Agent Based Response Time Simulation

Data Analytics Lab

January 2020

For the police response team, the most important KPI is response time. Police response time is related to the number of response units, their locations and the types of crew available. The simulation presented here utilises an agent based model (ABM) to investigate how the mix of single and double crews translates to achieved response times, fleet costs, and injury risks.

1 Table of Contents

2 Introduction

Response time is the elapsed time between receiving an emergency call and the arrival of an officer at the scene of an incident.

Police response time is related to the number of response units, their locations, the road conditions, the traffic volume, the types of crew available when a call for service to an incident is received, and the priority initially allocated to servicing the event.

Rapid police response to a call for service has a significant impact on the positive outcomes to police work (Kirchmaier and others (2015); Cihan, Zhang, and Hoover (2012); Coupe and Blake (2005)). Kirchmaier and others (2015) estimate a 10% increase in response time leads to a 4.6% decrease in the likelihood of detection. Also, police response time is critical to prevent possible injuries or deaths of victims in domestic violence cases or other incidents with a potential risk to life.

Additionally, rapid response has the collateral benefits of deterring criminal behaviour and fostering public security.

Hence, maintaining rapid response times justifies the expenditure of considerable sums of money on police officers, fleet, equipment, backroom systems, and supporting infrastructure.

Given the availability of fleet, preferring single crewed to double crewed units increases the pool of available vehicles to maintain a rapid response to operational objectives while reducing the number of incidents and area covered per crew. In addition to operational benefits, single crewed officers are more cost-effective; however, there is evidence of substantial officer welfare implications regarding mental wellbeing and risk of injury (M. Elliott-Davies et al., 2016).

Response time is complex, with spatial and temporal factors coming into play. The simulation presented here utilises an agent-based model (ABM) to investigate how the mix of single and double crews translates to achieved response times, fleet costs, and injury risks. The aim is to minimise response time by manipulating the number and mix of crew types while also not increasing or reducing costs and risks to injury. The ABM methodology used here has been notably successful in the modelling of emergency vehicle response time (Coelho and Pinto (2018)) and incorporating criminological concepts (Johnson and Groff (2014); Birks, Townsley, and Stewart (2014); Malleson and Birkin (2012)) .

3 Agent Based Model

The model framework is built in python, utilising objects representing resource allocation dispatchers (RADs), crews and calls for service.

The model represents the actions of police officers in terms of the movement and interactions of single and double crews. Crews are represented in the simulation as police vehicle agents. Vehicles move around the simulated environment in response to incidents allocated by a RAD, spending time assessing the scene, and potentially apprehending offenders.

The primary data for analysis were obtained from West Midlands Police incident recording system (OASIS), and the crime recording system for the period June 2018 to July 2019.

These sources contain information about calls for service. For simulation the most pertinent data are

- incident date & time
- location of the incident
- the initial priority allocated at first contact
- time spent on site by allocated crew(s)
- number of arrests made.

Cases that were deemed not to be an actual event requiring response were excluded. These were predominantly cases dealt with on first contact by the RAD.

Additional data on weekly shift patterns and absence was sourced from the human resource system (GRS). A sample of telematics data was used to study additional behaviours to determine when a crew becomes available.

Other assumptions that have been incorporated are:

- Each available officer has the same cost per day.
- There is no change in the number and location of custody blocks
- There is no change in shift pattern
- Supervisors are not included as response units.

3.1 Resource Allocation Dispatchers (RADs)

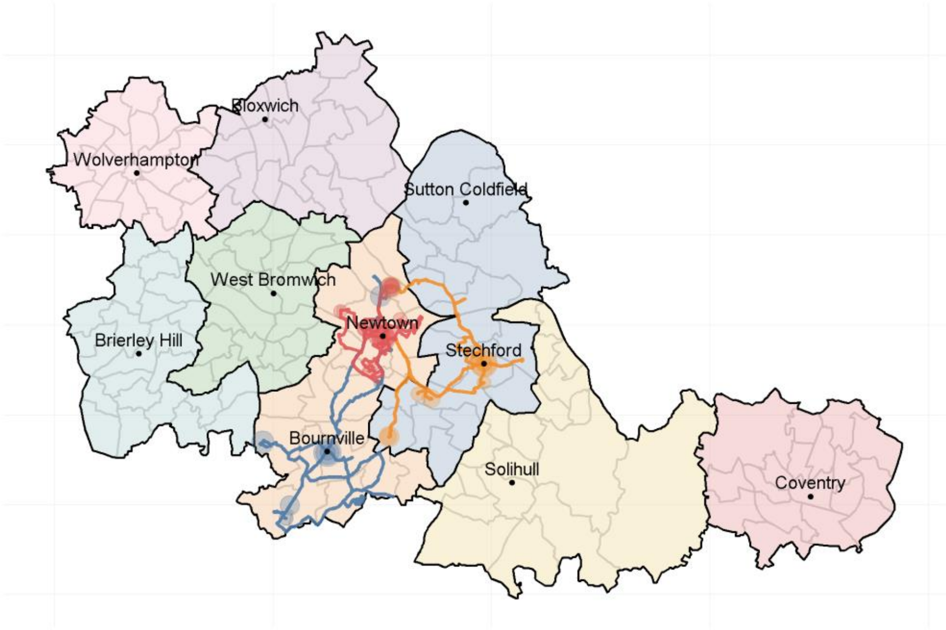
An initial emergency call is received via 999 or 101. The caller is then connected to a Contact Officer who triages the incident and allocates an initial priority. The RAD dispatch team then allocates the most appropriate police resource.

The West Midlands Police geography is divided into neighbourhood policing units (NPUs). Shown below are the NPUs and the stations housing response teams. There are three dispatch control rooms in the force area responsible for the allocation of work to response crews. The RADs and the NPUs they are responsible for are:

- Bournville: Birmingham West and Birmingham East
- Willenhall: Coventry and Solihull,

- Wednesfield: Sandwell, Dudley, Walsall and Wolverhampton

Station Locations within Neighbourhood Policing Units
Showing tours of duty for three vehicles



Source: WMP DAL 2019

In addition to the physical constraints of the geography, in the model crews are not shared between different RAD teams but will move between the NPUs controlled by the RAD depending on service requirements (examples of three event tours by one vehicle are shown on the above diagram). Crews are influenced in how they move by the incidents that are allocated to them by the dispatcher. The timing and location of these calls for service are drawn from the records of the West Midlands Police incident recording system (OASIS).

Incidents are "played back" during a simulation run. The RAD agent is responsible for allocating and directing crews to the real-world locations of calls for service at the appropriate times reproducing the resource pressures and constraints in the simulated setting.

If no crews are available, each RAD maintains a backlog of calls. Incidents in the backlog are allocated as crews become available. This is implemented as a priority queue, with events ordered by the incident priority and time required to service.

The RAD agent allocates incidents to available crews based on factors such as:

- if possible prefer a double crew for a P1 incident and single crews for lower priority incidents.
- prefer the nearest available crew not currently attending an incident or on a break. This can result in a crew in transit to a P2 incident being reallocated to a higher priority incident.
- Prefer crews with a home station in the same NPU as the incident.
- Prefer crews not nearing the end of their shift. Here there should be a reasonable expectation of being able to reach the location of an incident and perform a holding role until another crew becomes available.

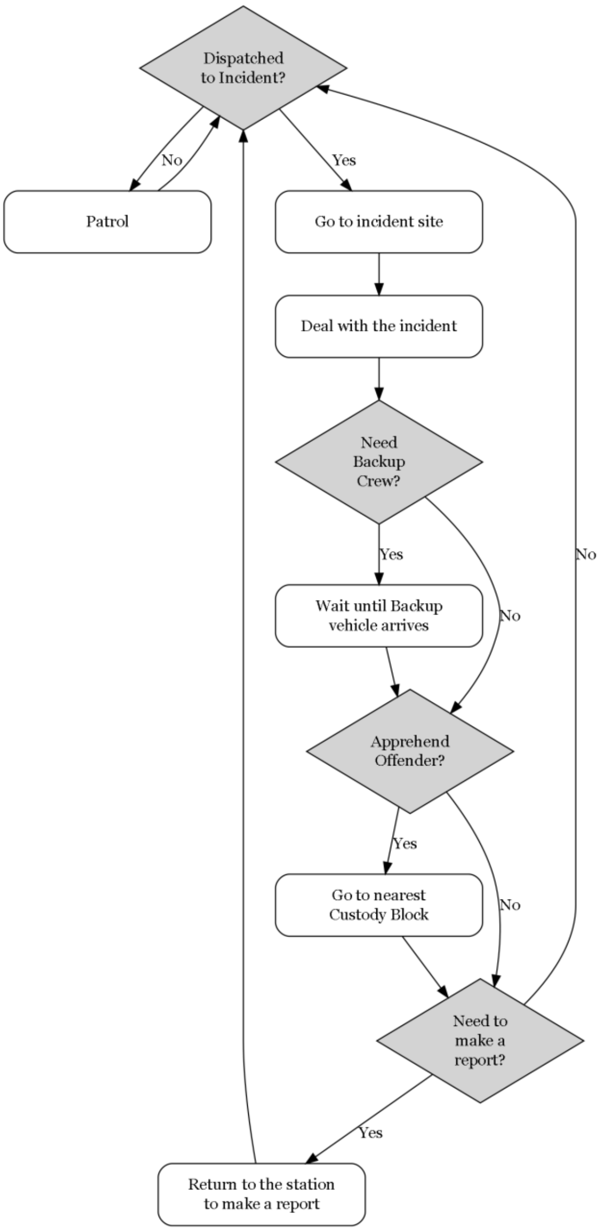
3.2 Crews

Crews have attributes that inform their actions. In particular, vehicles have a current location in space, a home station, a home RAD, a unique call sign, a current activity, a shift end time and a current status.

The crews' shift pattern and composition is sourced from the GRS system. Crew resources are made available to the RAD at the start of a shift, and are removed at the end of a shift. A crew will continue an in-progress task until relieved by another crew, or the task is completed. Where possible RADs do not allocate work to a crew nearing the end of a shift.

Crews move around the simulated environment in response to incidents allocated by a RAD, spending time dealing with the caller when it reaches the site of the event.

A basic flowchart for crew behaviour is:



Double crews are primarily responsible for attending urgent calls to service; high priority P1 incidents. Single crews are responsible principally for attending non-urgent P2+ calls unless no double crew is available (incidents range from P1 to P9 with P1 being the most urgent). Crews responding to a P1 incident move at a faster speed ignoring speed limits.

Simulated crews spend any unoccupied time patrolling, which is modelled in the ABM as a random patrol strategy where a crew targets any location within the home NPU This means for example that they do not undertake hotspot policing, which would be weighted toward areas with high historical risk of high priority events.

Simulated crews take a break of approximately 40 minutes around halfway through their shift. When a break is due, it is taken at the end of any current assignment.

Custody blocks are located in Oldbury, Wolverhampton, Perry Bar, and Coventry.

A double crew is required to detain a suspect. If a single crew attends an incident requiring an arrest, the crew is responsible for coordinating with the RAD to schedule a double crew to detain and transport a suspect to the nearest custody block. The single crew remains at the site of the incident until the double crew arrives.

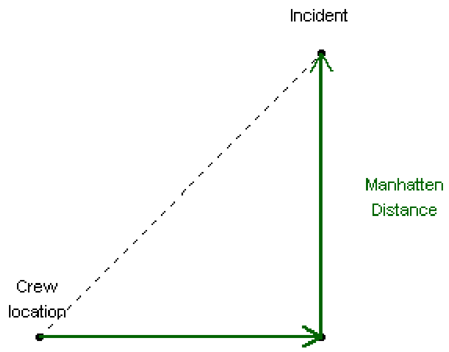
When multiple arrests are made, more double crews will be required to attend.

