The Effect of Police Response Time on Crime Detection*

Jordi Blanes i Vidal† and Tom Kirchmaier‡

July 14, 2015

Abstract

We study whether faster police response times help to combat crime by increasing the likelihood of an arrest. To identify causal effects, we exploit discontinuities in distance to the response station across locations next to each other, but on different sides of division boundaries. We find that faster response times lead to a large increase in the likelihood of detecting (i.e. 'clearing') crimes. The effects are stronger for higher priority calls. Part of the mechanism is the higher likelihood that a suspect will be named by a victim or witness when the police attend the scene more promptly.

JEL *classification*: D29, K40.
Keywords: Police, Crime, Organisational Performance.

*We thank Yona Rubinstein for his support and invaluable insights throughout this project. Special thanks also to DCC Ian Hopkins, Peter Langmead-Jones, Duncan Stokes and many others for making this project possible.

†Department of Management and Centre for Economic Performance, London School of Economics, London WC2A 2AE, United Kingdom. Email: j.blanes-i-vidal@lse.ac.uk.

‡Financial Markets Group, London School of Economics, London WC2A 2AE, United Kingdom. Email: t.kirchmaier@lse.ac.uk.

1 Introduction

The Question The likelihood that a crime is detected and its offender charged is a central component of the standard economic model of crime[1]. It is also an issue with significant scope for improvement, since, for instance, less than a quarter of crimes are cleared in the US. It is therefore unfortunate that the economics literature has devoted such scant attention to studying in detail the determinants of crime detection. A better understanding of the effectiveness of policing practices could help improve detection rates without the need for additional public resources. Catching a higher number of offenders would in turn allow for the decrease in criminal sanctions without making criminal activities more attractive (Becker, 1968).

In this paper we study the arguably most important and most controversial instrument used by police forces to apprehend criminals: responding rapidly when alerted to a crime. The effectiveness of rapid response policing seems self-evident. By arriving more quickly, police officers are able to arrest any suspect and/or question any witness at the scene, as well as prevent the destruction or contamination of physical evidence. Accordingly, rapid response is an integral part of the toolkit used by police forces to detect crimes (Bratton and Knobler 2009, Karn 2013). Practitioners' textbooks teach that the initial response is the most important part of any criminal investigation (Hess and Hess 2012, College of Policing 2013). In turn, police agencies devote vast resources to minimising response times; they track and publicise response time statistics; and they often include target response times as part of the core performance measures by which they are evaluated[2].

The effectiveness of rapid response policing has, however, long been questioned by criminologists. In his survey on the rise of evidence-based policing, Sherman (2013) argues, for instance, that *'there is no evidence that rapid response can make any difference in detection or crime rates and some indirect evidence that it cannot.'* This view is shared in the reviews of Walker (1994), Weisburd and Eck (2004), Katz and Walker (2012) and Siegel and Worrall

[1] We use the terminology of 'crime detection' to be true to the UK context where the empirical setting of this paper takes place. The concept of detection is equivalent to the concept of 'clearance' used by the US FBI. We define 'detection' formally in Section 3. The overwhelming majority of detected crimes result in a criminal charge.

[2] An illustration of the first point is that the response team comprises of 24% of the total number of police officers in the Greater Manchester Police. On the second point, Appendix A displays a list of links to response time statistics among major US police agencies. To demonstrate the third point, the Boston Police Department lists target average response within seven minutes for priority 1 calls as one of its three key performance indicators. Response times are also an important part of other police departments' performance goals, including Houston, Phoenix, Austin and San Diego.

(2014), among many others. Two complementary arguments are commonly put forward. Firstly, response time matters only within the first minute after a crime takes place, an unrealistically short interval for even the most efficient police organisation (Bayley, 1996). Secondly, the delay before the police are notified is typically so long that the speed of any subsequent police action becomes irrelevant (Sherman et al., 1997). The consensus among criminologists advocates a move away from rapid response policing and into other activities, such as hot spot targeting (Braga, 2001) and problem-oriented policing (Goldstein, 1990), for which there is substantial evidence of effectiveness.

As we argue in detail below, existing evidence on the effect of police response times on crime detection is far from satisfactory. This is unsurprising, since public-use crime-level datasets do not document police inputs, and therefore analysing response times requires the unlikely collaboration of a police department. Additionally, there is the problem of endogeneity in response times. Crimes assigned a higher priority could be those with an ex ante higher or lower detection difficulty, making the identification of causal effects challenging.

This Study We estimate the effect of police response time on crime detection using a uniquely rich dataset and a research design that exploits discontinuities in response times around the boundaries of police force divisions. Our dataset comprises of the 2008-2014 internal records of the Greater Manchester Police, which is the second largest force in the United Kingdom and oversees a population of 2.6 million. Our dataset contains information on crime characteristics, police inputs such as response time, and police outputs such as whether the crime was detected and, if so, how long that took. We first use OLS regressions to document a negative, albeit small, semi-elasticity between response time and detection likelihood.

To credibly identify causal effects, we first take advantage of a particular feature of our police force: the fact that, when a call for service is received, the responding officer often departs from the station where she is based, rather than from a random point along the patrolling route. Therefore, crime scenes closer to the response station are reached more quickly following a call for service.

Unobserved determinants of detection difficulty at the area level might, however, correlate with distance to the response station. To account for this, we exploit the partition of the Greater Manchester territory into 11 operationally distinct divisions. This implies that crime scenes within a small local area, but on different sides of a division boundary, are served by separate response stations, which may be at very different respective distances. In our empirical specifications we control for the 'local area' by introducing a large number of geographical cell indicators, each representing an area of .185 squared kilometres. Variation in distance to the division response station, which we use as an instrument for response time, is then largely due to crime scenes in the same geographical cell falling on separate sides of division boundaries.

We perform three separate balancing tests to confirm the identification assumption that the characteristics of a crime are uncorrelated with distance to the division response station, conditional on the geographical cell indicators.

Findings The estimated effect of response time on detection likelihood is negative, large and strongly significant. The 2SLS estimate is in fact larger than its OLS counterpart, which is consistent with response stations being endogenously located in areas of difficult-to-detect criminal activity. We also find an effect on the *intensive margin*: conditional on detecting a crime, the police take less time to do so if the initial response time was faster. The effects are larger for thefts than for violent crimes, although they are also large for the latter.

The richness of our dataset is such that it contains administrative information on the delay between the crime occurring and the victim or witness alerting the police. This allows us to examine the argument put forward by criminologists that a fast response time is more effective when the police is called more quickly. We find that this is indeed the case and conclude that police and citizens' celerity are strategic complements in the detection production function.

We also observe the priority assigned to an incident by the handler answering the 999 call. Grade 2 calls are defined as those without imminent threat of violence, but where witness or evidence may get lost if attendance is delayed beyond one hour. We find that the effect of response time is larger for Grade 2 calls than for lower-priority Grade 3 calls. We interpret this as evidence that the police is at least partially successful in assigning priority to the most time-critical incidents.

As we discussed above, there are several potential mechanisms through which the police could convert a faster response into a higher likelihood of detection. In our dataset, we observe whether a suspect was named to the police by a victim or witness to the crime. This is a particularly important factor for minor or difficult-to-solve crimes, which may not necessarily trigger a comprehensive police investigation. Arriving at the crime scene quickly should, however, allow an officer to find witnesses to the crime, question them before their recollections worsen and encourage witness and victim cooperation by signaling efficiency and dedication. Using our baseline empirical strategy, we find that the likelihood of having a suspect named decreases with response time. We conclude that this represents at least part of the mechanism through which lower response times increase the likelihood of an arrest.

Related Work Economists have devoted a lot of attention to examining the reduced form effect of police on crime. Typical approaches include studying whether police numbers (Levitt, 1997), police composition (McCrary 2007, Miller and Segal 2014) or high visibility patrolling (Di Tella and Schargrodsky 2004, Klick and Tabarrok 2005, Evans and Owens 2007 and Draca et al. 2011) are associated with lower crime rates. The implicit assumption is that a change in these variables can lead to higher chances of catching offenders, which has an immediate deterrence effect as well as an incapacitation effect over longer horizons. However, very little work has examined directly whether the police can actually increase the detection rate with either higher numbers or different practices.

Recent availability of some datasets documenting police inputs and outputs has allowed some progress in this area. Garicano and Heaton (2010), Soares and Viveiros (2010) and Mastrobuoni (2014) all study whether the adoption of information technology by police agencies allows them to detect more crimes. Adda et al. (2014) show that the depenalisation of cannabis possesion in a London borough allowed the police to reallocate effort and detect more non-drug related crimes.

While no study in economics has examined the relation between response time and crime detection, Mastrobuoni (2015) studies a related question: whether detection rates of commercial robberies in Milan are lower around the time during which police patrols change shifts[3]. He finds that the detection rate around these shift changes is 30% lower, and calculates that this is likely to lead to a decrease in crime through the incapacitation channel. While Mastrobuoni (2015) does not observe response times directly, a natural interpretation of his findings is that detection rates around shift changes are lower because the police takes longer to reach the crime scene at these times.

As we mentioned earlier, criminologists are fairly unanimous in their dismissiveness of rapid response policing. This consensus emerged as a result of the influential Kansas City Response

[3]An institutional peculiarity of the Milanese police is that these shift changes are likely to be particularly disruptive at regular and exogenous intervals, which allows the estimation of the causal effects of such shift changes.

Time Analysis Study (Pate et al. 1976, Kelling 1977). The Kansas City study examined a limited set of crimes in two neighbourhoods throughout one year, and found no correlation between police travel time (i.e. the time between an officer being asked to attend a scene and arrival at the scene) and the likelihood of an arrest. It was then concluded that the lack of a correlation was due to the fact that it took too long for the police to be alerted (see also Spelman and Brown, 1981).

The Kansas City study suffered from significant shortcomings, including the limited and highly non-random sample; the fact that only one component of total response time (i.e. travel time) was evaluated; the fact that even this component was measured with substantial measurement error; the fact that only on-scene arrests were measured, while ignoring later arrests; and perhaps most importantly, the lack of any attempt to identify causal effects. In addition to the deficiencies above, the relevance of the Kansas City study for modern times is limited by the vast organisational, technological and societal changes that have occured in the last 40 years. Despite these shortcomings, the matter has long been regarded as settled, with no major study in four decades revisiting the issue. Sherman et al. (1997), for instance, argue that '*the evidence is strong*' and that, while it is non-experimental, '*there is neither empirical nor theoretical justification for such an expensive (experimental) test*'. The fact that police agencies devote enormous resources to minimising response times (partly as a result of the general public demand) is therefore regarded as highly unfortunate and counter to evidence-based best practices.

