Detective Superintendent William Hodgkinson

Candidate Number: Pol-1942
Fitzwilliam College
Supervisor: Dr Barak Ariel
Year: 2021

Thesis Title: 'Comparing Two Different Alarm Systems for High-Risk Domestic Abuse Victims: A Randomised Controlled Trial'

Word Count: 17,964

Submitted in part fulfilment of the requirements for the Master's Degree in Applied Criminology and Police Management

RESEARCH CONTRACT

Key Research Question: Which system of personal alarms produces greater reductions in repeat victimisation and harm suffered and improves criminal justice outcomes when used for high-risk domestic abuse (DA) victims: a system of silent alarms with no sticker on the door and no audio recording of voices in the home, or an alarm system advertised on entry point stickers that includes an audio recording system sounds in the home?

Sub Research Questions:

A) How many activations of the alarm, on average across all victims, were there in the six months after alarm request?

B) How many DA crime reports were recorded in the six months prior to alarm request and the six months following the request for alarm installation?

C) What was the before-after difference in the six-month total number of police recorded DA incidents against the participating victims, and how does the mean of these totals compare between the two treatment groups?

D) What was the six-month total of Cambridge Crime Harm Index (CCHI) value for offences committed against each victim before and after the request for an alarm?

E) How many calls for service (CFR) were there for each victim in the six months after the request for an alarm?

F) What was the before-after difference in the six-month total number of CFR by caller type and how did this compare between the two treatment groups?

G) How many DA related arrests occurred in the six months post alarm request and how did this compare between the two treatment groups?

H) How many DA related charges occurred in the six months post alarm request and how did this compare between the two treatment groups?

I) What can be concluded about which alarm system is more effective at protecting victims and in improving criminal justice outcomes?

Research Design: Randomised Controlled Trial (RCT)

Unit of Measurement: High-risk DA victims.

Data and Methodology: The data to be used is:

A) Crime counts (all crime) and crime counts (DA only crime)

B) Crime harm (all crime) and crime harm (DA only crime). Crime harm is measured through use of the Cambridge Crime Harm Index (CCHI).

C) Calls for service (CFS)

D) Police units deployed (PUD)

E) Alarm activations

F) Arrests

G) Charges

Methodology Used: This is a simple trickle-flow RCT conducted across 13 London boroughs covering a population of 3,676,154 people. The identification of suitable cases begins when an officer working in one of the 13 boroughs identifies a high-risk DA victim they feel would benefit from a panic alarm. The officer contacts the victim to confirm that they were willing to accept a panic alarm. If supportive, an application to National Monitoring is made using their online platform. Upon receipt of the request and confirmation that the case is suitable the alarm request is randomised. The two possible outcomes are

A) Install an RDA2 - National standard alarm

B) Install an RDA3 – Audio recording alarm with visual warning sticker

Cases will be randomised in blocks of 10. Following randomisation, an appointment is then made with the victim and engineers attend to install the randomised alarm. Analysis covers the six months pre and post alarm request date.

Analytic Methods: The methods used to analyse the data will be a combination of descriptive statistics covering sum, range, mode, mean averages and standard deviations, T-Tests, P-Values and Cohen's D. Two types of T-Tests will be used. The first are two-sample T-Tests (between groups comparisons). The second are paired T-Tests (before-after comparisons).

Key Findings:

A) The between-group comparisons post-randomisation find no treatment effect for crime counts (All), crime counts (DA), crime harm (all), crime harm (DA), CFS, PUD and activations.

B) Following the identification of a high-risk DA victim, responses that incorporated a panic alarm achieved statistically significant reductions in harm and repeat victimisation.

C) The one area that could be considered as delivering a meaningful treatment effect is arrest. Differences between the RDA2 treatment and the RDA3 treatment are statistically significant at the 0.1 level. The RDA3 alarms generated double the arrest rate of RDA2 alarms, a mean of 0.2 arrests per victim compared to 0.1 arrests.

Policy Implications of the findings:

A) The findings support policies that allows for victims to have a choice of which alarm is installed.

B) Policies aiming to increase levels of arrest in high-risk DA victim cases could consider greater use of audio alarms.

C) Policies should seek to require the gathering of data to better understand whether one alarm is preferable over another is specific circumstances. The findings support a possibility of tailoring alarms to victims and achieving better reductions in harm and repeat victimisation.

D) Policies addressing which panic-alarm to use in DA should consider the intrusion into the private lives of victims.

ACKNOWLEDGEMENTS

This research would not have been possible if it were not for the support throughout. First and foremost, I formally thank my family. My wife, Laurel has shown enormous levels of patience and understanding and our children Flora, Celeste and Monty have lost time with their father during the last 2-year period. My parents-in-law, Michael and Fiona have provided a much-needed sanctuary in North Yorkshire to progress this work. Thank you for the Wi-Fi re-boots, procuring of Sue and Peters cottage and provision of an umbrella for daily grip and tracking calls, which were on occasion taken up a hill and surrounded by sheep in order to obtain phone reception. I could not have completed this work without your support.

Secondly, I wish to thank the Metropolitan Police Service (MPS) for funding this work and affording me the opportunity to challenge myself and my views on policing. This investment has led to an exposure to knowledge that has fundamentally changed the way I approach policing challenges. The MPS Intelligence Directorate (MO2) supported this research with time and can-do attitude. This helped deliver the trial and obtain the data required for the analysis. Detective Inspector Adam King was my deputy during this work and aside from his time keeping during the morning meeting, he was an absolute rock of consistency. Thank you for your humour, perseverance and tenacity.

Significant analytical support and manual data checks were needed in this research due to police software limitations. It was delivered with aplomb by my colleague and lead analyst Rachel Brown. Tragically whilst cycling to work to avoid public transport during the pandemic Rachel lost her life in a collision, a huge loss that rocked the lives of many. Rachel was one of best. I know her family are incredibly proud of her work focused on domestic abuse (DA). Without doubt this has helped so many of the most vulnerable. The new digital-audio alarm

created by National Monitoring (NM) is named RB25 as a small nod to her importance in this work. Rachel will never be forgotten.

Thirdly, I am very thankful to Bedfordshire Police for giving me time to complete this thesis. I am incredibly proud to now work in Bedfordshire and look forward to doing all I can for my community.

Fourthly, it is my view that partnerships work best when an outcome is jointly agreed upon, communication is effective, trust levels are high and partners are highly motivated. This would describe my relationship with NM. Without Paul, Emily F-N, Emily, Elaine, the engineer team, call handlers, graphics lead and manufacturing team this would not have been possible. In under two-years the team, took an idea to prototype, tested it and supported a randomised controlled trial (RCT). You are all epic. Working in partnership like this is incredibly exciting and immensely enjoyable. This work would not have been possible without you, you represent a key part of our policing partnership nationally.

Lastly, I want to say thank you to all at the University of Cambridge who have supported me and this research. The student journey at Cambridge is the richest I have experienced. Whilst the architecture and course content are things to behold, it is the people who shine the most. I hope I have not been the most troublesome student on the course, albeit I do probably hold the record for most extension requests. You have given me the gift of knowledge. This has opened up trains of thought and possibility that were not there before. Thank you.

Table of Contents

LIST OF TABLES

LIST OF GRAPHS

LIST OF IMAGES

LIST OF APPENDICES

1.0 INTRODUCTION

Domestic Abuse (DA) crime reports have been increasing year-on-year, resulting in 758,941 crime reports in the year ending March 2020 excluding Greater Manchester Police (ONS 2020). Within this total volume, the number of repeat victims is also increasing. In the same period, 124 victims of domestic homicide were recorded, a homicide every three days (ONS 2020). The cost of DA was estimated to be £66 billion in 2017 (Oliver et al 2019), with increases in crime reporting this is likely to now be nearer £80 billion a year. Given the startling figures associated with DA, it is surprising not to find a wealth of high-quality quantitative research focused on ensuring police practice is doing all it can to reduce levels of harm and repeat victimisation.

This research, as part of an MSt in Applied Criminology and Police Management may well be the first RCT to test technological tactics used to prevent high-risk DA. This research sets out to test a technological innovation in panic alarm technology against the national standard panic alarm system for England and Wales. This trial is set within London and the district of the Metropolitan Police Service (MPS) during 2020.

Results show that audio alarm technology complimented with visual warning stickers at the point(s) of entrance to the homes of high-risk DA victims, overall delivered no statistically significant treatment effect when compared to the national standard alarm. Following identification of high-risk DA victims both alarm systems as part of an array of police and partnership activity achieve statistically significant reductions in harm and repeat victimisation. Arrest rates are the only difference worthy of note, with audio alarms achieving statistically significant increased arrest rates at the 0.1 level.

2.0 LITERATURE REVIEW

The policing practice addressed in this research paper is DA, a widespread, pervasive problem (Sherman et al 2017) that is highly under-reported (Sherman et al 1997). DA is a challenge layered with complexity. This review tries to de-layer that complexity with an initial appraisal of what is known on targeting and testing. The second half of the literature review specifically focuses on this research and seeks to understand whether supplementing a panic alarm system for high-risk DA victims with audio recording capabilities causes a reduction in subsequent harm and repeat victimisation. The literature reviewed will focus on crime prevention theories, the evidence base and a specific focus on warning stickers, traceable liquids and visual cues, body-warn video (BWV), the use of alarms to deter unwanted behaviours and specifically panic alarms for DA.

2.1TARGETING AND RISK FACTORS

Literature focused on how to effectively target in DA is comparatively advanced (Sherman 2013), yet many facts are claimed about DA which are not supported by scientific evidence (Sherman et al 2017). Evidence shows that most first-time DA reports within a couple are the last (Bland and Ariel 2015). This is the case across all offence types including deadly harm. Efforts led by Thornton (2017) to identify risk factors that could be used to predict deadly harm identified that in the majority of cases serious assaults came 'out of the blue', with no prior police involvement. These findings have since been replicated (Chalkey and Strang 2017; Bridger et al 2017; Button et al 2017; Barnham et al 2017). The best risk factor for predicting future domestic homicide is an offender's prior suicide attempt, threat or ideation. The challenge is that organisations outside of policing typically hold this information and data-sharing is not frequent.

With most first-time DA victims not requiring a policing response for a second time and information on the best predictor of domestic homicide sitting outside of policing, our focus moves to the 'power-few' (Sherman 2007), those victims who suffer the highest levels of harm and/or repeat victimisation. Among offender's subject to three or more DA cases, an escalation in frequency and probability of re-offending is seen (Sherman et al 2017). This targeting insight applies to the MPS, the geographical area of focus of this research.

The MPS is a Constabulary area situated in the English capital, with a policing demand of over 140,000 DA reports a year (MPS Internal Report 2020). London is estimated to have 9.2 million residents (London Assembly 2020). Some 2,446 Londoners have reported DA on four or more occasions in the last year (MPS Op Athena Performance Pack: November 2020). At three instances the victim should be considered for referral to their local Multi-Agency Risk Assessment Conference (MARAC) along with cases that have not reached this volume threshold, but are considered 'high-risk' by professionals. The cohort of 2,446 victims represents just 0.027% of the London population. These victims are at the top end of the harm pyramid (Sherman 2018) and account for at least 7% of all demand, or 259 times more victimisation than if victimisation was spread equally across the population. This is not a new finding. A targeted approach to DA offers the possibility of achieving significant reductions in harm and repeat victimisation (Grove et al 2012). Concerningly whilst targeting occurs operationally across England and Wales and risk factors of homicide are better understood, the same cannot be said for testing. Understanding 'what works' at a tactical level is the area of focus for this research.

2.2 TESTING WHAT WORKS: A REVIEW OF RCTS

The methodological quality of evidence is measured using the Maryland Scientific Method Scale (SMS). At the top-end of the scale are RCTs that are able to draw a causal link between a treatment or practice and effect seen. At the bottom end of the scale is simple comparison research. This can suggest a relationship between two or more factors, but is unable to evidence 'causation'. This research is an RCT. As such the literature focuses on studies using this methodology to better understand 'what works' for DA.

2.2.1 - MINNEAPOLIS DA ARREST EXPERIMENT

In 1981 Sherman and Berk (1984) commenced an experiment to ascertain the deterrence effects of arrest on simple assault offences for DA in Minneapolis. This field experiment had three treatments. The first was arrest, the second advice and simple mediation and the third was an order to leave the premises for eight hours. The experiment had 314 accepted cases and delivered the treatment to a total of 264 cases, a fidelity rate of 84.1%. Randomization was achieved through use of randomly compiled coloured pads directing an outcome. Researchers observed that this approach led to some internal validity challenges. Sherman and Berk (1984) summarise that whilst it was unclear if any officers had subverted the design, a series of statistical methods were used to gain an accurate outcome looking at re-offending in the following six months. Reassurance as to internal validity was supported by victim self-reporting leaving the authors to summarise that internal validity was 'rather sound'. Results detailed that arrest was the most effective treatment at preventing repeat victimisation. Just 13% of those arrested committing a repeat assault in the six-month follow-up period. This compared to 26% for those who were separated. The experiment was able to show causation between arrest and a reduction in further reported incidents of DA.

Acknowledging that Minneapolis was not representative of all urban areas (Sherman and Berk 1984) this methodology was replicated across six different sites in the US. The same outcomes were not found (Wilcox and Hirschfield 2007). Whilst the external validity of the research is problematic, the outcomes documented are not. The research did show that arresting a perpetrator in Minneapolis led to a reduction in reports to police.

2.2.2 Minneapolis DA Arrest Experiment Re-visited

Some 23 years after the initial study the cohort were revisited to see what the longer-term implications of arrest had on victim morbidity (Sherman and Harris 2015). Research found that victims whose partners were arrested, were 64% more likely to have died of all causes than when the perpetrator was warned. The increase in morbidity was even greater amongst black victims. Within the African-American victim cohort an arrest increased chances of morbidity by 98% (Sherman and Harris 2015). Sherman and Harris (2013) also sought to understand the effect of arrest versus warning on perpetrators. Here suspects who had been arrested were almost three times more likely to have died from homicide. Whilst arrest did lead to fewer further allegations made to police in the six-month review period, longer-term harmful outcomes were documented.

2.2.3 Conditional Caution and Relationship ABUSE (CARA)

Between 2012 and 2015 in Hampshire, England the cautioning and relationship abuse (CARA) RCT was conducted. Eligibility criteria for this experiment was English speaking male perpetrators over 18 years of age, who had admitted to low-level, low-risk intimate partner violence and where the victim had not indicated that the use of conditional caution would put them at greater risk (Strang et al 2017). Both treatment and control groups accepted not

to have any repeat offending in the following four months. However, the treatment group received an additional two-day workshop. Workshops were five hours long and four weeks apart. The RCT achieved a fidelity rate of 91%, with 293 perpetrators in the study. Over the following 365 days the treatment group had a re-offending prevalence that was 35% lower and whose crime harm value was 27% lower than those re-arrested in the control group. Both repeat victimisation and harm levels were reduced. This study evidences that harm and repeat victimisation levels can be reduced by evidence-based policing. The limitation of this particular study is found within the offender type studied (Strang et al 2017). CARA does not focus on the power-few (Sherman 2007).

2.2.4 Integrated Case Management (ICM) Approach

In 2013, in Berkshire, England an RCT commenced seeking to address the causal factors of DA. Eligible couples had called police two or more times in the previous 12-months about DA. Offenders were male, victims were female. Both groups received the same set of interventions, but the treatment group received a far higher dosage of that intervention. Interventions included contact and engagement from Berkshire Women's Aid, one-to-one perpetrator counselling session and police and practitioner follow-up visits to the home address. In total, 180 couples were incorporated into the trial (Goosey et al 2017). Analysis was conducted in the two years after randomisation. Groups who received enhanced interventions achieved a mean reduction of 8.85 less harm compared to a 4.15 increase in harm in the control group. In relation to crime counts the treatment cohort generated 112 non-crime domestic events compared to 85 in the control. Greater repeat victimisation was recorded with the enhanced treatment. Delivering one outcome does not necessarily mean the other outcome is achieved. Limitations in the research methodology mean that it is

unclear how much intervention occurred in each case and which aspect of the interventions delivered the outcomes. This research does demonstrate another approach that has been tested with as much rigour as a new pharmaceutical (Sherman et al 2017) and shown to achieve a reduction in harm.

2.2.5 System-Initiated Warrants for Suspects of Misdemeanour Domestic Assault

Following the Sherman and Berk (1984) RCT testing police actions at the scene of a standard assault, Dunford (1990) identified that almost half of DA cases have no suspect present. Dunford proceeded to commence an RCT where suspects were not present. The control group followed the existing process of advising the victim that they could attend court and pay for a warrant for the suspects arrest. The treatment group had police apply for the warrant. Suspects were informed that warrants would be issued. Where police had applied successfully for a warrant the offender scored lower on both prevalence and frequency of offending than the control. This outcome was hypothesised to be down to a 'sword of Damocles' effect, with the suspect experiencing a constant threat of unknown penalties, leading them to behave.

It is accurate to say that there is a significant gap in knowledge as to 'what works' for achieving a harm and repeat victimisation reduction in high-risk DA. Sherman et al (2017) believe that the CARA and the ICM approach were the first two RCTs of policing DA in England and the first to show success in reducing harm to victims. These studies though do not focus on high-risk DA. CARA focused on low-level, low-risk offenders and ICM focused on couples who had contacted police two or more times in the last 12 months. In the US, Sherman and Berk (1984) tested the police actions at scene of low-level assault DA and Dunford (1990) tested DA

offending at the lower end of the spectrum. To this authors knowledge no RCTs have been completed specifically for high-risk DA cases anywhere in the world.