3.3 Time to incident.

The computation of routing to minimise the time between a response vehicle being dispatched and arrival at the location of an incident is undertaken by the response crew. Vehicle level telemetry data are currently not available to West Midlands Police as a bulk download so the ABM estimates time to incident based on a number of heuristics.

The ABM does not explicitly use information about the real-world road transport network and is unaware of hazards, roadworks, or traffic flows at different times of the day, or day of week. The calculated time to arrival depends only on the priority of the incident, the distance to the incident, the time of day and the day of week. Because the simulations are based on real events that occurred during real times of day and days of the week, etc. items such as 'average' traffic conditions, hazards, roadworks, etc. are implicitly incorporated into the modelling.

The ABM uses the Manhattan distance between the location of a crew and the incident as a proxy for the real road distance. Manhattan distance, also known as rectilinear distance, is the distance between two points measured along axes at right angles. It is calculated as the sum of the horizontal and vertical components of the pairs of points.



Several studies compare how Euclidean and Manhattan distances differ from real distance measures, based on data collected from vehicle routing on the UK transport network. The correlation between the measures is very high such that replacement of the real routing distance by a vehicle with the Euclidean or Manhattan distance is unlikely to impact analytical results. For UK roads, Cooper (1983) determined empirical detour factors of 1.2 - 1.6 times the "straight line" distance, which has had some acceptance and use in the statistical community (Boscoe, Henry, and Zdeb 2012). The Manhattan distance has a mean detour factor of 1.3 with a standard deviation of 0.12, which agrees well with these empirical results.

The ABM travel time calculation is based on the Manhattan distance and the estimated vehicle speed. The estimate for the vehicle speed is calculated algorithmically by sampling from historical speeds which are stratified to account for the priority of the incident, the distance to the incident, the time of day and the day of the week.

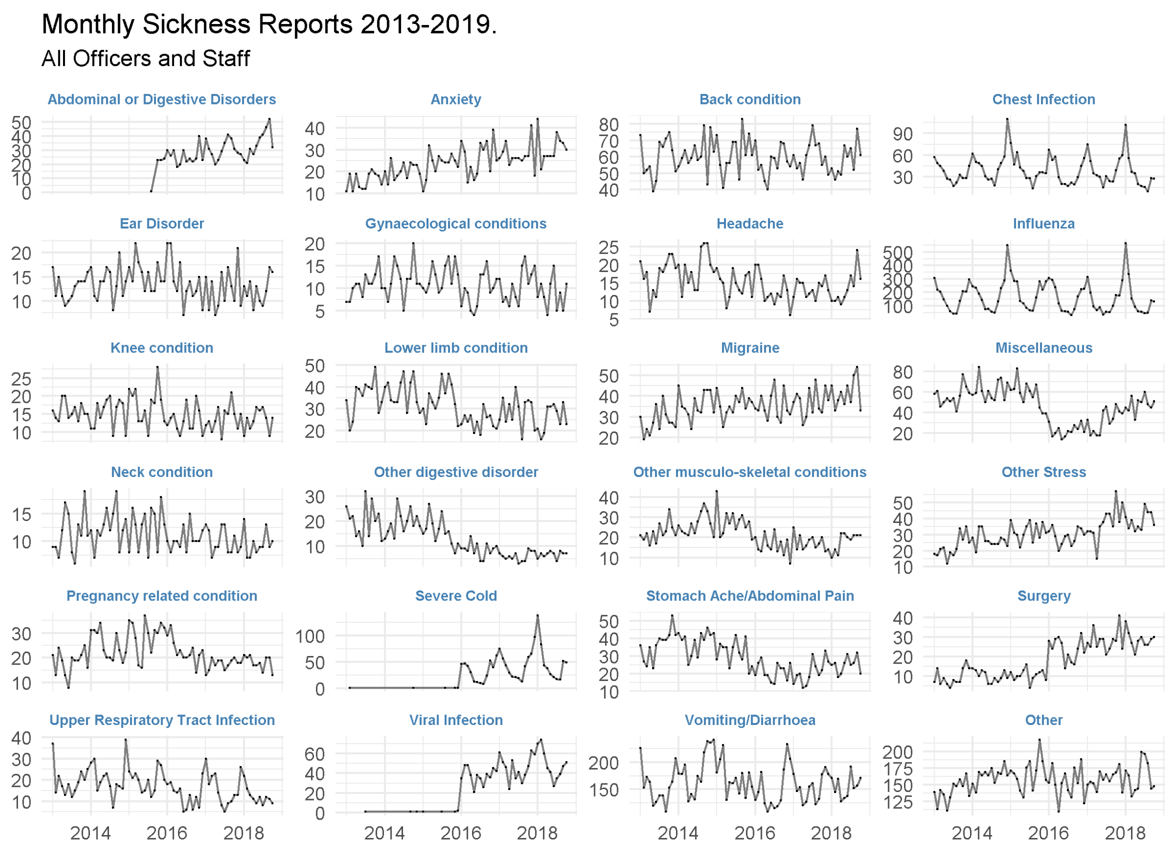
3.4 Injury and Sickness

In an extensive study of more than 11,000 officers by Houdmont, Elliott-Davies, and Donnelly (2019) into the difference in relative risk of violence and injury of single officer crews, approximately three-quarters of officers reported having been single crewed often or always in the 12 months under consideration.

Compared to those who were never single crewed, officers who were always single crewed were

- 82% more likely to be verbally insulted at least once per month.
- 58% more likely to be verbally threatened at least once per month.
- 41% more likely to report having been physically attacked at least once per month.
- 56% more likely to have suffered one or more injuries arising out of work-related violence in the preceding year.

This study agrees well with an analysis of absence records for the period January 2013 to July 2018 from the West Midlands Police HR system. An analysis shows a single crewed officer has increased relative risk of injury and sickness between 11-54% when attending a P1 event, and a similar level of increased risk attending P2 and P3 events.



Working in response carries inherent risks to officer safety. Over the analysis period compared to office staff, response officers were

- between 2.1 and 3.6 times more likely to suffer head or facial injury.
- between 1.5 and 2.8 times more likely to suffer skin injury.
- between 1.5 and 2.1 times more likely to suffer upper limb injuries.
- between 1.4 and 4.1 times more likely to suffer multiple injuries.
- have increased risk of musculo-skeletal injury, neck injury, broken or fractured bones or dislocations, burns or scalds, poisoning, shoulder injury, back injury, eye injury, and post-traumatic stress.

The HR system only includes injuries resulting in absence from work. The eSafety system includes additional minor injuries. This shows other minor injuries occur at an average of ~20 incidents per week (between 10 and 29 based on 10% and 90% quantiles).

The ABM calculates the risk of serious injury resulting in absence based on the type of incident attended, the priority and the crew size. The historic rates are

- 6.6 injuries per 10,000 incidents attended for a P1 attended by a single crew
- 5.1 injuries per 10,000 incidents attended for a P1 attended by a double crew
- 3.0 injuries per 10,000 incidents attended for a P2 attended by a single crew
- 2.4 injuries per 10,000 incidents attended for a P2 attended by a double crew

4 Costing

4.1 Injury costs

At pay point 4, which typically approximates to four years of service, the cost to West Midlands Police (salary, employers NI, and pension cost) for a single officer is around £33,700. Given the very low levels of recruitment over the last 10 years, we use the higher pay point 5 cost of £36,500.

Each injury is charged for 3 weeks to cover injury and restricted duty. There is no visibility of restricted duty through the GRS system. An overall recovery time of around 3 weeks is plausible based on the median found in the absence records. This assumption can be altered when more accurate information becomes available.

This leads to an estimated injury cost of £2,100 to the force of each injury leading to absence.

Clearly, there are other collateral considerations not considered. For example, although the likelihood of injury is low, any increase will also be associated with a rise in staff churn and the raised probability of a catastrophic event.

4.2 Fleet Costs

Since 2011/12, forces have spent £105.3m on maintenance such as annual services, replacement tyres and other wear and tear items, and a further £22.2m on accident repairs, like new bodywork panels and windscreens. Gloucestershire had the most expensive average bill. It spent £1.7m in 2013/14 on its 457 vehicles – that's an average of £3,722 per car. Nottinghamshire was just behind, spending £3,558 per car.

Procurement costs for a Compact High performance estate vehicle fully converted is estimated at £26,841. A vehicle typically remains part of the fleet for 3 years.

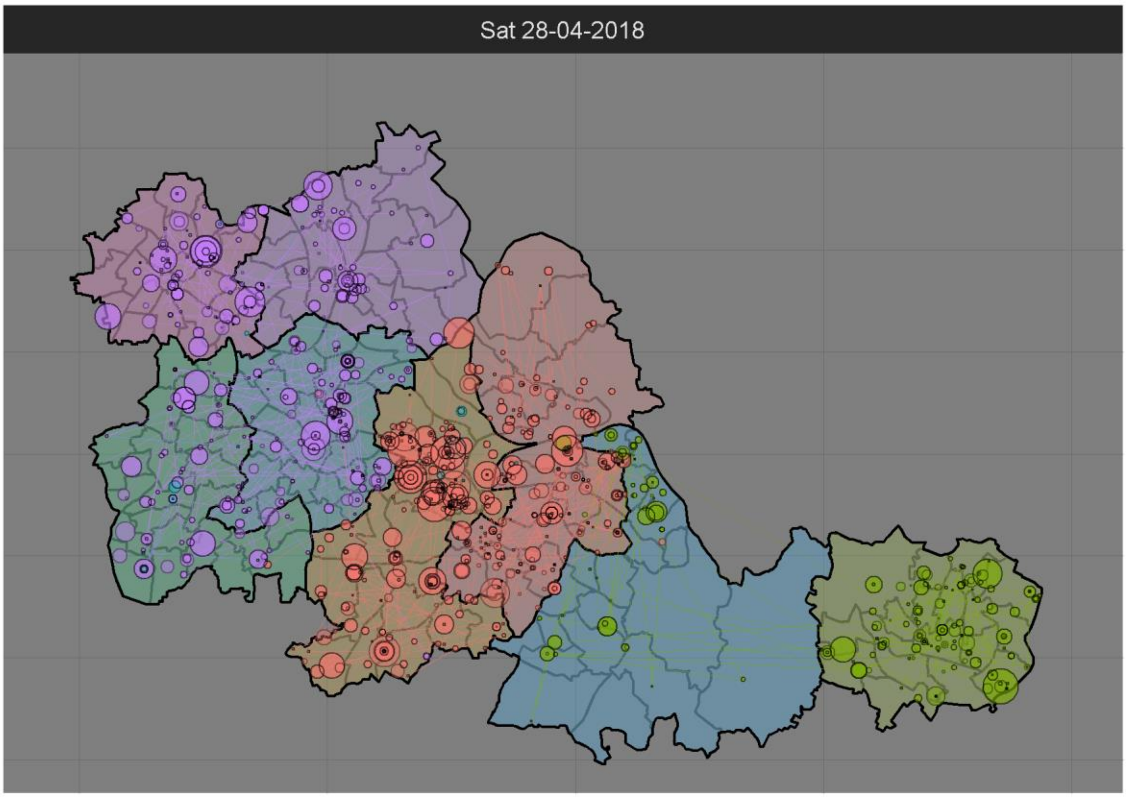
As at October 2018, there are 164 vehicles in the WMP response fleet. 526 vehicles overall (PSU, Vans, Dog car, Camera units etc.)

