

Environment for Development

Discussion Paper Series

July 2023 ■ EFD DP 23-10

Does Traffic Congestion pose Health Hazards?

Evidence from a Highly Congested and Polluted City

Kanishka Kacker, Ridhima Gupta, and Saif Ali



Central America
Research Program in Economics and Environment for Development in Central America Tropical Agricultural Research and Higher Education Center (CATIE)



Colombia
The Research Group on Environmental, Natural Resource and Applied Economics Studies (REES-CEDE), Universidad de los Andes, Colombia



India
Centre for Research on the Economics of Climate, Food, Energy, and Environment, (CECFEE), at Indian Statistical Institute, New Delhi, India



South Africa
Environmental Economics Policy Research Unit (EPRU)
University of Cape Town



Uganda
EFD-Mak, School of Economics and Department of Agribusiness and Natural Resource Economics, Makerere University, Kampala



MAKERERE UNIVERSITY

Chile
Research Nucleus on Environmental and Natural Resource Economics (NENRE)
Universidad de Concepción



Ethiopia
Environment and Climate Research Center (ECRC), Policy Studies Institute, Addis Ababa, Ethiopia



Kenya
School of Economics
University of Nairobi



Sweden
Environmental Economics Unit
University of Gothenburg



USA (Washington, DC)
Resources for the Future (RFF)



China
Environmental Economics Program in China (EEPC)
Peking University



Ghana
The Environment and Natural Resource Research Unit, Institute of Statistical, Social and Economic Research, University of Ghana, Accra



Nigeria
Resource and Environmental Policy Research Centre, University of Nigeria, Nsukka



Tanzania
Environment for Development Tanzania
University of Dar es Salaam



Vietnam
University of Economics
Ho Chi Minh City, Vietnam



Does Traffic Congestion pose Health Hazards? Evidence from a Highly Congested and Polluted City

Kanishka Kacker*, Ridhima Gupta, and Saif Ali

Abstract

Will reducing traffic congestion bring health benefits? We use high frequency data from Uber for Delhi – a city that experiences high levels of air pollution and traffic congestion - to answer this question. Exploiting information by time of day for every day of 2018 at the neighborhood level that covers over 16000 possible trips during each of these time periods, we employ an econometric framework that models wind direction together with day, month, time-of-day and trip fixed effects to remove important sources of unobserved heterogeneity. Congestion has a non-linear, dynamic impact on pollution raising it sharply by over a standard deviation. The pattern of response shown by the results is consistent with known information regarding vehicular emissions and ambient air pollution, suggesting bias in the estimates to be low. Simulations using parameters from epidemiological studies suggest congestion may be responsible for up to 40% of all premature deaths from pulmonary and heart disease in Delhi.

Keywords: air pollution, traffic congestion, vehicular regulation

JEL Codes: Q53, R41, O18, L91

* Corresponding Author: email kkacker@isid.ac.in. The authors gratefully acknowledge funding from the Environment for Development (EfD) network. Comments from E. Somanathan, participants at the 16th Annual EfD meeting in Kampala – Uganda, audience members at seminars held in the Indian Statistical Institute seminar and Ashoka University have helped to improve the draft of the paper.

Does Traffic Congestion pose Health Hazards? Evidence from a Highly Congested and Polluted City

Kanishka Kacker*

Indian Statistical Institute

Ridhima Gupta

South Asian University

Saif Ali

Indraprastha Institute of
Technology

30 June 2023

Abstract

Will reducing traffic congestion bring health benefits? We use high frequency data from Uber for Delhi – a city that experiences high levels of air pollution and traffic congestion - to answer this question. Exploiting information by time of day for every day of 2018 at the neighborhood level that covers over 16000 possible trips during each of these time periods, we employ an econometric framework that models wind direction together with day, month, time-of-day and trip fixed effects to remove important sources of unobserved heterogeneity. Congestion has a non-linear, dynamic impact on pollution raising it sharply by over a standard deviation. The pattern of response shown by the results is consistent with known information regarding vehicular emissions and ambient air pollution, suggesting bias in the estimates to be low. Simulations using parameters from epidemiological studies suggest congestion may be responsible for up to 40% of all premature deaths from pulmonary and heart disease in Delhi.

Keywords: air pollution, traffic congestion, vehicular regulation

JEL Codes: Q53, R41, O18, L91

* Corresponding Author: email kkacker@isid.ac.in. The authors gratefully acknowledge funding from the Environment for Development (EfD) network. Comments from E. Somanathan, participants at the 16th Annual EfD meeting in Kampala – Uganda, audience members at seminars held in the Indian Statistical Institute seminar and Ashoka University have helped to improve the draft of the paper.

Section I: Introduction

Traffic congestion imposes significant costs on society: Longer commutes waste time as people are forced to spend their hours on the road. Engines idle for longer at major road intersections with adverse effects on air-pollution and human health (Currie and Walker 2011; Knittel et al 2016; Barth and Boriboonsomsin 2008). According to the World Health Organization, 9 of the top 10 most polluted cities in 2016 are in India.¹ Three of the top 10 most congested cities in the world are also in India – Hyderabad, Delhi and Mumbai (Numbeo Traffic Index 2019). Private and public forms of transport contribute about 10% to 12% of the emissions of nitrogen oxides and non-methane volatile organic compounds in 2015 in India (Venkatraman et al 2018). At the same time, private ownership of cars in India is projected to increase: more than 85% of respondents in four major metropolitan Indian cities – Bangalore, Delhi, Hyderabad and Mumbai - plan to buy a car in the next five years (Boston Consulting Group 2018).

Yet, there is little rigorous analysis of the effects of traffic congestion in developing countries such as India: Chen et al (2020) which studies this relationship in Beijing is an important exception. Indian cities are not just congested and polluted they are also densely populated implying exposure to pollution is also amongst the highest in the world. Health costs from pollution exposure are thus large but it is unclear how much can be attributed to vehicular congestion. Indeed, recent epidemiological research indicates that the mortality-exposure relationship appears nonlinear and thus flatter at high levels of pollution (Pope et al 2011, Burnett et al 2018): mild reductions in pollution at high-pollution levels are thus unlikely to bring about substantial health gains.

We attempt to fill this gap by analyzing the effect of vehicular congestion on air pollution and health outcomes in the city of Delhi in India. Our measure of traffic congestion comes from an extremely high frequency dataset collected by Uber, the rideshare company. The data spans the year of 2018, with information on travel times collected by time of day for every day, at the level of local neighborhoods called wards. We define congestion as the amount of time required to go from point A to point B, normalized by distance. This is the inverse of speed

¹ <https://www.who.int/airpollution/data/cities-2016/en/>

(minutes per kilometer) and is a well-known measure of traffic flow (Mangrum and Molnar 2018). Our main outcome variable is PM 2.5 concentration, but we also look at PM 10, CO and NO_x.

Using these different pollutants as outcomes helps us in identification. CO and particularly NO_x are well-known precursors to particulate matter formation and are the main pollutants emitted by vehicles (Hodan and Barnard 2004). Therefore, if we find congestion has contemporaneous impacts on NO_x and CO, but lagged impacts on PM 2.5, it will mean our estimates arise from vehicular congestion and not any other source. For example, in Delhi, the other major source of air pollution is construction. There is no logical reason to believe construction will affect NO_x levels contemporaneously and PM 2.5 levels with a lag.

Our estimates imply congestion increases air pollution, in a complex manner. The impact is non-linear, congestion must build up before it has a large impact. Crucially, it is also dynamic: heavy congestion raises concentrations of CO and NO_x immediately, but PM 2.5 concentrations only after a few hours have passed. Such a response is sensible if NO_x emissions (for instance) re-combine in the atmosphere to form particulate matter, as this set of reactions take time to occur. At the same time, wind will blow these emissions so the areas where the emissions take place and where they are ultimately deposited as particulates will be different, which will be indicated by the direction of wind (Hodan and Barnard 2004). Using a rich set of fixed effects and incorporating the effects of wind, we are thus confident our OLS specifications reject the null hypothesis of zero impact of traffic congestion on pollution.

Congestion – defined in terms of travel times beyond the 75th percentile – raises PM 2.5 concentrations by 20 micrograms/m³. Simulations using epidemiological research suggest this impact contributes to annual premature deaths from chronic obstructive pulmonary disease and ischemic heart disease in Delhi by 19%.

To the best of our knowledge, this is the first study that looks at the impact of traffic congestion on air pollution and health in India. These estimates can be used by policymakers to assess policies that aim to relieve vehicular congestion. Our estimates suggest a potentially important role for policies such as congestion pricing that could reduce the occurrence of extreme congestion events.

In the next section (Section II) we describe briefly the institutional and policy background to tackling vehicular emissions in India. We then specify our econometric model and identifying strategy in Section III. Section IV describes our key variables, and how the data were put together. Main results are then shown in Section V and the health implications are discussed in Section VI. Section VII discusses the possible threats to identification considering our study design; in particular, what may plausibly confound our estimated relationship between congestion and pollution. Section VIII offers some preliminary conclusions.

Section II: Emissions from Transport in India

Policies aimed at cutting emissions from the transportation sector in India are increasingly becoming more aggressive. For example, the Delhi government introduced a quantity-based regulation of vehicular emissions through its “Odd-Even” program in a pilot from January 1 to 15, 2016. Under this program, cars whose license plates ended in an odd number could be used on odd numbered days, and similarly for license plates ending in an even number. This program has been periodically re-introduced and was last reintroduced from November 4th, 2019 to November 15th, 2019.² The Ministry of Petroleum and Natural Gas has brought forward the implementation of stricter fuel standards (Bharat Stage VI standards) by two years for the National Capital Territory: these standards were in effect from April 1st, 2018, whereas earlier they were to go into effect by April 1st, 2020.³ These norms are in line with European Union regulations on most classes of vehicles and represent a significant tightening of existing norms in India.

Such policies may be well intentioned, but it is unclear to what extent they are justified. While there is little doubt about the toxic levels of pollutants in most Indian cities - India is home to 14 of the top 15 cities with the worst PM 2.5 pollution out of 100 countries covered by WHO from 2011-2016 (WHO 2018) – how far transportation accounts for this is unclear. Current estimates of emissions and their sources in India largely focus on national level estimates (Guttikunda et al. (2018)). Such estimates typically use an all India geographic scale

² Greenstone et al (2017) describe various other measures recently adopted by the Delhi government to restrict vehicular emissions: increased charges for commercial vehicles, discontinued registration of diesel cars, and further restrictions on the entry of trucks.