Plan We describe the institutional setting in Section 2. We introduce the data in Section 3. We describe the empirical strategy in Section 4. We present the results in Section 5. Section 6 concludes.

2 Institutional Setting

In this section, we outline some of the key features of the institutional setting in which our study takes place.

Organisational Structure The Greater Manchester Police (henceforth GMP) employs approximately 6,200 officers to serve a metropolitan area with a population of 2.6 million people. Many important units such as those engaged in the investigation of organised crime are situated in the GMP central headquarters, in North Manchester. The neighbourhood patrolling and the incident response functions, however, fall under the responsibility of the 11 territorial divisions. Figure 1 displays the geographical areas served by each of the divisions[4].

Each division has its own headquarters and a number of additional police stations from which the neighbourhood and (sometimes) the response teams operate. While the organisational structures often differ across divisions, the response and the neighbourhood teams are always operationally and hierarchically separated. Figure 2 provides a simplified version of a typical division organisational chart. The two teams are supervised by their respective chief inspectors, who in turn report to different superintendents. The lines of authority only merge at the highest level, in the figure of the chief superintendent. A consequence of this operational independence is that, when an incident call is received, an officer in the response team will typically be assigned to it even if a neighbourhood officer happens to be patrolling a nearby location[5].

Call Handlers and Grade Allocation Every 999 call transferred to the GMP must be answered within a very short time by a specialised staff member, i.e. a call handler. The call handling team operates from a single central location in Manchester. Call handlers are not geographically specialised, i.e. every handler indistinctly receives calls from every area of Manchester. In answering a call, the handler questions the victim or witness, provides advice and support if necessary, records the information received in the internal system, and assigns an opening code and a grade level.

The GMP Graded Response Policy divides calls into three categories that require the prompt allocation of an officer. Calls allocated a grade level 1 (Emergency Response) require the attendance of an officer within 15 minutes of their receipt. The corresponding targets for grade levels 2 (Priority Response) and 3 (Routine Response) are 60 and 240 minutes respectively. The allocation of a grade level to an incident call is done by taking into account

[4]The division boundaries coincide with municipal boundaries other than for the city of Manchester, which is divided into North Manchester and South Manchester. Municipalities are called 'local authorities' in the United Kingdom. Our empirical strategy in Section 4 will separately identify local authority/division effects from the effect of response time.

[5]This is regarded as an efficient way to operate, for three reasons. Firstly, asking a neighbourhood officer to respond would obviously distract her from her main responsibility, i.e. patrolling. Secondly, neighbourhood officers operate mostly on foot, so it would often take longer for them to arrive at an incident scene, even if they start from a closer location. Thirdly, response officers have an array of legal powers and specialised training that other officers lack. For instance, neighbourhood officers typically do not carry guns while response officers typically do. For very serious incidents a response officer will be assigned even if a patrol officer is 'right outside the house'.

two main factors: (a) whether there is a danger to someone's safety or for serious damage to property, and (b) whether evidence or witnesses are likely to be lost if attendance is delayed. The decision rule that call handlers follow in practice is quite complex, as it involves a combination of written guidelines, unwritten but generally followed practices and their own experience. The GMP Graded Response Policy prescribes that a grade level 1 should typically be assigned when there is an imminent threat of violence or a crime in operation, while a grade level 2 is appropriate when there is no imminent threat but there may be a genuine concern for someone's safety. Calls where it is appreciated that witness or evidence is likely to be lost if attendance is delayed beyond one hour should also allocated a grade 2. Grade 3 calls are those which fall outside the other two categories but still require the presence of an officer.

Radio Operators Once the call handler has provided her input the incident becomes the responsibility of a radio operator. The radio operations team is also located centrally, but separately from the call handling team. Radio operators are geographically specialised, so when a call is received it will be the operator in charge of that area of Manchester who will be assigned to it. The radio operator uses the call handler's information, her own judgment and officer availability to assign response officers to incidents. Coordination between the radio operator in charge of a division and the local response officers is mostly direct, i.e. without involving the shift sergeant. Officers are in constant communication with the radio operators and inform them when they have reached the incident scene. The time elapsed between the call handler's creation of the incident and the officer arrival to the scene is the response time whose effect we will be estimating.

Response Stations Response officers could in principle spend their shifts moving from one incident scene to another, without ever setting foot in the station where their team is officially based. For two reasons officers are, however, often present in the station when they are asked to respond to a call. Firstly, they may have finished dealing with an incident and, in the absence of a new call for service, reported back to the station. The second and more important reason is that, to be processed, most incidents require the inputting of information into the internal systems, which are office-based. While we do not observe in our dataset where response officers travel from on an incident-by-incident basis, we will confirm empirically that the geodesic distance between an incident scene and the closest response station (in the division to which the incident scene belongs) represents a very strong predictor of response time.

Some divisions changed the location of their response stations during our sample period. These changes led to mechanical variation in distance to the response station within a local area and across time. Needless to say, the relocation of the stations may be correlated with the evolution of crime patterns in Manchester. For example, the headquarters of the North Division relocated in 2011 to a newly developed business park, as part of an expensive urban regeneration project that included the creation of a new tram link. It is conceivable, in this example and in others, that the detection difficulty of local crimes could have changed contemporaneously with the response station relocation. Therefore, we will be careful in Section 4 to isolate the cross-sectional variation in distance to the response station (on which our empirical strategy is based) from the potentially endogenous time variation caused by station relocations.

Response Time and the Technology of Crime Detection Upon arrival, response officers interrogate the victim and/or caller, question potential witnesses, undertake a preliminary investigation and report back to the radio operators. The importance of the crime will obviously determine the amount of resources the GMP devotes to its investigation. There are several mechanisms through which faster response times could translate into a higher likelihood of detection. If the crime is still ongoing, the officers could obviously apprehend the offender before he manages to flee. Even if the criminal has left the scene, a faster response may help coordinate the search for him before he has managed to get too far[6]. Secondly, the evidence could be improved when it is gathered more quickly. Physical evidence deteriorates over time, especially when located outdoors, so collecting it earlier will potentially improve its quality. Perhaps more importantly, when arriving more promptly, responding officers may be more likely to find witnesses to the crime and to interrogate them before their recollections worsen. A faster response also provides a strong signal to the victim and witnesses that the police is both competent and likely to take the offense seriously, which could improve their willingness to cooperate in the investigation.

[6] A senior leader associated with a different police organisation confided to us that street robberies and assaults are much more likely to result in an arrest if the responding officers do a 'drive-around'. A 'drive-around' consists of obtaining the description of the assailant and circling the vicinity of the crime scene in a police car looking for individuals matching the description.

3 Data and Descriptive Evidence

Our dataset contains every 999 call alerting the police to an ongoing or past crime in the period between April 2008 and August 2014. For every call we observe among other things the location of the incident, the police response time, the UK Home Office crime classification code and an indicator of whether the crime was detected[7]. We also obtained from the GMP the locations of the police stations where the response teams were based during our sample period, for each of the 11 divisions.

Summary Statistics Table 1 provides basic summary statistics for the main variables in our study. Around 32% of the crimes were detected, although as we show in Figure 4 this percentage varies considerably by Home Office classification code or by grade level. We can also see that the response time distribution is highly skewed, with a mean of more than five hours and a median of just 31 minutes. This skewness is confirmed in Figure 3, where we can see that the density of the response time distribution peaks at around 5 minutes and falls concavely after that.

Around 20% of calls are allocated a Grade 1 priority level, and the corresponding numbers for Grades 2 and 3 are 45% and 34%. Theft offences represent around half of all crimes (52%), followed by violent offences (20%) and criminal damage (15%).

OLS Estimates Table 2 displays linear probability models of detection on (the log of) response time, accounting for an increasingly richer set of controls. Our most exhaustive specification in Column 5 controls for the hour of day in which the call was received, the division where the crime occurred, the grade level assigned by the call handler, and the Home Office-classified crime type. We find that faster response times are associated with a higher likelihood of detection. The estimated effects are, however, not large: a 10 percent decrease in response time is associated with a .29 percentage points increase in the detection likelihood.

For several reasons we need to be cautious in giving the OLS estimates a causal interpretation.

[7]Its official definition is as follows: 'A sanctioned detection occurs when (1) a notifiable offence (crime) has been committed and recorded; (2) a suspect has been identified and is aware of the detection; (3) the Crime Prosecution Service evidential test is satisfied; (4) the victim has been informed that the offence has been detected, and; (5) the suspect has been charged, reported for summons, or cautioned, been issued with a penalty notice for disorder or the offence has been taken into consideration when an offender is sentenced.' See http://data.london.gov.uk/dataset/percentage-detected-and-sanctioned-offences-borough.

Firstly, note that the information recorded by the handler during the call will determine the priority it receives, and likely be correlated both with unobserved characteristics of the crime and with the difficulty of detection. Secondly, response time may be directly affected by the estimated likelihood of detection, in ways that are difficult to pin down. For instance, response times may be particularly slow when a burglar is reported to have left the scene long ago (low likelihood of detection), but also when a shoplifter has been detained by a security guard (high likelihood), since in both cases the marginal effect of faster response may be evaluated to be low. Lastly, response times may be correlated with other policing inputs, such as the ability and attention of the responding officer. On the one hand, officers who are less competent in other dimensions of the investigation may take longer to arrive at the scene. Alternatively, it may be officers those who are aware of their low ability in other dimensions that put more effort into responding quickly. It is difficult to draw conclusions regarding the causal relation between response time and detection without a credible source of exogenous variation in the former.

4 Empirical Strategy

In this section we explain in some detail the construction of our instrument for response time. We first describe our measure of distance, discuss its potential and explore its empirical variation. We then explain why, after controlling for a large number of small geographical cells, distance to the division response station could be regarded as a valid instrument for response time. We also discuss potential threats to the exclusion restriction. In the latter part of this section, we perform tests of exogeneity of our instrument and interpret the corresponding results.

Distance and its Variation As we argued in Section 2, a call for service from a Greater Manchester location needs to be responded by the division response team, i.e. the team in the GMP division to which that location belongs. Our measure of distance is the geodesic, or 'as the crow flies', distance between the latitude and longitude of the crime scene and the latitude and longitude of the closest division response station. Table 1 shows that this measure is skewed to the right, with a mean of 3.2km. and a much lower median of 2.3km. We find in Figure 5 that the relation between response time and distance is strongly positive and approximately linear. This finding provides a validation of our claim in Section 2 that response officers often depart from the division response station when called to attend an incident.

