-0.0004	0.0132	-0.0467	-0.0010
			(0.0430)	(0.0482)	(0.0634)	(0.0589)
Obs	20863	20863	20863	19883	20863	3812
No.of LSOAs	1663	1663	1663	1595	1663	277

Dependent variable is the natural log of traffic flow. **CCZ** (**WEZ**) is a binary variable denoting LSOAs in the CCZ(WEZ) after the charge is implemented. **Remove WEZ** is a binary variable denoting LSOAs in the WEZ after the WEZ is removed. Other unreported control variables include job density, % of population from 18 to 25 years old, total population size, gross annual income,hours worked collected at Local Authority Level at year t and mean wind speed, temperature, precipitation and relative humidity collected at LSOA level at year t. In column 1, we estimate the joint effects of the CCZ and the WEZ on traffic flow. In column 2, we separately estimate the effects of the CCZ and the WEZ on traffic flow. In column 3, we further estimate the impact of the removal of the WEZ on traffic conditions. In column 4. we omit from the analysis the CCZ. In column 5, we control for local area districts linear trends to allow traffic conditions to vary linearly across areas. In column 6, we limit our analysis to areas within 2km from the CCZ and WEZ boundary. Robust standard errors (in parenthesis) are clustered at LSOA. * p<0.10, ** p<0.05, *** p<0.01.

II. Additional Elasticity Estimates on Serious Injuries & Fatalities & Intensive Margin Estimates

Table A3: Elasticity estimates for Serious Injuries and Fatalities & Intensive Margin estimates

	(1) Serious Inj	(2) Fatalities	(3) $\frac{SeriousInj}{Accident}$	(4) $\frac{Fatalities}{Accident}$
Ln(Traffic)	-0.6006***	0.0140	-0.9440***	-1.7204
	(0.2109)	(0.7614)	(0.2922)	(1.7931)
Obs	20303	10580	19568	10405
Mean Dep Variable	4.16	0.20	0.14	0.01
No.of LSOAs	1574	759	1574	759

Dependent variable is the counts of various collision outcomes denoted in the column headers. **Ln(Traffic)** is the natural logarithm of annual average daily traffic flow. The sample size, mean dependent variable and number of LSOAs are consistent across panel A and B. All regressions are estimated with LSOA and year fixed effects. Other control variables include job density, % of population from 18 to 25 years old, total population size, gross annual income,hours worked collected at Local Authority Level at year t and mean wind speed, temperature, precipitation and relative humidity collected at LSOA level at year t, distance to CCZ-by-year and distance to CCZ square-by-year dummies. Mean Dep Variable is the average collision outcomes or daily traffic flow in the CCZ before the charge is implemented. Robust standard errors (in parenthesis) are clustered at LSOA. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

III. Non-monotonic elasticity estimates across traffic quartiles

Table A4 presents the estimates that corresponds to Figure 7, illustrating how elasticity estimates for different collision outcomes vary across different traffic flow quartiles.

Table A4: Non-monotonic elasticity estimates from Control Functions for LSOAs within 2.5km from the CCZ boundary

	(1) Accidents	(2) Slight	(3) Serious Inj & Fatalities
Ln(Traffic)* Q1	0.2024	0.3333	-0.5943
	(0.2899)	(0.3148)	(0.5377)
Ln(Traffic)* Q2	0.5614**	0.7091**	-0.4639
	(0.2457)	(0.2761)	(0.4172)
Ln(Traffic)* Q3	0.3613	0.4960*	-0.6018
	(0.2547)	(0.2856)	(0.4103)
Ln(Traffic)* Q4	0.2786	0.4914	-0.8131*
	(0.2717)	(0.3079)	(0.4215)
Obs	3158	3158	3142
Mean Dep Variable	28.21	27.40	4.31
No.of LSOAs	226	226	223

Dependent variable is the counts of various collision outcomes as denoted in the column headers. **Ln(Traffic)** is the natural logarithm of annual average daily traffic flow. Specification adopted is similar to Table 3. Refer to earlier tables for more information. We allow the elasticity measure to vary according across LSOAs with different levels of traffic density by interacting ln(Traffic) with traffic density quartiles. We classify LSOAs into different quartiles based on the average traffic flow across the sample period. Q1 and Q4 corresponds to the lowest and highest quartile of traffic flow respectively. In all regressions, we constrain our analysis to LSOAs within 5km from the charge boundary. Robust standard errors (in parenthesis) are clustered at LSOA. * $p<0.10$, ** $p<0.05$, *** $p<0.01$. Estimated coefficients are plotted in Figure 7 in main manuscript.

IV. Instrumental Variable Elasticity Estimates

In this section, we report elasticity estimates from OLS regressions. This include reduced form estimates and first stage regressions that capture the effects of the CCZ on various collision outcomes and traffic flow. Subsequently, we combine these estimates to compute instrumental variable (IV) estimates. In other words, we are now instrumenting local traffic flow ($\mathbf{T_{jt}}$) using the CCZ, exploiting the sharp variation in traffic conditions induced by the CCZ to measure how an additional car can affect the probability and severity of traffic collisions. The system of equations to be estimated includes:

$$T_{jt} = \lambda_j + \gamma \mathbf{CCZ_{jt}} + X'_{jt}\rho + \psi_t + \nu_{jt}, \tag{6}$$

$$Y_{jt} = \pi_j + \zeta \mathbf{CCZ_{jt}} + X'_{jt}\delta + \upsilon_t + \epsilon_{jt}, \tag{7}$$

$$Y_{jt} = \alpha_j^{IV} + \beta^{\mathbf{IV}} \widehat{\mathbf{T_{jt}}} + X'_{jt}\phi^{IV} + \tau_t^{IV} + \varepsilon_{jt}, \tag{8}$$