³ <http://pib.nic.in/newsite/PrintRelease.aspx?relid=173517>

which is less likely to capture detailed variation in traffic-related exposure within urban areas and near roads. Consequently, they are likely to be underestimates. According to the latest GBD estimates, the transportation sector accounted for about 2 percent (23,100) of deaths attributable to PM 2.5 pollution in India in 2015. Most city level studies on source apportionment of air pollution focus on Delhi, the national capital. Guttikunda et al. (2018) is an exception, since the authors provide estimates of sector-wise contribution to air pollution by 20 Indian cities in 2015. The marginal impact of reducing traffic congestion on air pollution, however, is not considered by this study.

Knowledge of the magnitude of the marginal impact of congestion on pollution is crucial for understanding the welfare loss associated with emissions from congestion. To see this, first consider the effect of congestion. Increasing congestion by itself involves a deadweight loss: this is simply a commons problem (Akbar and Duranton 2017). When making the decision to drive, drivers should account for their marginal impact on congestion which is the social cost of the drive, but instead make their decisions on the average impact which is the private cost. This leads to too much driving, with the extent of the deadweight loss given by the distance between the social and private cost curves.

But there is another source of divergence between social and private cost, the effect of emissions produced as a result of driving. Congestion due to driving may increase emissions, which raise ambient pollution concentrations, which in turn have adverse effects on health. Therefore, the deadweight loss associated with a given privately optimal level of driving will be larger than that due to congestion alone. An estimate of congestion's impact on pollution will tell us how large this additional loss would be, conditional on the effect of pollution on health. A small estimate would suggest that additional policies to curb vehicular emissions are unlikely to lead to sizeable welfare increases while a large estimate would suggest that such policies are needed, as long as increased pollution negatively impacts health outcomes. The latter is true because Greenstone et al. 2017 have shown that reducing pollution is likely to improve health outcomes by a large amount in India. Recent work suggests the deadweight loss from congestion itself – the effect of pollution is ignored in these studies - in highly congested cities is likely to be small (Kreindler 2018, Akbar and Duranton 2017). Our estimate of the effect of

congestion on pollution will therefore inform policy on whether there are any significant welfare losses from congestion.

To the best of our knowledge, the only study that estimates the impact of traffic congestion on air pollution is by Knittel et al. (2016). Here the authors' use hourly data on average speed and total flow of cars in California from the Freeway Management and Performance System (PeMS) maintained by the University of California, Berkeley Department of Electrical Engineering and Computer Sciences. They estimate the effect of congestion on ambient air pollution and infant mortality rates within postcodes in California from 2002-2007. The authors' employ an instrumental variables fixed effect model to estimate the impact of congestion on carbon monoxide (CO) and PM 10. Since pollution in a neighborhood can be correlated with the unobservables that influence the level of infant mortality in a neighborhood, the authors' instrument for pollution using a measure of car miles travelled interacted with weather variables. They find that a standard deviation increase in traffic results in a 0.2% of a standard deviation increase in infant deaths. However, they do not account for wind, nor do they consider the impact on PM 2.5, the most harmful by-product of vehicle emissions. Finally, infant deaths are only a short-run measure while we look at long term impacts which are arguably more relevant for policy.

Currie and Walker (2011) estimate the health effects of traffic congestion by examining the effect of a policy change that caused a sharp decline in congestion across places in the United States. Examining US cities, Duranton and Turner (2011) find that vehicle-kilometers travelled increases with roadway lane kilometers, with public transportation having little effect. In a similar vein, Mangrum and Molnar (2018) find increased supply of taxis worsened congestion in New York City. Contrary to these results, Anderson (2014) estimates an increase in delays, measured in minutes per mile, of commutes using private transit when public transport services stop in Los Angeles. Exceptions to the studies focused on developed countries are Akbar and Duranton (2017) and Kreindler (2018) who study traffic in Bogota and Bangalore respectively and find the welfare loss from congestion to be small. Both studies, however, ignore air pollution in their estimates of congestion costs.

Section III: Identifying the Effect of Congestion on Pollution and Health

Our main specification is written as:

$$p_{shdm} = \beta_0 + \beta_1 * d_{shdm} + \beta_2 * X_{shdm} + \lambda_s + \lambda_h + \lambda_d + \lambda_m + \varepsilon_{shdm} \quad (1)$$

Equation (1) above is our preferred specification. Here, p_{shdm} records the pollution level recorded at monitoring station s , during hour h of day d in month m .⁴ There are 36 monitoring stations in total, which are operated by the Central Pollution Control Board. The congestion measure d_{shdm} is calculated as follows. We construct a 90-degree cone using information on wind direction for all wards in Delhi. Using this cone, we can draw up a list of which wards are upwind for any chosen ward at every point in time. We then identify the specific ward containing monitoring station s , and all the wards upwind of this ward for all points in time.

Our measure of traffic congestion equals the amount of time trips take. Using the list of wards upwind of the ward containing the monitoring station, we calculate the average congestion levels for all trips originating from, and ending in, every upwind ward including the ward in which monitoring station s is located – this average is the congestion measure d_{shdm} . Trips from or to those wards lying outside the list of upwind wards are excluded from d_{shdm} . We carry out this exercise for all monitoring stations.

We split the congestion measure thus calculated in a categorical variable that takes on the following values: 1 if travel times are below the 25th percentile, 2 if travel times are between the 25th and the 50th percentile; 3 if travel times are between the 50th and the 75th percentile terms of percentiles; 4 if travel times are between the 75th and the 95th percentile; and finally 5 if travel times are above the 95th percentile. We use this definition to account for potential non-linearity in the relationship between pollution and congestion while imposing the least possible structure on functional form. Our reasons to use a non-linear specification come from two sources. One, we reject a linear relationship⁵, and two, previous work in Beijing (Chen et al 2020) also finds a non-linear relationship. The interpretation of such non-linearity is that congestion must accumulate beyond a certain point to affect pollution levels.

⁴ Uber's data collapses information into 5 hour blocks for every day: Early Morning (12 am – 7 am), AM Peak (7 am – 10 am), Midday (10 am – 4 pm), PM Peak (4 pm – 7 pm) and Evening (7 pm – 12 am).

⁵ These results are not reported but are available on request.

The other control variables - X_{shdm} - record weather information (temperature, precipitation, humidity and solar radiation) for the wards in which the monitoring stations are located.⁶ Specifying fixed effects at the monitoring station λ_s (or equivalently, at the ward in which the monitoring station is located) allows us to eliminate any time invariant characteristic of the monitoring station or the ward it is located in - these could be size and location of wards or location of monitoring stations. The other set of fixed effects eliminate important sources of unobserved heterogeneity with respect to time: hour-of-day, day-of-week, and month. These sources will be discussed in the next section.

Finally, we will compute the health costs associated with increased congestion by multiplying the health impacts of increased pollution from existing epidemiological research (Pope et al 2011, Burnett et al 2018) with our estimated impact of increased congestion (β_1). The health outcomes specifically considered are ischemic heart disease and cardiopulmonary disease. These studies have been selected to obtain long-run health costs from air pollution that are useful for two reasons. One, long-run health impacts are of deeper policy relevance than short-run impacts which are typically the focus of many air pollution studies in developing countries such as India (Gupta and Spears 2017, Barrows, Garg and Jha 2019). Two, these studies draw out the implications for health for various concentrations of pollutant: the relationship between health and pollution appears non-linear implying smaller marginal health changes for larger pollution levels. Since pollution levels for Delhi are very high, such non-linearity can potentially play a role in telling us the size of the health impact.

Section IV: Traffic Congestion, Pollution and Commute Shocks

In this section we briefly describe the sample we use to estimate our equations. Our dataset is assembled from three different sources, described as follows.

Air Pollution: Data on ambient air pollution concentration comes from 36 monitoring stations that are run by the Central Pollution Control Board. These stations record pollution at an hourly frequency, for every day. Latitude and longitude information for the monitoring

⁶ For these variables, we take our data from the Copernicus ERA5-Land dataset which contains gridded data on these variables for the city of Delhi. Each grid is of 8 km by 8 km dimension, for a total of 23 grids that overlay the city. As wards are irregularly shaped, they don't fall neatly into these grids, so we assign to each ward the value of the meteorological variable from a particular grid based on the ward's share of area that falls within that particular grid.

station enables us to locate the station within a ward - the lowest level of civic administration in Delhi – for which we have the latitude-longitude of the centroid of the ward. Wards are irregularly shaped polygons, so the location of a ward is identified from its centroid position.

Traffic Congestion: Congestion measures – which we refer to as travel times from here onward – are developed using Uber’s public release dataset, which records the length of trips between wards. These data are available at five time blocks for every day: Early Morning (12 am to 7 am), AM Peak (7 am – 10 am), Midday (10 am – 4 pm), PM Peak (4 pm – 7 pm) and Evening (7 pm to 12 am).

Weather: Weather data are taken from the ERA5-Land dataset from Copernicus Climate Change Service, which record data again at an hourly frequency for every day; in terms of space, each grid is defined to be 9 km in resolution. The centroid of the grid is used to identify the weather status for each ward for every hour of every day.

All three sources are merged using latitude-longitude information and time: hour, day, and month. Since the lowest common temporal frequency is given by the Uber defined hour blocks, we average all our other data – pollution, and weather – up to these hour blocks to complete the merge.