Distance has been used as a source of exogenous variation by Reinikka and Svensson (2005), Dittmar (2011), Dube et al. (2013) and Campante et al. (2014), among others. Its validity as an instrument will obviously depend on the specific setting that is being studied. In our setting caution is warranted, since crimes in different areas may differ in terms of their detection difficulty. This could be *organically*. For instance, it may be that response stations tend to be located in city centres, and that crimes in these areas are easier or more difficult to detect than crimes in suburban areas. The correlation between distance to the response station and detection difficulty could also be *by design*. In particular, it may be that police agencies choose to locate their response stations in high-crime, and high-difficulty-crime, areas, so that they can minimise response times for crimes in these areas.

Intuition of the Instrument Locations that neighbour each other but are on different sides of division boundaries are the responsibility of different response teams departing from stations that will typically be located at different respective distances[8]. Our instrument is based on the notion that, if we control with sufficient precision for the 'local area' where a crime occurs, the remaining variation in distance is mostly due to crime locations falling on different sides of division boundaries and can therefore provide a source of exogenous variation in response time.

In Figure 6, we clarify this intuition by displaying two geographical cells crossing the boundary of two divisions. As we can see, the two marked locations in Cell 1 differ significantly in the distance to their respective stations. And yet, these locations are next to each other, and crimes occurring in them should have the same average detection difficulty.

The grid that we select for our baseline specifications covers the entire Greater Manchester Area and it contains 5,152 unique cells of .005 by .005 decimal degrees (approximately .185 squared kilometres or 556 metres by 332 metres)[9]. On average, around 504 people reside in

[8]To investigate compliance with this rule, we merged the location of all the crimes in our dataset with GMP-supplied shape files detailing the division boundaries. We then compared the division of the team that responded to a call with the division that, according to our shape files, should have officially responded. We found that they coincide in 99.76% of the crimes in our dataset, indicating that the official assignment rule is followed in the vast majority of cases. The remaining .24% of cases could have been associated with deviations from the official rule or, alternatively, may have been affected by measurement error in either the location variables or the division variable.

[9]At the latitude of Greater Manchester, a movement of .005 decimal degrees towards the equator represents a higher number of metres than an equivalent movement in the direction of the Greenwich meridian,

each cell[10]. Figure 6 superimposes two realistically-sized cells on the map of Manchester and illustrates that only a handful of streets fit into a cell.

Accounting for Division Effects One reason that the two locations in Figure 6 differ is, of course, that they belong to different divisions. This fact may be of concern, for two reasons. Firstly, it may be that response teams in some divisions are simply better than others at detecting crime, and that this is correlated with the average distance to the station across divisions. For example, the response team of the (small) North Division may be differently effective at detecting crime than its counterpart in the (large) Wigan Division (see Figure 1). Secondly, division boundaries often coincide with political boundaries and this could have an independent effect on the difficulty of detection if the populations in different local authorities are affected by different types of crimes. For these two reasons, we need to ensure that any estimated effect of response time does not include the 'division effects' resulting from divisions (of potentially different size) having different types of crime or levels of policing efficiency.

Figure 6 illustrates that these division effects can be separately identified from the response time effect. Because the two marked locations in Division F vary in the distance to their response station, the addition of geographical cell indicators and division indicators does not exhaust all the sample variation in distance. Our empirical specifications will therefore always introduce a full set of division indicators.

Estimating Equations We use a 2SLS approach to estimate the effect of response time on detection. The first stage equation between (the log of) response time and (the log of) distance for crime i occurring in year $t(i)$ in a location belonging to cell $j(i)$ and division $d(i)$ is:

$$Response_i = \alpha_0 + \alpha_1 Distance_i + Cell_{j(i)} \times Year_{t(i)} + Division_{d(i)} + \mathbf{X}_i + \epsilon_i \qquad (1)$$

hence the rectangular shape of our cells. We chose the number of cells with the following procedure. We first constructed a grid consisting of cells of .0005 squared degrees. We then tested whether, controlling for these cell indicators, distance was still a strong predictor of response time. If the answer was yes, we divided the cells into 2^2 sub-cells, and estimated our first stage regression with the new set of cell indicators. We continued until reaching the highest possible number of cells, consistently with a strong first stage that could be used to vary response time significantly. Our findings are robust to using a smaller or larger number of cells, although the strength of the instrument obviously decreases if we increase the number of cells beyond 5,152.

[10] For comparison, US census block groups have an average of 1,400 inhabitants.

where $\mathbf{X}_i$ is a vector of controls such as hour of day, grade level and Home Office classification code indicators.

Note that the cell indicators are interacted with a set of year indicators. The reason for this is as follows. Remember from Section 2 that the location of the response stations changed across our sample period for some GMP divisions. These changes create time variation in distance to the division station for a fixed location, even after controlling for the cell indicators. For the reasons discussed in Section 2 this time variation is potentially correlated with changes in crime patterns. We can, however, separate this time variation from our preferred cross-sectional variation based on division boundaries by interacting the cell indicators with a set of year indicators. After doing that, any variation in distance is mostly due to crimes in the same cell/year falling on different sides of division boundaries and therefore at different distances of their (fixed within a cell/year) response stations.

The second stage equation is:

$$Detected_i = \beta_0 + \beta_1 \hat{Response}_i + Cell_{j(i)} \times Year_{t(i)} + Division_{d(i)} + \mathbf{X}_i + \upsilon_i \qquad (2)$$

where $Detected_i$ is a dummy variable that takes value one if the crime was detected and $\hat{Response}_i$ captures the fitted values from (1).

In Appendix B we compare our empirical strategy with the strategies in other papers such as Black (1999) and Doyle et al. (2015) that use boundary discontinuities for identification. In particular, we comment on the differences in approach and explain why our current strategy is well suited to the question and institutional setting of this paper.

Threats to Identification This empirical approach is subject to five main concerns. The first is the possibility that the geographical cells may not be small enough, in which case the assumption of homogeneity of locations within a cell will not be satisfied.

The second potential concern is the possibility of household sorting. For example, households concerned about crime may decide to locate themselves on the side of the division boundary with the lowest police response times, in the same way that educationally committed families have been shown to congregate in the catchment areas of good schools (Black 1999, Bayer et al. 2007).

The third concern is due to potential sorting by criminals. If sophisticated criminals target locations on the higher-distance side of a border, these locations will be associated with more crimes, and with crimes that are more difficult to solve, posing a threat to identification.

The fourth concern is more subtle, and it has to do with the mechanical effect that response times and the associated likelihood of arrest and imprisonment have on the composition of the criminal population. Namely, it may be that lower-distance locations are depleted of some of their local criminals over time, and that the remaining criminals commit crimes of higher or lower detection difficulty[11].

The fifth concern is that a lower response time may affect not only the likelihood of detection but also the nature of the crime itself. For instance, an immediate response may prevent an attempted murder from becoming a murder or an assault from turning into an aggravated assault. If different crime classifications require different standards for detection, a faster response time may be affecting detection indirectly (through its effect on the type of crime itself) as well as directly.

In the remainder of this section, we undertake three separate balancing tests to evaluate the empirical relevance of the concerns outlined above.

Balancing Test 1: Household Demographics Our first test examines the first two concerns outlined above. In particular, we want to examine empirically whether, controlling for the cell and division indicators that are at the core of our empirical strategy, there is any evidence that households located at different distances of their respective stations differ in their demographic characteristics. To do this, we create a dataset of the 8,683 Greater Manchester output areas, the smallest geographical areas in the 2011 UK census. Using the latitude and longitude of every output area geographical centre, we assign it to a division, and compute the distance to the 2014 closest response station. We also assign each output area to a geographical cell. We then regress distance on a set of demographic characteristics, controlling for the cell and division indicators. The estimated coefficients and confidence intervals can be found in Figure 7.

To illustrate the value of our empirical strategy, we also display the equivalent estimates in a regression omitting the baseline set of cell and division indicators. We find in Figure 7 that several demographic variables appear to be statistically significant predictors of distance to

[11] This is probably not a big concern, for two reasons. Firstly, it relies on criminals being extremely consistent in their location decisions, for instance by always committing their crimes at home. If instead they cross division boundaries with a positive likelihood, the differential effect on the composition of the local criminal populations on opposite sides of the boundaries will be milder. Secondly, the United Kingdom has relatively low incarceration rates, at least by U.S. standards. Differential incapacitation of criminals (and their associated effects on the composition of the local criminal populations) are therefore likely to be small.

the respective station at the output area level, when cell and division controls are not included in the regression. For example, households living further away from response stations are older, more likely to have children, and less likely to have no qualifications.

Controlling for cell and division indicators has two effects. Firstly, it dramatically reduces the residual variance in the dependent variable (the adjusted R-squared jumps from .06 to .98), leading to much lower standard errors. Secondly, we observe that the estimated coefficients are now indistinguishable from zero (the F-statistic of joint significance of the demographic variables decreases from 20.4 to 1.3).

We interpret the evidence in Figure 7 as indicating that different sides of a division border (but within a geographical cell) contain households of similar demographics. This implies that the geographical cells that we use are sufficiently small to ensure that locations within each cell are identical in their observables, and therefore most likely in their unobservables. It also indicates that the possibility of household sorting across division borders is unlikely.

Balancing Test 2: Total Number of Crimes Our second test evaluates jointly the empirical relevance of all five threats to identification. Every one of these hypotheses predicts that the level of crime will be correlated with distance to the response station, even after controlling for the cell and division indicators. To illustrate, consider the possibility of sorting by criminals. If some criminals are sophisticated and target locations with slow response times (including the high-distance side of division boundaries), then we should observe that locations further away from the station have more crime, both across and within geographical cells.

To study whether this is an empirically relevant issue, we create a panel dataset of census output areas and years, and compute the total number of crimes in each output area and year combination. Again, we assign each output area/year to a division and to a geographical cell. We then calculate the distance between the centre of each area and the closest response station in that year. In Table 3, we regress one on the other, with and without controlling for the cell/year and division indicators.

Column 1 shows that areas further away from the response station are associated with less crime (the elasticity is -10% and strongly significant). Note that this finding is inconsistent with the notion that criminals target areas with slower response times. The idea that criminals are sophisticated in their location decisions seems therefore to be contradicted by the evidence. On the other hand, a negative elasticity is consistent with the notion that the GMP choose to locate their response stations in high-crime areas[12].

Importantly for the purposes of evaluating the validity of our empirical strategy, note that the estimated elasticity decreases dramatically and becomes statistically insignificant after we control for the cell/year and division indicators. We therefore interpret the evidence in Table 3 as indicating that the five threats to identification discussed above do not seem empirically relevant.

Balancing Test 3: Crime Characteristics All five threats to identification predict that, conditional on a crime occurring, the *type* of crime should be correlated with distance to the response station, even after controlling for the cell indicators. For instance, if sorting by criminals is an empirically relevant issue, we would expect it to be more prevalent among property crimes such as thefts than among violent crimes. This is because thieves are widely regarded to be more sophisticated than violent criminals. Therefore, sorting by criminals predicts that crimes occurring further away from the response station should include a higher proportion of thefts, relative to violent offences.