From monthly billing, WMP spend approximately £250,000 on a monthly basis. This equates to a spend of around £500 in fuel per month per vehicle.

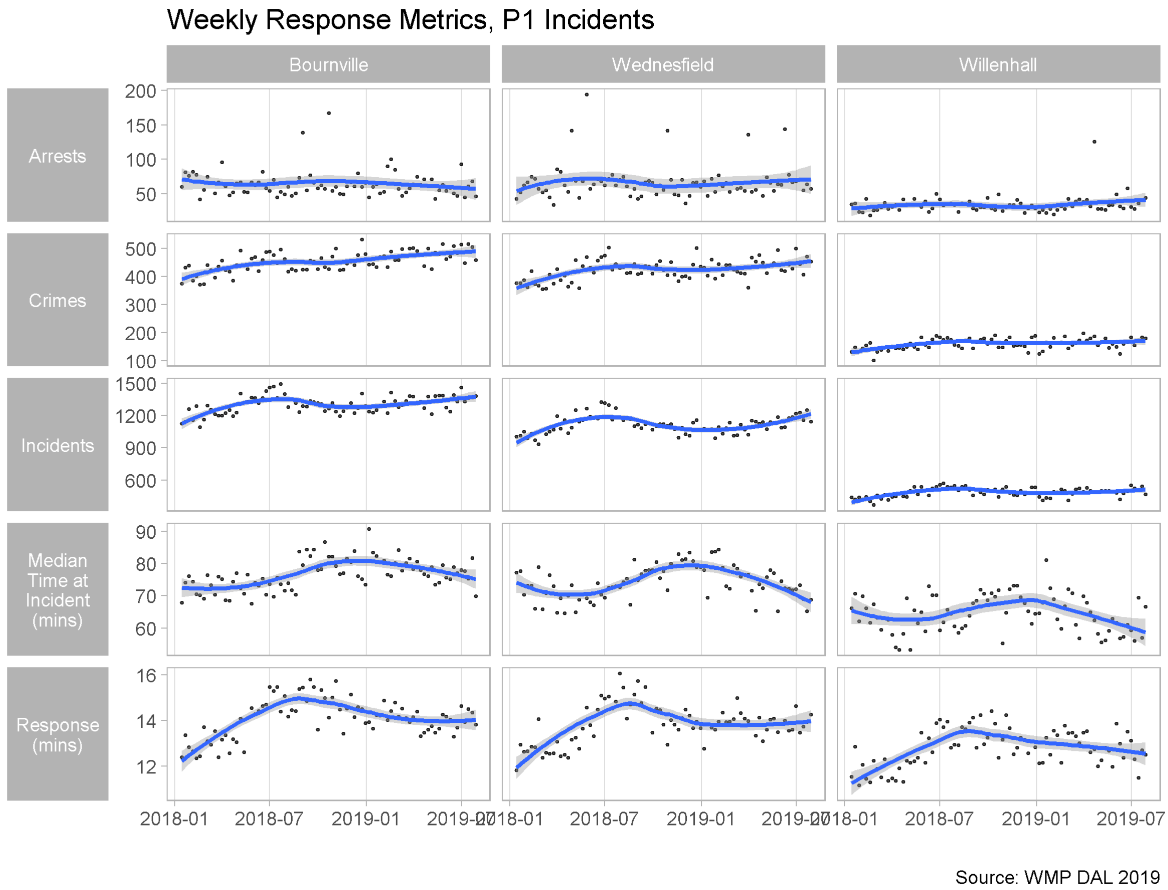
This leads to an estimated annual cost of ownership of £18,500 for each vehicle.

5 Response Metrics

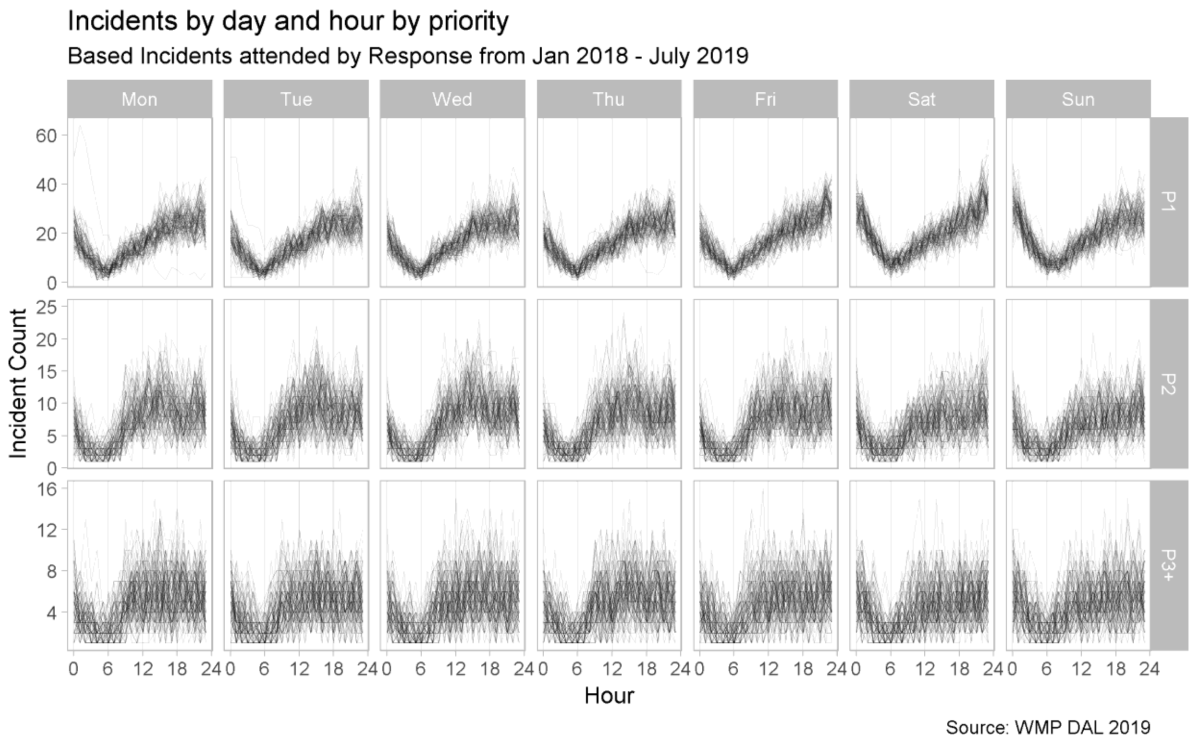
Below we show the geographic spread of incident locations for a single day. The size of points is normalised to indicate the length of time a crew attended.



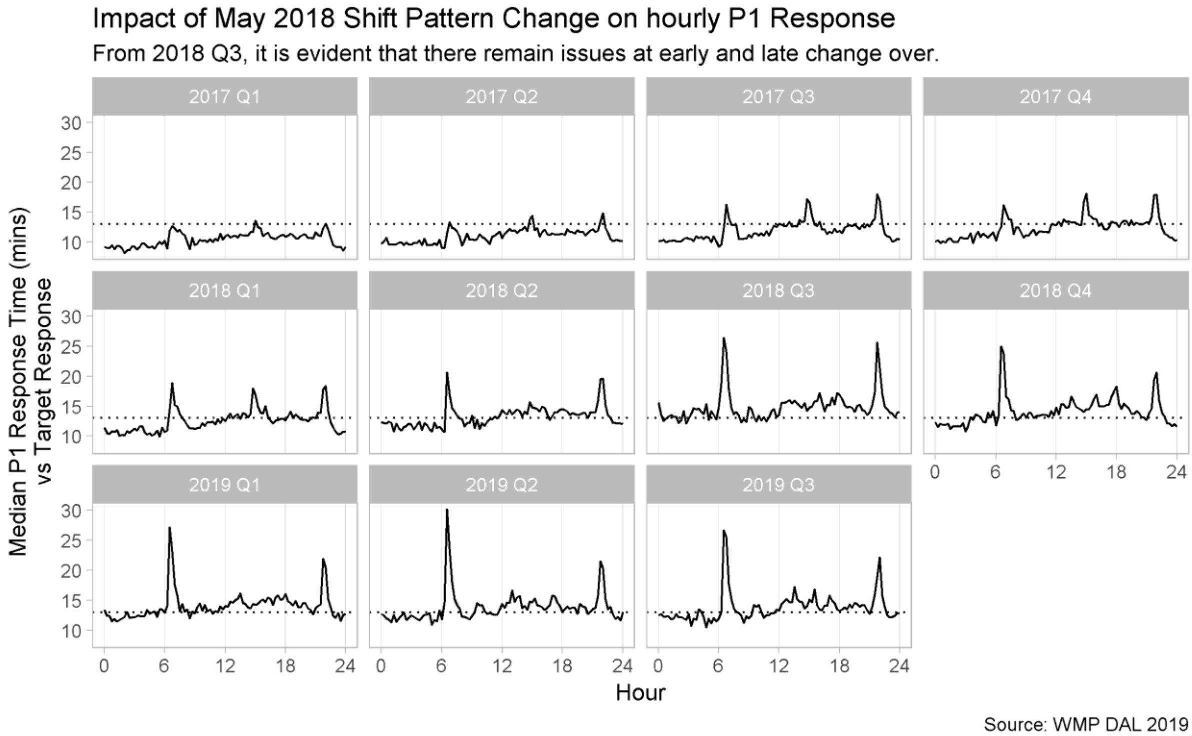
In the following, the ranges in brackets are the 10% and 90% quantiles: The median number of incidents per week in 2019 is 4,740 (4,542 - 4,812). Of these, ~63% are for P1 incidents with ~255 (228 - 317) leading to an arrest. On average a crew spends 75 minutes on site (12 - 265 minutes). The median P1 response time is 13.8 minutes (6 - 36 minutes). The response times have been trending downwards since July 2018.

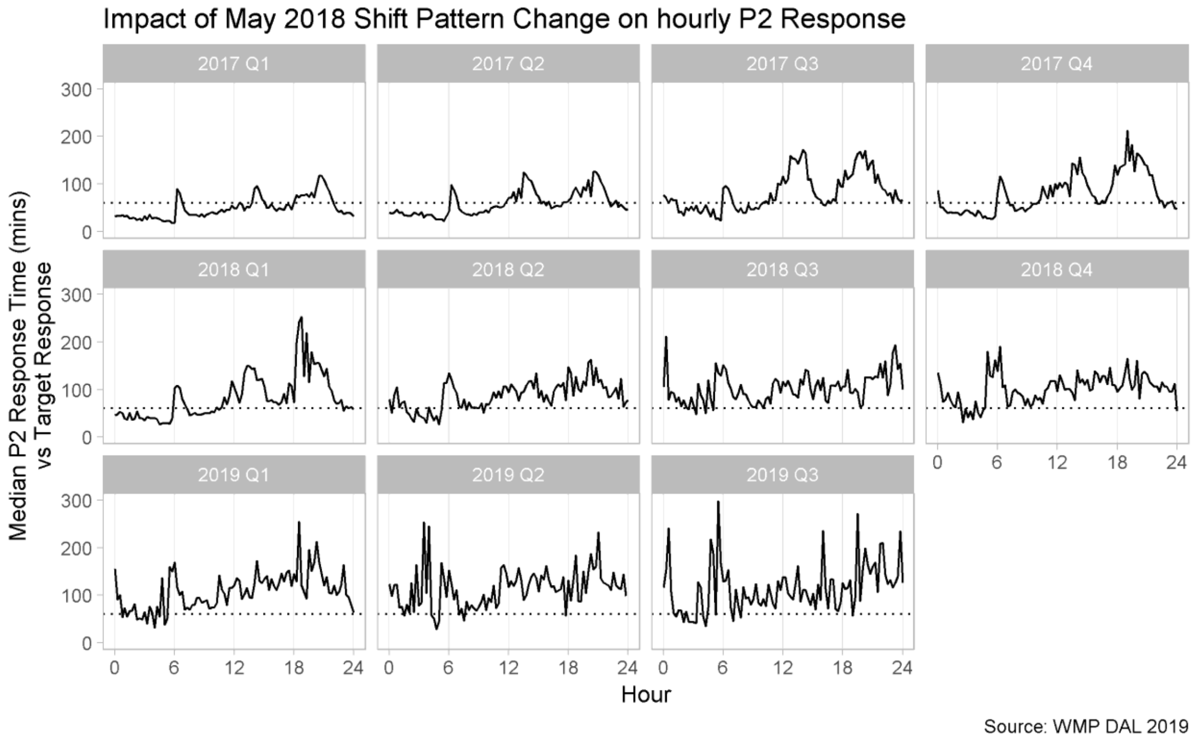


There is a clear daily pattern with demand dropping overnight being lowest around 6am and highest from 10pm to ~12pm.