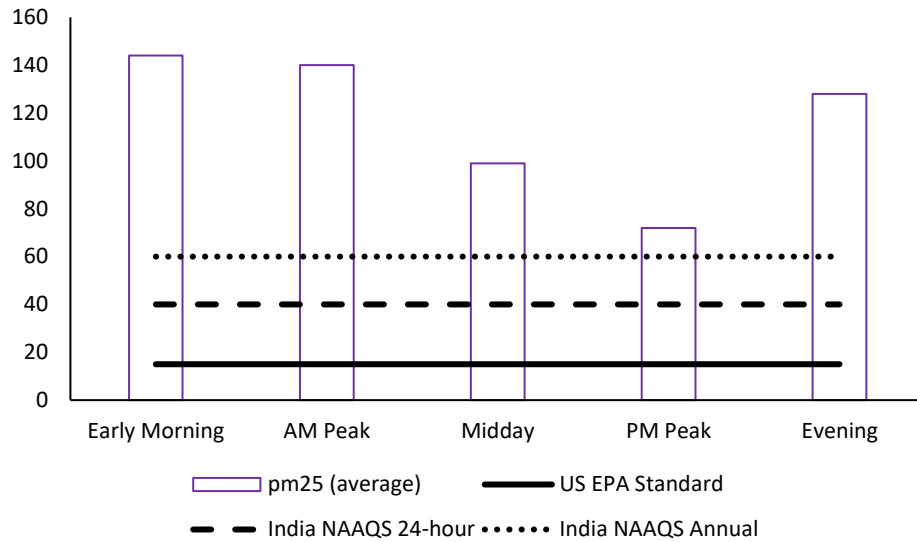


Figure 1: Average PM 2.5 Concentrations in Delhi, by hour of day, for 2018

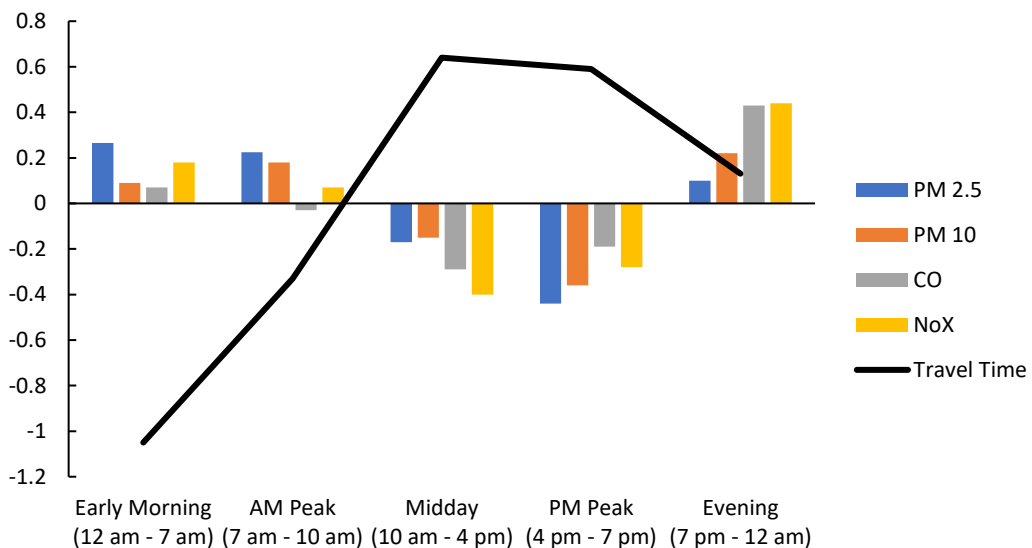


Figure 2: Average Standardized Pollution Concentrations and Travel Times in Delhi, by hour of day, for 2018

Pollution concentrations and travel times vary significantly within a day. This variation is shown in Figures 1 and 2. Figure 1 plots the average PM 2.5 concentrations across all monitoring stations by hour of day in purple rectangles. The dotted and dashed black lines indicate the Indian national standards for ambient PM 2.5 concentration, while the thick black line indicates the tighter US EPA standard for the same. We can see that there is no time during the day which meets *any* of these standards, a potent indication of how badly polluted Delhi's

air is. We also see that pollution is highly correlated with the time of day, as colder times are associated with higher pollution – the well-known temperature inversion effect being the reason. Similarly, colder months tend to see higher concentrations than warmer months.

The relationship between travel times and pollutant concentrations is shown in Figure 2. Here, we show standardized values to compare all these measures, as they are measured on different scales. All pollutants follow the same pattern over the course of a day, falling during the middle of the day and rising at night/early morning: precisely how we expect them to behave given the influence of temperature. Travel times, however, show a precisely opposite trend which is also sensible since most people need to travel during the day.

Since Figure 2 shows standardized values, the difference between the high and low values provides an estimate of how much variance exists. While PM 2.5 varies by roughly 0.6 standard deviations, NOx varies the most amongst the pollutants (0.9 standard deviations). Travel times vary the most, with the intra-day variation being 1.6 standard deviations. It is important to note that these figures are not adjusted for well-known confounders: seasonality over the year, within-week variation of travel demand, temperature, rainfall, humidity, and wind direction. Some caution should be exercised to avoid taking them too literally.

To get a sense of the spatial dimension of our data, in Figure 3 we show a heat map of travel times for an arbitrarily chosen day-hour combination and location for the city of Delhi: the green dot surrounded by black is our marker for an origin ward, arbitrarily chosen. These data are taken directly from Uber and form the basis for our main independent variable: d_{thd} in equation (3). Deeper blue colours indicate increased travel times and ward boundaries are shown by the thin white lines. Note the map will change by changing the location or day-hour combination, and thus provides an incredibly detailed look into travel conditions in the city.

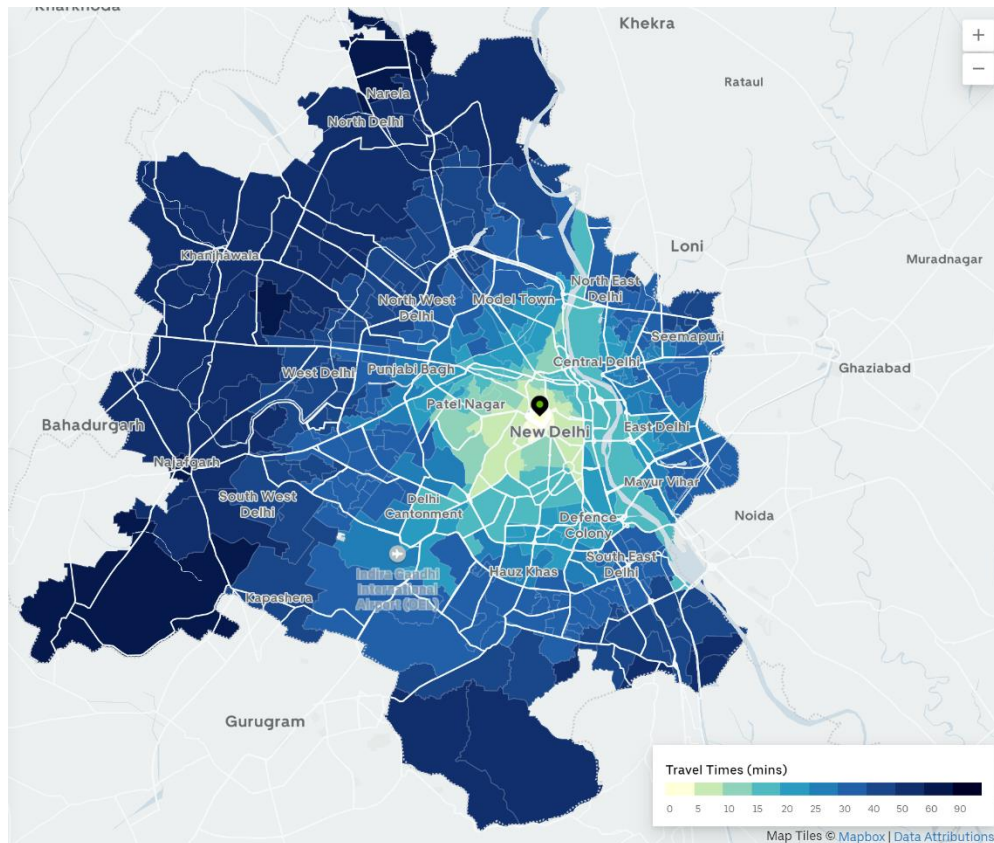


Figure 3: Heat Map for Travel Times for an arbitrary location and day-hour combination in Delhi (Source: Uber)

As mentioned in the section above, we only focus on wards upwind of monitoring stations. Monitoring station locations are shown in Figure 4 by red dots. We can see that there are many monitoring stations that sit at intersections of different wards. The ward whose centroid is closest to the monitoring station is recorded as containing the monitoring station. Figures 5 and 6 show spatial variation in average and standard deviations respectively of PM 2.5 recorded at each ward that has a monitoring station, adjusted for weather, hour-of-day, day-of-week, and month. In general, wards located in the east and south have both the highest concentration of PM 2.5 as well as the largest variation. Note the heat maps in figures 5 and 6 show pollution concentrations that cannot be explained by weather, hour-of-day, day-of-week, and month; the source of these differences both within and across stations therefore must arise from some other factor.

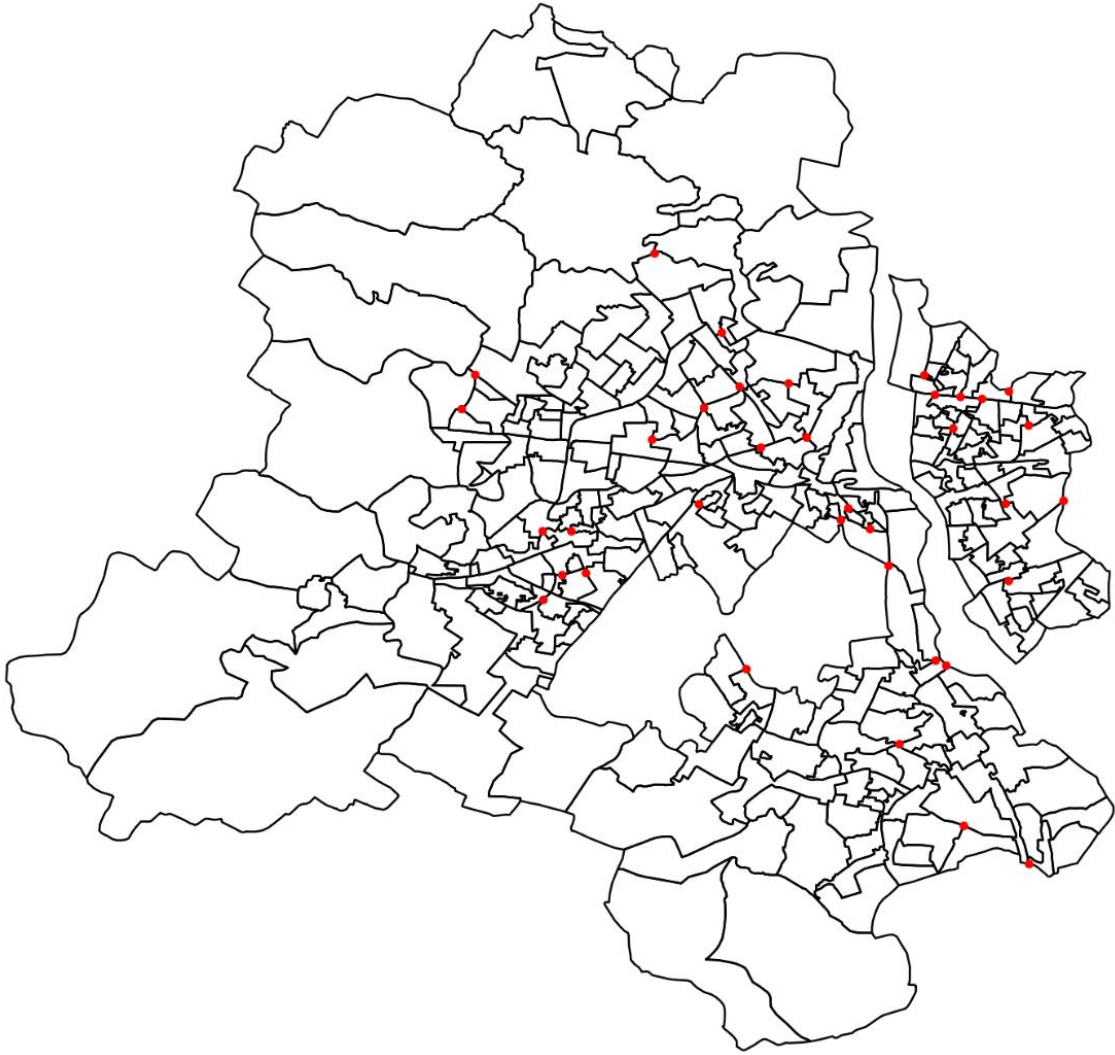


Figure 4: Location of Monitoring Stations, shown as red dots.