To examine whether this is the case, we estimate the following empirical model on our baseline dataset:

$$Distance_i = \pi_0 + Cell_{j(i)} \times Year_{t(i)} + Division_{d(i)} + \mathbf{X}_i + \epsilon_i \tag{3}$$

where $\mathbf{X}_i$ is our vector of interest, as it includes crime characteristics such as the UK Home Office crime classification code and the grade level which are strongly correlated with the likelihood of detection.

Figure 8 displays the coefficients and confidence intervals resulting from the estimation of (3), again with and without cell/year and division controls. As we can see, the classification code and grade level dummies are correlated with distance in the unconditional regression. It is interesting to note, however, that the correlations seem inconsistent with the notion that sophisticated criminals sort themselves away from the response station. In particular, theft offences are *less* numerous in locations further away from the response station, relative to violent offences.

[12] An important difference between criminals and police agencies is that the latter have much better information on which to base their decisions. For instance, they observe the locations of the response stations and can experiment with them. They also have access to a large set of experience and hard data regarding the relation between response time, distance and detection. Not even the most diligent criminals can match that level of knowledge. We would therefore expect police agencies to be more sophisticated than criminals in their location decisions.

Controlling for the cell/year indicators has the same two effects as in Figure 7. Firstly, the confidence intervals narrow significantly as a result of the decrease in the residual variance of the dependent variable (the adjusted R-squared of the regression jumps from .35 to .98). Secondly, the estimated coefficients become essentially zero. Despite the much narrower confidence intervals, only one classification code dummy is statistically different from zero at the 5% level, and even in this case the coefficient is much smaller than in the regression without the cell/year indicators[13]. We interpret the evidence in Figure 8 as supporting the identification strategy in this paper[14].

5 Results

In this section we present and discuss the main results of the paper.

Naive IV Estimates Before displaying our baseline estimates, we show in Table 4 'naive IV' coefficients, based on regressions that use distance to the response station as an instrument but do not control for the cell/year indicators. We label these estimates naive based on our findings from Figures 7 and 8 and Table 3 that distance is unlikely to be (unconditionally) orthogonal to the difficulty of detecting a crime. The reduced form estimate in Column 1 suggests that crimes occurring 10% further away from the response station are .37 percentage points less likely to be solved. In the second column we find the first stage estimate. Reassuringly, the correlation between distance and response time that we first identified in Figure 5 is robust to the inclusion of time, division and crime characteristics indicators. The second stage coefficient is -.223, much larger than the OLS estimates.

Baseline Estimates We display the main results of the paper in Table 5. A comparison with Table 4 shows that the reduced form estimate almost doubles in size, from -.037 to -.061, when we introduce the cell/year controls. The first stage estimate is very strong (the Kleibergen-Papp F statistic for weak identification is 45.07), and it suggests that a 10% increase in distance to the response station is on average associated with a 1.36% increase in response time. The second stage coefficient is -.451, approximately twice the size of its

[13]By any benchmark this estimated coefficient is minute in economic terms. It indicates that public order crimes are on average located .38% closer to the response station than the omitted group (violent offences).

[14]The F-statistic of a test of joint significance of the classification code and grade level dummies also decreases dramatically from 18 to 2.5, although given that the regression is based on almost half a million observations, the F-test still rejects the null hypothesis that the coefficients are jointly zero at the 5% level.

naive IV counterpart, and ten times larger than the OLS estimate.

The most natural interpretation for the large gap between the OLS and the IV estimates is that the bias associated with the OLS estimates is positive. In our setting, this implies that crimes that are inherently more difficult to detect are given higher priority by the GMP, and therefore benefit from faster response times[15]. This priority could also take the form of response stations being endogenously located in high-difficulty crime areas. This would be consistent with the evidence from Column 1 in Table 3 that the crime levels are higher in areas close to the response station.

An alternative explanation for the difference in the OLS and IV estimates is that the IV estimates are LATE, and capture the effect of response time on detection for the compliant sub-population. In our setting the compliant crimes are those that would have received a very fast response, had it not been for the fact that they are located far away from the station. If response times are allocated optimally, these 'constrained' crimes should be precisely the ones for which the marginal return to a fast response is higher. Hence, the large IV estimate may be the result of the fact that the LATE compliant crimes are those with the largest marginal return to a fast response.

Robustness to the Sequential Introduction of Controls Given that our instrument is supposed to induce exogenous variation in response time, the estimated coefficients should not be sensitive to the inclusion or exclusion of other control variables. In Tables 6 and 7 we examine whether this is the case, respectively for the reduced form and second stage estimates. We introduce sequentially the hour of the day in which the police was alerted to the crime, the call handler's grade level and the Home Office classification code. The estimates appear remarkably robust to the set of controls in the regression. This is consistent with our finding in Figure 8 that distance is (conditionally) uncorrelated with these controls. We interpret Tables 6 and 7 as providing reassurance with regards to our baseline estimates in Table 5.

Robustness to Using Only Variation from Boundary Cells Our second specification check is more subtle. The need for it is perhaps best understood by looking at Figure 9. Our identification strategy exploits variation in distance controlling for a large set of very small

[15] Our discussions with police officers in the GMP revealed that there is no career penalty for being associated with unsolved crimes, at least at the officer level. This would be consistent with officers not shying away from responding to seemingly difficult-to-solve crimes.

geographical cells. A large part of this variation is coming, as we have argued throughout this paper, from cells that cross division boundaries, such as Cell 1 in Figure 9. However, some variation in distance is also due to locations that are slightly apart from each other and in the same cell and division (i.e. the two locations pictured in Cell 3). Arguably, this last type of variation is very small in magnitude, given that the cells are very small. Nevertheless, we want to test the robustness of our baseline estimates to a specification that strictly uses only variation in distance across locations in 'boundary cells'.

This can be done in a 2SLS framework simply by interacting distance and response time with a 'boundary cell' dummy. In Table 8 we do this and find that the coefficients are similar to our baseline estimates. The reduced form estimate is -.04 and the first stage coefficient is almost identical at .141. The second stage estimate is large in absolute terms and statistically significant at the 5% level[16]. We interpret this evidence as reinforcing the main conclusion of the paper.

Heterogeneity We now examine how the effects of response time on detection differ across types of crime. To do this, we create sets of dummies that split the population of crimes along several dimensions and interact these dummies with response time in (2) and with distance in (1)[17]. While the exploration of heterogeneous effects is interesting per se, it can also be regarded as a robustness test of our main finding in Table 5. Arguably, if our estimated effects vary across crimes in ways that we have difficulty in understanding, we may want to revisit the confidence in our baseline estimate.

We start by examining how the estimated coefficients vary across grade level. Remember

[16] The second stage estimate is, however, different from its baseline regression counterpart in Table 5 (-.286 versus -.451). There are two potential explanations for this. Firstly, note that the strength of the instrument in Table 8 is relatively low (the Kleibergen-Papp F-statistic is 9.828), as it uses only a subset of the variation in distance. Therefore, this coefficient may be slightly biased. The second potential explanation is based on LATE, i.e. boundary and single-division cells are different in nature. For instance, the average distance to the response station for boundary cells is much higher than for single-division cells, a mechanical result of being in the outskirts of the division to which they belong. While our use of distance and response time in logarithmic form certainly allows for non-linearities in their effect on detection, it may be that the effect of response times for boundary cells is still different (in percentage terms) from the effect for single-division cells.

[17] Alternatively, we could run (1) and (2) separately for every type of crime. Due to the large number of control indicators and to the fact that variation in distance to the response station is based on a relatively small number of observations, this strategy leads to very weak instruments. We have therefore decided to impose the assumption that the control variables do not have a differential effect by type of crime. Every result in this section must be interpreted with this caveat in mind. Note also that, to preserve the strength of the instruments, we have divided the population of crimes into only three groups at a time. The three panels in Table 9 should therefore not be interpreted as the outcomes of independent tests.

from Section 2 that Grade 1 calls are typically those where there is an imminent threat to someone's safety or where a crime is taking place. Grade 2 calls are those where there is no imminent threat but key evidence or witnesses are likely to be lost if attendance is delayed (Police and Crime Commissioner, 2013). Because Grade 2 calls are explicitly defined as those for which a fast response is likely to matter most (in terms of crime detection), we would expect Grade 2 calls to be associated with a higher coefficient. This is indeed what we find in Panel A of Table 9. As we can see, Grade 2 calls have an estimated coefficient of -.542, which is slightly larger than the -.433 estimate for Grade 1 calls and much larger than the -.264 estimate for Grade 3 calls.

In Panel B we explore the heterogeneity of the coefficients in terms of the time delay between the crime occurring and the police being alerted. This exercise is motivated by Pate et al. (1976) and Spellman and Brown (1981), who argue that, because it takes a long time for the victims to call the police, any improvement in police response times is unlikely to make a difference.

We define the call delay as the number of minutes between the end point of the crime and the police receiving the call[18]. The first thing we notice is that, contrary to the findings in the aforementioned papers, the typical call delay does not seem to be excessively long. In our dataset, this variable has a median of around 10 minutes (see Table 1) and the 33th and 66th centiles are 3 and 27 minutes respectively. Furthermore, it takes negative values for approximately 10% of observations, indicating that the police was alerted even before the crime stopped taking place.

While it does not seem as if the call delay is typically high enough to preclude any effect from a faster response time, we agree with Pate et al. (1976) and Spellman and Brown (1981) that the estimates in our paper should be larger when the call delay is shorter. In Panel B of Table 9 we split the sample into thirds and study this question. We indeed find that police response time has a weaker effect on detection if the call delay is higher than 27 minutes, relative to crimes where the call delay is shorter.

In Panel C we split the sample by UK Home Office Classification Code and find that fast response times matter for every type of crime. While the estimates are larger for theft offences, they are still sizable for the relatively more serious violent offences.

[18]The GMP gathers information on the estimated start time and end time of every crime, with the purpose of aiding the crime investigation. These variables are probably subject to some measurement error, since they are typically based on the estimates and/or recollections of victims and/or witnesses.

To summarise, we interpret the findings in Table 9 as largely consistent with our expectations in terms of how the technology of crime detection works.

Effect on the Intensive Margin: Time to Detection One way to interpret the evidence above is that response times have an effect on the *extensive margin* of the police production function, since faster response times can turn potentially undetected crimes into detected crimes. We now proceed to examine whether response times also have an effect on the detection *intensive margin*, in terms of reducing the time that it takes to detect crimes. To do this, we restrict the sample to crimes that were detected by August 2014, and use our baseline empirical specification to study whether faster response times are associated with faster detection times[19].