Even given the low demand for service from 6am - 9am, there are issues with backlogging overnight leading to high P1 response times the following morning. It is also evident that the response time is greater than the 13 minute target from noon - 11pm.





6 Results

Each simulation covers a two week period. The first week is used for "warm up". This allows the simulation to mimic the "steady-state" conditions of the system. Statistics are only recorded for the 2nd week.

Simulations were repeated for 52 weeks with 15 replications for each combination of

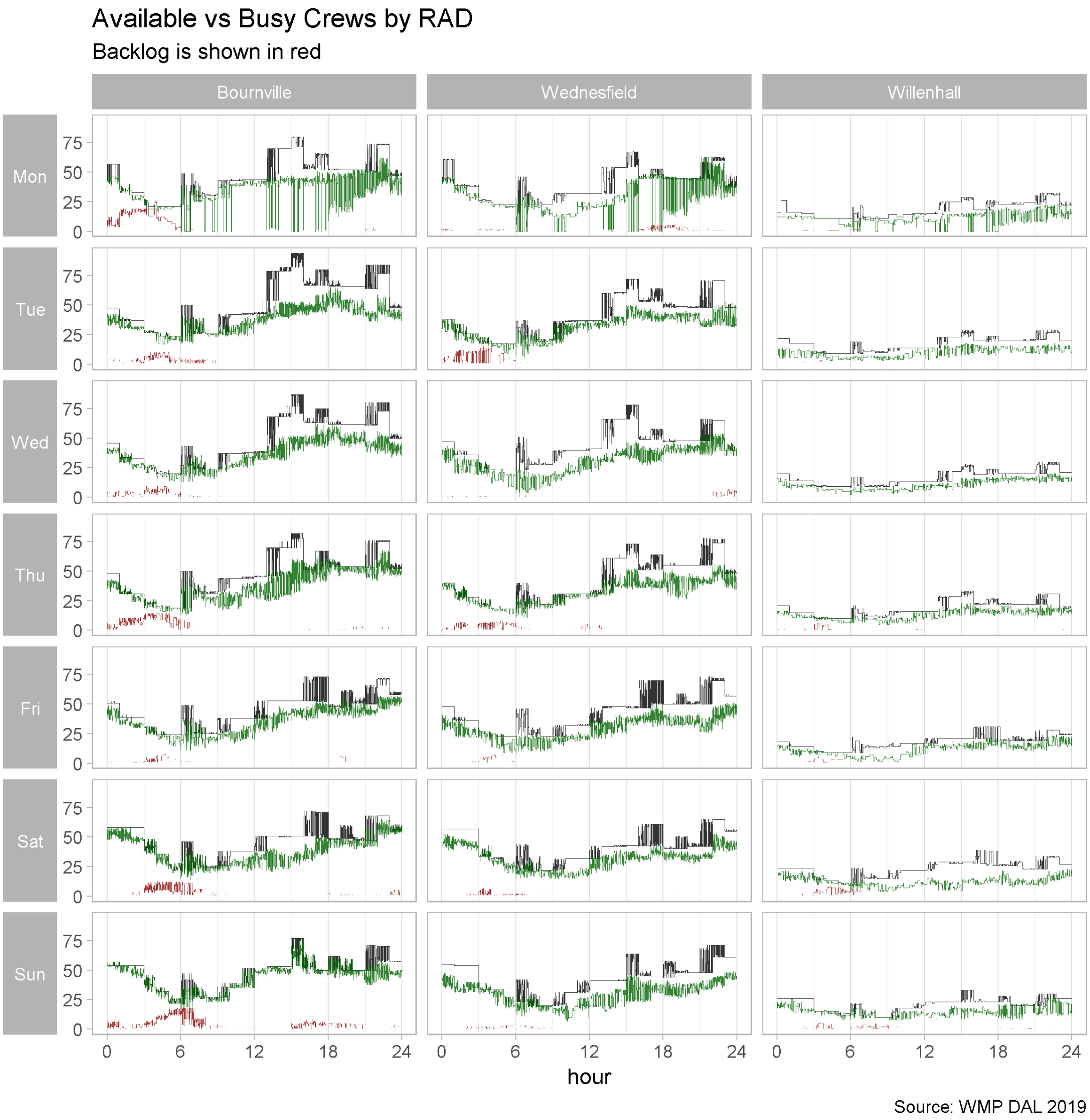
- additional P1 load (0% to 50% in increments of 10%) Additional load is added to the incoming incidents by combining a random sample from the incidents the week following the simulated week.
- proportions of double crews (10% to 100% in increments of 5%)
- number of crew (100% to 140% in increments of 10%)

Overall data were collected for 374,400 simulation runs.

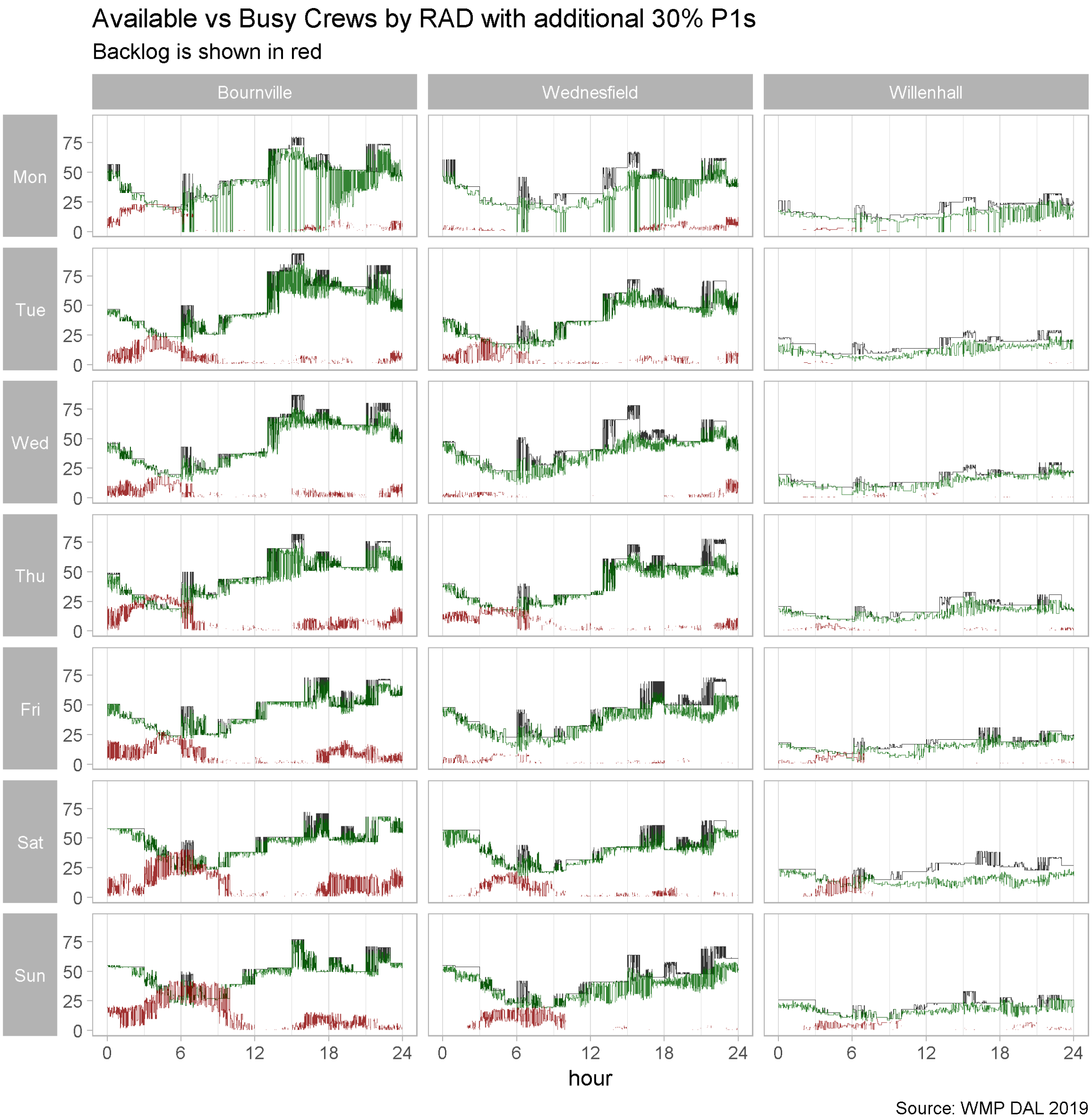
The simulation generates credible results (whereby current resources lead to simulations which reflect actual findings). The base model (the proportion of double crews based on the actual shift pattern (~58%), 100% crew, 0% additional load) calibrates well against observed response times.

6.1 Available vs Busy Crews.

This graph visualises how the crew availability develops over time. We see pinch points where there is a backlogging of work which agrees well with the increased response times noted during the early shift change over. (Busy shown as grey, available as green.)