2.3 Crime Prevention

With an ambition to build up the evidence base of 'what works' to reduce harm and repeat victimisation in high-risk DA, it is important to understand the difference between primary, secondary and tertiary crime prevention. Primary crime prevention is focused on the whole population. Secondary crime prevention is more targeted and focuses on those at greater risk. Tertiary crime prevention focuses on those who have already become a victim (Radford et al 2011). This research proposal and all prior DA RCTs discussed focus on tertiary crime prevention (Radford et al 2011).

Theoretical frameworks surround the concept of crime prevention. Routine activity theory (RAT) is one of these (Cohen and Felson 1979). RAT specifies that in order for a crime to occur a motivated offender is required along with a lack of guardian and a suitable target. All three need to converge at a moment in time for there to be a heightened chance of a crime occurring. Deterrence theory posits that when certainty of apprehension is high and outcomes for that behaviour are swift and unattractive the unacceptable acts (Von Hirsch et al 1999) are less likely to occur and people revert to the social norms (Ariel et al 2015). These theories will form the framework when discussing the findings of this research.

2.4 Warning Stickers, Traceable Liquid and Visual Cues

A common 'preventative' tactic used for primary, secondary and tertiary crime prevention is that of warning stickers. These are an economic tactic designed to alert offenders of a heightened risk of apprehension. Stickers have been widely used to publicise neighbourhood watch schemes, the presence of burglar alarms or more recently the use of property marking liquids (Raphael 2015).

2.4.1 Traceable Liquid and Warning Stickers

Using a before and after trial methodology, Raphael (2015) concluded that the use of traceable property marking liquid was 'highly effective' at reducing burglaries. Here households were offered property marking by a uniformed officer and visual warning stickers were attached to the premises, so that they could be seen from the road. However, significant implementation challenges were documented and the causation claimed is challenged. The MOPAC Annual Report 2014-15 (MOPAC 2015) details that London achieved its lowest level of burglary since 1974. During this mayoral tenure seven key crimes, including burglary were identified with an ambition to reduce volumes by 20%. Significant focus was put on achieving this crime reduction with daily, weekly and monthly tracking. A claim that traceable liquid was 'highly effective' is a leap too far, due to the methodology used and implementation challenges.

A more advanced methodological approach was seen in Aarhus, Denmark where an RCT was conducted across a sample size of 6,603 households that had experienced one or more burglary in the prior 4 years. Control households were not contacted. Treatment households all received a letter encouraging them to sign up register for the experiment and request a

property labelling kit and warning stickers. Only 32% of households requested to be part of the experiment and successfully followed instructions on property marking and application of stickers. Results from this experiment found that even sending the letter had a positive impact on becoming a repeat burglary victim in the following 15.5 months with 21% fewer burglaries reported. This finding was statistically significant (P = 0.02). The burglary reduction was experienced more greatly amongst the 32% of the treatment who received the full treatment. Those who took up the full treatment were documented as having an intrinsic concern for burglary prevention. The drop-in burglary rates were reported as lasting for a limited period of time only. This research supports that simply having a targeted letter after suffering a burglary is more effective than no contact in reducing future repeat victimisation. (Kyvsgaard and Sorensen 2020).

What would be accurate to say about traceable liquid and warning stickers is that evidence supports that targeting households that have been burgled with information on crime prevention reduces levels of repeat victimisation. Evidence also supports that individuals who actively engage in crime prevention experience even lower levels of repeat victimisation. A view is held that this is because those individuals place greater importance on crime prevention and may have greater socio-economic wealth (Tilley and Webb 1994; Kyvsgaard and Sorensen 2020; Gaines and Bichler 2007). The specific effectiveness of warning stickers and traceable liquid remains unproven.

2.4.2 Operation Identification

Far earlier research in Sweden carried out by Knutsson (1984) under Operation Identification also sought to establish whether burglary could be prevented through the use of visual warning stickers, property marking and burglary alarms. Visual warning stickers were placed

at the point of entry to properties. An area of approximately 3,500 households was chosen on the outskirts of Stockholm and invisible ink pens were used to write unique identification numbers on valuable property. These numbers were recorded on a policing database to assist identification of stolen goods. Here the treatment area saw an increase in burglary compared to non-study areas. The trial concluded that property marking had no effect on crime reduction. A low take-up rate of the treatment was cited as a reason for this. Qualitative feedback from burglars found that whilst most had seen the warning stickers, they did not know what it meant. The area where an effect was found was in households where burglar alarms had been installed. Here a chi-squared test identified that households with a burglar alarm had just a 2.6% likelihood of being burgled in the four years after the Operation Identification treatment compared to 8.3% with no treatment and 7.9% in households with just the property marking and warning stickers. Whilst evidence supports that alarm systems may well carry a deterrent effect, the methodology used in the trial does not allow for a causal effect to be claimed.

2.4.3 Warning Stickers and Visual Cues

Doubts surrounding stickers as an effective deterrent in themselves are not just confined to policing (Mackinnon 1995). Research identified that warning stickers on prescription medicines often failed to attract the attention of the older population (Guy 2012). However, work within the healthcare setting focusing on the efficacy of visual cues (Nevo et al 2010) lauded the positive effect of large warning signs. This finding was documented during an RCT aimed at establishing whether increased hand-hygiene could be achieved by changing the visual cues. When large warning signs stipulating a requirement to wash hands were present upon the door of a patient's room, a 93.3% compliance rate in hand sanitisation was recorded.

This compared to a 33.3% compliance rate in the control setting. A review of the methodology casts doubts as to the treatment efficacy. The sticker was not a treatment in isolation. Prior to healthcare professionals entering the room a researcher stopped the individual at the door, pointed out the warning sign and read it out aloud to them. This warning sign detailed a threat of being reported and explained that the room contained an alarm that would be triggered if the individual did not sanitise their hands. It is perhaps surprising that given the level of supervision and threat that 6.7% of healthcare professionals still failed to wash their hands. It is clear that the difference in treatment efficacy was caused by more than the sticker. The work does though offer some insight into visual cues and deterrence theory. For visuals cues to be effective they require an ability to convey a hazard or threat with significant personal impact, the sign needs to be visually attention grabbing and at eye level, should be in close proximity to the location of the threat and needs to be viewed by a reasoned individual to have the desired impact (Nevo et al 2010). Additional insight is offered by Jones and Nowell (1973) who identify a link between the colour of visual cues and physiological responses. The colour red is identified as conveying a message of danger in the natural world (Brown et al 2000). Additionally, Cohen and Sloan (2016) document that strong visual contrast is needed in visual cues because one's gaze is drawn and attracted to stimulus change. The literature supports that whilst stickers have been widely used to deter offenders, no high-quality evidence exists to support this working as a tactic in isolation and no evidence can be found to see whether they work in DA settings.

2.5 BODY WARN VIDEO (BWV)

Invoking a behavioural change as a result of technology and visual cues was the ambition of BWV and its roll out across policing. A meta-analysis of BWV RCT's found no clear or consistent

effects on most citizen or officer behaviour (Lum et al 2020). The rapid roll out had been completed without clarity as to what would be achieved. Whilst viewed as a valuable evidence gathering tool the aim of BWV was to increase police accountability, reduce complaints, use of force, increase the levels of confidence in policing and reduce attrition of criminal cases in the criminal justice (CJ) system (Drover and Ariel 2015). The systematic review concluded that BWV did achieve a mean relative reduction in complaints of 17% (Lum et al 2020). The reason(s) behind this though were unclear. Self-awareness of being watched was identified by Ariel et al (2015) as one reason for why behaviour changed in both citizens and officers. This was linked to deterrence theory. The combination of video capture increasing the likelihood of apprehension and subsequent sanction led to a behavioural change. In order for this effect to take place, visual cues are needed to alert individuals to the presence of BWV. Even small cues indicating that someone may be watching (Ariel et al 2015), the passive presence of close circuit television cameras or presence of mirrors may invoke a reaction leading to more socially acceptable behaviours. There is compelling evidence that awareness of being watched or observed, leads to greater socially acceptable behaviours. The literature now examines whether the same effect is seen where individuals are aware that an alarm is present.

2.6 BURGLAR ALARMS

Personal and property alarms are a widely used tactic aimed at reducing levels of offending or harm. In the US approximately 17 million intruder and fire alarms have been installed (Hakim et al 1995). However, their effectiveness has been called into question particularly in the US where 8-25% of all calls for service (CFS) generated from burglar alarms are for false alarm activations. This is estimated to generate 36 million false alarm activations costing $1.8

billion, or the equivalent of 35,000 police officers every year (Gaines and Bichler 2007). Burglar alarms are cited as having two key functions. The first is to deter and prevent burglaries. The second is to facilitate detection and apprehension of burglars and other criminals. With an arrest for burglary being made in just 0.08% of police responses (Gaines and Bichler 2007) the evidence supports that alarms here are not successfully achieving the second aspect of their purpose. The first purpose of alarm systems is prevention or deterrence. The deterrent effect of burglar alarms was cited as being evidenced in data from Operation Identification (Knutsson 1984). Yet, it is unclear whether this was due to the deterrent effect of the alarm or the type of individual who would have a burglar alarm. Doubt surrounding the efficacy of burglar alarms is increased in the work of Buck et al (1993). After two and a half years of observations the research concluded that burglars first choose a neighbourhood to burgle based on familiarity and proximity to a thoroughfare, then a street and then a property. Whilst a burglar alarm is cited as reducing the likelihood of being burgled if all other things are equal, the property is the third stage in decision making (Buck et al 1993). It is possible that often cited reductions in burglary caused by burglar alarms are simply due to the type of property that would have a burglar alarm. These being properties with a higher socio-economic wealth and where existing target hardening has been completed. What is clear is that alarms have created an industry that costs billions (Hakim et al 1995).

2.7 Panic Alarms in Policing

Personal panic alarms, both static and mobile have been used widely in England and Wales in the policing of DA for nearly three decades. Use of panic alarms for DA was first documented by Farrell and Pease (1993) as a basic system that worked with landlines and 'alarm pendants'. When the alarm was triggered an automated call was made to the local police station and

police were despatched to the address. Farrell and Pease (1993) believed that an effective way of targeting crime prevention was to focus on the most powerless in society, those repeat victims of DA. Theoretical discussion focused on how the installation and response to panic alarms could reduce repeat victimisation. However, no scientific evidence is found to support panic alarms reducing repeat victimisation. Farrell et al (1993) explore in greater detail what Sherman (2007) labelled the 'power few'. They identify that a large proportion of the total DA demand came from a small number of households. Effective targeting data was present as early as 1993. The challenge centres on the lack of high-quality testing of tactics or treatments.

The work of Walker (2001) is the only study identified that sets out to establish the effectiveness of panic alarms as a tactic for preventing repeat victimisation and harm for DA. Whilst this study identifies that alarms have the potential to deter perpetrators it overreaches in its conclusions and lacks evidential rigour. Only four alarms were installed during a six-month trial. This does not constitute the high-quality evidence needed to assess whether alarms operate as an effective deterrent to perpetrators. Further doubt is cast as to the effectiveness of panic alarms in victim feedback (Radford et al 2011). Here a victim articulates their doubt of alarms, stating that she thought the police would get to her just as fast if she dialled 999.

There has been limited change to alarm functionality in some three decades. In that time over 150,000 alarms have been installed by NM on behalf of 19 police force areas across England and Wales. This tactic, targeted against high-risk DA victims needs to be better informed by scientific evidence. The literature review has found that alarms with or without visual warning stickers share commonalities. Both are widely used and aimed primarily at deterring

offending, but neither's use is supported by high-quality scientific evidence. In both cases evidence exists that casts doubt as to their effectiveness.

2.7.1 Mobile Panic Alarms

Mobile alert systems represent a similar technology to panic alarms, the primary difference is that they operate outside of the home and use the global positioning system (GPS) to inform officers where the victim is located. This technology is used internationally (Natarajan 2016; Dowling et al 2018). Across England and Wales most forces use a system called Technical SOS (TecSOS). Natarajan (2016) sets out to evaluate the effectiveness of TecSOS, but conclusions made are not supported by scientific evidence. Instead individual incidents are used to support evidence lacking claims that are in at least one instance contradictory to prior 'advantages' listed. In this case the device is detailed as being discreet so that the offender will not know of its presence. Later on, it is suggested that the low device usage is because perpetrators were deterred by the presence of the device. Fundamentally this is contradictory to all that is known about deterrence theory. Behavioural change occurs when an individual is aware of being observed. Evaluations of such tactics require more rigorous methodologies than the one used in this research. The testing of technology-based interventions requires the gold standard use of RCTs (Ariel et al 2015; Natarajan 2016) to understand the causal effects it has. These have to carefully consider the safeguards needed for victims (Dowling et al 2018).

2.8 Panic Alarms Beyond Policing

Panic alarms are used more widely than just policing and crime. Establishing an evidence base for panic alarms installed in wider contexts and settings may better frame understanding. One such setting is within the health service (Perkins et al 2017). Here panic alarms have been

used to try and reduce assaults on staff. Perkins et al (2017) complete a systematic review focusing on the effect of alarms and identify that whilst alarms are used widely in 85% of trust areas, there has been a lack of attention to establish whether they work to reduce levels of assault. The lack of a clear evidence base leads Perkins et al (2017) to identify the need for high-quality experiments and specifically RCTs, to understand whether alarms work to deter and prevent violence.

2.9 Literature Summary

Whilst DA is a challenge with significant complexity, targeting research allows for interventions to take place against those most likely to suffer repeat victimisation again. Whilst the targeting of those more likely to go on and commit a domestic homicide requires partnership information, the ability to identify the power-few (Sherman 2007) is established.

A number of high-quality RCTs have been completed for DA. The first sought to understand what police action at the scene of a low-level assault worked best for reducing repeat victimisation (Sherman and Berk 1984). This successfully identified that arrest was more effective than mediation and the temporary banning of an offender from the household. It shared key learning points for effective implementation and identified the need for repeat studies to build external validity. Replication of this methodology did not achieve the same results. Professor Lawrence Sherman further evidenced the importance of follow-up research. Whilst arrest worked to reduce repeat victimisation in the following six-months, over the next two decades victims and offenders who received the arrest treatment experienced statistically significant increases in morbidity.

The CARA study successfully completed an RCT targeting low level, low risk offending to prove that two sessions of perpetrator counselling was more effective at reducing both harm and

repeat victimisation in the following year. The ICM study successfully completed an RCT targeting DA couples who had called police on two or more occasions in the previous year. This study successfully showed that an integrated partnership approach with enhanced levels of contact reduced harm suffered, but also increased crime counts and CFS. Lastly, the use of police funded warrants for the arrest of perpetrators of misdemeanour DA was seen to reduce both frequency and prevalence of offending. An outcome hypothesised as a 'sword of Damocles' effect, the notion that one shows complaint behaviour when the threat is very real. None of the RCTs identified have provided an evidence base as to what works for high-risk DA. This research appears to be the first.

DA can be prevented. The theories framing this idea support that additional layers of deterrence can move a person's behaviour further towards socially acceptable behaviour. Warning stickers are widely used to communicate an increased likelihood of being apprehended. Research testing whether warning stickers work has over-reached in its conclusions and been unable to isolate warning stickers as a treatment by itself. However, the simple sending of a letter to encourage burglary victims to take up crime prevention was shown to reduce repeat victimisation (Kyvsgaard and Sorensen 2020). Small benefits were seen when burglar alarms were combined with stickers (Knutsson 1984), but the ability to link this to the presence of a sticker is doubtful. Perpetrators did not understand what the sticker meant. Outside of policing, research focused on use of warning stickers in prescription medicine and within a healthcare setting adds further doubt as to efficacy of this treatment to promote favourable behaviour. It is suggested that if warning stickers are to work, they need to be at eye level, visually striking and that the use of red in stickers has inherent links to danger across the natural world.

A visual cue and deterrent that has been seen to achieve a positive outcome is BWV. A systematic review (Lum et al 2020) identifies that BWV achieves a 17% reduction in complaints. Greater self-awareness of being watched or observed may positively alter behaviour. A second piece of technology aimed at deterring socially unacceptable behaviour is burglar alarms. Burglar alarms have two primary functions, to deter criminals and to raise the alarm more quickly to support their apprehension. This tactic is widely used globally at the cost of billions of pounds to policing with false alarm activations a significant problem. Apprehension rates were shown to be very low, whilst some evidence exists as to the potential deterrence effect of burglar alarms. However, a commonly held view is that other factors may be more important in reducing the likelihood of being burgled.

Alarms are not just confined to the protection of property. For three decades, panic alarms have been used in policing DA. Yet, there have been no high-quality experiments to establish the effectiveness of panic alarms or like technology across both sector and international boundaries, despite research identifying the need for rigorous research and use of RCTs to establish the causal effect of an alarm. The research proposed will be the first to do so and will close the knowledge gap. In this sense the proposal follows a similar path to the research of Ariel et al (2015) seeking to establish whether BWV works.

3. RESEARCH METHODOLOGY

3.1 RESEARCH SETTING

The setting for this research is London and the MPS Constabulary area. The population of London was 8,904,081 in 2018 (London Councils 2018). London is one of the most diverse global cities in the world with 1,981,000 people not of UK birth, 22 % of the population. The

MPS contains 32 Local Authority (LA) areas. This research takes place within 13 of the 32 LA areas and within five of the MPS's 12 Basic Command Unit (BCU) areas. The trial area contains 3,676,154 people (Table 1.0), 41.29% of London's population (London Councils 2018).

Table 1.0 – Research setting by BCU, LA Area and Population (London Councils 2018).