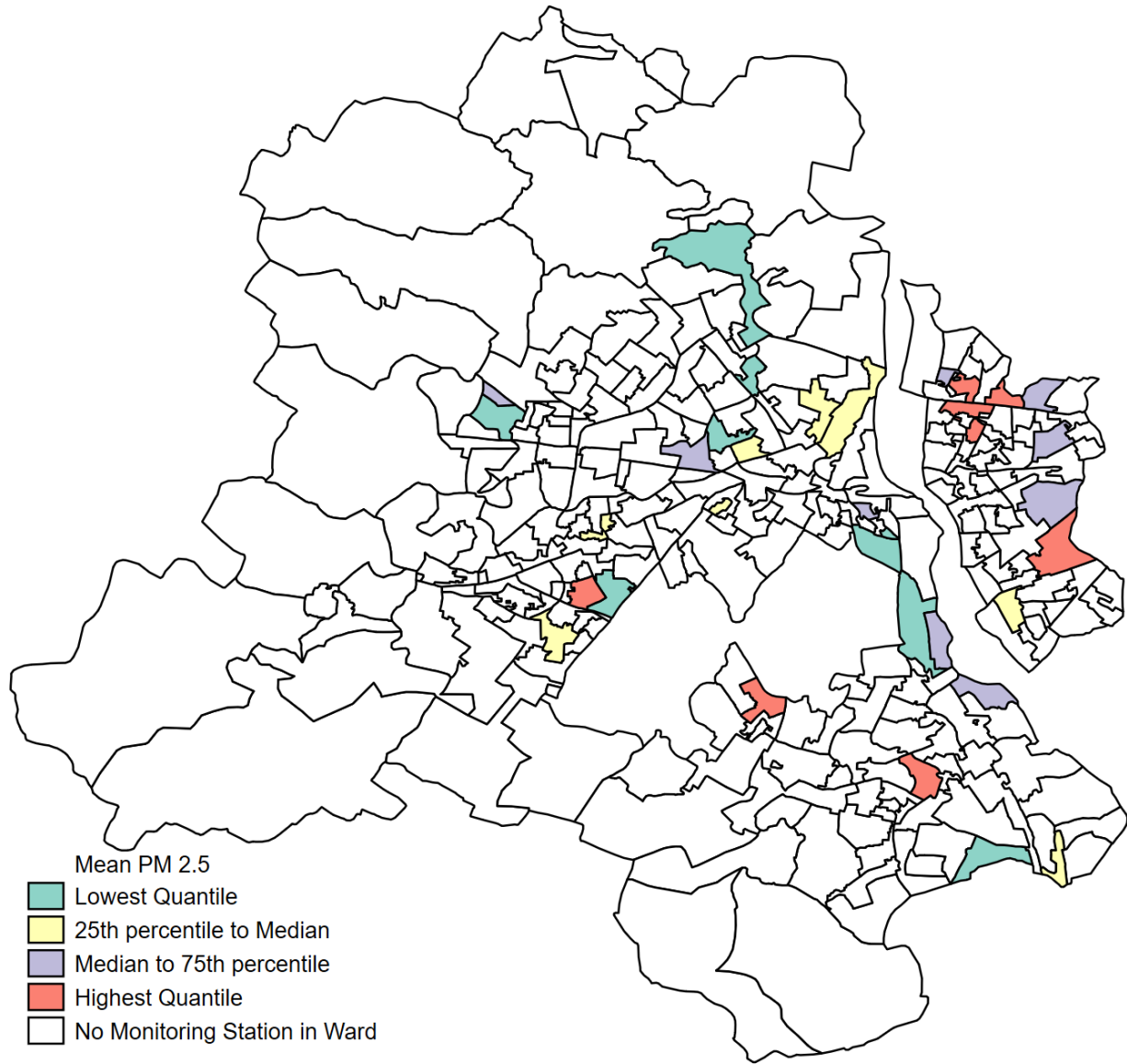


Figure 5: Heat Map of Mean PM 2.5 concentrations, adjusted for weather, hour-of-day, day-of-week and month.

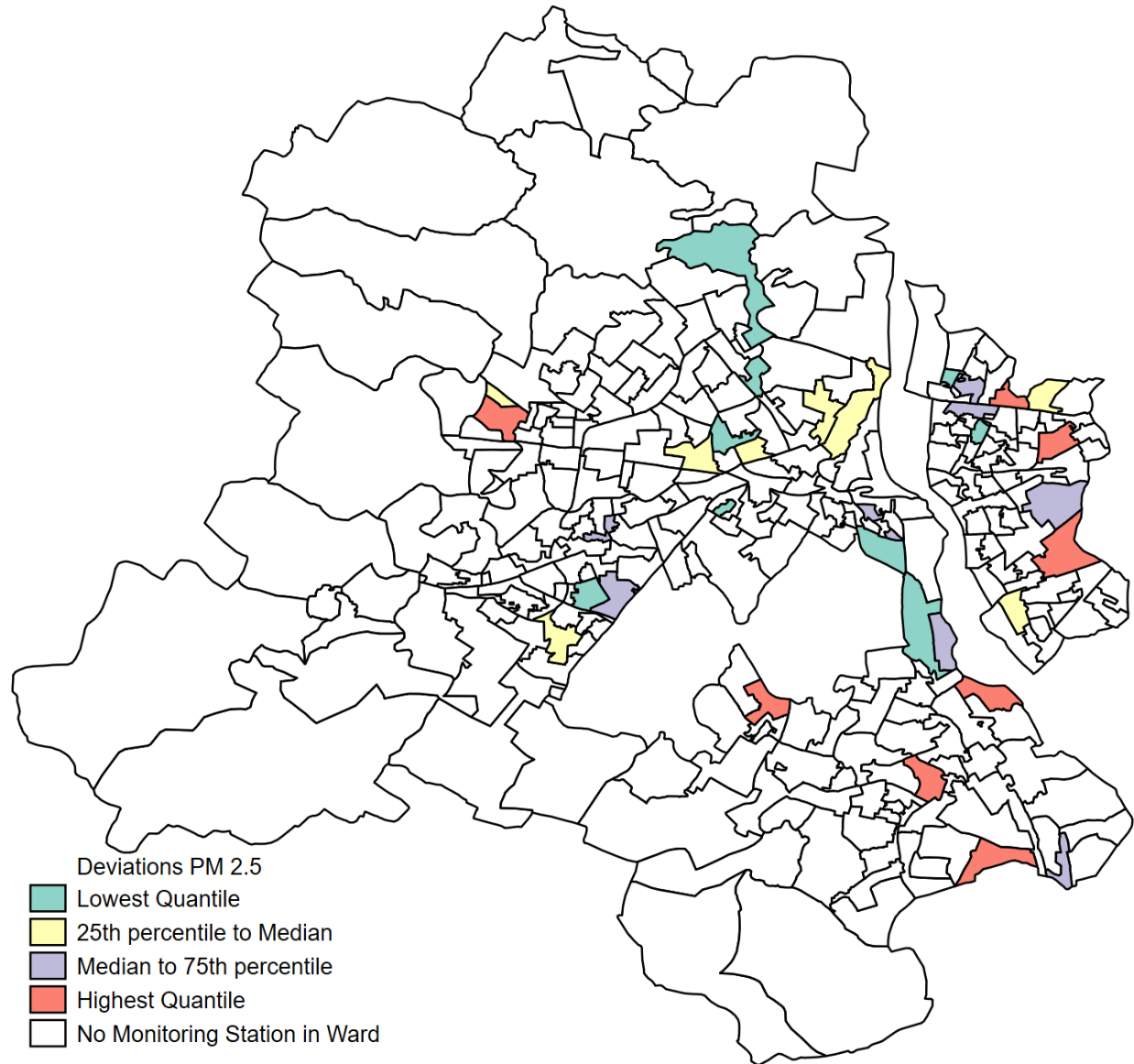


Figure 6: Heat Map of Standard Deviations of PM 2.5 concentrations, adjusted for weather, hour-of-day, day-of-week and month.

Summary statistics for all variables used, including sources used, are shown in Appendix Table 4.

Section V: Estimates of the Pollution-Congestion Relationship

OLS estimates of equation (1) are shown in Table 1. The first two columns show the impacts on NO_x and CO, with contemporaneous travel times. The next two columns show PM 2.5 and PM 10, with lagged travel times. We use the congestion measure d_{shdm} – which include travel times for trips within all wards upwind of the monitoring station – to define a categorical

variable that takes on one of five mutually exclusive values. The base category is whether the travel time lies below the 25th percentile, and the other categories are as shown in Table 1.

The pattern of results highlights very clearly the complexity of the relationship between congestion and ambient pollution concentration. Increased congestion leads to higher pollution levels, but this impact is driven differently given the percentile range in which congestion falls: pollution is therefore non-linear in congestion. Congestion over the 25th percentile, but less than the 50th percentile, adds NO_x to the ambient air – but note the other coefficients are all mostly of the same size, it is only the higher standard error that does not allow us to reject a null hypothesis of zero effect of congestion. Similar results show up for CO, except this time the impact comes when congestion moves over the 50th percentile. Consistent with the notion of emissions mixing in the air to form particulate matter which then drifts downwind after a length of time, PM 2.5 concentrations rise with a lag, when travel times go up beyond the 75th percentile. PM 10 concentrations are unaffected. NO_x emissions indeed are the primary source of this pattern of vehicular contribution to PM 2.5 emissions (Hodan and Barnard 2004), lending support to the identification of traffic congestion being the source of air pollution in these estimates.