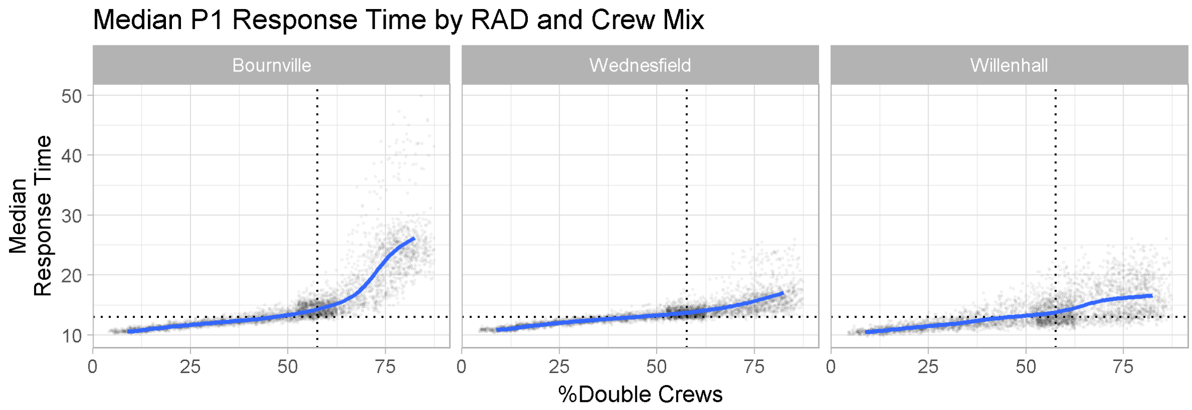


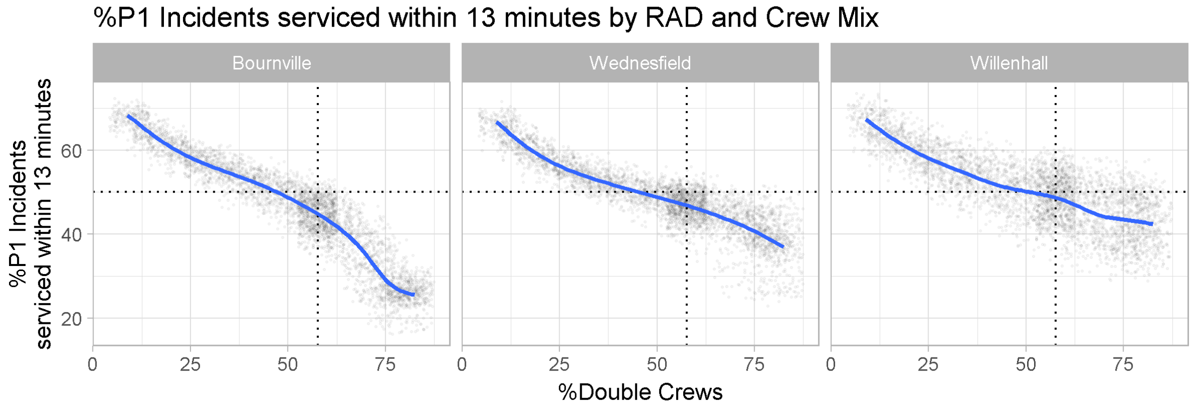
Adding an additional 30% P1s highlights further the pinch points overnight and early morning and leads to more backlogging throughout the day.



As can be seen from the charts below, with current resources, if the proportion of double crews increases above the current levels, median response time will increase, and the proportion of P1 incidents reached within 13 minutes will decrease.

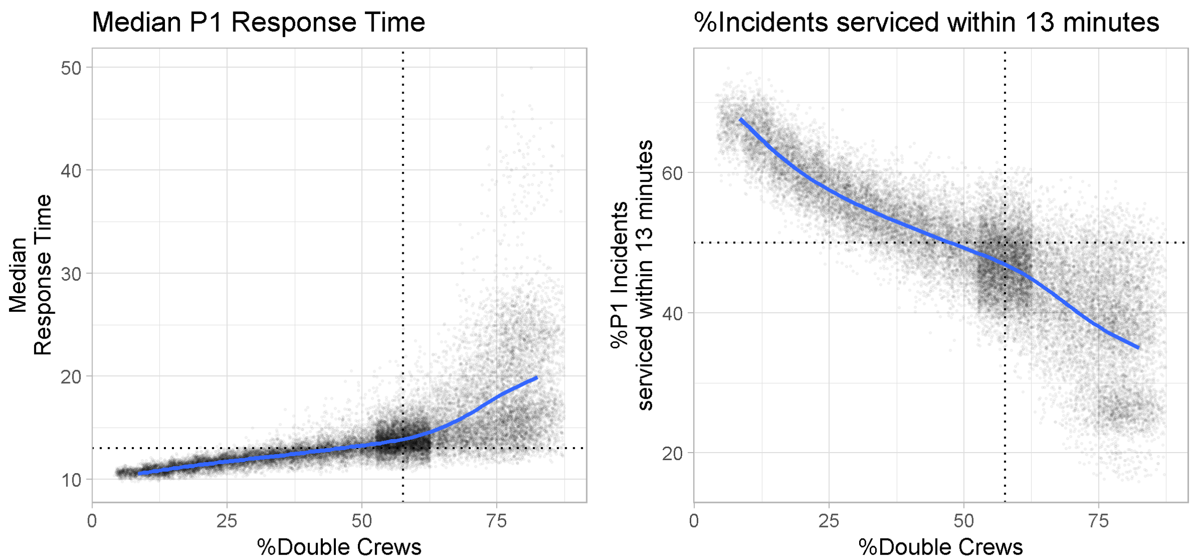
For Bournville, an increase in double crews to circa 60% leads to a rapid degradation in response times. This is indicative that there are not enough resources to service the demand.





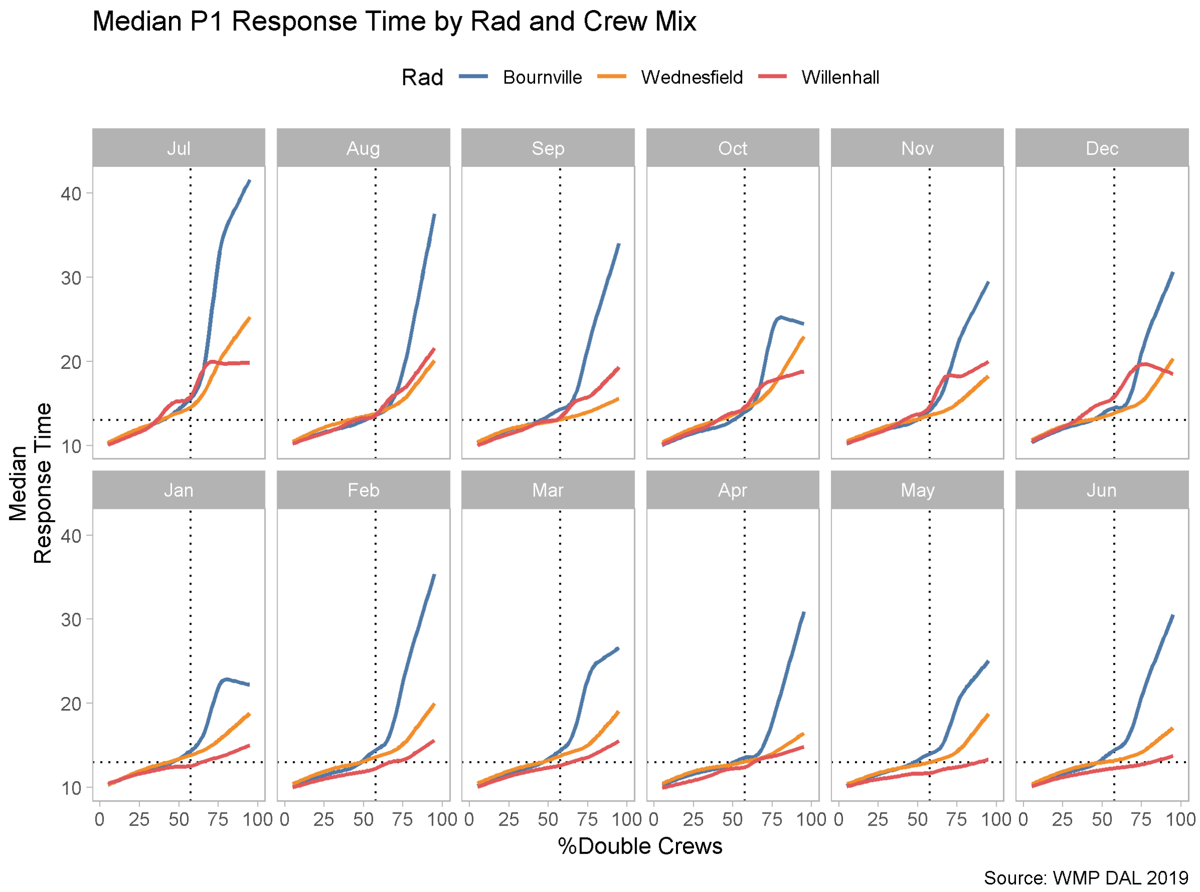
Source: WMP DAL 2019

The pattern is similar for the force area as a whole.

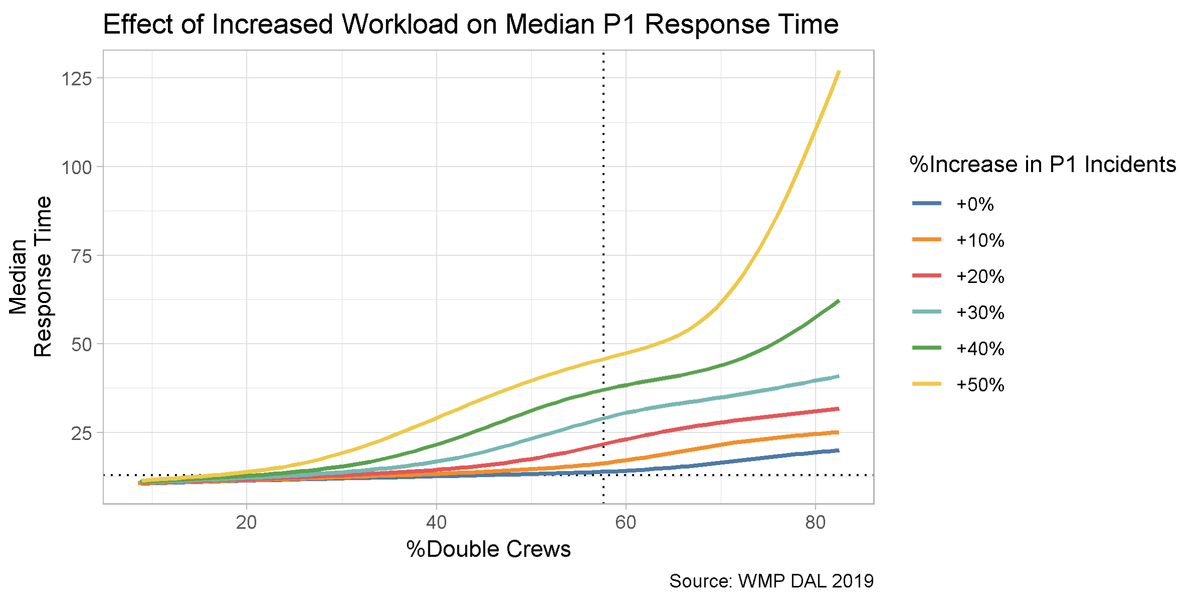


Source: WMP DAL 2019

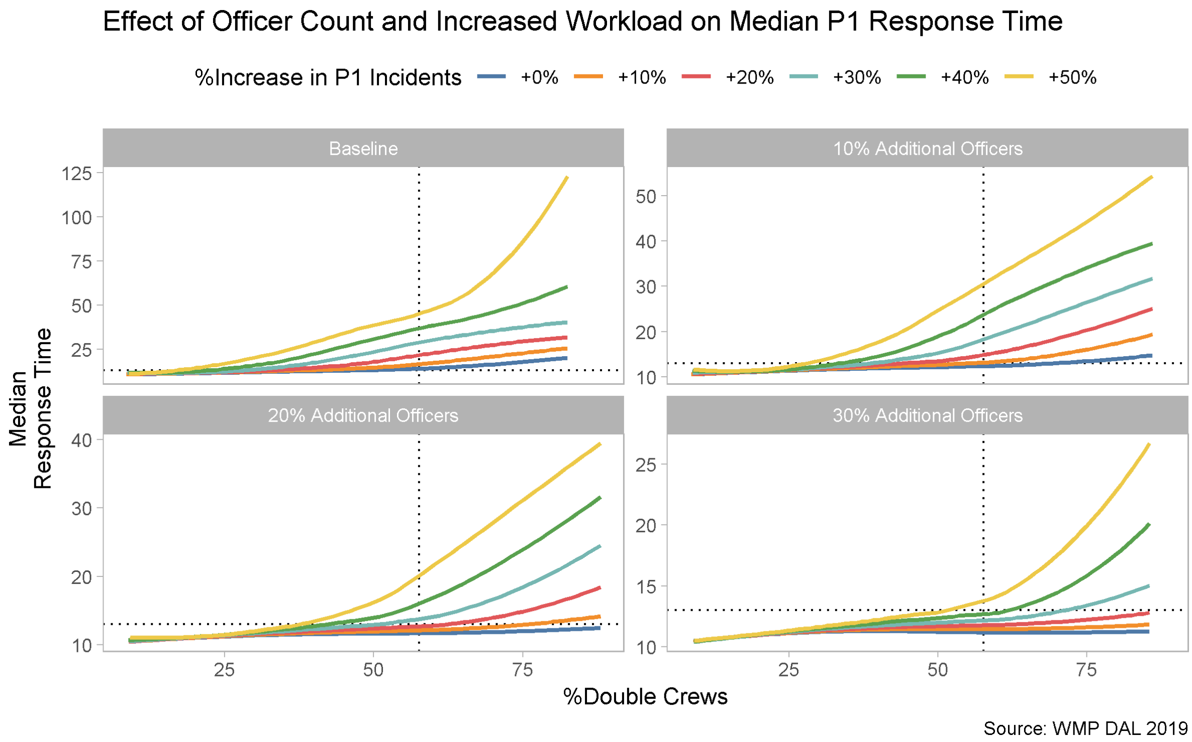
Some months pose a bigger likelihood of increased response time than others if the % of double crews increases. Again, Bournville shows acute sensitivity.