MPS BCU	**LA Area**	**Population**
East Area (EA)	Havering (KD)	257,810
East Area (EA)	Barking and Dagenham (KG)	211,998
East Area (EA)	Redbridge (JI)	303,858
Central North (CN)	Camden (EK)	262,226
Central North (CN)	Islington (NI)	239,142
South East (SE)	Lewisham (PL)	303,536
South East (SE)	Greenwich (RG)	286,186
South East (SE)	Bexley (RY)	247,258
Area South (AS)	Lambeth (LX)	325,917
Area South (AS)	Southwark (MD)	317,256
South Area (SN)	Croydon (ZD)	385,346
South Area (SN)	Bromley (PY)	331,096
South Area (SN)	Sutton (ZT)	204,525
	Total	3,676,154

Geographically across London the research setting has a blend of inner and outer LA areas, with eight of the boroughs selected coming from south London and five from north London (Image 1.0). The selected areas cover some of the most demanding areas to police.

Image 1.0 – BCU Area Map of London.

3.2 RESEARCH TIMELINE

In April 2019 the MPS and NM started to work in partnership to innovate with new technology under a shared aspiration of better protecting DA victims across London. On December 16th, 2019 a pilot phase commenced to test a new audio alarm system named, the RDA3. The pilot ran until February 19th 2020. The pilot area was Lewisham (PL). It was selected due to prior working and existing relationships in that geographical area. The pilot stage identified and overcame a significant number of challenges (Appendix A). After 65 days the audio alarm system was running effectively, with audio recordings available to police for review within four minutes of their arrival at the alarms location. A headline summary of the pilot is provided in (Table 2.0).

Table 2.0 – Headline Summary of Pilot.

Length of Pilot = 65 days
Total Alarms Requested = 59
Total Alarm Activations = 10
Total Successful Alarm Installations = 50 or 84.7%
Total Failed Alarm Installations = 9 or 15.3%

The full trial commenced on February 20th 2020 and ran until May 27th 2020, when case 300 was achieved. Analysis of DA victimisation focused on the six-month period prior and post alarm request. This finished on November the 27th 2020.

3.3 TRIAL DESIGN

This is an RCT using a simple trickle-flow process. Police officers operating across the geographical areas (Table 1.0) identified cases of DA that they consider to be high-risk. The terminology of high-risk here is based upon the officer's professional judgement. It is not based on risk assessment frameworks such as Domestic Abuse, Stalking and Harassment (DASH) or Domestic Abuse Risk Assessment (DARA).

Police officers assessed whether a panic alarm would assist in managing the risk to the victim. If they considered that use of an alarm would mitigate or reduce risk the officer contacted the victim to ascertain whether they would support an alarm being installed. If this was agreed, the officer applied for an alarm using the NM Online platform.

The platform was configured to identify only cases detailed as 'Domestic Abuse' and where the officer making the submission was working within one of the five BCU trial areas. There was no limit to the number of cases allowed across the five BCU areas. Officers applying for the alarm received a trial awareness input, but the system was designed to operate on the basis that the officer applying did not need to be aware. Importantly, the trial design operated

through automation. From the officer's perspective nothing changed to the way they operated when applying for an alarm. The officer applying did not get to select whether their application was included in the trial and they could not influence the alarm installed. During the trial no changes were made to trial design.

3.5 Random Assignment

Upon receipt of an alarm request partners within the control rooms of NM would see a prompt appear upon their computer screen to alert them to a 'trial case'. All NM operatives had received a briefing in person from NM leadership and were supported by three team leaders who worked across the full spectrum of shifts. The alert or prompt can be seen in Image 2.0.

Image 2.0 – NM Randomizer prompt

With a trial case successfully identified, the NM operative then logs into TheRandomiser using www.therandomiser.co.uk and is asked two qualifying questions. The first to check whether this was a DA case and the second was to confirm that the alarm was to be installed in the trial area. The crime reference served as a unique reference number.

The crime reference number was generated from the MPS Crime Reporting Information System (CRIS). The operative then submits the case for randomisation and receives one of two responses:

A. Install RDA3 (Audible Alarm)

B. Install RDA2 (Standard Alarm)

3.6 Type of Randomisation

TheRandomiser is a widely used randomisation platform and follows on from the Cambridge Randomiser. TheRandomiser randomly assigns cases in blocks (Linton and Ariel 2020) and works with a MySQL function "ORDER BY RAND()". This methodology ensures that neither the researcher nor the treatment provider can influence the outcome. The block size of the randomisation is 10. This means that in every 10 randomisations performed sequentially, five will be control and five will be treatment (Linton and Ariel 2020). Additional features within TheRandomiser ensure that the same case cannot be entered twice. The creators of TheRandomiser detail that at the time of publishing, 5,778 Randomisations have been completed with no software errors identified.

3.7 Interventions

Following successful randomisation, the case is moved onto a 'job sheet'. The NM control room team contact the victim using contact detailed provided by the requesting officer to

arrange a date and a time for installation. When a time has been agreed upon the case gets moved to an engineer. Across London, a team of engineers operate, supported by a team leader. The engineer is instructed to install either an RDA2 or an RDA3. Installations typically occur within 24 hours.

3.8 RDA2 – NATIONAL STANDARD ALARM (CONTROL)

The RDA2 standard alarm is a piece of hardware (Appendix B) installed within the home of a victim. The engineer plugs the device into the mains electricity, tests and explains to the victim how the technology works. If mains electricity ceases for more than an hour the system sends an alert to NM. Operatives attempt to contact the victim. If the alarm is without mains electricity for three hours the case escalates. At this stage, the officer in case (OIC) is contacted and then the Force Control Room (FCR) who arrange for a welfare check to take place. The battery lasts approximately 12 hours and then the device shuts down.

Method of Activation: A simple handheld 'fob' is given to the victim. The fob contains two buttons. Both of these buttons need to be pressed down simultaneously for two seconds (Appendix D).

Policing Response: At the point of activation NM operatives based in a control room (Appendix H) receive an alert. This is complimented by the lighting on top of their workstations flashing blue to signal a panic alarm activation. Following an activation, the pre-recorded details of the location, victim, case type are automatically sent into the MPS control room. This generates a Computer Aided Despatch (CAD) reference. The officers responding are directed to turn on their BWV. The CAD is passed to a despatcher who operates within a defined BCU area. The despatcher communicates the need for an immediate response to the pre-assigned location and details that this is a high-risk DA alarm activation. A police unit is

despatched to the location on an immediate grade (I Grade) call. This has a 15-minute response time attached.

3.9 RDA3 – AUDIBLE ALARM (TREATMENT)

The RDA3 audio alarm is a piece of hardware (Appendix C). In this trial it was accompanied by an RDA2. The reason for this was to ensure that if the new technology within the RDA3 failed, the victim would still receive an immediate response. The RDA2 alarm acted as a back-up alarm. Only if the RDA3 technology failed did the RDA2 technology operate. The RDA3 also required mains electricity to power it and defaulted to a battery pack in emergencies. The battery lasted for four hours, but due to the presence of an RDA2 system the same protocol was followed when notification of a mains failure had occurred.

At the point of an RDA3 installation, the engineer is required to complete two different activities to allow for installation. The first is to complete a consent form (Appendix F). The engineer explains that the RDA3 will always be recording what it can hear within the address. The audio recording operates on a loop, meaning that only the last five minutes of audio will be stored at any one time. The second part of the audio recording is for a live audio link to be established with the control room to assist in understanding the risk and recording evidentially what is happening. The second activity is to apply visual warning stickers to doors that enter into the property (Appendix G). The stickers are required to warn persons entering the property that the address contains audio recording technology. Following victim feedback testing four different stickers within the pilot phase, only red stickers were offered. Every sticker was photographed, retained as a record and checked for positioning compliance. Stickers were required to be either on the door or directly to the side of the door and be in a highly visible location, ideally at eye level. Due to the design of certain doors, eye level was

not always achievable. In this case the next best available location was selected. Following this agreement, the engineer explained how to activate the alarm. The alarm was then tested to ensure that it was operating without issue.

Method of Activation: The same handheld 'fob' was given to the victim. Method of activation was exactly the same (Appendix D).

Policing Response: Activation of the audio alarm generates the same immediate despatch of a police unit to the pre-established location. However, the call type detailed on the message changes to 'Audible Panic Alarm Activation' opposed to 'Panic Alarm'. The officer is told on the CAD message that this alarm activation will mean that there will be an audio recording available which they are required to review at the scene. The officer is told to turn on their BWV.

Audio Technology: At the point of an alarm activation, a live audio link is established using the fourth generation of mobile phone technology (4G) network. Following the blue flashing alert into the NM control room (Appendix H) an operative listens to the live link. When anything clear and urgent was heard, the NM operative would call into the MPS control room to alert them to the risk. This would generate an additional CFS. When officers are heard to arrive, the NM operatives again call into the MPS control room to confirm their attendance. When confirmed NM close the live audio link. This triggers the RDA3 to start sending the post-activation audio recording. The audio arrives within four minutes into an accessible folder location that can be found within four clicks of the MPS Intranet page. This trial used Box, a secure cloud-based storage system pre-approved for use within the MPS. The route to finding the audio is detailed on the CAD. The CAD is further updated to inform the officers of the availability of audio evidence and a requirement for them to listen to it. Following the sending of the post-activation audio the five-minute pre-activation audio is then sent from the RDA3.

This file arrives in the same location as the post-activation audio. Both are recorded under the CAD reference number for ease of identification. In total around 20 minutes of audio is captured for each RDA3 activation.

Every morning at 08:10 hours the last 24 hours of alarm activity is reviewed. For RDA3 activations this also means a reviewing of the audio. Following this review a member of the research team dials into the BCU's Safeguarding Unit morning management meeting to outline what was heard, check that the recordings have been reviewed and ask for an update on the investigation. Interventions took place in cases where audio reviews clearly challenged the view of initial responding officers that no crime had taken place. When this happened, the Safeguarding Unit reviewed the audio and re-engaged with the victim to establish what had happened. Where an order existed that prohibited the attendance of an individual such as through a Domestic Violence Protection Order (DVPO) the Safeguarding Unit checked the BWV and the audio to ascertain whether anyone present was prohibited from being there.

There are five differences in the RDA3 treatment compared to the RDA2 treatment. Firstly, the capturing of audio recording within the address from before the alarm activation. Secondly, the on-going recording and review until police arrival. Thirdly, the notification to police that audio recording is taking place and available for review. Fourthly, the presence of deterrent stickers at points of access to the property. Lastly, governance to ensure that investigating officers were aware of the presence of audio evidence.

3.10 Data Collection

Data collection was on high-risk DA victims. Data captured came from three primary sources. The first was CRIS, a system that has been used within the MPS since 1992. This was used to capture data on the victim, the offending against them, arrest status of named suspects and

outcomes of the investigations. The second was from NM who maintained a record of the randomisation result, alarm unit reference numbers, date of request, date of installation, sticker placement, any alarm installation refusals and the reason for refusals. The randomisation result was also stored within TheRandomiser. This was used to cross-check and ensure accuracy. The third was the MPS Call Management System (CMS). CMS is the system prior to police despatch via CAD. CMS captures all call demand into the MPS and breaks down demand based on caller type.

The data collected from CRIS proved a challenge due to a lack of accuracy within records. The same victim was commonly inputted with slightly different spellings. This meant accurate data could not easily be retrieved from a simple search. To overcome this, searches were completed against victim forename, surname, date of birth, address and the same corresponding categories for the offender. The data search parameters covered offending against the victim in the six months prior to alarm request and the six months after the alarm request. Stalking, Harassment and Controlling and Coercive Behaviour crime reports were treated differently due to national crime recording standards. When one of these crime reports was identified with on-going and continuing offending in the trial period, but created before the six-month pre-alarm request period this was included.

Data collection was supported by a team of analysts and researchers working within the MPS Intelligence Command. Search parameters for the data collection were retained and dip sampled for accuracy by a Lead Analyst. Data was merged from CRIS and NM to form a single master Microsoft Excel document. In order to obtain CFS data from the CMS data extraction a significant amount of manual work was required. System limitations meant that in order to understand CFS accurately, data needed to be abstracted through the address location

detailed on CMS. In order to capture every possible CFS the search was purposively wide. For an address of New Scotland Yard, 35 Victoria Embankment, London, England, SW1A 2JL the search was made to bring up all calls with 'Victoria' in the location within the relevant ward area. The MPS has over 600 ward areas. This search was required because the MPS first contact staff member does not utilise a structured address search look-up tool. This generated over 60,000 records which needed to be filtered and cleansed. Data was cleansed by matching this broad data set against other common aspects of the address. Following this data cleansing a manual process was completed. This checked the addresses listed against the recorded address for the victim from the master crime sheet and the caller to ensure CFS were recorded accurately where an incident occurred outside of the premises for instance. The same search parameters and methodology of data cleansing and checking was applied to both treatment and control. As a result of the data cleansing just over 2,500 CFS were identified. Following this and due to the process implemented by NM to clarify police arrival and relay urgent messages a further data cleanse was required against the caller to police. This cleansing removed the telephone numbers used by NM as well as a manual search for callers who identified themselves as working for NM. The second cleansing by caller name was required because a large number of records had no confirmed telephone number recorded against them. This achieved a data set of 2,267 CFS for 300 victims.

3.11 Data Analysis

This research sets out to understand which alarm system is more effective. Effectiveness is understood in relation to outcomes, outputs and system use. A key measurement for outcomes is harm reduction. The basic premise of a harm measure is that all crime is not created equally with some crime more harmful to victims than others. The two main methodologies used are the Cambridge Crime Harm Index (CCHI) which utilises sentencing

guidelines and converts the minimum sentence for a first-time offence into a score based on the number of imprisonment days detailed. The second is the Crime Severity Score (CSS). This harm index is used by the Office of National Statistics (ONS) and uses a mean score for the actual outcome at court to generate a harm score. There are benefits and limitations to both methodologies, but the CCHI will be used here due to its consistency in scoring which is relevant here for longer term analysis. The crime data collected will be converted into total crime reports and total crime harm score. A common practice across policing is to record domestic incidents that did not amount to a criminal offence such as a verbal argument as a 'non-crime domestic incident'. If a crime investigation was detailed as a 'non-crime domestic incident' this is recorded on CRIS. Here this crime report type will receive a harm score of zero, but is added into the crime report counts.

The methods used to analyse the data will use a combination of descriptive statistics, T-Tests, P-Values and Cohen's D. Descriptive statistics will cover the sum, the range, mean, mode and medium averages. Visual presentations of the data will be used to assist understanding. Two sample T-tests will establish whether the outcomes between the two different treatment groups (RDA2 and RDA3) are similar or not. Paired T-Tests will establish what the statistical difference is between the same group i.e. RDA2 at two different time points, which in this case is the six months pre-test and six months post-test. Analysis will be formed on the basis of intention to treat. P-Value measures will provide clarity as to whether differences observed between the six groups are statistically significant or down to random chance. Here statistical significance will be set at P=0.05. Where P is $\leq$0.05 the null hypothesis will be rejected. Cohen's D will be used to measure the effect size. This will establish if the effect is very small (d=0.01), small (d=0.2), medium (d=0.5) or large (d=0.8). Supporting this analysis will be the standard deviation for each measure, confidence intervals and degrees of freedom.

3.12 DATA PROTECTION

In order to complete this trial a Data Processing Agreement (DPA) was signed by the Commissioner of Police of the Metropolis and AVR Group Limited who operate and trade under National Monitoring Online (NM). The DPA enabled personal data to be securely stored and retained by NM. This was required in order to fulfil obligations under a national procurement agreement. The data stored by NM was the data inputted by the applying officers.

All data abstracted from CRIS and CHS and stored on a master Microsoft Excel Sheet was saved within a secure and auditable location within the MPS. Individuals could only access this information if invited to the folder area.

All data recoded within TheRandomiser was against the crime reference number. This unique reference number allowed for data look-up and checking within secured data sets in the MPS.

3.13 QUALITY OF RESEARCH DESIGN AND LIMITATIONS

A review of the literature has identified an inability across areas of business and international boundaries to provide clarity as to whether panic alarms and visual deterrence through warning stickers can 'work'. The term 'work' here has been defined as reducing levels of repeat victimisation and harm. The area researched is DA and the victim type are those considered by police officers to be high-risk. Distilled down, the purpose of a research design is to devise a strategy for finding something out (Maxfield and Babbie 2015). In order to understand whether panic alarms and warning stickers work the 'golden standard' for treatment outcome research has been used (Shadish et al 2002).

The golden standard of treatment outcome research is that of an RCT, a true experiment that has the ability to "isolate, define and explain the relationship between key variables in order to predict and understand the nature of reality" (Hagan 2006:78). True experiments require three key features. The first is to have two or more comparison groups receiving different treatments/interventions. The second is random assignment. The third is the ability to assess change in outcomes in the post-test stage within the two groups. Use of a true experiment methodology allows for "greater confidence in the validity of causal conclusions than is possible with other research designs" (Bachman and Schutt 2017: 178). For this reason, the research design has some significant strengths. When a true experiment methodology and an RCT is conducted in real world settings it is considered a 'field experiment' (Bachman and Schutt 2017:181).

The strengths of this research design are now discussed. Firstly, the method of assigning victims into treatment and control groups is through random assignment. This "eliminates the potential for systematic bias in selecting subjects" (Maxfield and Babbie 2015:175), creating statistically equivalent groups.

The second strength is fidelity of treatment. Fixsen et al (2005) detail the challenges of effective implementation, a challenge seen in a prior RCT looking to understand the effectiveness of Domestic Violence Protection Notices (DVPN's) (Gwillim 2019). Kirby (2013) believes that effective implementation is one of the biggest challenges that policing faces. Implementation becomes more challenging depending on the number of people engaged and the area of focus (Fixsen et al 2005). This research design tries to overcome this challenge through automation where possible. Linked to fidelity is tight tracking. Here, daily phone calls and twice-weekly data runs allowed for a zero-error data set. Every case that was applicable was randomised because of the structured data input. Every case that was randomised was

inputted onto the master data sheet and on every occasion an NM operative tried to deliver the treatment. Where this was not possible, the reason was recorded. Similarly, every time an RDA3 treatment was issued the placement of the sticker was photographed, retained and reviewed by the author. Any sticker incorrectly placed was re-tasked and a further photograph was taken. To ensure that the sticker was still present all available BWV was reviewed when police attended the address and the sticker visibility was checked. Where a sticker was not seen, a new sticker was put on the door and the victim was spoken to by NM operatives.