We have examined the robustness of these estimates by running the same specification on a sample that excludes all upwind wards except for the ward containing the monitoring station. All these results become either statistically insignificant or of illogical sign.⁷ From this we can infer that excluding the influence of wind shuts down the necessary mechanism of atmospheric mixing of vehicular emissions produced during high congestion times together with downwind drift, required to produce the estimates in Table 1.

The spatial pattern of PM 2.5 concentration adjusted for weather, hour-of-day, day-of-week, and month as shown in Figures 5 and 6 suggested heavier concentrations in the east and south. The main direction in which wind blows across Delhi is from the north and west to the east and south (Sangomla 2018) so this eastern/southern concentration is likely a result of wind. Our results are consistent with this spatial spread as well, making us confident that we are indeed identifying the impact of congestion.

⁷ These results are not shown but are available on request.

Table 1: Congestion and Pollution: OLS Estimates

	(1)	(2)	(3)	(4)
	NOx	CO	PM 2.5	PM 10
<i>Contemporaneous:</i>				
Travel Time between 25 th and 50 th percentile	3.68** (1.40)	0.04 (0.04)		
Travel Time between 50 th and 75 th percentile	3.44* (1.85)	0.10** (0.04)		
Travel Time between 75 th and 95 th percentile	0.00 (2.24)	0.10* (0.06)		
Travel Time above 95 th percentile	3.89 (2.72)	0.13 (0.08)		
<i>2-period Lag:</i>				
Travel Time between 25 th and 50 th percentile			1.46 (2.10)	-2.16 (4.46)
Travel Time between 50 th and 75 th percentile			4.02 (2.70)	-3.69 (5.15)
Travel Time between 75 th and 95 th percentile			7.24*** (2.63)	3.42 (5.38)
Travel Time above 95 th percentile			13.63*** (3.92)	6.77 (5.77)
Weather Controls	Y	Y	Y	Y
Trip Fixed Effect	Y	Y	Y	Y
Time of Day Fixed Effect	Y	Y	Y	Y
Day of Week Fixed Effect	Y	Y	Y	Y
Month Fixed Effect	Y	Y	Y	Y
Observations	27,538	26,467	23,989	22,104
R-squared	0.24	0.14	0.43	0.38
# Monitoring Stations	34	34	34	31
Outcome Mean	63	1.39	117	241
Outcome Standard Deviation	69	1.57	101	157

Standard errors in parentheses, clustered by monitoring station. ** $p < 0.05$, *** $p < 0.01$. Weather controls include rainfall, temperature, surface radiation and dew point temperature (as a measure of humidity).

How do these estimates compare with existing estimates of congestion on ambient pollution levels? Knittel, Miller and Sanders (2016) report (in figure 2 of their paper) that congestion raises CO concentration by about 0.1 on average, which is equal to what we find when congestion rises above the 50th percentile. For PM 10, they find an average effect of about 8 (in figure 3 of their paper); our estimates suggest that traffic congestion raises PM10 by about 7 once congestion crosses the 95th percentile although it is statistically significant. An important difference between our estimates and Knittel, Miller, Sanders (2016) is that we allow for a non-linear relationship between pollution and congestion. Chen et al (2020) report on page 343: “vehicular emissions within the Fifth Ring Road contribute to 38.8%, 43.4%, 57.1%, and 54.1% of PM2.5, PM10, NO2, and CO pollution in Beijing”. Our estimates suggest the following percentages for PM 2.5, NOx and CO: 17%, 6% and 7%. We exclude PM 10 since our estimates were statistically insignificant.

Chen et al (2020) also find an inverse relationship between traffic congestion and air pollution when these are averaged out at various points during a day, like we show in Figure 2. Arguing that this perhaps represents reverse causation – people prefer to travel when pollution levels are low – they then use an IV strategy to estimate the impact of congestion.

In our case, given that it was well known by the public in 2018 that air quality will worsen with the coming of winter, due to crop fires upwind necessitated by the sowing of the winter crop and thermal inversion induced by colder temperatures, we can test for whether such reverse causation might be operational. To do this, we restrict the sample to only the summer months. If reverse causation is indeed the reason why we estimate the results we do, then by focusing on the summer months when air quality is better, we should find smaller effects. The results of such a sample restriction are, however, like the unrestricted sample used for the results in Table 1. We can therefore rule out reverse causation as a potential confounder.⁸

Section VI: Health Impacts from Reduced Congestion

We use our estimates in Tables 1 to study their implications for long-run health outcomes, relying primarily on the results in Burnett et al (2018). This paper summarizes the exposure-response relationship for PM 2.5 using data on 15 cohort level studies from across the globe,

⁸ These results are not shown but are available on request.

including one from China, a country that also has high PM2.5 concentrations like India. A further advantage of the results in this paper is that it documents the relationship between exposure to PM 2.5 and the resulting mortality implications at various levels of exposure with minimal assumptions over toxicity per inhaled dose. We can only study the implications of PM 2.5 reduction and not for the other pollutants. Since PM 2.5 is the deadliest of all air pollutants, this is not a major problem.

Burnett et al (2018) provides a set of estimates for the relative mortality risk from ischemic heart disease, and cardiopulmonary disease. Using our estimates on the impact of travel times on PM 2.5 concentrations from Table 1, we simulate the impact on mortality from a policy aimed at a reduction in the probability of extreme congestion events. To get at the impact of congestion on PM 2.5, we add the coefficient on trips with travel times between the 75th and 95th percentile (7.2) together with the coefficient on trips with travel times above the 95th percentile (13.6); doing so implies congestion raises PM 2.5 concentrations by 20 micrograms per cubic meter.

Table 2 shows the resulting implications for mortality risks from chronic obstructive pulmonary and ischemic heart disease, showing the premature deaths from congestion. We look at these outcomes because Burnett et al (2018) report specifications of the hazard rate for populations above the age of 25. The other two major diseases from particulate concentration are lung cancer and lower respiratory infections: of these the former is still not very prevalent in India while the latter is more of a concern for children.⁹ The total deaths avoided from transportation congestion reductions will then be bounded between the 4 numbers shown, as deaths can take place due to complications induced by both diseases or just one of them. Because all the point estimates used are bounded away from zero, these mortality reductions will also be bounded away from zero.

⁹ We calculate premature deaths avoided by multiplying the attributable fraction (calculated from the relative risk, evaluated using the parameter estimates for the hazard function given in the Burnett et al (2018) appendix), population incidence of these diseases and the exposed population (see Ostro 2004). We take the population of Delhi to be 19,438,678 and use the 2011 Census to calculate the percentage of the population above the age of 25 in Delhi. Information on the population incidence of these diseases is taken from India specific data on the Global Health Data Exchange (<http://ghdx.healthdata.org/gbd-results-tool>). Appendix II has more details on how these deaths were calculated.

Table 2: Premature Deaths from Congestion

	Point Estimate of the impact of congestion on PM 2.5: 20
Implied mortality:	
Chronic Obstructive Pulmonary Disease	3149
Ischemic Heart Disease	5417

Ideally, we would only use data on the population within the grid cells where we have information on traffic flow and pollution concentration. As we lack such spatially fine data, we calculate deaths from all of Delhi's population implying our estimates are an upper bound of the actual deaths avoided. It is also possible that since we cannot account for all the ways congestion could impact deaths, our estimates could be potentially a lower bound.

We can see these mortality reductions are large: to put these numbers in context, Pandey et al (2021) find that in 2019 a total of 16,595 premature deaths in Delhi could be attributed to ambient particulate matter pollution. Therefore, extreme congestion events lead to premature deaths by somewhere between 19% and 33% of the total deaths from particulate matter.¹⁰ Pricing congestion explicitly is therefore likely to bring significant gains.

Section VII: Possible Threats to Identification

Although we are confident of our identification of PM 2.5 concentrations arising from traffic congestion, there is a subtlety to the interpretation of our results. It is not clear whether our results pin down traffic congestion resulting from a squeeze on the existing road transport infrastructure, or traffic congestion resulting from other reasons. Prominent among these other reasons is construction activity. This distinction is important. Construction of the scale required to severely interrupt vehicle movement is likely to be a one-time event whereas inadequate road infrastructure is a more long-term concern, leading to questions of possible substitution toward public transport or increased expansion of the current network. If construction projects are really the underlying reason for traffic congestion, our estimates do not imply a long-term mortality cost.

¹⁰ We include the entire population of Delhi in our calculations, so we assume extreme congestion events have similar impacts regardless of location within the city.

To get past this problem, we use an instrumental variables strategy. We construct our instruments for travel times in the following way. The Delhi Traffic Police runs a twitter account where they make regular announcements of travel alerts. Using machine learning techniques, which are described in Appendix I, we use the text within the tweets to classify the tweet as a type I or type II congestion event, with a type II event denoting higher congestion. If either the origin or destination ward is affected by a type I congestion event, the instrument takes a value of one; if both origin and destination ward is affected by a type I congestion event the instrument takes a value of 2; and the instrument equals 0 otherwise. A similar definition uses type II congestion events.

Figure 7 below plots a time series graph, showing the total number of trips that had at least one type I congestion event on any given day. Because a single ward can enter multiple trips as either an origin or destination ward, a type I congestion event in one ward can affect several trips. We can see two big spikes – in the months of July and August as well as in January. January 26th happens to be Republic Day in India, a major national holiday with many roads closed partially or wholly in the center of the city in the weeks preceding a parade held in the heart of the city. The other major national holiday is August 15th, the day India became an independent country. Monsoon rains hit Delhi during July and August, causing roads to flood, and slowing traffic. Figure 7 tells us there is significant variation in traffic congestion events across time.