The specification adopted is similar to earlier regressions. CCZ_{jt} is an indicator variable that **takes the value of 1 if LSOA j is located in the Congestion Charge Zone (CCZ) after charge is implemented in 2003**. $\alpha_j^{IV}; \lambda_j; \pi_j$ represents LSOA fixed effects that partial out time-invariant unobservables. τ_t^{IV} , υ_t and ψ_t represent year fixed effects that control for general trends in traffic flow and collisions across areas over time. X'_{jt} represents a vector of neighbourhood characteristics that could be correlated with traffic flow and affect collision outcomes. The main difference is that the dependent variable is the natural logarithm of various collision outcomes. Because a significant proportion of our observations is zero, especially for serious injuries/fatalities, we take log(0.5+various collision outcomes) to make sure that these observations are not dropped out from the analysis. [37] [38]

Equation 6 is the *first stage* regression that estimates the effectiveness of the CCZ in reducing local traffic flow surrounding each property. The dependent variable, T_{ijkt}, is the natural logarithm of the average daily road traffic flow from vehicles with four or more wheels. The efficacy of the charge is captured by γ that measures the percentage change in the traffic flow. Equation

[37] Approximately 4% of the observations have no accidents or slight injuries in a particular. This percentage goes up to 40% for serious injuries, and 95% for fatalities.

[38] We also do alternative transformations such as ln(0.5 + accident outcomes) and ln(accident outcomes) as a form of robustness. These unreported estimates remain fairly consistent across the different transformations and are available upon request.

7 measures the impact of the CCZ on collision outcomes and ζ captures this effect. If the implementation of the CCZ reduces traffic flow within the charge zone, and roads become safer the charge is enforced, we expect γ to be <0 and ζ to be >0. Equation 6 and 7 combine to form the *instrumental variable regression* in equation 8 that identifies the causal effect of traffic flow on various collision outcomes. The main results of this paper come from the estimation of β_{IV}, which measures the direct elasticity of traffic flow and accidents. $\widehat{T_{jt}}$ denotes the traffic conditions instrumented with CCZ_{jt}. Since β_{IV}, is exactly identified, it is simply the ratio of the two reduced form parameters ($\beta_{IV} = \frac{\zeta}{\gamma}$). For the instrumental variable estimator to provide a consistent estimator of the elasticity between traffic flow and accidents, CCZ_{jt} must not only affect local traffic conditions (relevance), but they must influence collision outcomes through changes in traffic flow only (exclusion restriction).

Table A5: Reduced form, IV and OLS estimates

	(1) Accidents	(2) Slight	(3) Serious Inj & Fatalities	(4) Traffic
Panel A: Reduced Form & First Stage Regressions				
CCZ	-0.0976***	-0.1178***	-0.0462	-0.1505***
	(0.0279)	(0.0306)	(0.0402)	(0.0237)
Obs	20863	20863	20863	20863
R2	0.80	0.76	0.49	0.96
Mean Dep Variable	28.21	27.40	4.25	19007.53
No.of LSOAs	1663	1663	1663	1663
Panel B: IV Regressions				
Ln(Traffic)	0.5012***	0.6051***	0.2382	
	(0.1381)	(0.1594)	(0.2083)	
1st Stage F-Statistics	54.87	54.87	54.87	

Dependent variable is the natural logarithm of various collision outcomes in LSOA j at year t for Columns 1-3 and the natural logarithm of annual average daily traffic flow for vehicles with 4 wheels or traffic flow in Column 4. All regressions are estimated with LSOA and year fixed effects. Other control variables include job density, % of population from 18 to 25 years old, total population size, gross annual income,hours worked collected at Local Authority Level at year t, and mean wind speed, temperature, precipitation and relative humidity collected at LSOA level at year t. Mean Dep Variable is the average collision outcomes or daily traffic flow in the CCZ before the charge is implemented. 1st Stage F-stats reported is the Kleibergen-Paap rk Wald F statistic from first stage regressions. Robust standard errors (in parenthesis) are clustered at LSOA. * p<0.10, ** p<0.05, *** p<0.01.

Table A5 summarizes results from the estimation of equation 6 and 7 in Panel A and equation 8 in Panel B. Overall, we document results fairly consistent with our baseline estimates in Table 2 and 3 at least for Accidents and Slight Injuries. In particular, our reduced form estimates suggest that traffic flow is around 14.0% lower and this attributed to a 9.3% and 11.1% reduction in accidents and slight injuries respectively. The strength of the CCZ as an instrumental variable is further reflected by the first stage F-statistic that is larger than 10. Putting these estimates together, a 1% increase in traffic correspond to a 0.50% and 0.61% increase in accidents and slight injuries, indicating that an additional driver leads to a less than proportional increase in accident risk for other road users. These estimates are comparable to that reported in Table 3.

A notable difference of our IV estimates is that we no longer observe any significant spikes in the number of serious injuries/fatalities. A plausible explanation for the disparity is due to the transformation of our dependent variables in these IV regressions. As serious injuries/fatalities rarely occur, there are many zeros in our dependent variable and taking the natural logarithm of 1 plus these variables could materially affect our measure of serious injuries/fatalities. Hence, we prefer estimates from our Poisson regressions.