The third strength is measurement validity, the accuracy of measuring what you intend to measure (Newcomer et al 2015). The data abstraction for crime is based not only on the victim surname, but forename, address and date of birth and the same for the corresponding offender. Each search is recorded and a dip sampling process was implemented to check measurement validity. Similarly, for CFS a broad data set was obtained containing some 60,000 records. This was then cleansed to ensure that the right address was identified. When there was any uncertainty the record was checked against caller name and telephone number. Every data row was manually checked to achieve a total of 2,267 CFS.

The fourth strength is internal validity, which is achieved when combining the first three points. The research design of an effectively implemented RCT allows for examination of the effect of an independent variable, which in this case is the alarm system on a dependent variable. In this case the dependent variable is a high-risk victim of DA (Maxfield and Babbie 2015; Hagan 2014). This allows for a high level of confidence in the cause-and-effect relationships identified and should be able to identify what effect the treatment has compared to the control on high-risk DA victims. Simply, RCTs have the ability to 'illuminate' causal inference. X causes Y.

Whilst an RCT is considered the 'gold treatment' there are still a number of limitations and challenges for the research design. These areas of challenge are now discussed.

Firstly, warnings are provided that the very act of studying something may change it (Maxfield and Babbie 2015; Wilcox and Hirschfield 2007). If the subject is aware that they are being studied, their behaviour may change, impacting the internal validity of the results seen. In order to mitigate for this, the experiment was designed so that the victim was unaware that they were part of the experiment.

The second warning is in relation to those administering the treatment. Wilcox and Hirschfield (2007) warn that the person administering the treatment should be unaware of the allocation and blinded to on-going trial results. Here officers requesting an alarm were not informed of the subsequent alarm selection. The only way of establishing which alarm has been issued would be for that officer to physically attend the victims address and check. There were no reports of this happening. On a single occasion an officer contacted the author directly asking to receive the audio alarm. The officer was informed that their request would have no bearing on the alarm issued.

The fourth limitation surrounds external validity, the ability to generalise the findings to other settings across both time and place (Wilcox and Hirschfield 2007). Whilst field experiments conducted under 'real-world' conditions make them more likely to be valid in other like settings (Maxfield and Babbie 2015) this is not guaranteed. External validity challenges are continuous in social sciences. Even where something is likely to be externally valid, this claim is limited to that point in time and rapidly becomes history (Shadish et al 2002). The external validity of this experiment is likely to be limited. Whilst the trial covers a population of some 3,676,154 people, equivalent to 41.29% of London's population there are some significant limitations. The first is the time when the experiment occurred. The experiment commenced

on February the 20th 2020 and the first national lockdown in response to the global Covid-19 pandemic occurred on March 23rd 2020 lasting into July 2020. The experiment continued into November 2020 during a period where a second national lockdown had commenced. During this period of national lockdown considerable concerns were raised about the impact of lockdown on DA. The Communities Secretary, Robert Jenrick announced during lockdown that an extra £76 million would be made available to help trapped victims. The second limitation to external validity is the place researches setting. London is the capital of England and a major global city with high population density. Any deterrence effect from an audio alarm and warning sticker in a far more rural setting may well differ to London's urban setting. Results of this trial may not be generalisable to other settings (Wilcox and Hirschfield 2007).

The fifth limitation surrounds the number of victims within the trial. A total number of 300 individuals is likely to present limitations with statistical power. Ideally a greater number of individuals would have been incorporated within the trial to address this. Practical time constraints to deliver an experiment and analysis during this MSt qualification have meant that this could not be prevented. In order to mitigate this challenge, the experiment continued until six-months post the last case on May 27th. Between May 27th and November 27th 2020, the experiment obtained another 588 high-risk DA victims to allow for future analysis of a cohort of 888 individuals.

The sixth limitation concerns the data gathering methodology. Only offending recorded within the MPS was captured. This is due to the use of different systems across policing areas. Both EA and SN BCU areas are likely to be affected more greatly here due to bordering with the force areas of Essex and Surrey.

4.0 RESEARCH FINDINGS

The research findings are broken down into five distinct sections. The first section looks at the pre-test to understand how successful the randomisation was in forming two statistically similar groups. The second is the research outcomes measured by crime reports generated, crime harm, CFS and Police units deployed. The third section is focused on outputs within the CJ process with findings on arrest and charge. The fourth section is focused on system use and the number of alarm activations across both treatment groups. The fifth section details the results of the statistical tests used. All the results are summarised in Table 21.0 through to Table 25.0 at the end of the chapter.

On May 27th 2020, some 97 days after the RCT started on February 20th 2020 the 300th alarm had been requested and successfully randomised. The post-test period was for six months from the date of each alarm request. This last day of analysis was November 27th 2020. The implementation of this research method was successful. The high levels of tracking, effective leadership and feedback loop led to an overall fidelity rate of 84.7% or 254 individuals successfully receiving the treatment out of 300 (Table 3.0). In every case the alarm engineer tried to fit the correct randomised alarm system. The reduced fidelity of treatment in the RDA3 cohort was largely down to a view from the victim that the technology was too intrusive or not needed. This was the reason given in 14 of the 37 failures, or 38% of failed cases. The experimental process is visualised in the below consort diagram.

4.1 CONSORT DIAGRAM

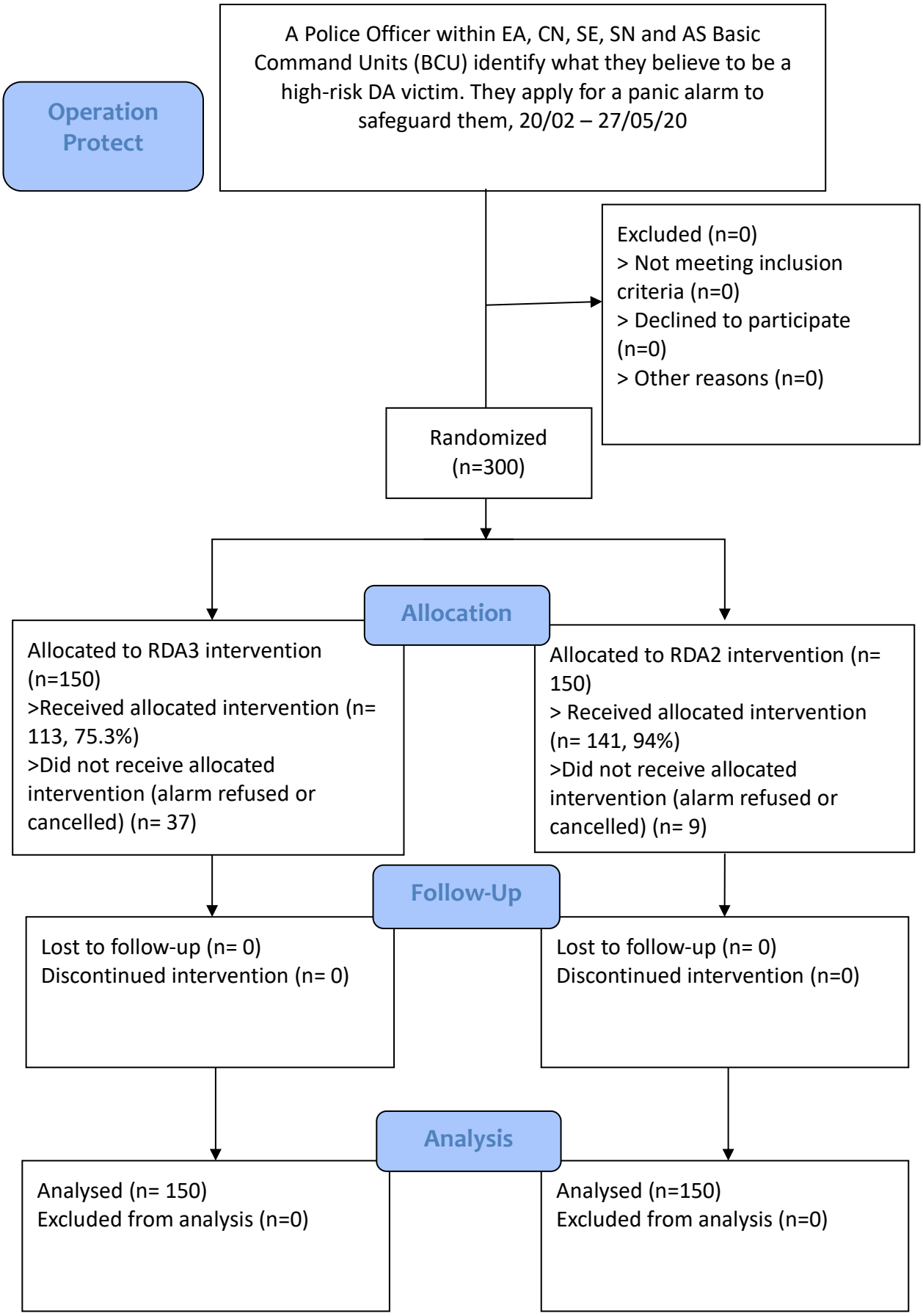


Table 3.0 – Fidelity of Treatment

	RDA2	RDA3	Total
Received allocated intervention	141 (94%)	113 (75.3%)	254 (84.7%)
Did not receive allocated intervention	9 (6%)	37 (24.7%)	46 (15.3%)
Total	150	150	300 (100%)

4.2. PRE-TEST CRIME REPORTS (ALL)

The total number of crimes recorded within the six months prior to the date of the alarm being requested was 648 crime reports. The total pre-crime count for the RDA2 intervention was 320. The total pre-crime count for the RDA3 intervention was 328. The mean was 2.16 crime records per victim overall, 2.13 for the RDA2 intervention and 2.19 for the RDA3 intervention. The range, standard deviation, mode and median for the treatment groups were also similar (Table 4.0). This distribution is visualised in Graph 1.0. A two sample T-Test identifies that there is no statistically significant difference observed between groups when reviewing by pre-test crime reports (t=-.27, p=.79) (Table 22.0).

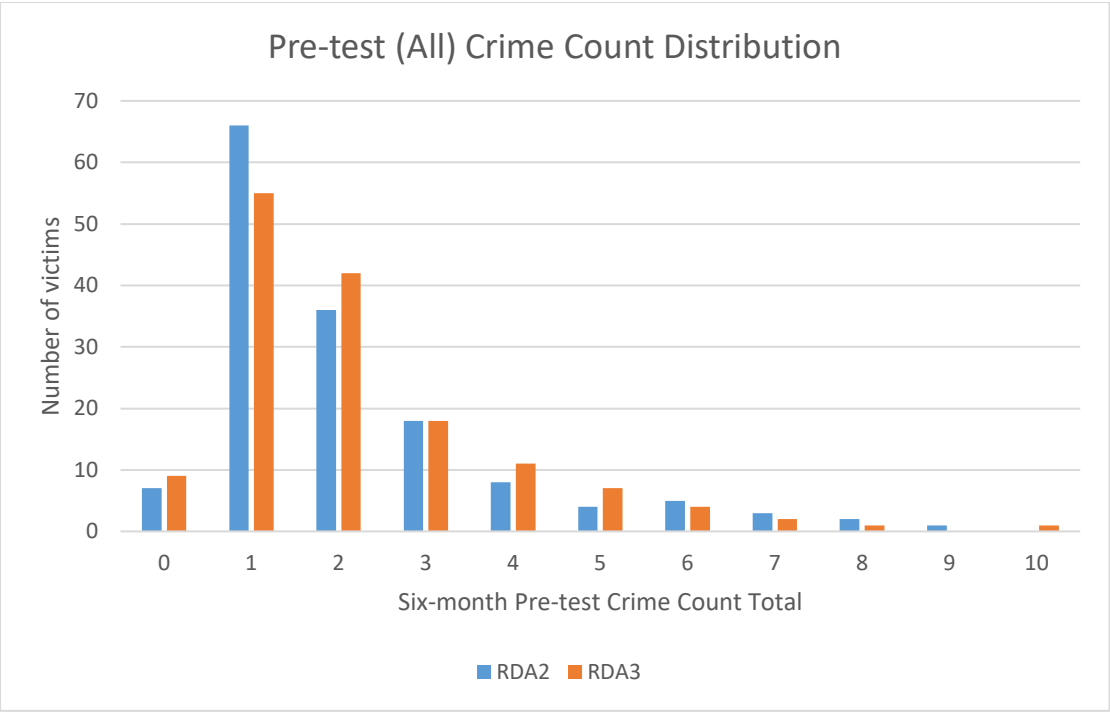
4.3. PRE-TEST CRIME REPORTS (DA)

There were 578 DA flagged crimes in the six months prior to the date of the alarm request, 70 fewer crime reports than when counting all crime. In this sub-category of crime count the total number, range, standard deviation, mean, mode and median average are broadly comparable (Table 4.0). The RDA2 treatment group accounted for 286 DA crime reports or 1.91 per victim in the six months prior to alarm request. The RDA3 treatment group accounted for 292 DA crime reports or 1.95 per victim in the six months prior to alarm request . A two sample T-Test identifies that there is no statistically significant difference observed between groups when reviewing by pre-test crime reports (t=-.23, p=.82) (Table 22.0). Visually the

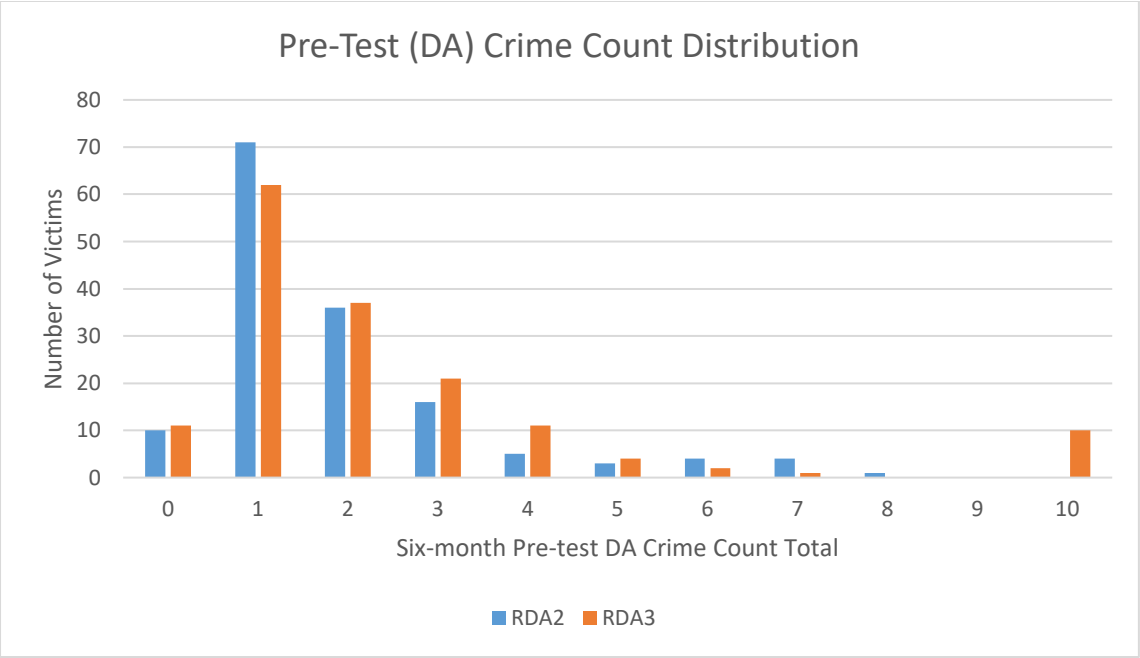
similarities can be seen in the distribution of cases within both the RDA2 and RDA3 cohort (Graph 2.0).

Table 4.0 – Pre-test crime count data.

	All Crime Reports			DA Only Crime Reports		
Six-month Pre-test Crime Count	**RDA2**	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
0	7	9	16	10	11	21
1	66	55	121	71	62	133
2	36	42	78	36	37	73
3	18	18	36	16	21	37
4	8	11	19	5	11	16
5	4	7	11	3	4	7
6	5	4	9	4	2	6
7	3	2	5	4	1	5
8	2	1	3	1	0	1
9	1	0	1	0	0	0
10	0	1	1	0	10	1
Total Six-month Pre-test Crime Count	320	328	648	286	292	578
Mean	2.13	2.19	2.16	1.91	1.95	1.93
Standard Deviation (SD)	1.74	1.69	1.71	1.58	1.48	1.53
Median	2	2	2	1	1	1
Mode	1	1	1	1	1	1
Range	9 - 0	10 - 0	10 - 0	8 - 0	10 - 0	10 - 0



Graph 1.0 – Pre-test (all) crime count distribution.



Graph 2.0 – Pre-test (DA) crime count distribution.

4.4. Pre-Test Crime Harm (All)

There was a total of 119,233.5 CCHI points in the six-month pre-test period for all crime. This was a mean of 397.45 CCHI points overall per victim. This is worth more harm than an offence of robbery (Home Office Classification 034/01) or arson with endangering of life (Home Office Classification 056/01). The RDA2 treatment group had a mean of 426.95 CCHI points in the six-month pre-test period whilst the RDA3 treatment group had a mean of 367.94 CCHI points. The range of harm identified that some victims experienced no crime harm in the pre-test period. The median and mode crime harm score was 10 (Table 5.0). The most common crime associated to this score was assault occasioning actual bodily harm (ABH) (Home Office Classification 008/06). There was no statistically significant difference observed between groups when reviewing by pre-test crime harm (t=.62, p = .53) (Table 22.0).