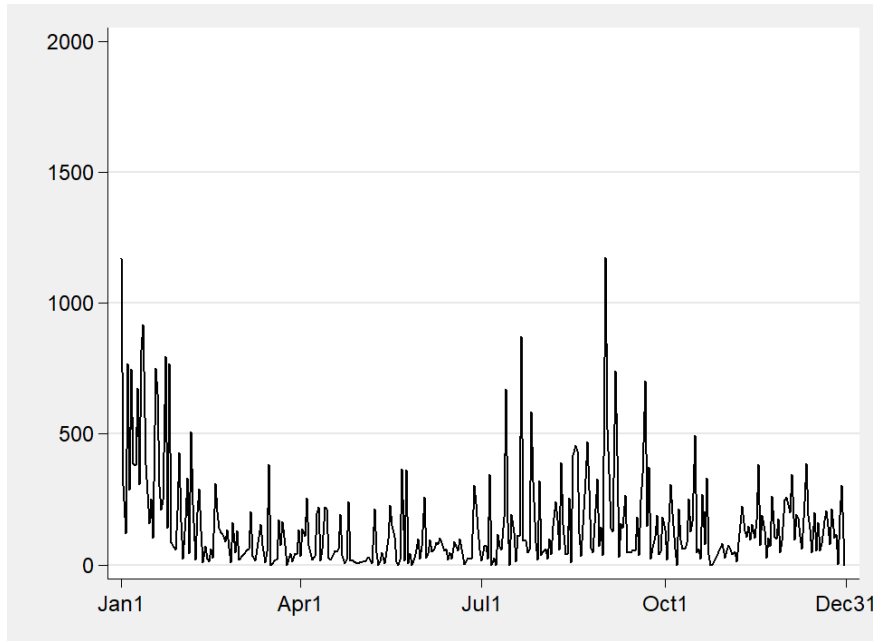


Figure 7: Tweet Activity over Time (Source: Delhi Traffic Police Twitter account, own calculations)

Spatial variation in congestion events is shown in Figure 8. Red dots refer to type I congestion events, green dots refer to type II congestion events and yellow represent those events we could not classify. These dots are reasonably spread out over the entire city suggesting significant variation in the probability of congestion events across origin and destination wards. Table 3 below shows examples of these three types of congestion events.

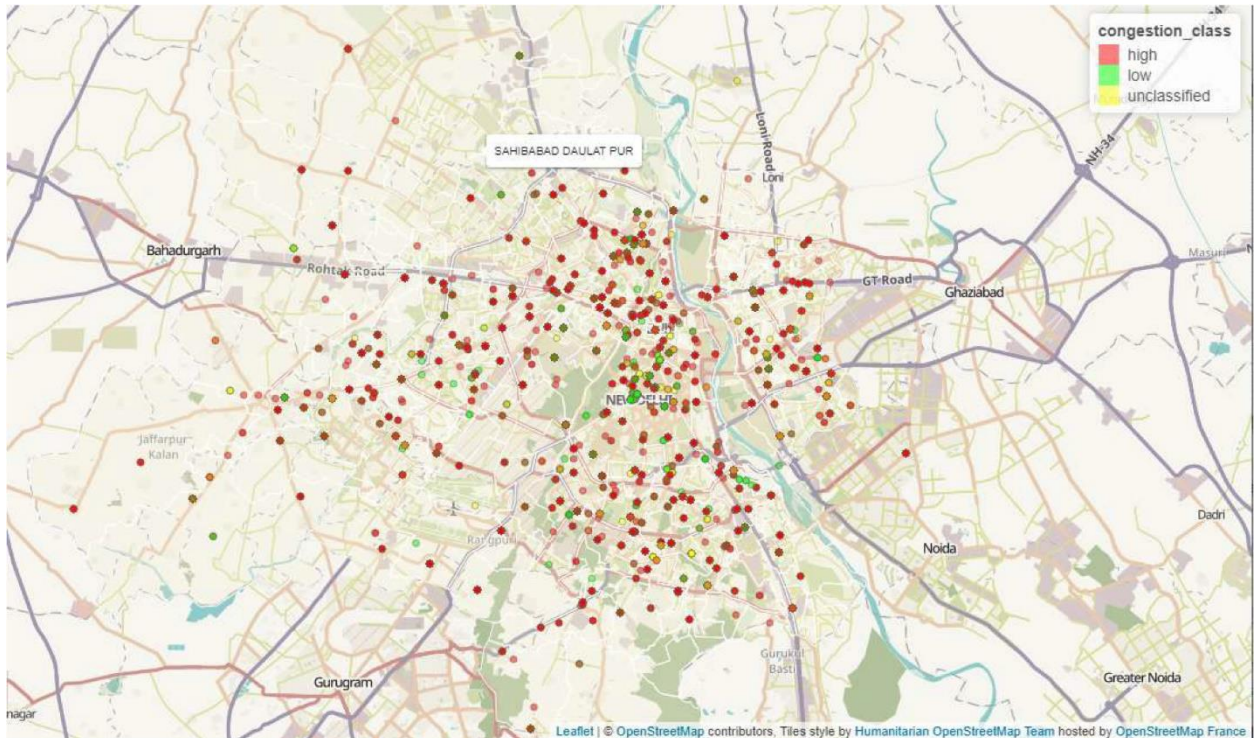


Figure 8: Spatial Variation in Tweets, by Congestion Event (Source: Delhi Traffic Police Twitter account, Own Calculations)

Table 3: The Three Types of Congestion Events

Tweet	Congestion Event
“Due to the work of Metro, traffic from Shyam Lal College to Bihari Colony (Road No 57) will be high, sorry for the inconvenience.”	Type I
“In front of Lajpat Nagar police station, MTNL Vallo has dug a hole. There is a hindrance in traffic.”	Unclassified
“Break down DTC bus No. DL1PC8733 has been removed from Kalkaji near foot over bridge. Traffic is normal now.”	Type II

Source: Delhi Traffic Police Twitter Account, Own Calculation

We are currently working on defining these instruments for upwind wards.

Section VIII: Preliminary Conclusions

In this paper, we estimate the relationship between pollution and traffic congestion. Taking travel times as our measure of congestion, which is equivalent to inverse speed – a common measure of congestion – when trip fixed effects are included, we find pollution is statistically significantly affected by congestion. Further the impact is both non-linear – rising

sharply for extreme congestion events - and dynamic, sustaining for many hours after an extreme congestion event. Health benefits in the long term from increased congestion regulation are large. Our preferred interpretation of these estimates of congestion is that they represent the number and type of vehicles interacting with existing transport infrastructure.

Given these results, a congestion pricing scheme would appear to have significant benefits. To the extent that some commuters can substitute between time of day when deciding their commutes, if the appropriate congestion price is put in place, it will encourage such substitution. The health benefits alone appear to be large enough to justify further work examining what a congestion pricing scheme might look like. Note we have explicitly ignored the loss in welfare that comes from waiting in traffic. Incorporating this concern will only further make the case for congestion pricing.

A potential concern with congestion pricing, which should be considered when designing such a scheme, is how easy would it be to evade paying the price. For instance, a toll road can be avoided by taking alternative, if longer, routes. In theory, one could track each car and legally mandate each car owner to have such a device. Technologically this is feasible but raises important issues of privacy. It might perhaps be more effective to subsidise traveling during certain times rather than raise prices during congested times.

References

1. Akbar, Prottoy, and Gilles Duranton. "Measuring the cost of congestion in highly congested city: Bogota." (2017).
2. Anderson, Michael L. "Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion." *American Economic Review* 104, no. 9 (2014): 2763-96.
3. Barth, Matthew, and Kanok Boriboonsomsin. "Real-world carbon dioxide impacts of traffic congestion." *Transportation Research Record* 2058, no. 1 (2008): 163-171.
4. Barrows, Geoffrey, Teevrat Garg, and Akshaya Jha. "The Health Costs of Coal-Fired Power Plants in India." (2019).
5. Burnett, Richard, Hong Chen, Mieczysław Szyszkowicz, Neal Fann, Bryan Hubbell, C. Arden Pope, Joshua S. Apte et al. "Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter." *Proceedings of the National Academy of Sciences* 115, no. 38 (2018): 9592-9597.
6. Chen, S., Qin, P., Tan-Soo, J. S., Xu, J., & Yang, J. (2020). "An econometric approach toward identifying the relationship between vehicular traffic and air quality in Beijing." *Land Economics*, 96(3), 333-348.
7. Currie, Janet, and Reed Walker. "Traffic congestion and infant health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3, no. 1 (2011): 65-90.
8. Duranton, Gilles, and Matthew A. Turner. "The fundamental law of road congestion: Evidence from US cities." *American Economic Review* 101, no. 6 (2011): 2616-52.
9. Frechette, Guillaume R., Alessandro Lizzeri, and Tobias Salz. "Frictions in a competitive, regulated market: Evidence from taxis." No. w24921. *National Bureau of Economic Research*, 2018.
10. GBD MAPS Working Group. "Burden of disease attributable to major air pollution sources in India." *Special report* 21 (2018).
11. Gupta, Aashish, and Dean Spears. "Health externalities of India's expansion of coal plants: Evidence from a national panel of 40,000 households." *Journal of environmental economics and management* 86 (2017): 262-276.
12. Guttikunda, Sarath K., K. A. Nishadh, and Puja Jawahar. "Air pollution knowledge assessments (APnA) for 20 Indian cities." *Urban Climate* 27 (2019): 124-141.

13. Hodan, W. M., & Barnard, W. R. (2004). Evaluating the contribution of PM_{2.5} precursor gases and re-entrained road emissions to mobile source PM_{2.5} particulate matter emissions." *MACTEC Federal Programs, Research Triangle Park, NC*.
14. Holland, Stephen P., Jonathan E. Hughes, Christopher R. Knittel, and Nathan C. Parker. "Some inconvenient truths about climate change policy: The distributional impacts of transportation policies." *Review of Economics and Statistics* 97, no. 5 (2015): 1052-1069.
15. Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders. "Caution, drivers! Children present: Traffic, pollution, and infant health." *Review of Economics and Statistics* 98, no. 2 (2016): 350-366.
16. Kreindler, Gabriel E. "Driving Delhi? Behavioral Responses to Driving Restrictions." Unpublished paper (2016).
17. Kreindler, Gabriel E. "The welfare effect of road congestion pricing: Experimental evidence and equilibrium implications." Unpublished paper (2018).
18. Mangrum, Daniel and Alejandro Molnar. "The marginal congestion of a taxi in New York City", Unpublished paper under review at the *American Economic Review* (2019)
19. Pope III, C. Arden, Richard T. Burnett, Michelle C. Turner, Aaron Cohen, Daniel Krewski, Michael Jerrett, Susan M. Gapstur, and Michael J. Thun. "Lung cancer and cardiovascular disease mortality associated with ambient air pollution and cigarette smoke: shape of the exposure–response relationships." *Environmental health perspectives* 119, no. 11 (2011): 1616-1621.
20. Sangomla, 2018, November 14, "Long-term wind patterns over Delhi have changed, say scientists", *Down to Earth*, <https://www.downtoearth.org.in/news/air/long-term-wind-patterns-over-delhi-have-changed-say-scientists-62126>
21. Venkataraman, Chandra, Michael Brauer, Kushal Tibrewal, Pankaj Sadavarte, Qiao Ma, Aaron Cohen, Sreelekha Chaliyakunnel et al. "Source influence on emission pathways and ambient PM_{2.5} pollution over India (2015–2050)." *Atmospheric Chemistry and Physics* 18, no. 11 (2018): 8017-8039.
22. Ward Jr, Joe H. "Hierarchical grouping to optimize an objective function.", *Journal of the American Statistical Association*, 58 no. 301 (1963): 236 - 244
23. WHO, 2018. WHO Global Ambient Air Quality Database (Update 2018), *World Health Organisation*, Geneva, Switzerland.