The second column of Table 10 shows that the estimated relation between response time and distance is very similar in the subsample of detected crimes relative to that in the overall sample (13.3% versus 13.6%). The smaller size of the sample implies, however, that the instrument is much weaker than in the baseline regressions of Table 5 (Kleibergen-Papp F-statistic of 10.35).

The reduced form estimate in Column 1 and the second stage estimate in Column 3 both indicate that a crime will be solved more quickly (conditional on it being solved at all) if the police reach the crime scene more promptly after being alerted. The estimated elasticity in Column 3 is economically large: a 10% increase in response time will lead to a 6.85% increase in the time that it takes to detect a crime (in addition to the possibility that it may never be detected at all). If 'justice delayed is justice denied', we should regard the extra time to detection as an additional detrimental effect of reaching a crime scene too late.

Mechanism In Section 2 we discussed potential mechanisms through which fast response times could make a difference. To reiterate, we mentioned that the police could arrest the offender either at the scene of the crime or in its vicinity; they could collect physical evidence before it is contaminated or destroyed; they could interrogate witnesses before they have left the scene and they could encourage victim or witness cooperation by signaling efficiency and dedication. We now explore empirically the importance of the last two mechanisms: victim or witness cooperation in terms of naming a suspect. In our dataset, we have information

[19] Because detection of a crime typically takes place within a relatively short horizon (i.e. a maximum of a couple of months), we have decided to ignore issues of censoring. Similar estimates are found if we drop the year 2014.

on whether a suspect was named to the police[20]. In Table 1 we find that this occurs for 21% of crimes. Unsurprisingly, the likelihood of detection is much higher when there is a named suspect (63%) than when there is not (23%).

Our hypothesis is that a suspect will be named relatively more often when the police is faster in attending the scene. To examine this, we display in Table 11 a version of (2) where the dependent variable is a dummy indicating whether a suspect was named to the police (the first stage equation is unchanged). We indeed find that faster response times lead to a higher likelihood of this type of cooperation by a victim or witness. We conclude that this is indeed a mechanism through which faster response times have an effect on crime detection.

6 Discussion

In this paper we have provided robust evidence of a causal effect of police response time on crime detection. The estimated effects are large and strongly significant. They hold on the extensive margin (likelihood of detection) as well as on the intensive margin (time to detection). We find stronger effects for thefts than for violent offenses, although the effects are large for every type of crime. Reassuringly, the effects are larger for incident calls that are given higher priority by the GMP on the grounds that the likelihood of detection may decrease if response time is slow (i.e. Grade 2 crimes). We also find that the effects are larger when the victims or witnesses take less time to inform the police. Lastly, we provide some evidence on one of the mechanisms through which police response time operates: the likelihood that a victim or witness will name a suspect to the police.

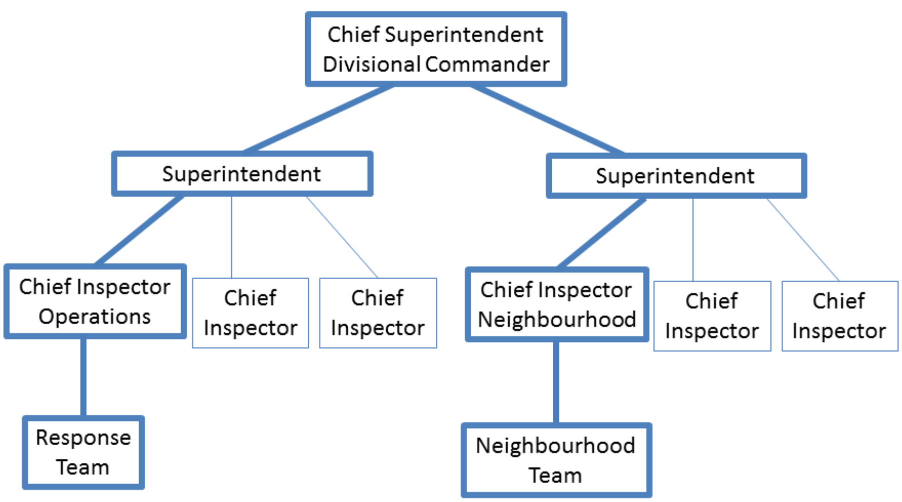
Our findings contradict long-standing consensus among criminologists regarding the effectiveness of rapid response policing. As Sherman (1997) indicates, one reason for the failure to question that consensus with new work is that field experiments on response policing are both expensive and ethically challenging. This paper illustrates the potential for using natural variation in policing inputs to study the effectiveness of policing practices, in settings where experimental variation is unfeasible.

[20]Unfortunately, we do not observe who named such a suspect or whether the suspect was confirmed as the person responsible for the crime. We do not observe either the variables necessary to explore the other potential mechanisms, such as the quality of the physical evidence gathered or the accuracy of witness recollections.

FIGURES

Figure 1: Map of Manchester Divisions

Figure 2: Typical Division Organisational Structure



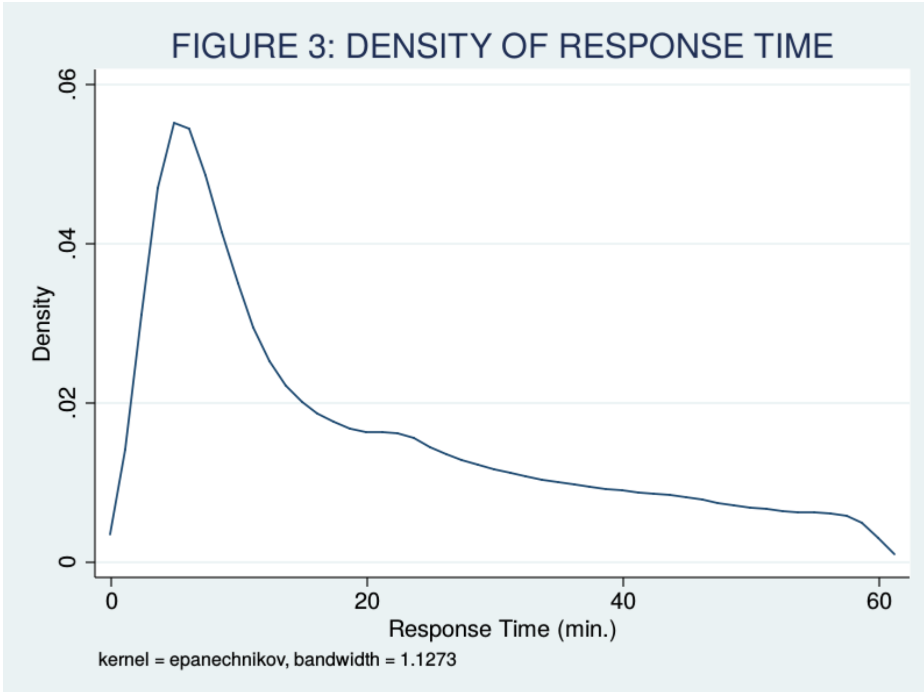


FIGURE 3: DENSITY OF RESPONSE TIME
Density
0
.02
.04
.06
0
20
40
60
Response Time (min.)
kernel = epanechnikov, bandwidth = 1.1273

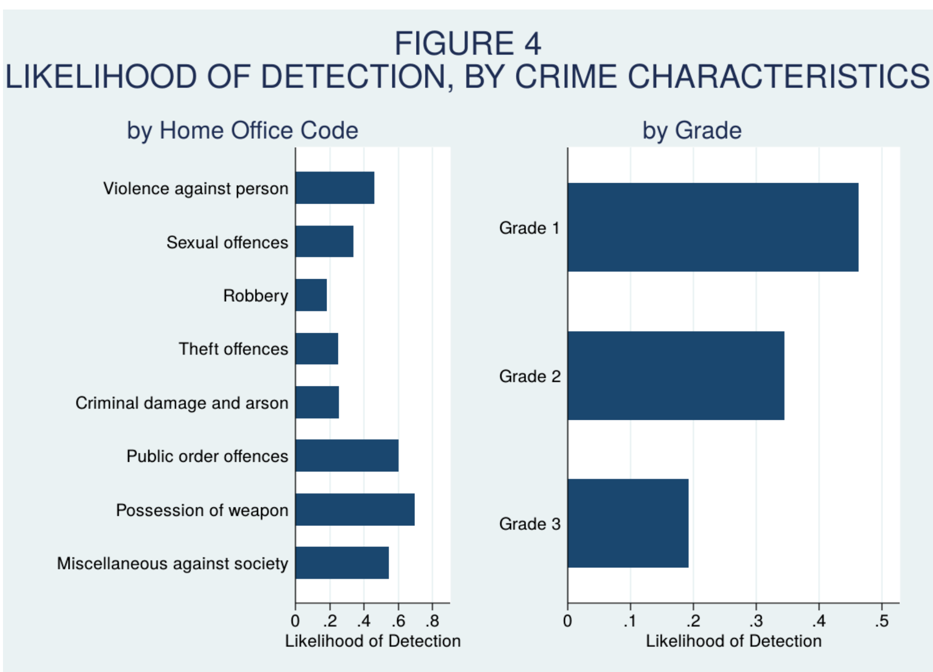


FIGURE 4
LIKELIHOOD OF DETECTION, BY CRIME CHARACTERISTICS
by Home Office Code
Violence against person
Sexual offences
Robbery
Theft offences
Criminal damage and arson
Public order offences
Possession of weapon
Miscellaneous against society
0 .2 .4 .6 .8
Likelihood of Detection
by Grade
Grade 1
Grade 2
Grade 3
0 .1 .2 .3 .4 .5
Likelihood of Detection