Table 5.0 – Pre-test crime harm data.

	Crime Harm (All)			Crime Harm (DA)		
Six-month Pre-Test Crime Harm	**RDA2**	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
Total Six-month Pre-Test Crime Harm	64,042.75	55,190.75	119,233.5	62,828.25	54,964.5	117,792.75
Mean	426.95	367.94	397.45	418.86	366.43	392.64
Standard Deviation (SD)	901.23	729.45	819.01	898.05	729.24	817.06
Median	10	12	11	10	11	10
Mode	10	10	10	10	10	10
Range	0 – 4,757	0 – 3,297	0 – 4,757	0 – 4,755	0 -3,297	0 – 4,755

4.5. Pre-Test Crime Harm (DA)

There was a total of 117,792.75 CCHI points in the six-month pre-test period for DA specific crime, a mean of 392.64 harm points per victim (Table 5.0). The RDA2 treatment group had a mean of 418.86 CCHI points in the six-month pre-test period whilst the RDA3 treatment group had a mean of 366.43 CCHI points. The range of harm was similarly broad as within the all pre-crime harm analysis. The mode harm was the same. The median harm for the RDA3 was one crime harm point higher than the RDA2 group. There was no statistically significant difference observed between groups when reviewing by pre-test DA crime harm (t=.56, p =.58) (Table 22.0).

4.6. Pre-Test Calls for Service

There was a total of 1,247 CFS in the six-months pre-test period, a mean of 4.16 calls per victim (Table 6.0). The RDA2 treatment group had a mean of 3.99 CFS per victim whilst the RDA3 treatment group had a mean of 4.32 CFS per victim. The mode is the same, with zero CFS. The median is one call different between groups. The range is two CFS different. There was no statistically significant difference observed between groups when reviewing by pre-test CFS (t=.53, p =.60) (Table 22.0).

Table 6.0 – Pre-test CFS data.

	RDA2	RDA3	Total
Total Pre-Test CFS	599	648	1,247
Mean (Calls per Victim)	3.99	4.32	4.16
Standard Deviation (SD)	5.61	5.13	5.37
Median	2	3	2
Mode	0	0	0
Range	0 - 30	0 - 32	0 - 32

4.7. PRE-TEST POLICE UNITS DEPLOYED (PUD)

The final pre-test covers the number of PUD according to CHS records. In the six-month pre-test period there were 1,559 PUD from 1,247 CFS. For the RDA2 group there were 702 PUD from 599 CFS, a police unit for every 1.35 CFS and a mean of 4.68 PUD per victim (Table 7.0). In 245 CFS there were no PUD. For the RDA3 group there were 857 PUD from 648 CFS, a mean of 5.71 PUD per victim. In 245 CFS there were no PUD, the same number as the RDA2 group. The median and mode PUD were the same. There was no statistically significant difference observed between groups when reviewing pre-test PUD (t= -1.34, p=.19) (Table 22.0). The pre-test results for crime counts, crime harm, CFS and PUD support that randomisation was effective.

Table 7.0 – Pre-test PUD data.

	RDA2	RDA3	Total
Total Pre-Test PUD	702	857	1,559
Mean (PUD per Victim)	4.68	5.71	5.20
Standard Deviation (SD)	1.49	1.78	1.65
Median	1	1	1
Mode	0	0	0
Range	0 -14	0 - 20	0 - 20

4.8. OUTCOMES: POST-TEST CRIME REPORTS (ALL)

The total crime count for the six months post alarm request was 336. This was split equally between the two interventions, with 168 crime counts in both treatment groups. This outcome is supported by a shared mean of 1.12, shared mode and median of zero crime reports. The range of crime counts in the post-test did differ slightly with the RDA2 treatment group recording between 0-14 crime reports and the RDA3 treatment group between 0-8 crime reports (Table 8.0). This can be explained by two outlier cases in the RDA2 cohort at 11 and 14 crime counts for two victims (Graph 3.0).

The before-after difference in the six-month total number of police recorded crime reports was a reduction of 312 crime reports overall or -1.04 crime reports per victim. Victims in the RDA2 treatment group have a before-after difference of 152 fewer crime reports or -1.01 reports per victim. Victims in the RDA3 treatment group have a before-after difference of 160 fewer crime reports or -1.07 reports per victim (Table 8.0). The RDA2 treatment group experienced a 47.5% reduction in crime reports. The RDA3 treatment group experienced a 48.8% reduction in crime reports. Both the treatment groups experienced statistically significant reductions in crime counts ($t= -6.4$, $p=<0.001$) (Table 23.0), but no between-group differences. The effect size differed slightly for all crime. For the RDA2 group $d= -0.54$ and the RDA3 group $d= - 0.64$, both were medium effects.

4.9. OUTCOMES: POST-TEST CRIME REPORTS (DA)

The similarity in distribution is also seen in the DA only crime reports (Graph 4.0). The before-after difference in the six-month total number of police recorded DA crime reports was a reduction of 308 DA crime reports overall or -1.03 DA crime reports per victim. Victims in the RDA2 treatment group have a before-after difference of 153 fewer DA crime reports or -1.02

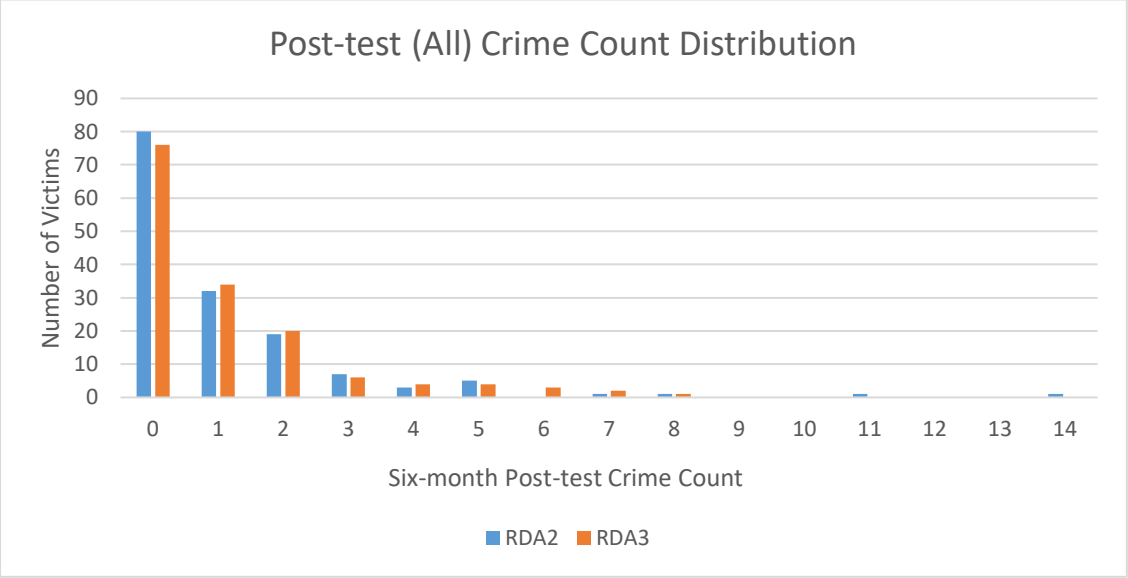
reports per victim. Victims in the RDA3 treatment group have a before-after difference of 155 fewer DA crime reports or -1.03 reports per victim (Table 8.0). The RDA2 treatment group experienced a 46.5% reduction in DA crime reports. The RDA3 treatment group experienced a 46.9% reduction in DA crime reports. Both treatment groups experienced statistically significant DA crime report reductions in the before-after comparisons (p=<0.001) (Table 23.0) with a medium effect reduction in number (RDA2: d= – 0.61, RDA3: d= -0.71). However, there is no relative change comparing the two treatment conditions.

Table 8.0 – Post-test crime count data.

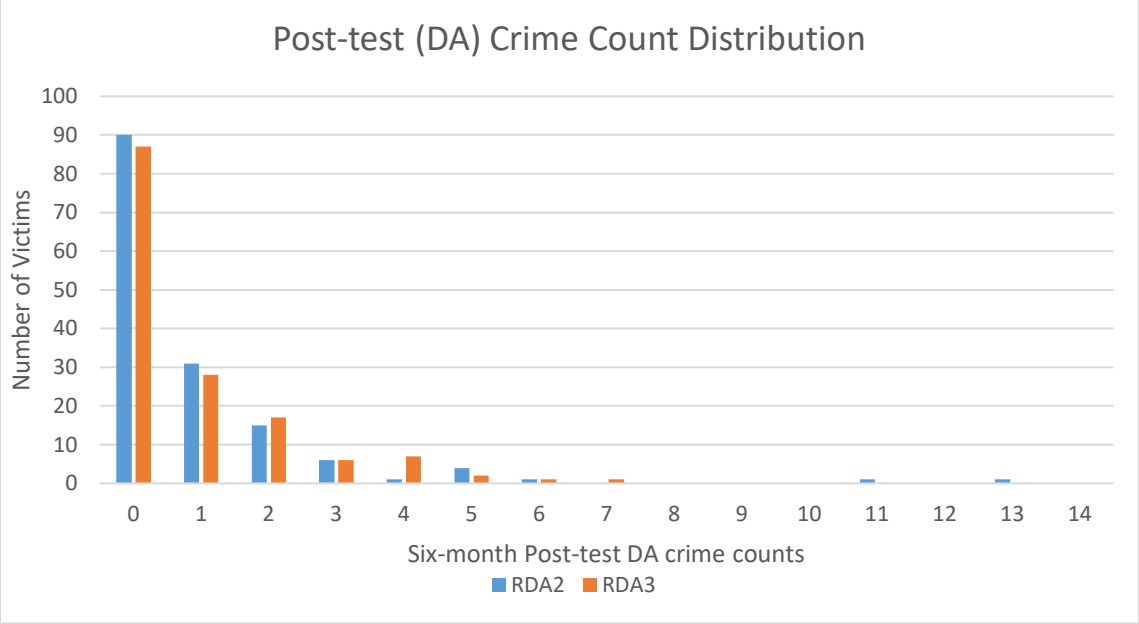
	All Crime Reports			**DA Crime Reports**		
Six-month Post-test Crime Count	**RDA2**	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
0	80	76	156	90	87	177
1	32	34	66	31	28	59
2	19	20	39	15	17	32
3	7	6	13	6	6	12
4	3	4	7	1	7	8
5	5	4	9	4	2	6
6	0	3	3	1	1	3
7	1	2	3	0	1	1
8	1	1	2	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	1	0	1	1	0	1
12	0	0	0	0	0	0
13	0	0	0	1	0	1
DA14	1	0	1	0	0	0
Total Six-month Post-test Crime Count	168	168	336	133	137	270
Mean	1.12	1.12	1.12	0.89	0.91	0.90
Standard Deviation (SD)	1.98	1.67	1.83	1.77	1.49	1.62
Median	0	0	0	0	0	0
Mode	0	0	0	0	0	0
Range	0 -14	0 - 8	0 - 14	0 - 13	0 - 7	0 -13

Table 9.0 – Post-test difference in crime count.

	Sum			Mean ($\bar{x}$)		
	RDA2	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
Difference in Count (all)	-152	-160	-312	-1.01 (SD=1.98)	-1.07 (SD=1.67)	-1.04 (SD=1.83)
Difference in Count (DA)	-153	-155	-308	-1.02 (SD=1.77)	-1.03 (SD=1.49)	-1.03 (SD=1.62)



Graph 3.0 – Post-test (all) crime count distribution.



Graph 4.0 – Post-test (DA) crime count distribution.

4.10 OUTCOMES: CRIME HARM (ALL)

The before-after difference for all crime harm was a reduction of 96,371.75 CCHI points. This was a mean reduction in crime harm of 321.24 CCHI points per victim or an 80.83% reduction in harm for both treatment groups. Victims in the RDA2 treatment group have a before-after difference of -54,493.25 CCHI points, -362.34 CCHI points per victim or an 85.1% reduction in harm. Victims in the RDA3 treatment group have a before-after difference of -41,878.5 CCHI points, -279.19 CCHI points per victim or an 75.9% reduction in harm (Table 11.0 and Graph 5.0). This equated to a 22.95% greater harm reduction achieved by the RDA2 alarm treatment or 83.15 fewer CCHI points suffered than the RDA3 treatment group. This is roughly 12 weeks imprisonment time, the same minimum custodial sentence for a first-time offence of Stalking Involving Fear of Violence, (Home Office Classification 008/65) or equivalent of four sexual assault sentences for a first-time offender (Home Office Classification 020/05). This left a mean harm score across both groups of 76.21 CCHI points (Table 10.0). The mean harm score for victims in the RDA2 treatment group was 63.66 CCHI points and 88.75 CCHI points in the RDA3 treatment group.

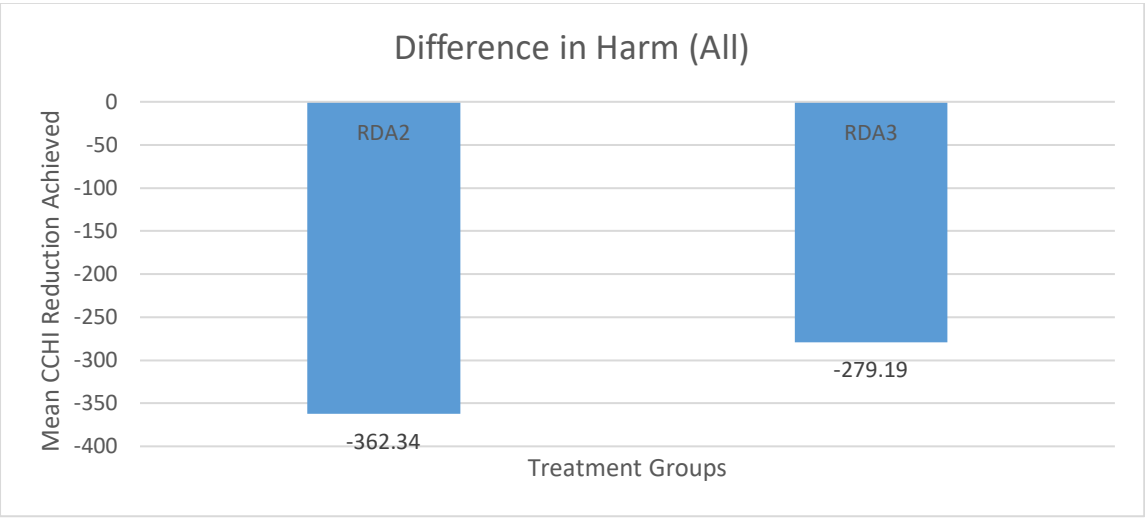
The results of paired T-Tests for the RDA2 treatment group show a statistically significant reduction in crime harm (t=-5.1, p=<0.001) with a medium effect size (d = -0.5). For the RDA3 treatment group the crime harm reduction is also statistically significant (t= -4.12, p=<0.001), but with a small effect size (d = - 0.48) (Table 24.0).

Table 10.0 – Post-test crime harm data.

	RDA2	RDA3	Total
Total Post-test Crime Harm	9,549.5	13,312.25	22,861.75
Mean Harm Score	63.66	88.75	76.21
Standard Deviation (SD)	474.46	370.75	425.25
Median	0	0	0
Mode	0	0	0
Range	0 - 5,505.75	0 - 3,285	0 - 5,505.75
Total Post-test DA Crime Harm	7,211.5	12,700.05	19,911.55
Mean Harm Score	48.08	85.24	66.60
Standard Deviation (SD)	336.74	500.82	358.21
Median	0	0	0
Mode	0	0	0
Range	0 - 3,680.75	0 - 3,285	0 - 3,680.75

Table 11.0 – Post-test difference in crime harm data.