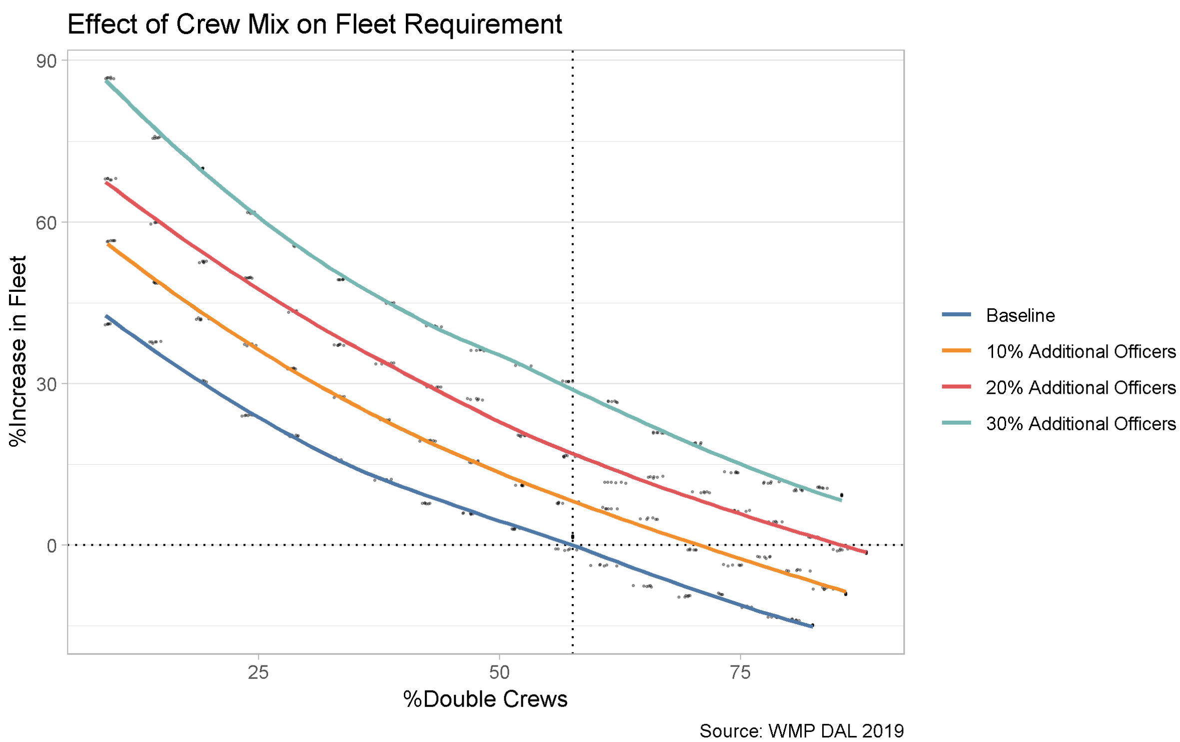
Should demand increase, this can be more easily met with an increase in the proportion of single crews.

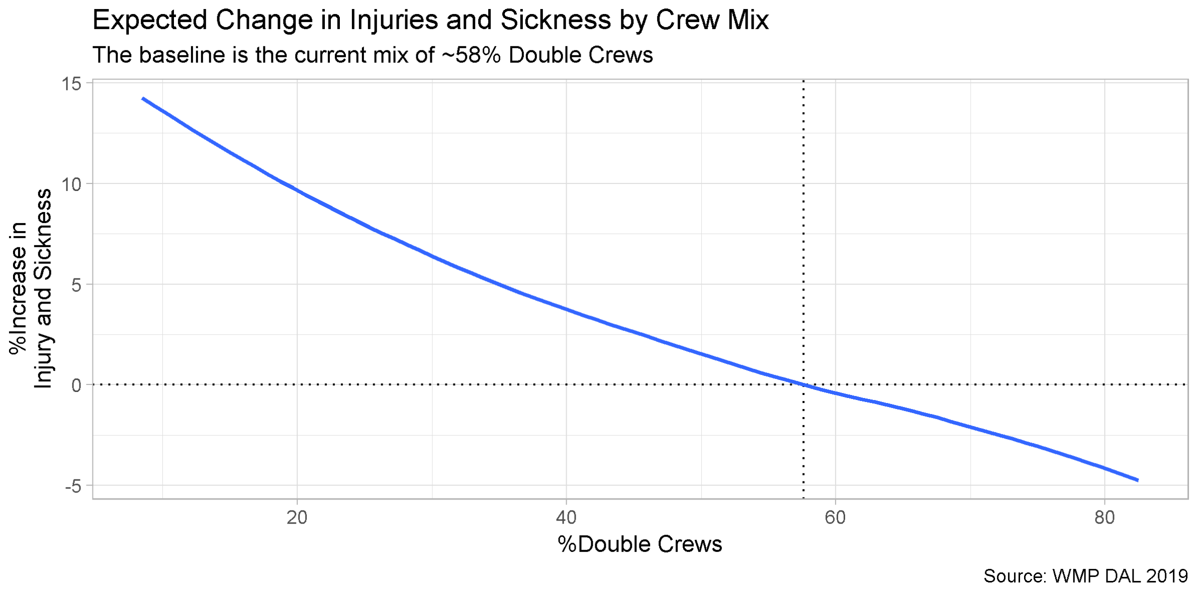


A 20% increase in officers would allow more flexibility and allow a far higher proportion of double crews whilst also allowing for fast median response times.



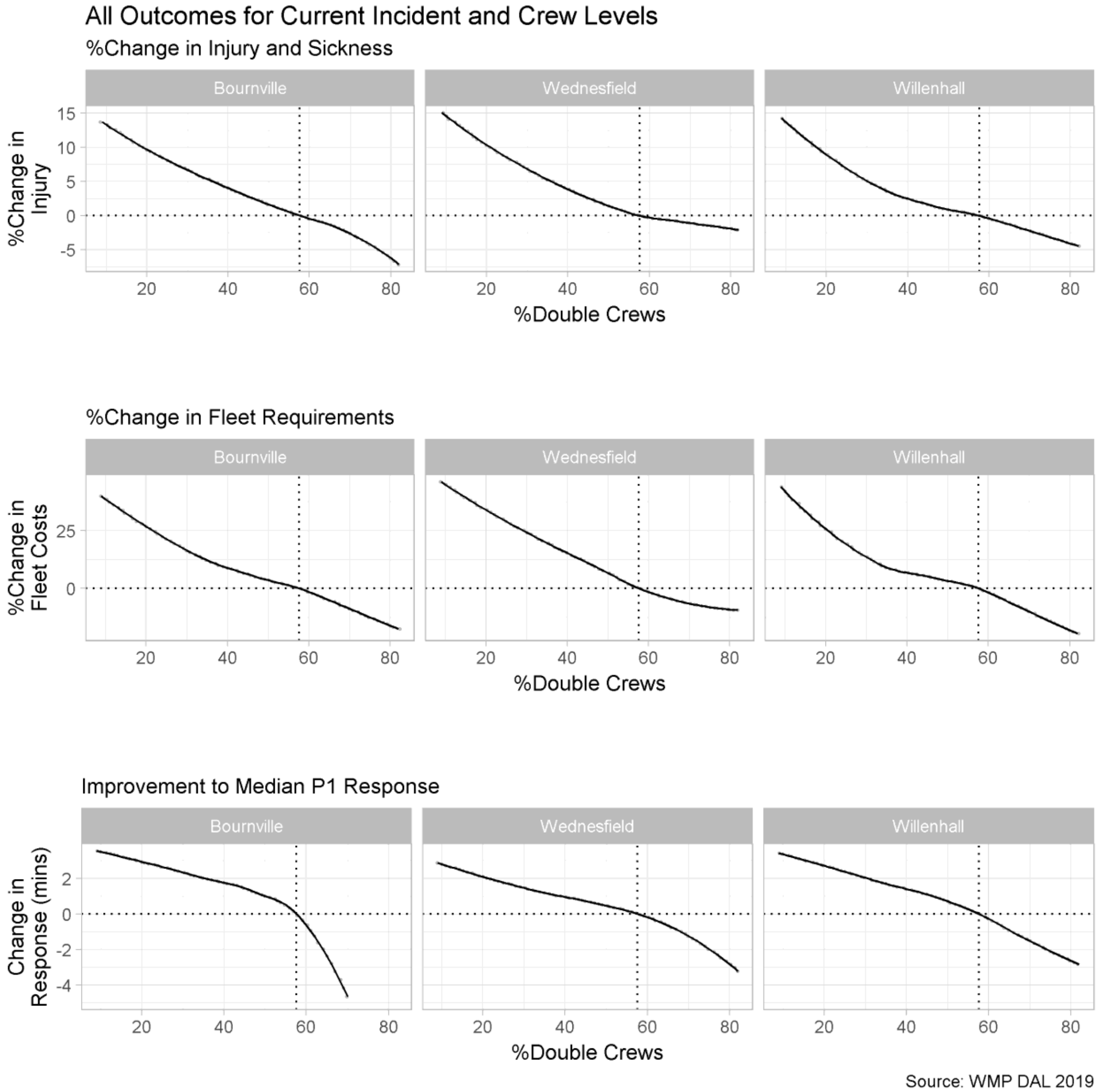
Of course, more single crews come at a cost in terms of increased fleet and injury.