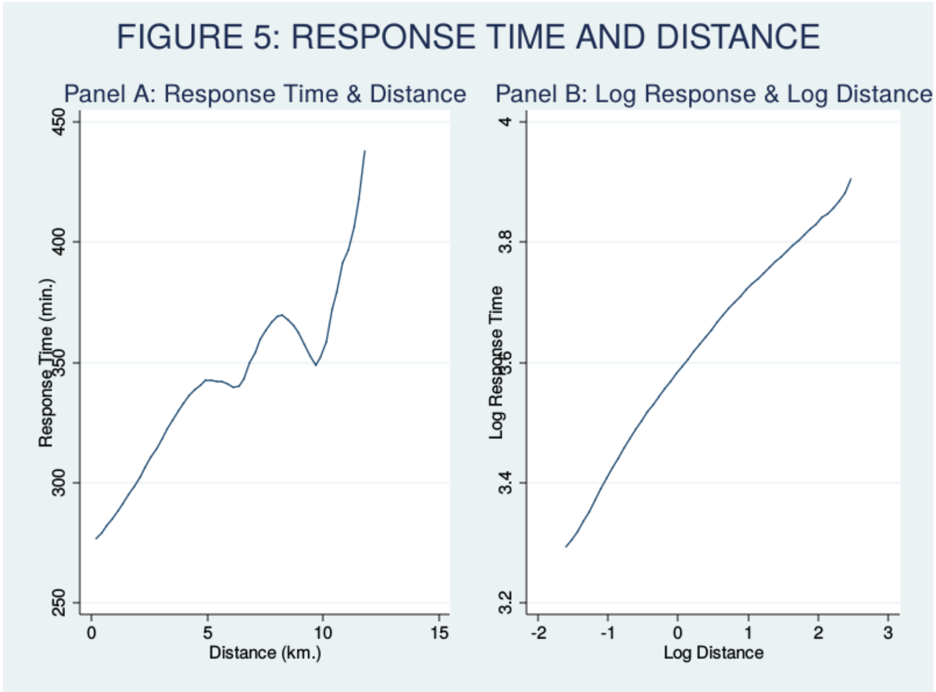


FIGURE 5: RESPONSE TIME AND DISTANCE

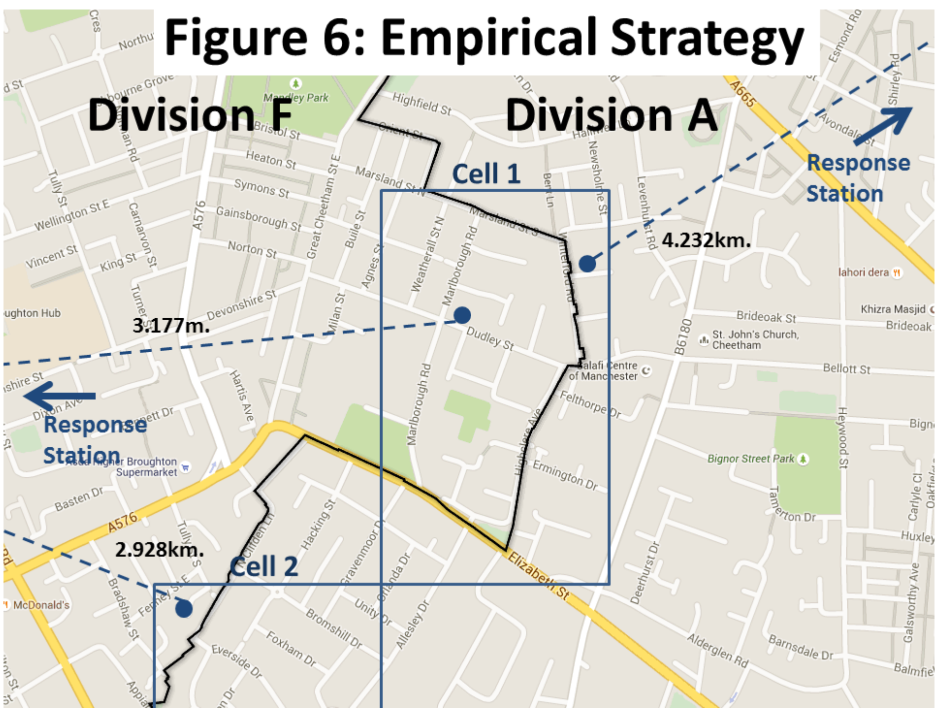
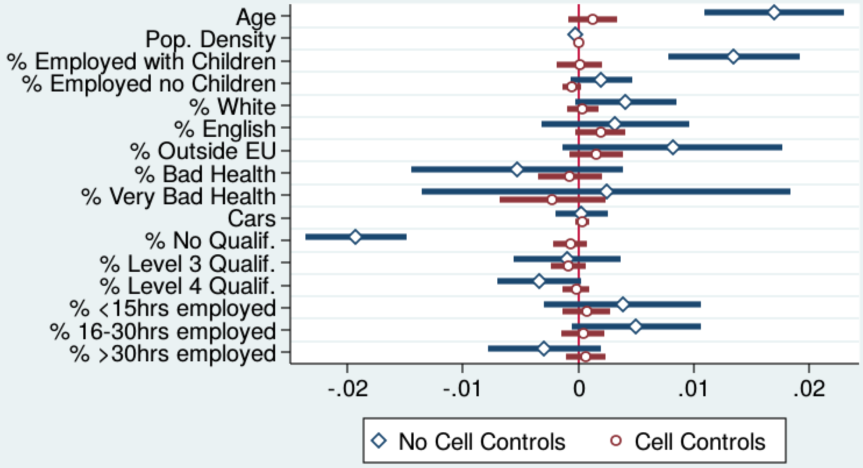


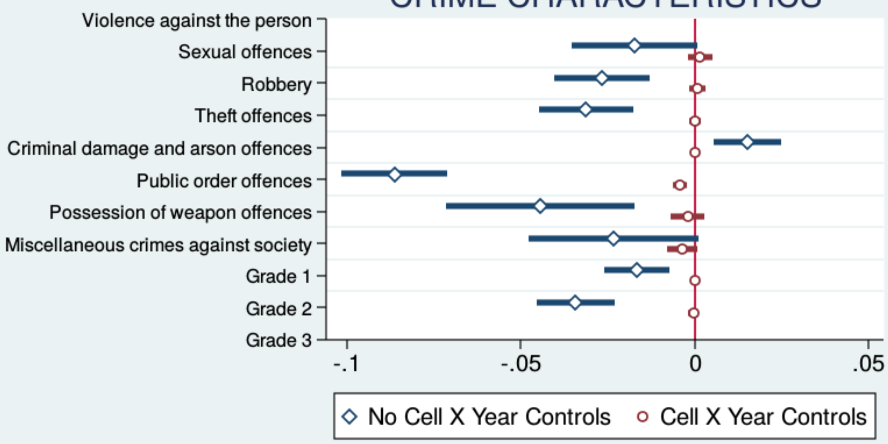
Figure 6: Empirical Strategy

FIGURE 7: BALANCING TEST 1
HOUSEHOLD DEMOGRAPHICS



This Figure displays the estimated coefficients and confidence intervals of two separate regressions. An observation is a 2011 UK census Greater Manchester output area. The number of observations is 8,683. The dependent variable is Log Distance. The independent variables are displayed in the vertical axis. Standard errors clustered at the Cell level in both regressions. The F-statistic of joint significance of the independent variables is 20.4 for the regression without cell indicators and 1.3 for the regression with cell indicators.

FIGURE 8: BALANCING TEST 3
CRIME CHARACTERISTICS



This Figure displays the estimated coefficient and confidence intervals of two separate regressions. We use the baseline dataset of crimes between April 2008 and August 2014. The number of observations is 469470. The dependent variable is Log Distance. The independent variables are displayed in the vertical axis. Violence Against the Person and Grade 3 are the Omitted Groups for the Home Office Code and Grade Categories, respectively. Both regressions control for Division and Hour of Day. Standard errors clustered at the Cell X Year. The F-statistic of joint significance of the independent variables is 18 for the regression without cell indicators and 2.5 for the regression with cell indicators.

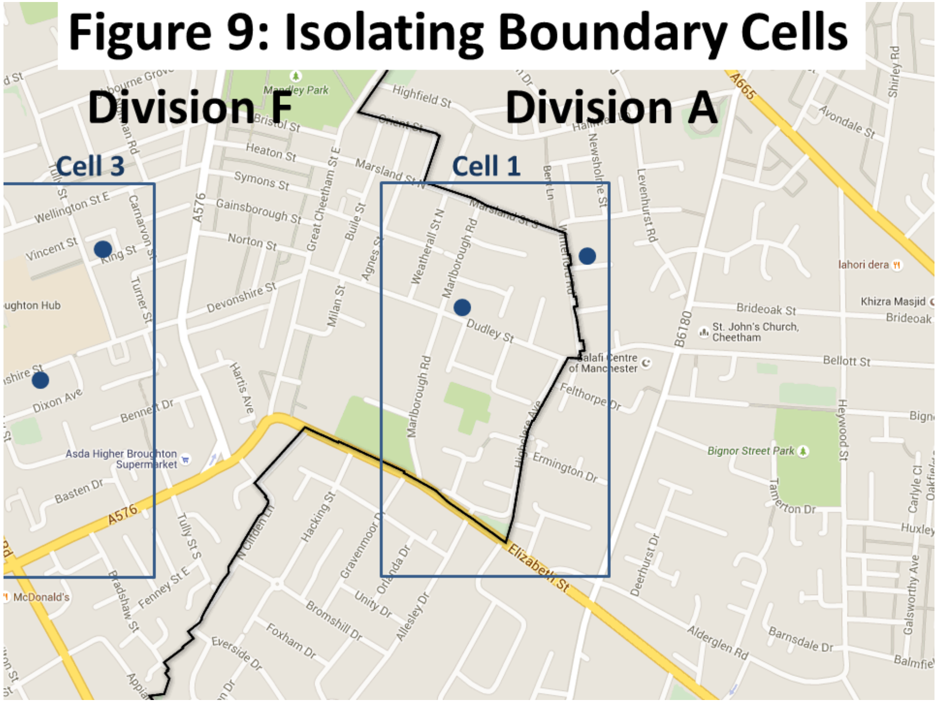


Figure 9: Isolating Boundary Cells

TABLES

Table 1: **SUMMARY STATISTICS**

Variable	Obs	Mean	Std. Dev.	Min	Max	P50
Detected	469470	.317	.465	0	1	0
Suspect Named	469470	.21	.408	0	1	0
Time to Detection (days)	143094	37.56	112.736	0	2255	5
Response Time (min.)	469470	304.777	1022.898	.05	39488	30.817
Distance to Station (km.)	469470	3.207	3.538	.004	18.999	2.286
Call Delay (min.)	468761	9407.808	322249.7	-39733.93	3.38e+07	10.039
Grade 1	469470	.207	.405	0	1	0
Grade 2	469470	.457	.498	0	1	0
Grade 3	469470	.335	.472	0	1	0
Violent Offenses	469470	.203	.402	0	1	0
Sexual Offenses	469470	.015	.122	0	1	0
Robbery	469470	.041	.199	0	1	0
Theft Offenses	469470	.52	.5	0	1	1
Criminal Damage/Arson	469470	.147	.355	0	1	0
Public Order Offenses	469470	.057	.231	0	1	0
Possession of Weapon	469470	.007	.082	0	1	0
Miscellaneous against Society	469470	.01	.101	0	1	0

TABLE 2: OLS ESTIMATES

Detected and Response Time

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Detected	Detected	Detected	Detected	Detected
Log Response Time	-0.048***	-0.051***	-0.028***	-0.029***	-0.029***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HourOfDay	No	Yes	Yes	Yes	Yes
Grade	No	No	Yes	Yes	Yes
HomeOfficeCode	No	No	No	Yes	Yes
Division	No	No	No	No	Yes
r2	0.0385	0.0527	0.0660	0.115	0.119
N	469470	469470	469470	469470	469470

Robust Standard Errors in Parentheses.