	Sum			Mean ($\bar{x}$)		
	RDA2	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
Difference in Harm (All)	-54,493.25	-41,878.5	-96,371.75	-362.34 (SD=474.46)	-279.19 (SD=370.75)	-321.24 (SD=425.25)
Difference in Harm (DA)	-55,616.75	-42,264.45	-97,881.2	-370.78 (SD=336.74)	-281.19 (SD=500.82)	-326.04 (SD=358.21)

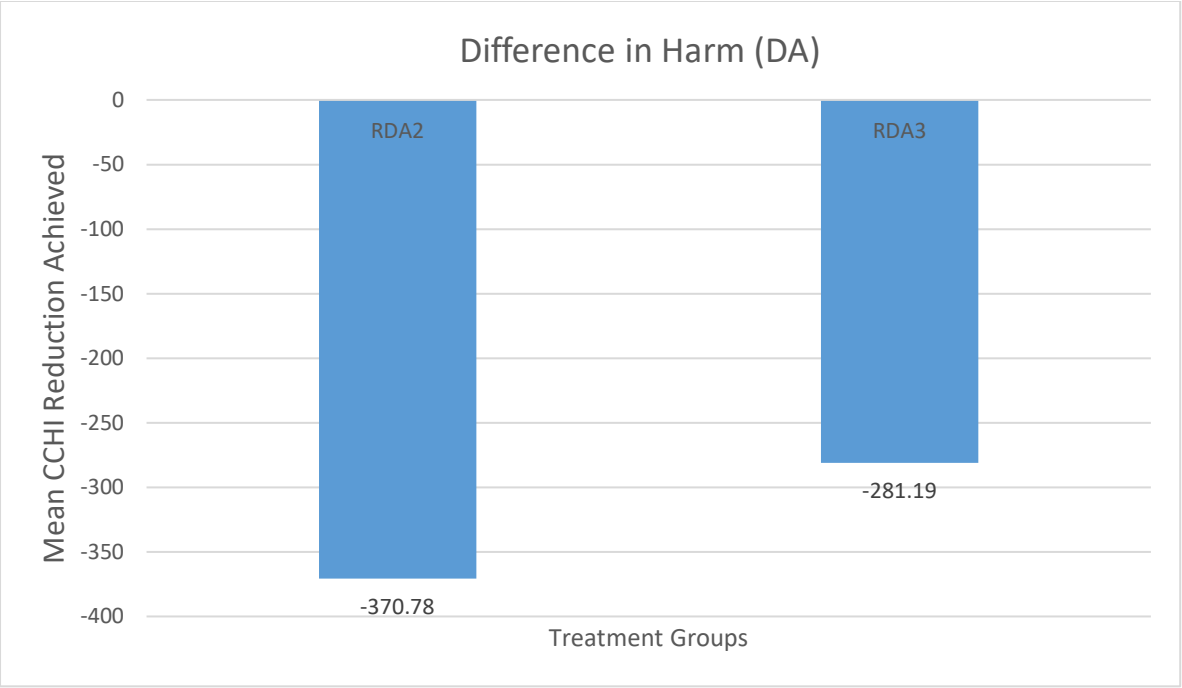


Graph 5.0 – Difference in harm (all)

4.11 OUTCOMES: CRIME HARM (DA)

The before-after difference for crime harm attributed to DA was a total reduction of 97,881.2 CCHI points, -326.04 CCHI points per victim and an 83.1% reduction in harm. For the RDA2 treatment group the difference was a total reduction of 55,616.72 CCHI points, -370.78 CCHI points per victim and an 88.52% reduction in harm. For the RDA3 treatment group the difference was a total reduction of 42,264.45 CCHI points, -281.19 CCHI points per victim and an 76.89% reduction in harm (Table 11.0 and Graph 6.0). Both treatments saw greater mean harm reductions when analysing just DA crime harm. However, the gap in effectiveness in reducing harm grew with 89.59 fewer CCHI points reduced in the RDA3 cohort. The RDA3 treatment was 24.16% less effective in reducing DA Harm when compared to the RDA2 treatment. The difference in harm reduction equates to roughly nine counts of Threats to Kill (Home Office Classification 003/01).

The before-after paired T-Test results for the RDA2 treatment group found that the reduction in DA harm in the six months post alarm request was statistically significant (t= -5.32, p=<0.001) with a medium effect size (d = -0.55) (Table 4.V). The reduction in harm seen in the RDA3 group were also statistically significant (t= -4.11, p=<0.001) with a medium effect size (d = -0.49). There was no statistically significant difference in the between-group comparisons.



Graph 6.0 – Difference in harm (DA).

4.12. Outcomes: Calls For Service (CFS)

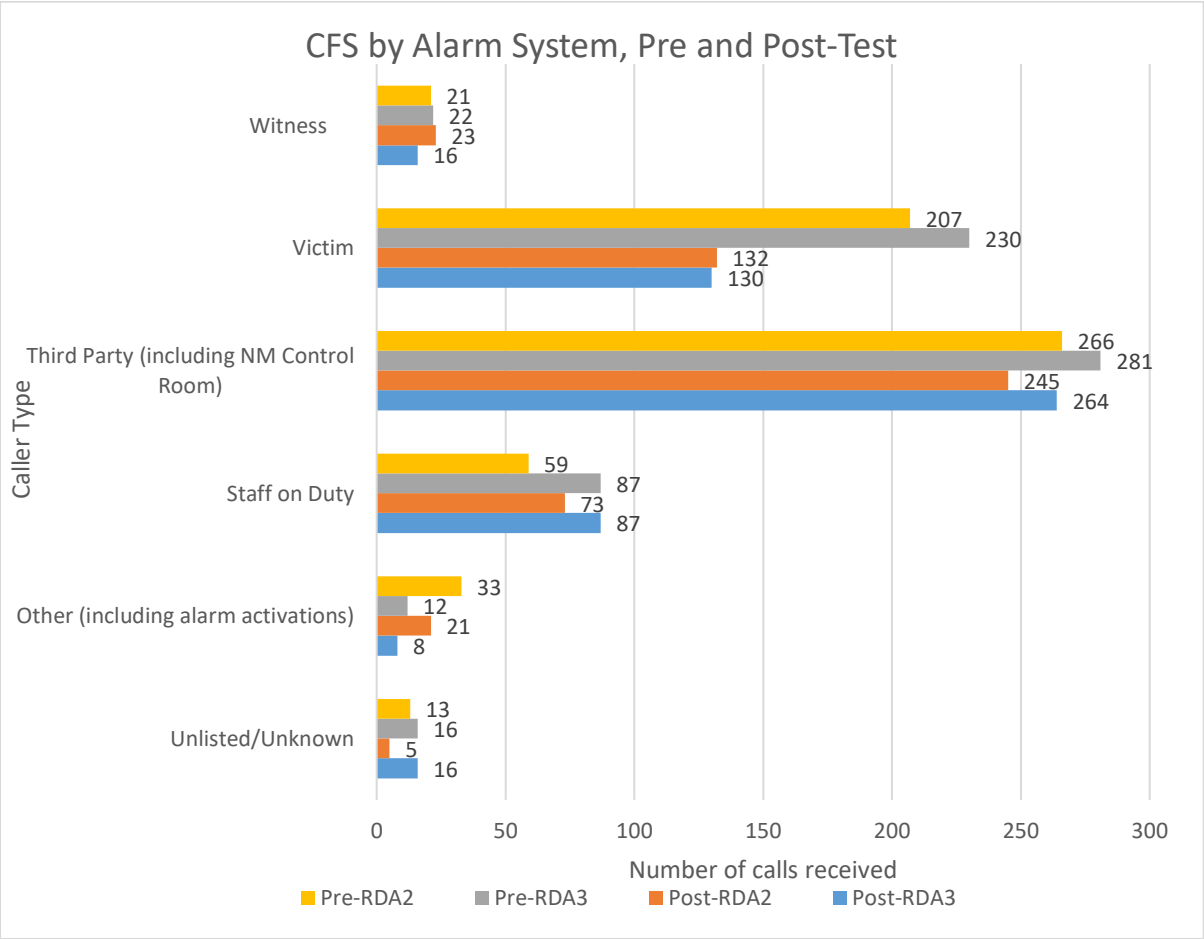
Calls for service (CFS) are broken into six categories within the MPS (Table 12.0). Every caller type is decided manually by a call handler in one of three MPS first contact centres. Not every CFS results in a PUD to the caller or location of concern. A number may simply be advised or re-directed to another agency. The first category of caller is witness. This is an individual who is not the victim of crime, but someone calling who has witnessed something they think may be a crime. The second is victim, the individual who is alleging that they have been a victim of crime. The third is listed as third party and this includes the alarm provider supporting this research, other police services and local authorities. However, this list is substantial and does include victims of crime. The fourth category is staff on duty. These are mainly telephone calls from within the police service. The fifth is other and this is the category that should be used for alarm activations. The last category is unknown caller. A review of the data at a granular level shows a high degree of inaccuracy with the caller type labels. This inaccuracy should be

shared equally across the treatment groups due to the research method of randomisation. The data shows that CFS generated from victims of crime reduced in both categories. RDA2 victim type callers reduced by 37.2%, 207 CFS to 132. RDA3 victim type callers reduced by 42.6%, 230 to 132. In both groups the number of caller types associated to alarm activations increased dramatically. This supports that victims changed their method of contacting police. In both treatment groups victim-based contact including both alarm activations and CFS from victims increased. In the RDA2 group this was an increase from 240 CFS to 251. In the RDA3 group this was an increase from 242 to 262.

The before-after difference for CFS was a total increase of 72 in the six months post alarm request, a total of 1,319 CFS (Table 14.0). The RDA3 treatment group increased by 51 calls, a mean increase of +0.34 CFS per victim. The RDA2 treatment groups increased by 21 CFS, a mean increase of +0.14 CFS per victim (Graph 8.0). This led to a post-test mean of 4.13 CFS in the RDA2 group and 4.66 CFS in the RDA3 group (Table 13.0). The changes to CFS volume were not statistically significant for either the RDA2 treatment group (t=.33, p= .74) or the RDA3 treatment group (t=.72, p=.47) in a before-after comparison (Table 25.0). A between-group comparison identified that the differences seen were not statistically significant (t=-.73, p=.46).

Table 12.0 – CFS by caller type.

	Pre-RDA2	Post-RDA2	Pre-RDA3	Post-RDA3
Unlisted/Unknown	13 (2.17%)	5 (0.81%)	16 (2.47%)	16 (2.29%)
Other	33 (5.51%)	119 (19.19%)	12 (1.85%)	130 (18.6%)
Staff on Duty	59 (9.85%)	75 (12.1%)	87 (13.43%)	87 (12.45%)
Third Party	266 (44.41%)	265 (42.74%)	281 (43.36%)	318 (45.49%)
Victim	207 (34.56%)	132 (21.29%)	230 (35.49%)	132 (18.88%)
Witness	21 (3.51%)	24 (3.87%)	22 (3.40%)	16 (2.29%)
Total	599 (100%)	620 (100%)	648 (100%)	699 (100%)



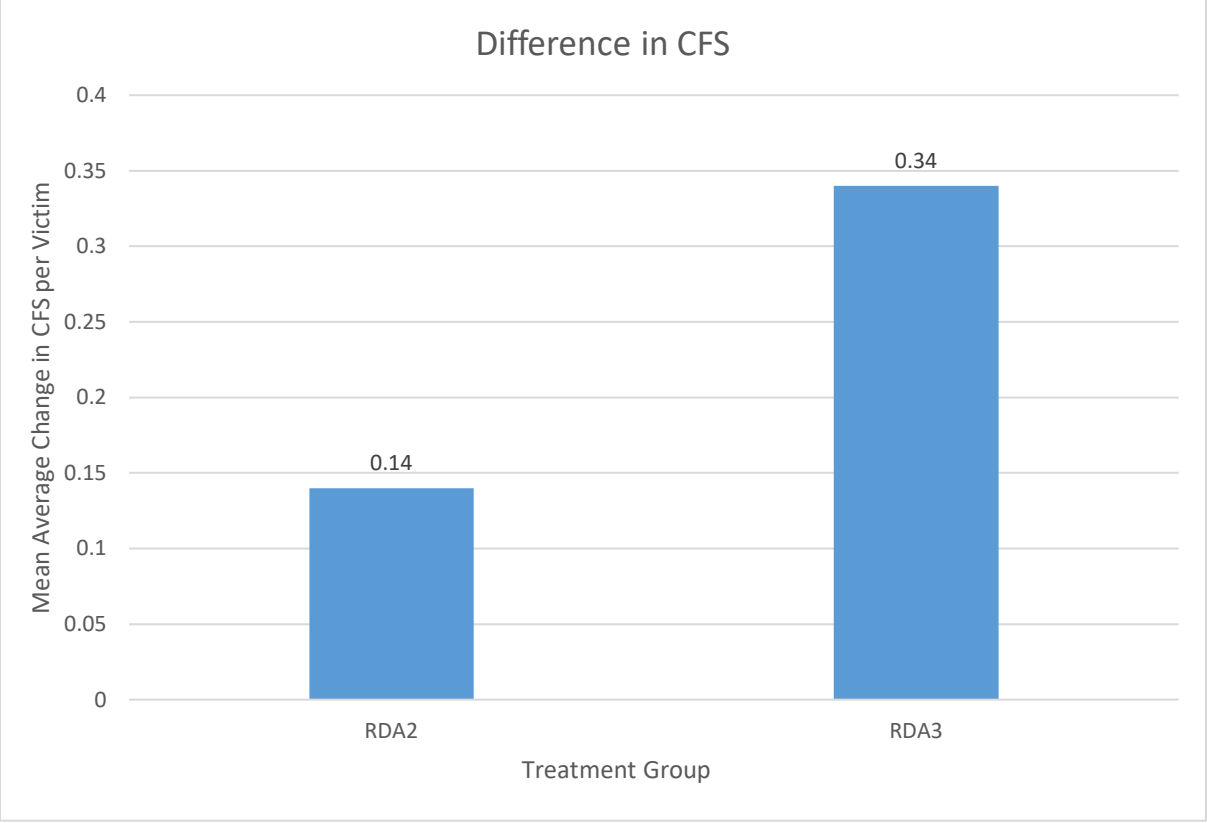
Graph 7.0 – CFS by alarm system, pre and post-test.

Table 13.0 – CFS data.

	RDA2	RDA3	Total
Total Pre-Test CFS	599	648	1,247
Mean (Calls per Victim)	3.99	4.32	4.16
Standard Deviation (SD)	5.61	5.13	5.37
Median	2	3	2
Mode	0	0	0
Range	0 - 30	0 - 32	0 - 32
Total Post-Test CFS	620	699	1,319
Mean (Calls per Victim)	4.13	4.66	4.40
Standard Deviation (SD)	6.07	6.37	6.22
Median	2	3	2
Mode	0	0	0
Range	0 - 50	0 -40	0 -50

Table 14.0 – Difference in CFS.

	Sum			Mean ($\bar{x}$)		
	RDA2	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
Difference in CFS	+21	+51	+72	+0.14	+0.34	+0.24



Graph 8.0 – Difference in CFS.

4.13. OUTCOMES: POLICE UNITS DEPLOYED (PUD)

Police units deployed (PUD) are a measure of call sign and can include vehicles as well as individual officers. The unit may well be a vehicle call sign such as 'SE21' which would be a rapid response vehicle operating in the South East (SE) BCU. Equally, it could include an officer on foot patrol. The most common PUD is a vehicle typically containing two officers. The RCT research method ensures that inaccuracy is equally distributed across both treatment groups. The mean number of PUD per victim across both treatment groups increased by 0.13 units per victim in the post-test period (Table 15.0). Overall there was an increase of 40 PUD, a total of 1,559 PUD in the post-test. The increase is not seen across both treatment groups.

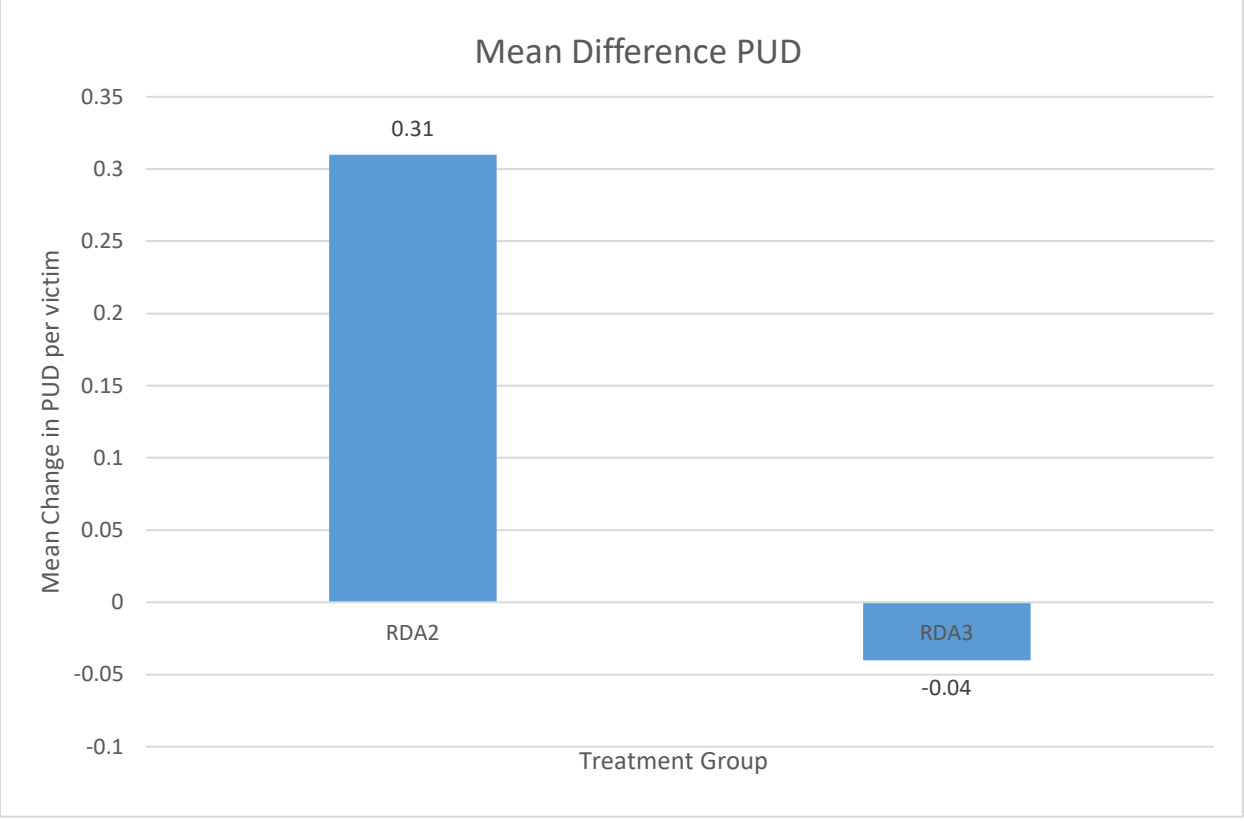
The RDA3 treatment group saw a small mean reduction in PUD of 0.04 units per victim. The RDA2 treatment group on the other hand saw an increase of 0.31 PUD per victim (Graph 9.0). This equated to 52 fewer PUD to the RDA3 treatment group compared to the RDA2 treatment group in the post-test period (Table 16.0). The before-after comparisons show that the changes seen in both RDA2 (t=.6, p=.56) and RDA3 (t=-.06, p=.95) were not statistically significant (Table 25.0). A between-group comparison identified that the differences seen were not statistically significant (t=-.71, p=.48) (Table 22.0).