24. Wikipedia contributors, "Ward's method," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Ward%27s_method&oldid=912897676 (accessed August 19, 2020).

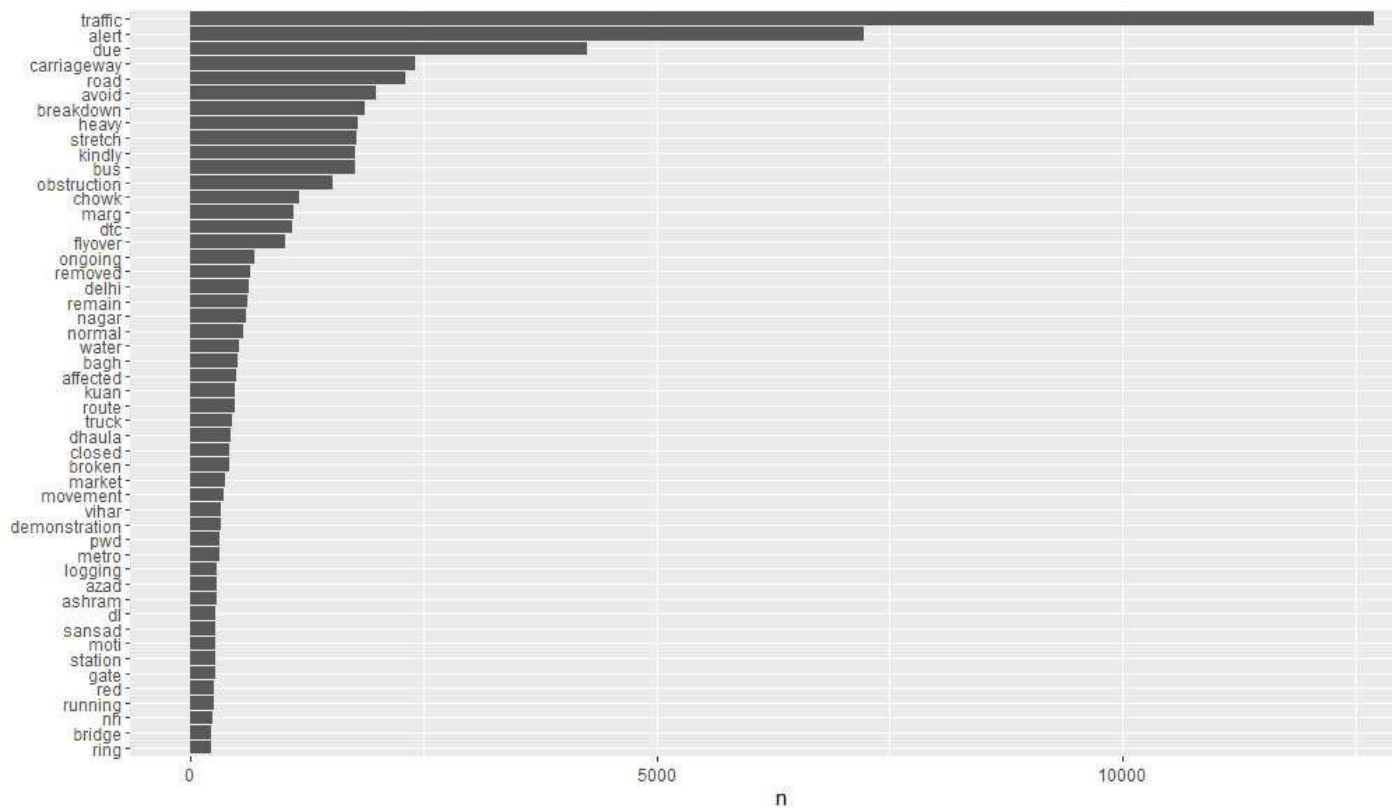
Appendix I: Tweet Classification Using Hierarchical Clustering

The document corpus consisted of the translated text of 6690 tweets posted by the Delhi Traffic Police using a Twitter handle called "dtptraffic" which is dedicated to traffic alerts. The tweets were pre-processed by removing punctuations, special characters, urls and numbers.

The goal of clustering was to partition the tweet corpus into classes each of which maps to either "type I" or "type II". We take a supervised classification approach in which parameters were tweaked until clusters were achieved which match a priori knowledge about the dataset.

Manual Extraction of Categories

We examined the 50 most frequent words in the corpus - having removed punctuations and special characters - which is shown in Appendix Figure 1 below.



Appendix Figure 1: 50 most frequent words

We then manually extracted the following categories from the list of frequent words.

1. **Extremely frequent words** (traffic, alert, due): words that occur more than 10000 times in the corpus
2. **Place words** (road, market, carriageway, station, gate, bridge, route, stretch, flyover): words that tie an alert to a generic place in a given location.
3. **Address words** (chowk, marg, sansad, metro, moti, bagh, ashram, delhi, nagar, azad, vihar. dhoula, kuan): words that locate the tweet at a specific addressable location. These are the geocoding key words.
4. **Event words** (breakdown, obstruction, water, logging, removed, demonstration, truck, broken, pwd, bus, ongoing): words that describe events that potentially affect traffic.
5. **Traffic Condition words** (closed, affected, heavy, normal, running, remain): words that describe traffic conditions

Appendix Table 1 summarizes what kind of information each of these categories of words conveys.

Appendix Table 1: Word categories

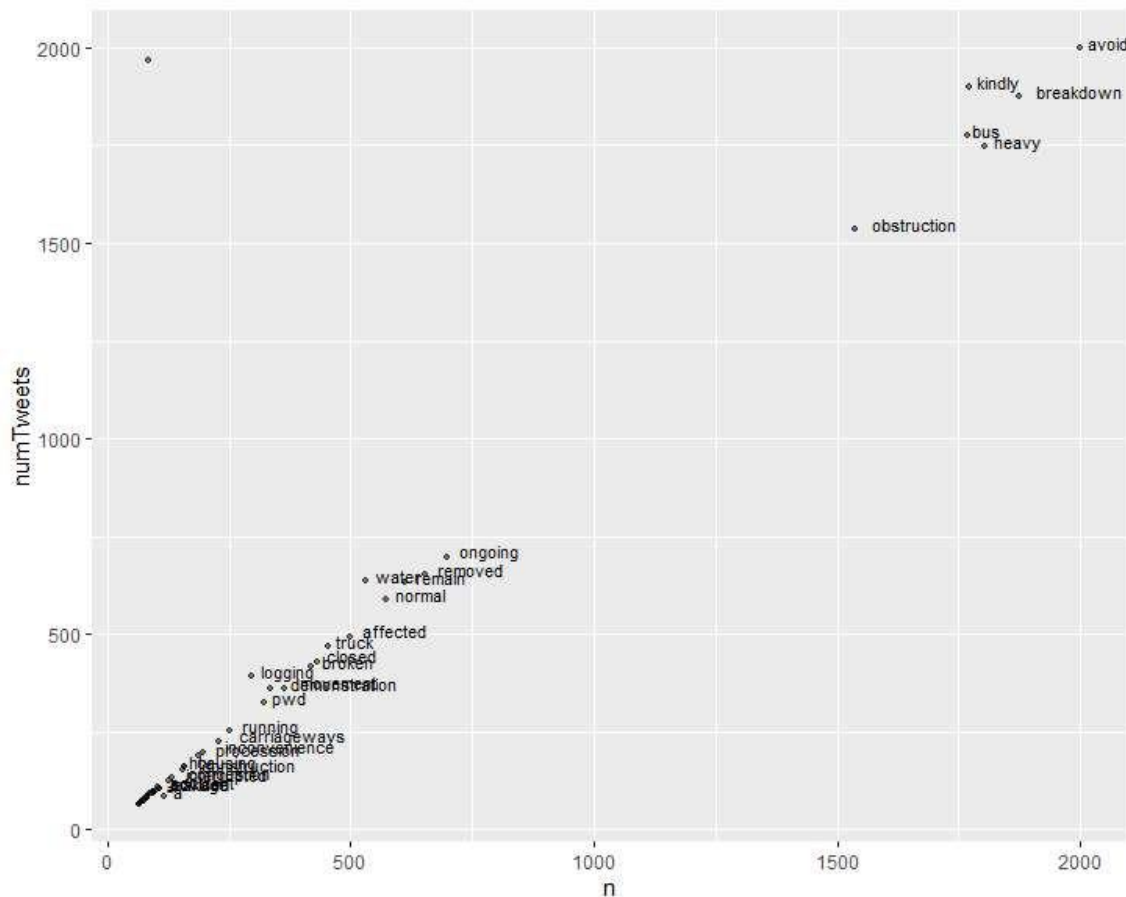
Category	What it tells us
Extremely Frequent Words	Not too meaningful since they occur in almost every tweet
Place words	Where is the tweet located? (generic location)
Address words	Where is the tweet located? (specific location)
Event words	What kind of traffic event has occurred or is occurring?
Traffic condition words	What is traffic like right now or will be in the near future?

From the categories manually extracted, extremely frequent words like "traffic" and "alert" do not convey meaningful information since they occur in almost every tweet. Place and location words convey information about *where* the tweet is located and therefore, they are not meaningful in the context of trying to find out *what* the tweet is reporting regarding the traffic condition. To know what traffic is like we have to look at "traffic condition words" like "heavy", "affected", "normal" etc. These are the most meaningful words for the purpose of clustering tweets into high and low congestion classes. Additionally, event words are also meaningful because they contain the causes of traffic

conditions. We therefore removed extremely frequent words, place words and address words from the corpus and kept only event words and traffic condition words in order to perform clustering.

Term Frequency vs Term-Document Frequency

The term frequency of a given term is the number of times it is used in the corpus and the term-document frequency is the number of documents (tweets, in our case) in which it appears. Appendix Figure 2 shows the term frequency (n) of common terms in the corpus plotted against its term-document frequency (numTweets).



Appendix Figure 2: Term-Document Frequency (numTweets) against Term Frequency (n)

Each term appears roughly once per tweet since its frequency is approximately the same as its term-document frequency. That means that we are able to characterize tweets by single words or combinations of single words. Next, I manually classified the words in the two selected categories - traffic condition words and event words - depending on whether they are associated with high congestion or low/normal congestion. The results

are shown in Appendix Table 2.