TABLE 3: BALANCING TEST 2 (TOTAL CRIMES)

Log Number of Crimes on Log Distance

VARIABLES	(1) LogCrimes	(2) LogCrimes
(Log) Distance	-0.103***	-0.002
	(0.005)	(0.006)
CellXYear	No	Yes
Division	No	Yes
r2	0.00790	0.645
N	55808	55808

An observation is a UK census output area and year combination.

Standard Errors in Parentheses Clustered by Cell X Year.

TABLE 4: NAIVE IV ESTIMATES

Instrument = Distance (Without Cell X Year Indicators)

VARIABLES	(1) ReducedForm Detected	(2) FirstStage Log Response Time	(3) SecondStage Detected
Log Distance	-0.037***	0.168***	
	(0.001)	(0.003)	
Log Response Time			-0.223***
			(0.006)
N	469470	469470	469470
widstat			3085

All Regressions Control for HourOfDay, Year, Division, Grade and Home Office Code.

Standard Errors clustered by Day X Hour.

widstat= Kleinbergen-Papp F statistic for weak identification

TABLE 5: BASELINE IV ESTIMATES

Instrument = Log Distance (With Cell X Year Indicators)

	(1)	(2)	(3)
	ReducedForm	FirstStage	SecondStage
VARIABLES	Detected	Log Response Time	Detected
Log Distance	-0.061***	0.136***	
	(0.011)	(0.020)	
Log Response Time			-0.451***
			(0.093)
N	469470	469470	469470
widstat			45.07

All Regressions Control for Cells X Year, HourOfDay, Division, Grade and Home Office Code.

Standard Errors Clustered by Cell X Year.

widstat= Kleinbergen-Papp F statistic for weak identification

TABLE 6: ROBUSTNESS OF REDUCED FORM REGRESSIONS

Detected on Distance (With Cell X Year Indicators)

	(1)	(2)	(3)	(4)
VARIABLES	Detected	Detected	Detected	Detected
Log Distance	-0.068***	-0.067***	-0.067***	-0.061***
	(0.011)	(0.011)	(0.010)	(0.011)
HourOfDay	No	Yes	Yes	Yes
Grade	No	No	Yes	Yes
HomeOfficeCode	No	No	No	Yes
r2	0.150	0.155	0.198	0.242
N	469470	469470	469470	469470

All regressions control for Division and Cells X Year.

Standard Errors in Parentheses Clustered by Cells X Year.

TABLE 7: ROBUSTNESS OF IV REGRESSIONS

Instrument = Distance (With Cell X Year Indicators)

VARIABLES	(1) Detected	(2) Detected	(3) Detected	(4) Detected
Log Response Time	-0.529***	-0.463***	-0.474***	-0.451***
	(0.117)	(0.086)	(0.089)	(0.093)
HourOfDay	No	Yes	Yes	Yes
Grade	No	No	Yes	Yes
HomeOfficeCode	No	No	No	Yes
N	469470	469470	469470	469470
widstat	24.33	28.71	47.12	45.07

All regressions control for Division and Cells X Year.

Standard Errors in Parentheses Clustered by Cells X Year.

widstat= Kleinbergen-Papp F statistic for weak identification

TABLE 8: IV ESTIMATES (ONLY BOUNDARY CELLS)

Instrument = Distance X Boundary (With Cell X Year Indicators)

VARIABLES	(1) ReducedForm Detected	(2) FirstStage Log Response Time	(3) SecondStage Detected
Log Distance X Boundary	-0.040**	0.141***	
	(0.016)	(0.045)	
Log Response Time			-0.286**
			(0.131)
N	469470	469470	469470
widstat			9.828

All Regressions Control for Log Distance X Single-Division Cell, Cell X Year, Hour of Day, Division, Grade and Home Office Code. Standard Errors Clustered by Cell X Year.

widstat= Kleinbergen-Papp F statistic for weak identification

TABLE 9: HETEROGENEITY OF BASELINE IV ESTIMATES

Panel A: By Grade	(1) Grade 1	(2) Grade 2	(3) Grade 3
	-.433***	-.542***	-.264***
	(.107)	(.086)	(.088)
P-value $\neq$ (1)		.021	0
P-value $\neq$ (2)			0
widstat		15.4	

Panel B: By Call Delay	(1)$\leq$ 3 min.	(2) 3-27 min.	(3)$\geq$ 27 min.
	-.511***	-.559***	-.386***
	(.103)	(.104)	(.134)
P-value $\neq$ (1)		.235	.027
P-value $\neq$ (2)			.006
widstat		11.38	

Panel C: By Code	(1) Violent	(2) Theft	(3) Other
	-.326***	-.596***	-.323***
	(.125)	(.083)	(.084)
P-value $\neq$ (1)		0	.955
P-value $\neq$ (2)			0
widstat		14.71	

Every Panel displays the estimated coefficients from a single regression. In every regression the population of crimes has been divided into three groups. The displayed coefficients are for the effect of response time on detection for the group described in the column. In every regression the three instruments are the interaction of distance with the three group dummies. All regressions control for Cells X Year, Hour of Day, Division, Grade and Home Office Code. None of the control variables are interacted with the group dummies. Standard Errors Clustered by Cell X Year. widstat= Kleibergen-Papp F statistic for weak identification.

TABLE 10: IV ESTIMATES (DEP. = TIME TO DETECTION)

Instrument = Distance (With Cell X Year Indicators)

VARIABLES	(1) ReducedForm LogTimeToDetection	(2) FirstStage Log Response Time	(3) SecondStage LogTimeToDetection
Log Distance	0.091**	0.133***	
	(0.045)	(0.041)	
Log Response Time			0.685*
			(0.380)
N	137270	137270	137270
widstat			10.35

All Regressions Control for Cells X Year, HourOfDay, Division, Grade and Home Office Code.
Sample Contains Only Detected Crimes. Standard Errors Clustered by Cell X Year.
widstat= Kleinbergen-Papp F statistic for weak identification

TABLE 11: IV ESTIMATES (DEP. = SUSPECT NAMED)

Instrument = Distance (With Cell X Year Indicators)

VARIABLES	(1) ReducedForm S.Named	(2) FirstStage Log Response Time	(3) SecondStage S.Named
Log Distance	-0.020**	0.136***	
	(0.008)	(0.020)	
Log Response Time			-0.146**
			(0.060)
N	469470	469470	469470
widstat			45.07

All Regressions Control for Cells X Year, HourOfDay, Division, Grade and Home Office Code.
Standard Errors Clustered by Cell X Year.
widstat= Kleinbergen-Papp F statistic for weak identification

REFERENCES

Adda, J., McConnell, B., and Rasul, I. (2015), "Crime and the Depenalization of Cannabis Possession: Evidence from a Policing Experiment", *Journal of Political Economy*, 122(5), 1130-1202.

Bayer, P., Ferreira, F., and McMillan, R. (2007), "A unified framework for measuring preferences for schools and neighborhoods (No. w13236)", *National Bureau of Economic Research*.

Bayley, D. H. (1996), "Police for the Future", New York: Oxford University Press.

Becker, C. (1968), "Punishment: An Economic Approach", *76 Journal of Political Economy*, 169(10.2307), 1830482169.

Benson, B. L., and Rasmussen, D. W. (1991), "Relationship between illicit drug enforcement policy and property crimes", *Contemporary Economic Policy*, 9(4), 106-115.

Black, S. E. (1999), "Do better schools matter? Parental valuation of elementary education", *Quarterly Journal of Economics*, 577-599.

Blundell, J. (2007), "Policing a liberal society", *Economic Affairs*, 27(4), 4-11.

Braga, A. A. (2001), "The effects of hot spots policing on crime", *The ANNALS of the American Academy of Political and Social Science*, 578(1), 104-125.

Bratton, W., and Knobler, P. (2009), "The turnaround: How America's top cop reversed the crime epidemic", London: Random House.

Campante, F. R. (2014), "Isolated Capital Cities, Accountability, and Corruption: Evidence from US States", *The American Economic Review*, 104(8), 2456-2481.

Clawson, C. and Chang, S. (1975), "Impact of response delay on arrest rates", Inspectional Services Division, Seattle Police Department

College of Policing (2013), "Investigation: Investigation Process", available at:
https://www.app.college.police.uk/app-content/investigations/investigation-process/

DeAngelo, G, and Owens E. G. (2014), "Learning the Ropes: Task Specific Experience and the Output of Idaho State Troopers", Working Paper.

Di Tella, R., and Schargrodsky, E. (2004), "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack", *The American Economic Review*, 94(1): 115-133.

Dittmar, J. E. (2011), "Information technology and economic change: the impact of the printing press", *The Quarterly Journal of Economics*, 126(3): 1133-1172.

Doyle, J., Graves, J., Gruber, J., and Kleiner, S. (2015), "Measuring returns to hospital care: Evidence from ambulance referral patterns", *The Journal of Political Economy*, 123(1), 170.

Draca, M., Machin, S., and Witt, R. (2011), "Panic on the streets of london: Police, crime, and the july 2005 terror attacks", *The American Economic Review*, 2157-2181.

Dube, A., Dube, O., and Garca-Ponce, O. (2013), "Cross-border spillover: US gun laws and violence in MexicoContracts", *American Political Science Review*, 107(03): 397-417.

Evans, W. N., and Owens, E. G. (2007), "COPS and Crime", *Journal of Public Economics*, 91(1): 181-201.

Garicano, L., and Heaton, P. (2010), "Information technology, organization, and productivity in the public sector: evidence from police departments", *Journal of Labor*

Economics, 28(1): 167-201.

Goldstein, H. (1990), "Problem-oriented policing", New York: McGraw-Hill.

Hess, K. and Hess Orthmann, C. (2012), "Criminal investigation", 10th edition, Clifton Park: Cengage Learning.

Hoxby, C. M. (2000), "Does competition among public schools benefit students and taxpayers? ", *American Economic Review*, 1209-1238.

Isaacs, H. (1967), "A study of communications, crimes, and arrests in a metropolitan police department", Task Force report: science and technology, 88-106.

Karn, J. (2013), "Policing and crime reduction: The evidence and its implications for practice ", Washington DC: Police Foundation.

Katz, C. M. and Walker, S. (2012), "Police in America", New York: McGraw-Hill.