Table 15.0 – PUD data.

	RDA2	RDA3	Total
Total Pre-Test PUD	702	857	1,559
Mean (PUD per Victim)	4.68	5.71	5.20
Standard Deviation (SD)	1.49	1.78	1.65
Median	1	1	1
Mode	0	0	0
Range	0 -14	0 - 20	0 - 20
Total Post-Test PUD	748	851	1,599
Mean (PUD per Victim)	4.99	5.67	5.33
Standard Deviation (SD)	1.6	1.77	1.69
Median	1	1	1
Mode	0	0	0
Range	0 -16	0 - 16	0 - 16

Table 16.0 – Difference in PUD.

	Sum			Mean ($\bar{x}$)		
	RDA2	**RDA3**	**Total**	**RDA2**	**RDA3**	**Total**
Difference PUD	+46	-6	+40	+0.31 (SD=1.6)	-0.04 (SD=1.77)	+0.13 (SD=1.69)



Graph 9.0 – Mean difference PUD.

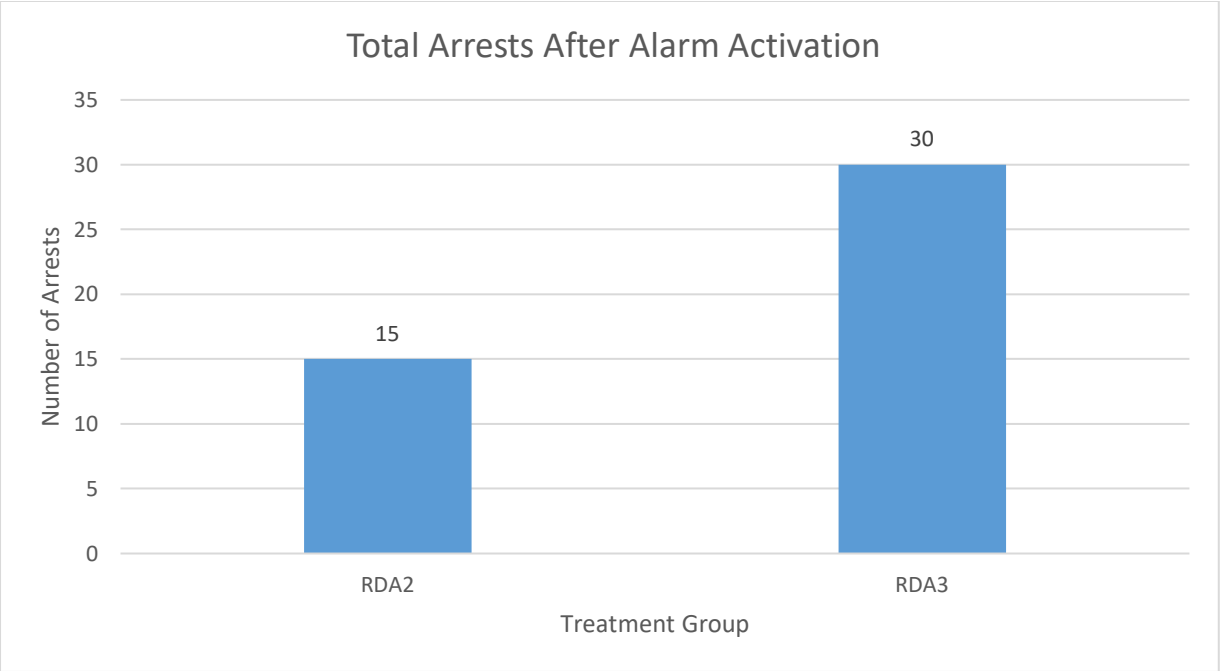
4.14. OUTPUTS: ARREST RATE

Arrest rate is a key measure of policing enforcement, a tactic used to support the CJ process and reduce risk. In the six-month post-test period a total of 45 arrests were generated from alarm activations (Table 17.0). The RDA2 treatment group experienced a mean of 0.1 arrests per victim. Victims in the RDA3 treatment group had an arrest rate that was double the RDA2 group, with a mean of 0.2 arrests per victim, a total of 30 arrests (Graph 10.0). The RDA3 activations led to an arrest rate double to that of the standard RDA2 alarm. The 30 arrests were generated from 17 unique victims (Graph 11.0).

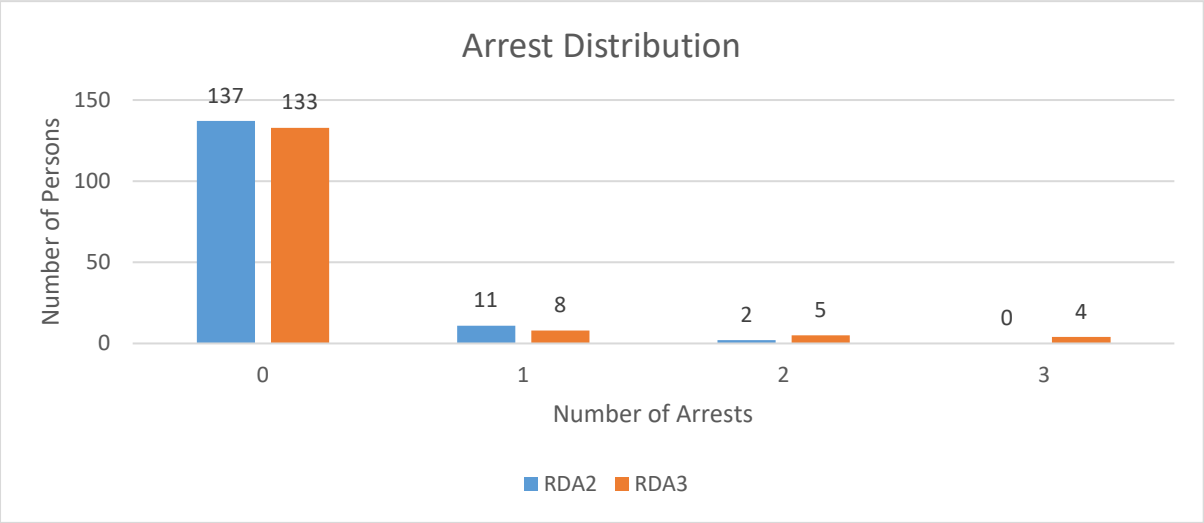
The likelihood of an arrest taking place from an RDA3 alarm activation was 32.3%. The likelihood of an arrest taking place from an RDA2 alarm activation was 25%. A between-group comparison T-Test result showed that the difference in outcome seen was approaching statistical significance ($t = 1.73$; $p = .08$). The effect size of here is small $d = -0.2$ (Table 21.0). The null hypothesis can be rejected at the .1 level.

Table 17.0 – Post-test arrest data.

Six-month Post-test Arrests	**RDA2**	**RDA3**	**Total**
0	137	133	270
1	11	8	19
2	2	5	7
3	0	4	4
Total Six-month Post-test Arrests	15	30	45
Mean (Arrests per victim)	0.1	0.2	0.15
Standard Deviation (SD)	0.34	0.62	0.5
Median	0	0	0
Mode	0	0	0
Range	**0 - 2**	**0 - 3**	**0 - 3**



Graph 10.0 – Total arrests after alarm activation.



Graph 11.0 – Arrest distribution.

4.15. Outputs: Charge Rate

The first part to a successful CJ outcome is to have gathered enough evidence to support a charge. A charge means that the Crown Prosecution Service (CPS) have been engaged and a decision has been made that the evidence supports a realistic possibility of conviction on an evidential weighting that is 'beyond reasonable doubt'. Alarm activations in this trial led to 28

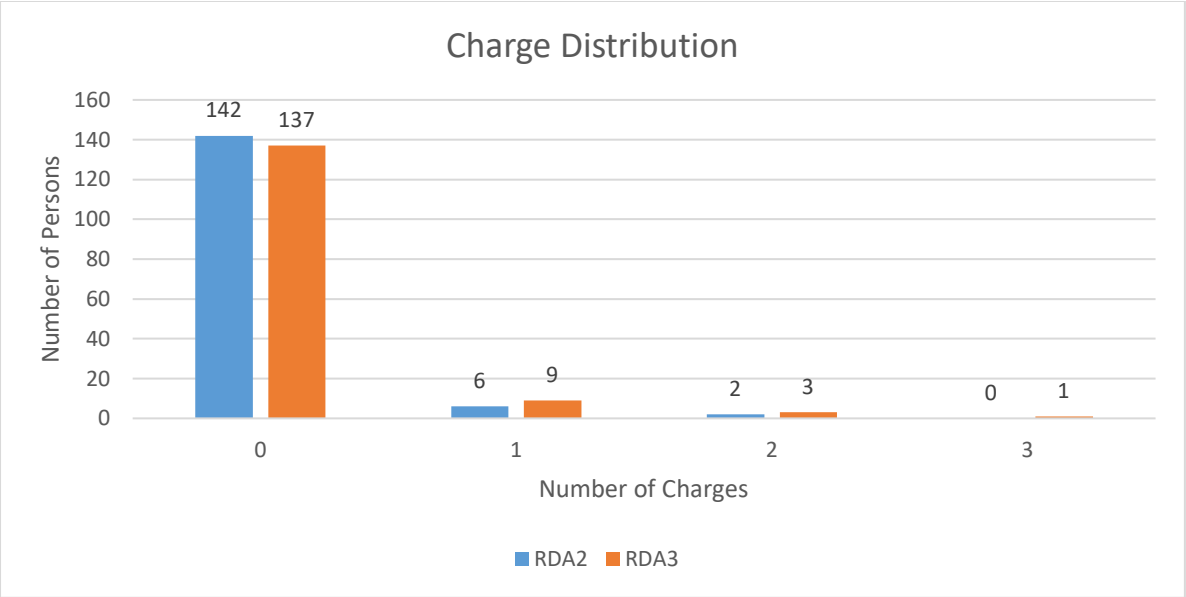
charges. The RDA2 cohorts alarm activations generated 10 charges and the RDA3 cohorts alarm activations generated 18 charges (Table 18.0).

When measuring the effect of alarm type on post-activation CJ outcomes the RDA3 treatment group had 80% more charges than that of the RDA2 treatment group, a mean of 0.12 charges per victim compared to 0.07 charges per victim. In total eight victims in the RDA2 treatment group activated their alarm leading to a charge and 13 victims in the RDA3 treatment group activated their alarm leading to a charge (Graph 12.0).

The between-groups two sample T-Test results support that the outcomes seen are not statistically significant (t=1.24, p=.2) and the effect size is very small (d= -.14) (Table 21.0).

Table 18.0 – Post-test charge data.

	All Crime Reports		
Six-month Post-test Charges	**RDA2**	**RDA3**	**Total**
0	142	137	279
1	6	9	15
2	2	3	5
3	0	1	1
Total Six-month Post-test Charges	10	18	28
Mean (Charges per victim)	0.07	0.12	0.09
Standard Deviation (SD)	0.30	0.43	0.37
Median	0	0	0
Mode	0	0	0
Range	0 - 2	0 - 3	0 - 3



Graph 12.0 – Charge distribution.

4.16. System Use: Alarm Activations

In total the trial generated 153 alarm activations from 300 victims. Out of the 300 victims, 77 used the alarm that had been issued to them, 223 did not or 74.3%. A total of 93 alarm activations came from the RDA3 treatment group and 60 came from the RDA2 treatment group (Table 19.0). This resulted in roughly a 60/40 divide in activations, with 60.8% of activations coming from the RDA3 group and 39.2% from the RDA2 group. Victims randomised to receive an RDA3 alarm generated 55.8% more activations than the RDA2 cohort. The mean number of activations for the RDA3 treatment group was 0.62 per victim compared to 0.4 activations per victim in the RDA2 treatment group. The range of use for victims in the RDA3 group was 0-15 activations. For the RDA2 group the range was between 0-4 activations. The RDA3 cohort contained two outliers, with eight and 15 activations each, worth 23 activations or 24.7% of all RDA3 activations.

The total number of victims who activated the alarm was very similar, with 38 victims in the RDA3 treatment group using the alarm compared to 39 victims in the RDA2 treatment group (Table 20.0). The likelihood of a victim using the alarm at least once was 25.3% in the RDA3

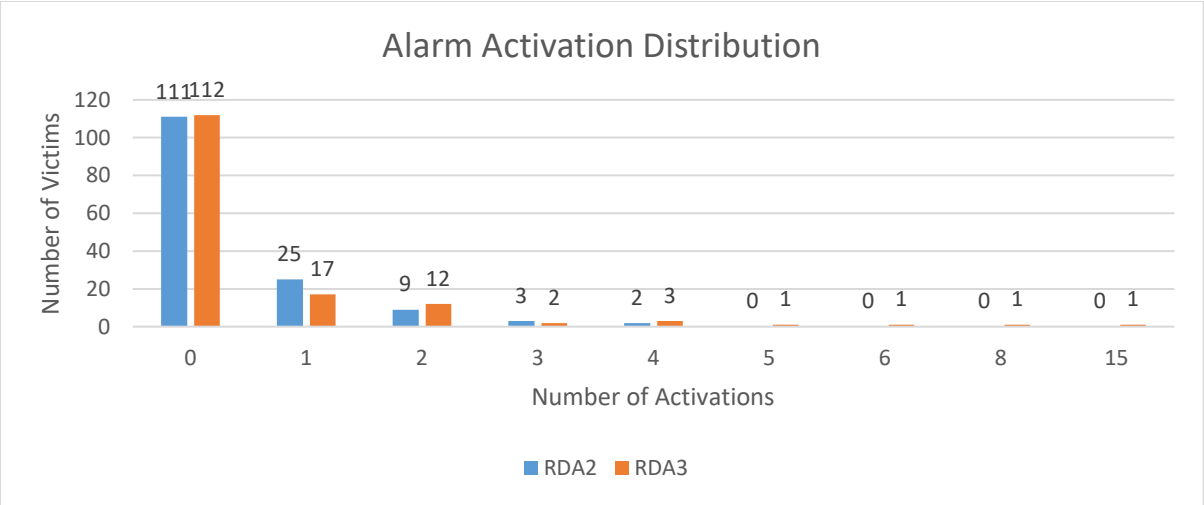
cohort and 26% in the RDA2 cohort. The difference in total alarm activations results from 21 victims activating their RDA3 audio alarm system more than once, compared to just 14 victims in the RDA2 treatment group (Graph 13.0). The RDA3 treatment group had 50% more repeat alarm users. The between-groups two sample T-Test results support that the outcomes seen are not statistically significant (t=-.27, p= .79) (Table 21.0).

Table 19.0 – Alarm activation data.

Number of Alarm Activations	**RDA2**	**RDA3**	**Total**
0	111	112	223
1	25	17	42
2	9	12	21
3	3	2	5
4	2	3	5
5	0	1	1
6	0	1	1
7	0	0	0
8	0	1	1
15	0	1	1
Total Number of Alarm Activations	60	93	153
Mean	0.4	0.62	0.51
Standard Deviation (SD)	0.80	1.66	1.30
Median	0	0	0
Mode	0	0	0
Range	0 - 4	0 - 15	0 - 15

Table 20.0 – Victim alarm use data.

	RDA2	**RDA3**
Number of Victims who used the alarm	39	38
Number of Victims who did not use the alarm	111	112



Graph 13.0 - Alarm activation distribution.

4.17 STATISTICAL TESTS

Table 21.0 – Two sample T-Tests: Arrests, Charges and Activations.

Table: Two Sample T-Tests (between groups comparisons)								
	N	**Mean**	**t**	**CI Lower**	**CI Upper**	**DF**	**p-value**	**Cohen's D**
Arrests								
RDA2	15	0.1	-1.72	-0.21	0.01	298	0.09	-0.20
RDA3	30	0.2						
Charges								
RDA2	10	0.07	1.24	-0.14	0.03	298	0.2	-0.14
RDA3	18	0.12						
Activations								
RDA2	60	0.4	-0.27	-0.44	0.34	298	0.79	-0.17
RDA3	93	80.62						

Table 22.0 – Two sample T-Tests: Pre-test crime counts (All), Pre-test crime counts (DA), Pre-test crime harm (All), Pre-test crime harm (DA), Pre-test CFS, Post-test CFS, Pre-test PUD and Post-test PUD.

Table: Two Sample T-Tests (between groups comparisons)								
	N	**Mean**	**t**	**CI Lower**	**CI Upper**	**DF**	**p-value**	**Cohen's D**
Crime Counts (Pre-Test (All))								
RDA2	320	2.13	-0.27	-0.44	0.34	298	0.79	-0.03
RDA3	328	2.19						
Crime Counts (Pre-Test (DA))								
RDA2	286	1.91	-0.23	-0.39	0.31	298	0.82	-0.03
RDA3	292	1.95						
Crime Harm (Pre-Test (All))								
RDA2	64042.75	426.95	0.62	-127.3	245.32	298	0.53	0.07
RDA3	55190.75	367.94						
Crime Harm (Pre-Test (DA))								
RDA2	62828.25	418.86	0.56	-133.46	238.31	298	0.58	0.06
RDA3	54964.5	366.43						
CFS (Pre-Test)								
RDA2	599	3.99	0.53	-1.55	0.89	298	0.60	-0.06
RDA3	648	4.32						
CFS (Post-Test)								
RDA2	620	413	-0.73	-1.94	0.89	298	0.46	-0.08
RDA3	699	4.66						
PUD (Pre-Test)								
RDA2	702	4.68	-1.34	-2.56	0.49	298	0.19	-0.15
RDA3	857	5.71						
PUD (Post-Test)								
RDA2	748	4.99	-0.71	-2.59	1.21	298	0.48	-0.08
RDA3	851	5.67						

Table 23.0 – Paired T-Tests: Crime counts RDA2 (All), Crime counts RDA3 (All), Crime counts RDA2 (DA) and Crime counts RDA3 (DA).