Appendix Table 2: Word associations with Type I and Type II congestion

Category	Type I Congestion	Type II Congestion
Traffic Condition Words	heavy, affected, inconvenience, congestion, interrupted, reported, obstruction	normal, diverted
Event Words	breakdown, bus, ongoing, water, truck, broken, demonstration, pwd, logging, procession, htv, construction, leakage, accident, waterlogging, sewer, pipeline, break, underway, elevated	removed, closed

Term-Document Matrix

We construct a term-document matrix each row of which corresponds to a term and each column to a tweet. If term i is present in tweet j then the element (i, j) of the term-document matrix is 1 and it is zero if term i is not present in tweet j . We then trim the matrix down to remove terms that occur in less than 1% of the tweets which is roughly 7 tweets. This step is important because it reduces dimensionality of the term-document matrix but also may lose some important terms which will cause misclassification or non-classification of certain tweets. A distance matrix is computed from the term document matrix which contains in the $(i, j)^{th}$ element, a similarity measured based on Euclidean distance between the terms i and j . This distance matrix is used as a basis for Ward's agglomerative hierarchical clustering method using a minimum variance method (Ward, 1963) that clusters terms together at each iteration by minimizing the error sum of squares.

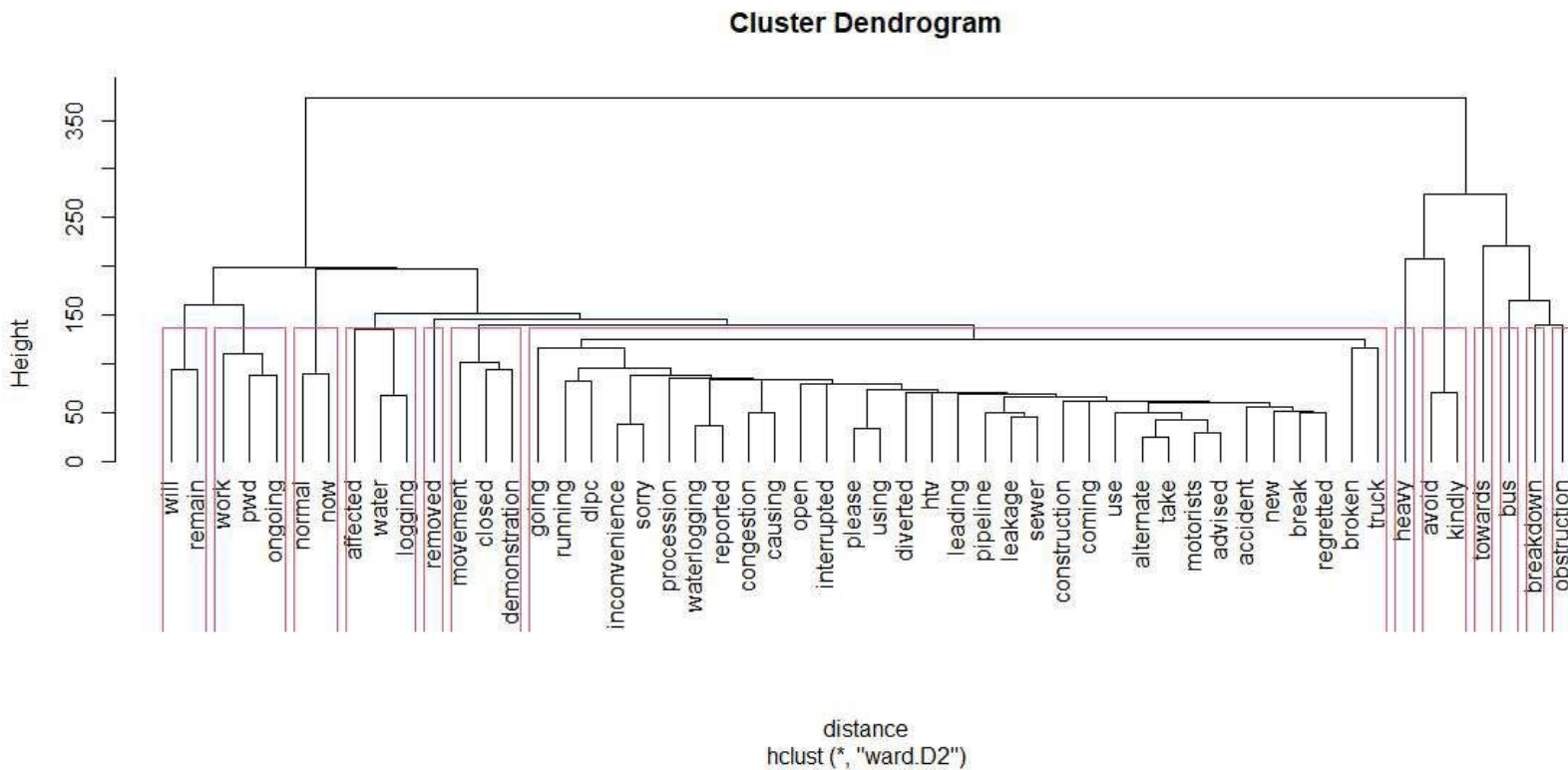
Initially, each tweet is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. At each stage distances between clusters are recomputed by the Lance-Williams dissimilarity update. Ward's minimum variance criterion minimizes the total

within-cluster variance. At each step it finds the pair of clusters that leads to minimum increase in total within-cluster variance after merging. This increase is a weighted squared distance between cluster centers. At the initial step, all clusters are singletons (clusters containing a single point). To apply a recursive algorithm under this objective function, the initial distance between individual objects must be (proportional to) squared Euclidean distance. The initial cluster distances in Ward's minimum variance method are therefore defined to be the squared Euclidean distance between points (Wikipedia, 2020):

$$d_{ij} = d(\{X_i\}, \{X_j\}) = \|X_i - X_j\|^2$$

Appendix Figure 3 shows the term clusters obtained with the following parameters. The sparsity threshold is set 0.01 (that is, remove terms which occur in less than 1% of the tweets) and the total number of clusters, $k = 13$. These parameters were manually tweaked to arrive at a set of meaningful clusters.

Appendix Figure 3: Term clusters, Sparsity Threshold = 0.01, $k = 13$, Ward's Clustering Method



The above clusters were used along with the word association matrix to arrive at the classifications shown in Appendix Table 3. Not all clusters are meaningful.

Appendix Table 3: Clusters and Congestion Event Classification

Cluster Number	Terms Contained in Cluster	Interpretation	Congestion
1	will, remain	-	-
2	work, ongoing, pwd	Ongoing public works	Type I
3	normal, now	Traffic is normal now	Type II
4	affected, water logging	Traffic affected due to water logging	Type I
5	removed	Traffic obstruction removed	Type II
6	movement, closed, demonstration	Traffic movement is closed due to a demonstration	Type II
7	going, running, dlpc, inconvenience, sorry, procession, waterlogging, reported, congestion, causing, open, interrupted, please, using, diverted, htv, leading, pipeline, leakage, sewer, construction, coming, use, alternate, take, motorists, advised, accident, new, break, regretted, broken, truck	A variety of traffic events that cause congestion or interruptions	Type I
8	heavy	Traffic is heavy	Type I
9	avoid, kindly	Motorists advised to avoid route due to congestion	-
10	towards	-	-
11	Bus	Traffic events related to buses	-
12	breakdown	A breakdown that is impacting traffic	Type I
13	obstruction	Presence of a traffic obstruction	Type I

The above table can be further simplified by removing common words like "now", "towards" and "please. Once that is done, the simplified table then leads to the simple classification algorithm below:

1. Do for all tweets
2. If the tweet contains any of the groups of words from clusters 3,5,6, classify it as type II
3. If the tweet does not contain any of the word groups from clusters 3,5,6 and it contains any of the words from the remaining clusters, classify it as type I

Appendix Table 4: Summary Statistics for the Main Sample

	Source	Observations	Mean	Standard Deviation	Min	Max
Trip Time (Seconds)	Uber	37,367	1471	500	157	4706
<i>Pollutants</i>						
PM 2.5 (Micrograms/M ³)	CPCB*	31,140	117	102	0	4084
PM 10 (Micrograms/M ³)	CPCB	28,875	241	157	0	1324
NOx (Micrograms/M ³)	CPCB	30,618	1.4	1.6	0	48.2
CO (Miligrams/M ³)	CPCB	31,668	63	69	0	960
<i>Weather</i>						
Rain (mm)	ERA5 ⁺	16,188	0.6	2.1	0	22.8
Temperature (Celcius)	ERA5	16,188	24	8	6	43
Net Solar Radiation (Joules/M ²)	ERA5	16,188	1.3E+07	5.4E+06	7.6E+05	2.4E+07
Dew Point (Celcius)	ERA5	16,188	13	7	0	27

Estimates for the year 2018, for the city of Delhi. Sample restricted to wards and times where data on tweets from the Delhi Traffic Police are non-missing.

*: CPCB refers to the Central Pollution Control Board. +: ERA5 refers to the ERA5-Land dataset from Copernicus Climate Change Service.

Appendix II: Calculating deaths avoided

Ostro (2004) suggests the following method to calculate deaths from air pollution:

$$D = AF * PI * EP$$

Where D equals deaths, AF is the attributable fraction from a specific mortality cause, PI is the population incidence of the mortality cause and EP is the exposed population.

To calculate the attributable fraction, we use the transformation:

$$AF = (RR - 1)/RR$$

Where RR is the relative risk from a mortality cause. This is evaluated using the hazard ratio formulation from Burnett et al (2018):

Hazard from a mortality cause = $exp \frac{\theta \log(z/\alpha+1)}{(1+\exp\{-(z-\mu)/\nu\})}$ where z equals the reduction in pollution from a transport congestion policy, θ , α , μ and ν are parameters that control the shape of the hazard function. We take the parameters specific to the cohort data that includes the Chinese male cohort for the age group 25 and over from Burnett et al (2018) appendix Table S2.

The population incidence of the two mortality causes is taken from India specific data on the Global Health Data Exchange (<http://ghdx.healthdata.org/gbd-results-tool>).

To calculate the exposed population we take the total population of Delhi in 2018, and multiply this by the proportion of the population over the age of 25 (this percentage is taken from the 2011 Census and equals 52%). Note that the eventual exposed population is smaller than this number because ideally, we should only use the population that lies in the downwind cells – however we lack data at this spatial level. Therefore, our calculation of deaths avoided is an upper bound.