Kelling, G. L., Pate, T., Dieckman, D., and Brown, C. E. (1974), "The Kansas City preventive patrol experiment: A summary report (Vol. 1015)", Washington DC: Police Foundation.

Kennedy, D. M. (1993), "The strategic management of police resources", US Department of Justice, Office of Justice Programs, National Institute of Justice.

Klick, J., and Tabarrok, A. (2005), "Using Terror Alert Levels to Estimate the Effect of Police on Crime", *Journal of Law and Economics*, 48(1): 267-279.

Levitt, S. D. (1997), "Using electoral cycles in police hiring to estimate the effect of police on crime", *The American Economic Review*,270-290.

Mastrobuoni, G. (2015), "Police and clearance rates: Evidence from recurrent redeployments within a city", working paper.

McCrary, J. (2002), "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime: Comment", *The American Economic Review*,92(4): 1236-1243.

McCrary, J. (2007), "The effect of court-ordered hiring quotas on the composition and quality of police", *The American Economic Review*, 318-353.

Miller, A. R., and Segal, C. (2014), "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence escalation. Effects on Crime Reporting and Domestic Violence Escalation", *working paper*, (September 16, 2014).

Pate, T., Ferrara, A., Bowers, R. A., and Lorence, T. (1976), "Police response time: its determinants and effects Police Foundation", Washington DC: Police Foundation.

Police and Crime Commissioner for Greater Manchester (2013), "GM Office of the Police and Crime Commissioner Briefing Paper: Operational Communications Briefing", available at
https://meetings.gmpcc.org.uk/documents/s6161/PCC%20OCB%20Briefing%20December %202013%20v0.6.pdf

Reinikka, R., and Svensson, J. (2005), "Fighting corruption to improve schooling: Evidence from a newspaper campaign in Uganda", *Journal of the European Economic Association*, 3(23): 259-267.

Sherman, L. W. (2013), "The rise of evidence-based policing: Targeting, testing, and tracking", *Crime and Justice*,42(1): 377-451.

Sherman, L. W., and Eck, J. E. (2002), "Policing for crime prevention", in *Evidence-based crime prevention* London: Routeledge.

Sherman, L. W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., and Bushway, S. (1997), "Preventing crime: What works, what doesnt, whats promising: A report to the United States Congress prepared for the National Institute of Justice ", Washington, DC.

Siegel, L., and Worrall, J. (2014), "Essentials of criminal justice", Clifton Park: Cengage Learning.

Soares, R. R., and Viveiros I. (2010), "Organization and information in the fight against crime: An evaluation of the integration of police forces in the state of minas gerais, brazil", *working paper*.

Spelman, W., and Brown, D. K. (1981), "Calling the police: Citizen reporting of serious crime", Washington DC: Police Executive Research Forum.

Visher, C. A. and Weisburd, D. (1998), "FIdentifying What Works: Recent Trends in Crime Prevention", *Crime, Law and Social Change*, 28:223-42.

Walker, S. (1994), "Sense and nonsense about crime and drugs: A policy guide", Belmont, CA: Wadsworth Publishing Company.

Weisburd, D., and Eck, J. E. (2004), "What can police do to reduce crime, disorder, and fear? ", *The Annals of the American Academy of Political and Social Science*, 593(1), 42-65.

APPENDIX A:
List of Links to Response Time Statistics

London:
http://www.met.police.uk/foi/pdfs/disclosure_2013/october_2013/2013010001787.pdf
Manchester:
http://www.gmp.police.uk/content/section.html?readforms=8F71EF5EC4028BC980257A000049B74A
Leeds:
http://www.westyorkshire.police.uk/help-advice/factfinder/contacting-police
Glasglow:
http://www.scotland.gov.uk/Publications/2013/11/6914/2
Bristol:
http://www.bristolpost.co.uk/Emergency-response-times-20/story-19533573-detail/story.html
Edinburgh:
http://www.gov.scot/Publications/2011/10/25105642/8
New York:
http://www.nyc.gov/html/911reporting/html/reports/trend.shtml
http://www.metro.us/local/police-response-time-to-emergency-calls-slows/tmWlit—64qIqw45j4J1Q/
http://www.nyc.gov/html/ops/downloads/pdf/mmr2013/nypd.pdf
Houston:
http://www.houstontx.gov/police/department_reports/goals_at_a_glance/FY144QGoal-at-a-Glance.pdf
http://performance.houstontx.gov/sites/default/files/FY15%20Q1%20Performance%20Insight_0.pdf
San Diego:
http://voiceofsandiego.org/2010/08/23/fact-check-how-quickly-do-police-arrive/
San Jose:
http://www.sanjoseca.gov/DocumentCenter/View/30154
http://www3.sanjoseca.gov/clerk/Agenda/20120918/20120918_sspres.pdf
Indianapolis:
http://ftpcontent.worldnow.com/wthr/PDF/impdreportexec.pdf
Jacksonville:
https://click2gov.ci.jacksonville.nc.us/ftp/pd/2010JPDAnnualFinalScreen.pdf
San Francisco:
http://www.sfcontroller.org/ftp/uploadedfiles/controller/wcm_controller/community_indicators
/publicsafety/policeresponsetime/policeresponsetimebigchart.htm
http://sf-police.org/Modules/ShowDocument.aspx?documentid=12928
Fort Worth:
http://www.fortworthpd.com/docmgmt/2012-Annual-Report-FINAL.pdf
http://www.fortworthpd.com/docmgmt/2011-Annual-Report-FINAL.pdf
Detroit:
http://www.detroitmi.gov/Portals/0/docs/EM/Reports/City%20of%20Detroit%20Proposal%20
for%20Creditors1.pdf
El Paso:
http://www.elpasotimes.com/ci_16861086

Memphis:
http://www.memphispolice.org/2010%20MPD%20Annual%20Report%20for%20web.pdf
Denver:
http://www.denvergov.org/Portals/741/documents/Audits%202014/Police_Response_Time_Audit_Report_06-19-14.pdf
Washington:
http://mpdc.dc.gov/sites/default/files/dc/sites/mpdc/publication/attachments/MPD%20Annual%20Report%202013_lowres.pdf
Boston:
http://www.cityofboston.gov/images_documents/Police%20-%20FY11%20Q1%20Web_ver2_tcm3-21707.pdf
Nashville:
http://swingrightrudie.blogspot.co.uk/2011/01/average-nashville-metro-police-response.htm
Portland:
http://www.portlandoregon.gov/dashboard/66479
http://www.portlandoregon.gov/dashboard/66479
Las Vegas:
http://www.jrn.com/ktnv/news/148660495.html
Milwaukee:
http://www.jsonline.com/watchdog/126286388.html
Minneapolis:
http://www.ci.minneapolis.mn.us/results/ps/policeresponse
Ventura:
http://www.cityofventura.net/pd/performancemeasures
Sydney:
https://www.police.nsw.gov.au/__data/assets/pdf_file/0009/214974/Annual_Report_Section_Three.pdf
http://www.audit.nsw.gov.au/ArticleDocuments/130/47_Police_Response_To_Calls.pdf.aspx?Embed=Y
Toronto:
http://www.torontopolice.on.ca/publications/files/brochures/2013business_plan.pdf
Montreal:
http://www.spvm.qc.ca/upload/documentations/statistiques_2010_EN.pdf
http://spvm.qc.ca/AnnualReport/2013/files/inc/f148490571.pdf
Winnipeg:
http://www.winnipeg.ca/audit/pdfs/reports/wps_cc_report.pdf
Hamilton:
http://www.hamiltonpolice.on.ca/NR/rdonlyres/3A874CD9-61E3-43A0-8C54-CBD76C9D1954/0/WorkloadStudy20011to2013_draft4_reduced.pdf

APPENDIX B: Comparison with other Empirical Strategies Using Boundary Discontinuities

Boundary discontinuities have been at the core of the empirical strategies in a number of important papers. For instance, Black (1999) uses variation at the boundaries of school districts to estimate the effect of school quality on house prices. The second empirical strategy in Doyle et al. (2015) compares patients that live close to each other but are served by different hospitals as a result of falling on different sides of ambulance dispatch area boundaries. A typical reduced form equation in these papers is:

$$y_i = \phi + z_{d(i)} + \lambda_{b(i)} + \mathbf{X}_i + \theta_i \tag{4}$$

where y_i is an outcome variable for unit i, $z_{d(i)}$ is the independent variable of interest for the district d to which unit i belongs, and $\lambda_{b(i)}$ is a 'boundary dummy' linking units that are close to each other but may fall on different sides of district boundaries.

Note that in the empirical settings of the papers above z varies exclusively at the district level. In other words, every unit in a district has the same level of z. For instance, all the households in district $d(i)$ receive the same exposure to the district school. All the patients in ambulance dispatch area $d(i)$ receive the same treatment from the local hospital in the area. The effect of z can therefore not be separately identified from other effects (if any) that may be changing discontinuously at the district borders.

Because our question and institutional setting are different from the ones in Black (1999) and Doyle et al. (2015), their empirical strategy is not well suited to our purposes. To illustrate this, consider Figure A1, where we display six hypothetical crime scenes belonging to two divisions. Remember that our main variation of interest, distance, occurs both across and within districts. Estimating equation (4) in our setting would therefore exploit variation in distance across locations (such as A1 and A3) that are far from each other, and for which the identification assumption of balance on the unobservables is unlikely to hold. On the other hand, note that in the example of Figure A1 the average distance does not differ across the two divisions.

The identification strategy in this paper instead produces a figure like A2. Dividing the Manchester geography into small geographical cells uses discontinuous variation across the border while minimising the heterogeneity in the local areas that are being compared to each other.

Figure A1: Boundary Indicator Strategy

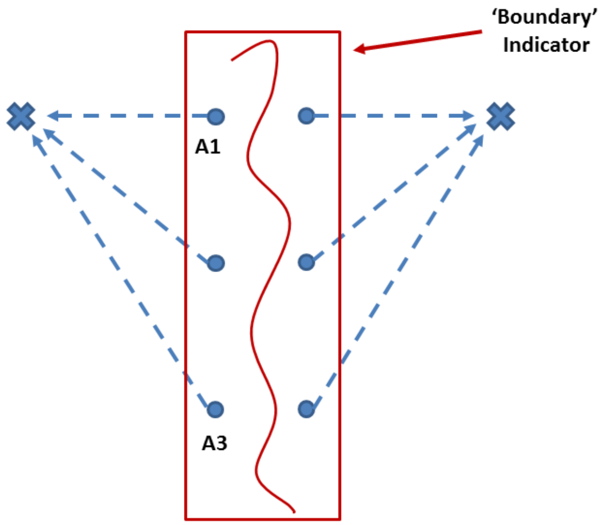


Figure A2: Cell Indicator Strategy