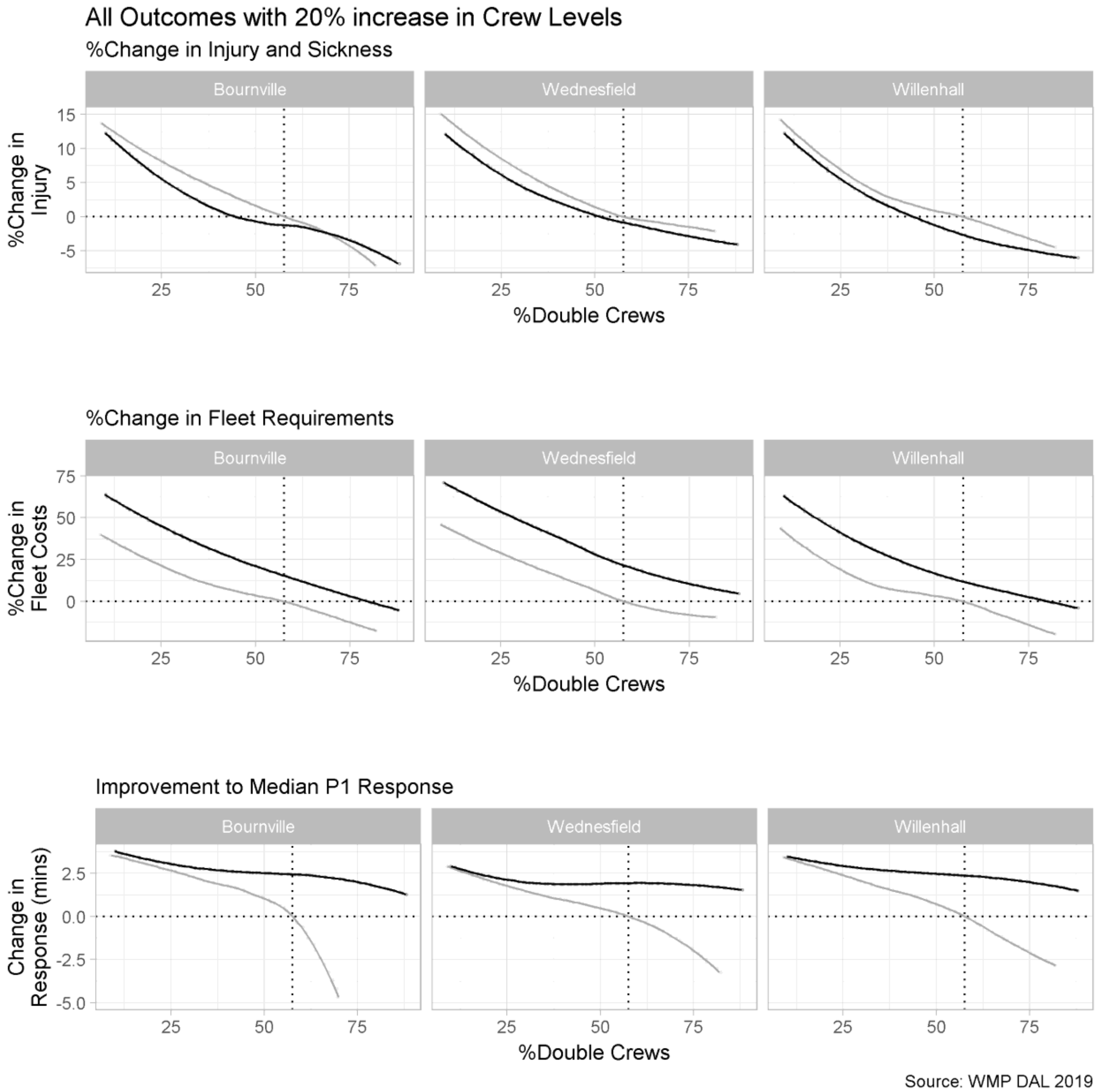

Expected Change in Injuries and Sickness by Crew Mix
The baseline is the current mix of ~58% Double Crews
15
10
5
0
-5
%Increase in
Injury and Sickness
20
40
60
80
%Double Crews
Source: WMP DAL 2019

7 Summary

Overall response times could be improved with a higher proportion of single crews, but this would come at a cost in terms of fleet and leave resulting from an increased likelihood of injury (note that in the last panel, a negative number means adding to the response time).

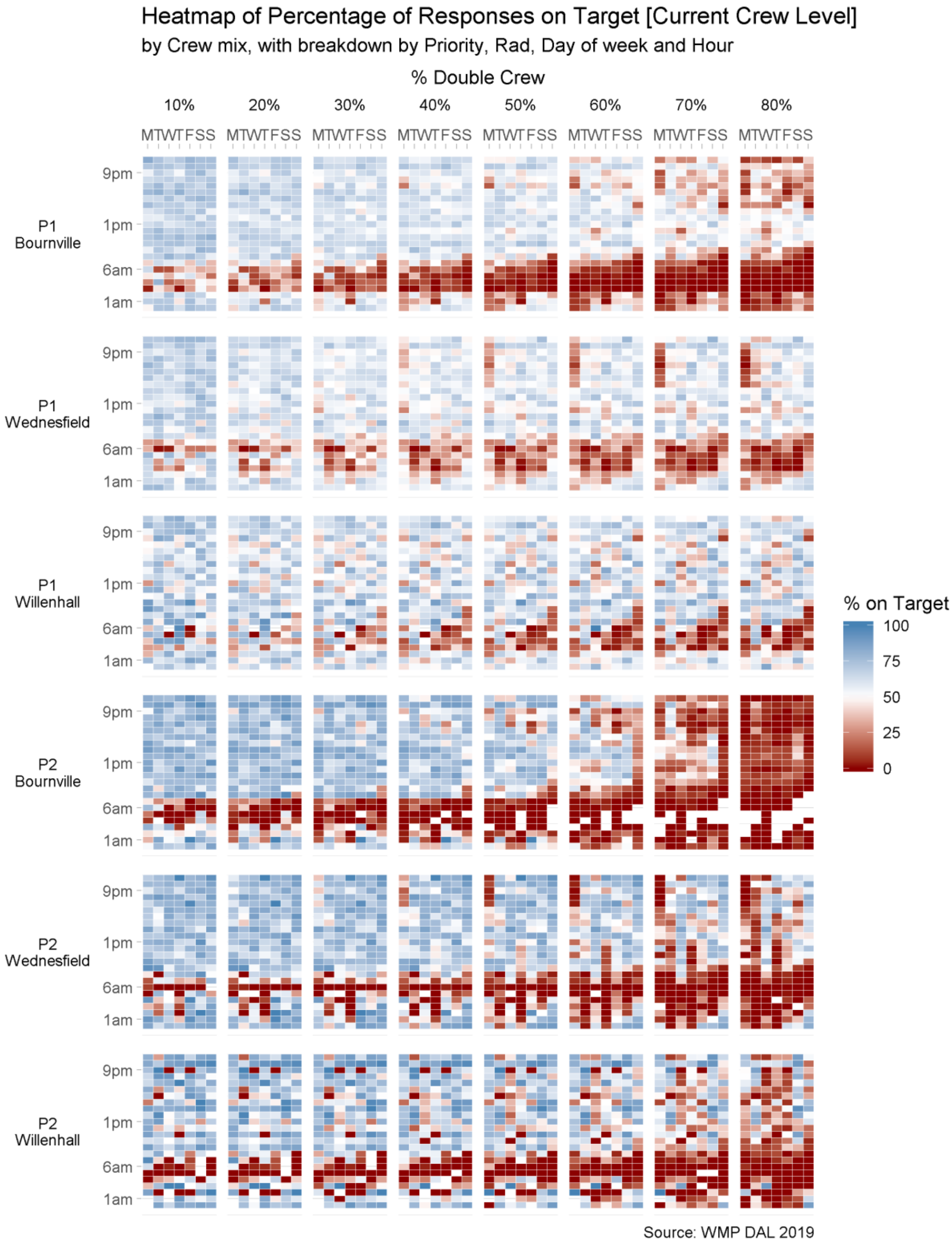


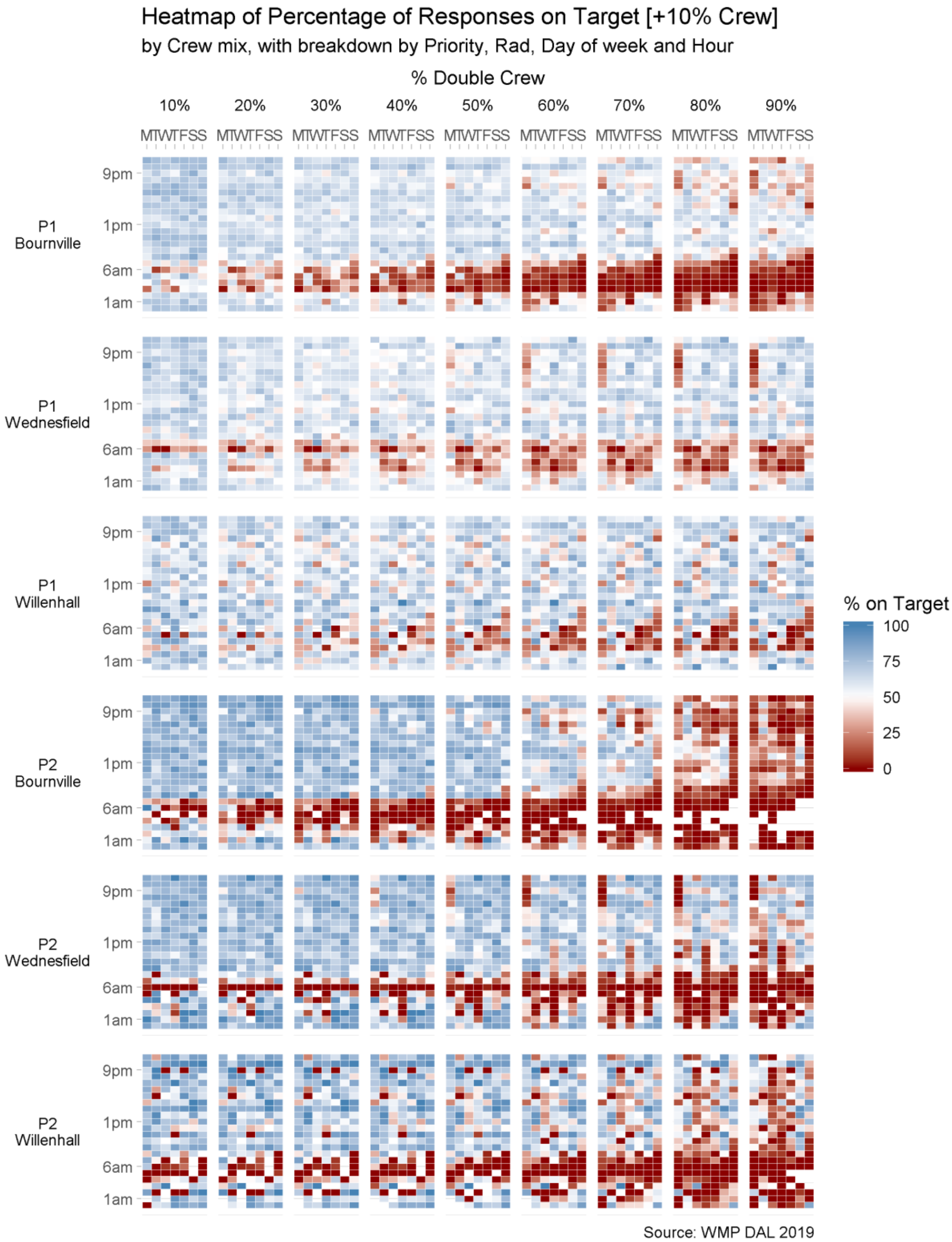
The tradeoff changes with the number of resources. A 20% increase in resource would lead to a much shallower response time curve (the dark line) and so enable changes to the % double crew and a reduction in response time without incurring increases in costs of injuries.

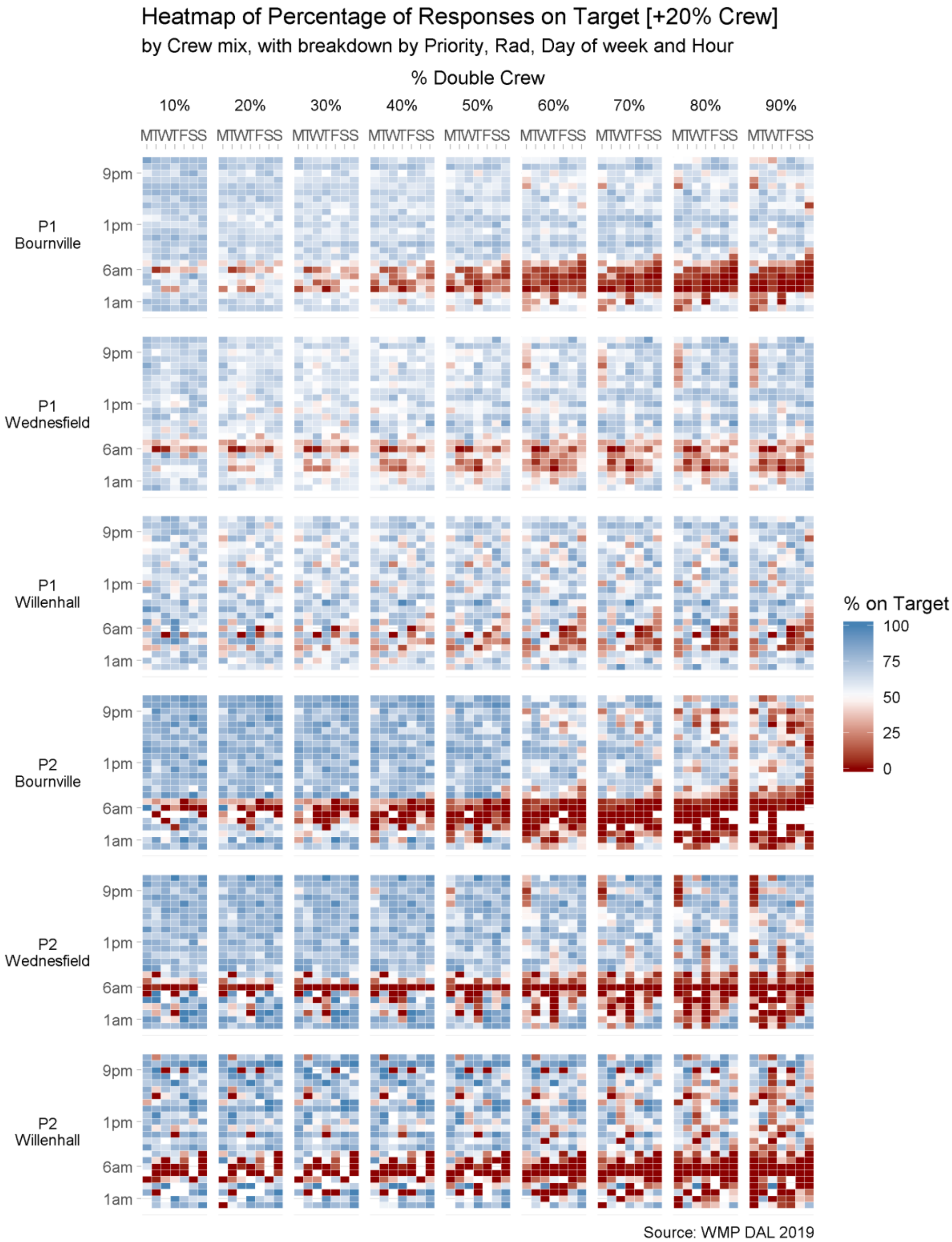


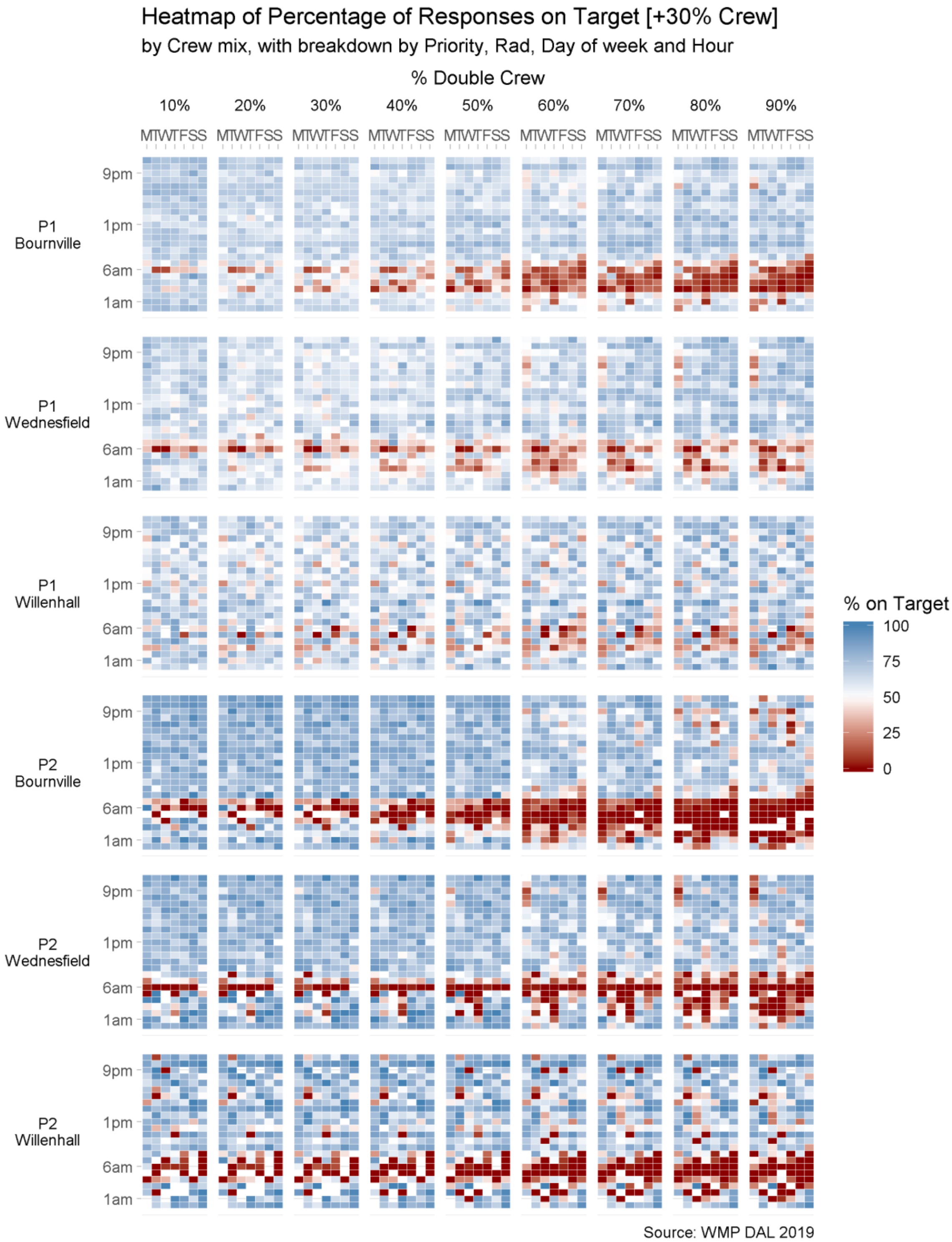
8 Appendix

8.1 Breakdown of results by Day of Week and Hour




Heatmap of Percentage of Responses on Target [+10% Crew]
by Crew mix, with breakdown by Priority, Rad, Day of week and Hour
% Double Crew
10%
20%
30%
40%
50%
60%
70%
80%
90%
MTWTFSS
9pm
1pm
6am
1am
P1 Bournville
P1 Wednesfield
P1 Willenhall
P2 Bournville
P2 Wednesfield
P2 Willenhall
% on Target
100
75
50
25
0
Source: WMP DAL 2019


Heatmap of Percentage of Responses on Target [+20% Crew]
by Crew mix, with breakdown by Priority, Rad, Day of week and Hour
% Double Crew
10%
20%
30%
40%
50%
60%
70%
80%
90%
MTWTFSS
P1 Bournville
P1 Wednesfield
P1 Willenhall
P2 Bournville
P2 Wednesfield
P2 Willenhall
9pm
1pm
6am
1am
% on Target
100
75
50
25
0
Source: WMP DAL 2019


Heatmap of Percentage of Responses on Target [+30% Crew]
by Crew mix, with breakdown by Priority, Rad, Day of week and Hour
% Double Crew
10%
20%
30%
40%
50%
60%
70%
80%
90%
MTWTFSS
P1 Bournville
P1 Wednesfield
P1 Willenhall
P2 Bournville
P2 Wednesfield
P2 Willenhall
9pm
1pm
6am
1am
% on Target
100
75
50
25
0
Source: WMP DAL 2019

9 References

Birks, Daniel, Michael Townsley, and Anna Stewart. 2014. "Emergent Regularities of Interpersonal Victimization: An Agent-Based Investigation." *Journal of Research in Crime and Delinquency* 51 (1). Sage Publications Sage CA: Los Angeles, CA: 119–40.

Boscoe, Francis P, Kevin A Henry, and Michael S Zdeb. 2012. "A Nationwide Comparison of Driving Distance Versus Straight-Line Distance to Hospitals." *The Professional Geographer* 64 (2). Taylor & Francis: 188–96.

Cihan, Abdullah, Yan Zhang, and Larry Hoover. 2012. "Police Response Time to in-Progress Burglary: A Multilevel Analysis." *Police Quarterly* 15 (3). Sage Publications Sage CA: Los Angeles, CA: 308–27.

Coelho, Guilherme F, and Luiz R Pinto. 2018. "Kriging-Based Simulation Optimization: An Emergency Medical System Application." *Journal of the Operational Research Society* 69 (12). Taylor & Francis: 2006–20.

Cooper, JC. 1983. "The Use of Straight Line Distances in Solutions to the Vehicle Scheduling Problem." *Journal of the Operational Research Society* 34 (5). Taylor & Francis: 419–24.

Coupe, Richard Timothy, and Laurence Blake. 2005. "The Effects of Patrol Workloads and Response Strength on Arrests at Burglary Emergencies." *Journal of Criminal Justice* 33 (3). Elsevier: 239–55.

Elliott-Davies, M, J Donnelly, F Boag-Munroe, and D Van Mechelen. 2016. "'Getting a Battering' the Perceived Impact of Demand and Capacity Imbalance Within the Police Service of England and Wales: A Qualitative Review." *The Police Journal* 89 (2). SAGE Publications Sage UK: London, England: 93–116.

Houdmont, Jonathan, Mary Elliott-Davies, and Joan Donnelly. 2019. "Single Crewing in English and Welsh Policing: Frequency and Associations with Violence Towards and Injuries in Officers." *Policing and Society* 29 (7). Taylor & Francis: 820–33.

Johnson, Shane D, and Elizabeth R Groff. 2014. "Strengthening Theoretical Testing in Criminology Using Agent-Based Modeling." *Journal of Research in Crime and Delinquency* 51 (4). Sage Publications Sage CA: Los Angeles, CA: 509–25.

Kirchmaier, Tom, and others. 2015. "The Effect of Police Response Time on Crime Detection." Centre for Economic Performance, LSE.

Malleson, Nick, and Mark Birkin. 2012. "Estimating Individual Behaviour from Massive Social Data for an Urban Agent-Based Model." *Modeling Social Phenomena in Spatial Context*. Lit Verlag Berlin, 23–30.