Table: Paired T-Tests (before-after comparisons)								
	N	**Mean**	**t**	**CI Lower**	**CI Upper**	**DF**	**p-value**	**Cohen's D**
Crime Counts (All) RDA2								
Pre-Test	320	2.13	-6.4	-1.32	-0.7	149	<0.001	-0.54
Post -Test	168	1.12						
Crime Counts (All) RDA3								
Pre-Test	328	2.19	-6.4	-1.4	0.74	149	<0.001	-0.64
Post -Test	168	1.12						
Crime Counts (DA) RDA2								
Pre-Test	286	1.91	-7.0	-1.31	-0.73	149	<0.001	-0.61
Post-Test	133	0.89						
Crime Counts (DA) RDA3								
Pre-Test	292	1.95	-7.4	-1.31	-0.76	149	<0.001	-0.71
Post-Test	137	0.91						

Table 24.0 – Paired T-Tests: Crime harm RDA2 (All), Crime harm RDA3 (All), Crime harm RDA2 (DA) and Crime harm RDA3 (DA).

Table: Paired T-Tests (before-after comparisons)								
	N	**Mean**	**t**	**CI Lower**	**CI Upper**	**DF**	**p-value**	**Cohen's D**
Crime Harm (All) RDA2								
Pre-Test	64042.75	426.95	-5.1	-504.1	-222.5	149	<0.001	-0.5
Post -Test	9549.5	63.66						
Crime Harm (All) RDA3								
Pre-Test	55190.75	367.94	-4.12	-413.0	-145.4	149	<0.001	-0.48
Post -Test	13312.25	88.75						
Crime Harm (DA) RDA2								
Pre-Test	62828.25	418.86	-5.32	-508.62	-232.94	149	<0.001	-0.55
Post-Test	7211.5	48.08						
Crime Harm (DA) RDA3								
Pre-Test	54964.5	366.43	-4.11	-417.07	-146.46	149	<0.001	-0.49
Post-Test	12700.05	85.24						

Table 25.0 – Paired T-Tests: CFS RDA2, CFS RDA3, PUD RDA2 and PUD RDA3.

Table: Paired T-Tests (before-after comparisons)								
	N	**Mean**	**t**	**CI Lower**	**CI Upper**	**DF**	**p-value**	**Cohen's D**
CFS RDA2								
Pre-Test	599	3.99	0.33	-0.7	-0.98	149	0.74	0.02
Post -Test	620	4.13						
CFS RDA3								
Pre-Test	648	4.32	0.72	0.6	1.28	149	0.47	0.06
Post -Test	699	4.66						
PUD RDA2								
Pre-Test	702	4.68	0.6	-0.73	1.35	149	0.56	0.04
Post-Test	748	4.99						
PUD RDA3								
Pre-Test	857	5.71	-0.06	-1.36	1.28	149	0.95	-0.005
Post-Test	851	5.67						

4.18 FINDINGS SUMMARY

Overall this experiment has found that there is no treatment effect. The between-group comparisons post-randomisation find no treatment effect for crime counts (All), crime counts (DA), crime harm (all), crime harm (DA), CFS, PUD and activations. The one area which could be considered as delivering a meaningful treatment effect is arrest. Here the differences between the RDA2 treatment and the RDA3 treatment are statistically significant at the 0.1 level. However, due to the sample size of 300 cases there is a risk that this could be just statistical noise. Equally, when reviewing the observed difference in arrest and then charge, the argument can be levied in the opposing position. Statistical significance is primarily driven by sample size, a larger sample size may well have achieved statistically significant outcomes at the accepted threshold. What is clear is that this is an area of research to explore further.

The facts are that the difference between groups following randomisation for arrest were not statistically significant at the accepted threshold and the effect size was small. Using one alarm system over another system does not achieve greater outcomes for high-risk DA victims who were assigned panic alarms. This is a major finding for at least three reasons. A) using one system over the other does not lead to a backfiring effect. B) for this particular population of DA victims, a system that incorporates a notification to the offender that the premises have a recording device does not lead to a backfiring effect. C) this line of research identifies additional research questions for the future focused on the slight increases in arrest and subsequently charges for the audible alarm system over the national standard alarm system without audio. Greater understanding is needed to investigate which cases have led to an arrest compared to those that have not. This may offer the possibility of forming an evidence-based tailoring of alarm system intervention to best support high-risk DA victims.

5.0 DISCUSSION

The discussion moves to outline the summary of findings, the implications of these on theory and knowledge-base, policy implications, research implications and limitations of this work.

5.1 SUMMARY OF FINDINGS

Use of an RCT methodology for high-risk DA victims was successful, with 84.7% fidelity of treatment seen. Police data and systems allowed for a comprehensive evaluation framework to be formed. This consisted of repeat victimisation, harm, CFS, PUD, activations, arrests and charges. Using this framework, the research finds that overall there is no treatment effect of an audible alarm with A6 warning stickers attached to points of entry to the premises, compared to the national standard alarm system and no warning sticker. Arrest rate differences are however noteworthy. The doubling of arrests with the RDA3 treatment was statistically significant at the 0.1 level. This corresponded with 80% more charges. However, the small sample size (N=300) creates a vulnerability that this is statistical noise. More research is needed with larger and more diverse populations. Still, the findings are informative in terms of theory, practice and future research.

5.2 IMPLICATIONS FOR THEORY AND KNOWLEDGE BASE

The literature review identified significant gaps in knowledge. The first of these was an understanding of 'what works' to prevent harm and repeat victimisation for high-risk DA victims. The second was understanding whether visual warning cues detailing audio recording technology deterred offenders. The third was whether panic alarms work. The term 'work' is double-faceted, based on reductions of harm and repeat victimisation.

Firstly, addressing the knowledge gap as to what works, this research adds to a growing number of RCTs targeted at high-risk DA victims. RCTs can be successfully conducted for high-risk DA victims with a broad range of output and outcome measures achievable.

Secondly, in seeking to understand whether panic alarms work for DA, this study identifies that following a police officer identifying an individual they consider to be a high-risk DA victim, responses that utilise panic alarms achieved statistically significant reductions in harm and repeat victimisation. This research was unable to isolate panic-alarms as causing that reduction.

Thirdly, addressing whether warning stickers deter offenders, this dissertation did not resolve this completely. Warning stickers were not tested in isolation from audio alarms systems. No evidence was found to support visual warning stickers having an effect on reducing harm and repeat victimisation to high-risk DA victims. Overall, no treatment effect was seen when the national standard alarm system was compared against an innovative audio alarm system with visual deterrent stickers.

Addressing the implications for theory it is posited that crime can be prevented when the chances of apprehension are high, the outcomes are swift and unattractive (Von Hirsch et al 1999). Facing this situation, an offender changes their behaviour in line with more acceptable social norms (Ariel et al 2015). An audio recording alarm system that uses visible red warning stickers at the point of entry to the homes of high-risk DA victims and achieves double the arrest rate did not achieve an increased deterrence effect. Two possible reasons for this outcome are now detailed.

The first reason is linked to a common, but incorrectly held belief that arrest is a punishment per se. Arrest is merely a tactical option at a relatively early stage of the CJ process. Nothing in the legislation surrounding arrest necessity talks about punishment. Punishment is

something that occurs upon conviction. A key part of the British Government's response to the pandemic was to introduce restrictions on physical contact. This exasperated existing delays in the CJ process. Punishment received from any subsequent conviction may not have been swift enough or suitably unattractive to positively impact the behaviour of offenders.

Secondly, deterrence theory requires an offender to perceive apprehension risk as being high (Dunford 1990). The only visible difference between the RDA2 treatment and the RDA3 treatment were A6 warning notices attached to entry points. It may well be that these signs were ignored, not understood (Knutsson 1984) or discounted. More research is needed, preferably through interviews with offenders exposed to the treatment effect, to understand the extent to which these offenders have or not have not noticed the sticker.

5.3 Policy Implications

Four policy implications have been identified by this research. The first is for a policy of affording high-risk DA victims their choice of panic-alarm. This choice is afforded due to both alarm systems achieving statistically significant reductions in harm and repeat victimisation in the pre to post comparisons.

Secondly, policies seeking to increase arrest rate for high-risk DA victims could consider the RDA3 alarm system. Use of an audible-alarm in this study generated statistically significant increases in arrests at the 0.1 level with the audible alarm system treatment having an arrest rate double that of the national standard alarm.

Thirdly, a policy implication to invest in research to better understand the situations where a specific alarm is preferable is advised. This follows the identification that the results seen offer the potential to tailor the panic-alarm system to individual victims. It may be that one alarm

is more effective than another depending on prior harm, levels of victimisation, presence of an order or the type of property where a victim resides.

Lastly, this research identified a number of victims who turned down audio-alarms because of an intrusion into their privacy (Victims 6, 32, 60, 82, 110, 158, 190, 197 and 203). Concerns as to the intrusion into one's private life need to be well-thought-out when considering which police tactics would work best to prevent harm and repeat victimisation. This also raises important questions as to the role of policing in a family setting.

5.4 Research Implications

This is the first piece of research to use an RCT methodology to test panic-alarms. The first research implication addresses the gap in testing what works to best protect high-risk DA victims. The void can be closed using the gold-standard research methodology of RCTs (Ariel et al 2015; Natarajan 2016) to show causal relationships between police practice and levels of harm and repeat victimisation suffered for high-risk DA. Far more research needs to focus in this space.

Secondly, this research was conducted on a small sample of 300 victims across 13 London boroughs, during a global pandemic. There are obvious limitations to this research's external validity and statistical power. This is illustrated in the analysis on arrest. Whilst the RDA3 system achieved 0.2 arrests per victim compared to 0.1 per victim in the RDA2 system treatment, a larger sample size is required to better understand whether arrest rate is increased by the audio capture. This would also allow for a better understanding of theoretical implications generated. The findings presented are limited by this context. Replication of this research design is essential (Hagan 2006). To support this, the methodology detailed continued to 888 cases with higher levels of fidelity. This increased sample-size needs

to be analysed with consideration of a longer period of evaluation. This may well change our understanding of the effectiveness of this technological innovation on high-risk victims of DA (Sherman and Harris 2015).

Thirdly, the previously discussed opportunity to tailor alarms to best suit victims opens up a research requirement to analyse sub-samples such as ethnicity and age. This is not something this research has completed.

Fourthly, the research considers there to be just one victim of DA. Children who reside are not included within the definition of victim here. A broader understanding of alarm effectiveness should look to understand the impact on children within a DA setting.

Fifthly, there was evidence within the research of the audible alarm being effective as a method of gathering evidence to help support the enforcement of orders and bail. This is demonstrated by case 138. Here the audio recording captured the offenders voice at the prohibited location which supported a conviction for breach of DVPO. It may be that audio alarms are an effective tool in the enforcement of orders specifically. The RDA3 alarm system should be further researched to understand whether audio alarms combined with the presence of an order reduce harm and repeat victimisation compared to either no alarm or the national standard alarm.

Lastly, this research has shown that the evaluation criteria for testing what works can successfully measure a number of outputs and outcomes using existing police systems. Future research should consider equally broad evaluation criteria to fully understand the treatment effect of that intervention.

5.5 LIMITATIONS OF STUDY

This research has extended the knowledge base on tested police practice for DA and specifically what works for reducing harm and repeat victimisation for high-risk victims. However, there are three clear limitations of this study. These are now detailed.

The first limitation focuses on the population from which the sample of 300 high-risk victims were taken from and size of that sample. Whilst this study covered 13 London boroughs, some 3,676,154 individuals (London Councils 2018) or 41.29 % of London's population, this is still a study of less than half of an incredibly diverse major global city. The external validity of these findings is likely to be limited both to the wider London population and alternative populations such as those in rural settings. Similarly, the 300-case sample within this trial was selected based on predicted case-flow, an ambition to review outputs and outcomes in the six months post alarm request and deadlines set by the MSt course. A larger sample size with greater statistical power and longer post-test period would have been preferable.

The second limitation is the time-period in which this study occurred. Unforeseen circumstances meant that this trial commenced during a global pandemic. During the trial period, two major periods of national lockdown were directed by law. This resulted in significant concern about the under-reporting of DA. Lockdown law exasperated delays in the CJ system. This included DA prosecutions. It is unclear whether and to what extent behaviour seen within this trial was affected by the national lockdown and impact of the pandemic.

The last limitation is the methodology chosen. Use of an RCT allows for cause and effect relationships to be understood and is a golden standard for quantitative research. However, this methodology does not address 'why' any effect may be seen. In social science this is a valuable knowledge to have. Future research should seek to enrich the knowledge base of

panic alarms in a DA setting through qualitative methodologies. These should focus on both the victim and perpetrator.

6.0 CONCLUSION

The scientific approach that underpins evidence-based policing has been replicated here to test police practice. This approach forms part of a continued movement towards policing practice being underpinned by the best available scientific evidence (Sherman 1998) and an approach termed 'triple-t', targeting, testing and tracking (Sherman 2013). This research assists in narrowing the knowledge gaps that exist in DA, tertiary crime prevention, visual warning cues and panic alarm technology. The research outlined within this thesis is one of few RCTs to focus on high-risk victims. This study has evidenced a number of key findings.

Firstly, use of RCT's as a golden standard research methodology is achievable in policing and by pracademics. Secondly, high fidelity rates are achievable. In this research an 84.7% fidelity rate was achieved through tight tracking, quick feedback loops, great partnership working and a focus on automation where possible. Thirdly, within this data set, overall findings identify that new innovations in audio alarm technology supported by visual warnings have no treatment effect compared to the national standard alarm system. Fourthly, increased arrest rates and charge rates seen in this study should be further explored within the larger 888 cases that were successfully randomised, but unable to be analysed due to limitations of time within this MSt course. Fifthly, available research and literature discussed within this study identified cases of exaggerated findings, inconsistent explanations and contradictory conclusions. Continued investment in police officer knowledge such as the University of Cambridge MSt in Applied Criminology and Police Management is required to ensure police decision makers are able to critically review available literature and make informed decisions

that affect service delivery to the public. DA is a significant area of challenge for policing and partners. An evidence-led approach to DA within policing can achieve improved reductions in harm and repeat victimisation. This focus must remain a priority for policing and academic partners.

APPENDICES

Appendix A: Pilot Details

In order to test the functionality of the system a pilot was conducted. This started on December 16th 2019 in the borough of Lewisham. Lewisham was selected due to high alarm use volume as well as both project leads having previously worked there with the operational teams. The first alarms were installed between the 19th of December and the 20th of December.

Pilot Activations:

The first activation occurred on the 3rd of January 2020. This did trigger an emergency response, but the audio did not send from the device. NM thought it theoretically possible to obtain the audio from the device remotely and were successful in doing so. The audio obtained did capture evidence to support a possible criminal damage occurring and detailed that the offender had left the premises with at least one kitchen knife. Upon police arrival the victim did not relay this information to police.

The second audio alarm activation occurred on the 19th of January. The audio this time successfully left the device and went to NM, but failed to get through the MPS systems due to file size limits. These were subsequently expanded. The victim in the second activation triggered the alarm because she thought that the perpetrator, who was on police bail was at her door. This case turned out to be another individual and not the perpetrator.

The third activation achieved audio landing into the MPS, but due to a manual requirement for the MPS Command and Control function to move the audio recording into a folder available to Emergency Response Teams, this failed to reach the attending officer. This is resolved with the creation of distinct folder for audio alarm recordings with an automated

pathway. The pre-activation audio recording again failed to send from the device. A 'bug' in the system was identified as the problem and fixed. This audio activation had a Spanish speaking victim who control room staff thought was Italian. This challenge was overcome with use of 'Language Line', the MPS solution to the need for interpreters.

On the 27th of January 2020 the fourth audio activation achieved evidential capture of a breach of DVPO. Here, the offender was heard saying "I am just going to have to go back to prison". In this case the offender was found outside of the property, a location not in breach of the order. However, the audio recording evidenced that the order had been breached and the individual was arrested and charged.

The pre-trial finished on February 19th 2020. Numerous challenged were identified and overcome. In total the pilot delivered the following:

Length of Pre-Trial = 65 days
Total Alarms Requested = 59
Total Activations = 10 or 1 per week on average
Total Successful Alarm Installs = 50 or 84.7%
Total Failed Alarm Installs = 9 or 15.3%

At the end of the trial the process was working consistently.

<u>Some of the Challenges Identified:</u>

- Victims not being at home when agreed upon to allow the engineer to install the alarm.
- Victim consent. A victim with learning difficulties was considered as not being able to consent and phone number of the person who could consent was incorrectly captured.

- Daily feedback from officers in the pre-trial site was one of more prisoners than officers on duty in Safeguarding Unit. This resulted in longer shifts, heightened stress and a more challenging pilot than initially considered.
- Local leaders at Detective Sergeant level and Detective Inspector level fed back a lack of awareness of the trial.
- Refusal of stickers on the basis of it concerning the neighbours.
- Offender opened the door to an engineer pretending to be the victim. When challenged the offender ran from the address.
- The placing of audio alarms next the television was identified as very challenging when trying to hear what is happening or has happened. Fed back to engineer team.
- The flow of alarms was suggestive of far more being needed that initially anticipated. Further parts were ordered from China.
- Delay in the fulfilment of an order for parts that would allow for 450 alarms to be built from China due to Chinese New Year celebrations.
- Following Chinese New Year celebrations, China experiences Covid-19 crisis and lockdown. The global pandemic follows. Production in China's factories reduces significantly.
- Rotation of staff on shift and different individuals chairing the morning meeting in Lewisham is resulting in slow updates to cases raised.
- Engineers raise a concern about possible damage caused to front door paint when stickers are removed and who would be liable. Sticker adhesive levels tested on various paint surfaces for reassurance.

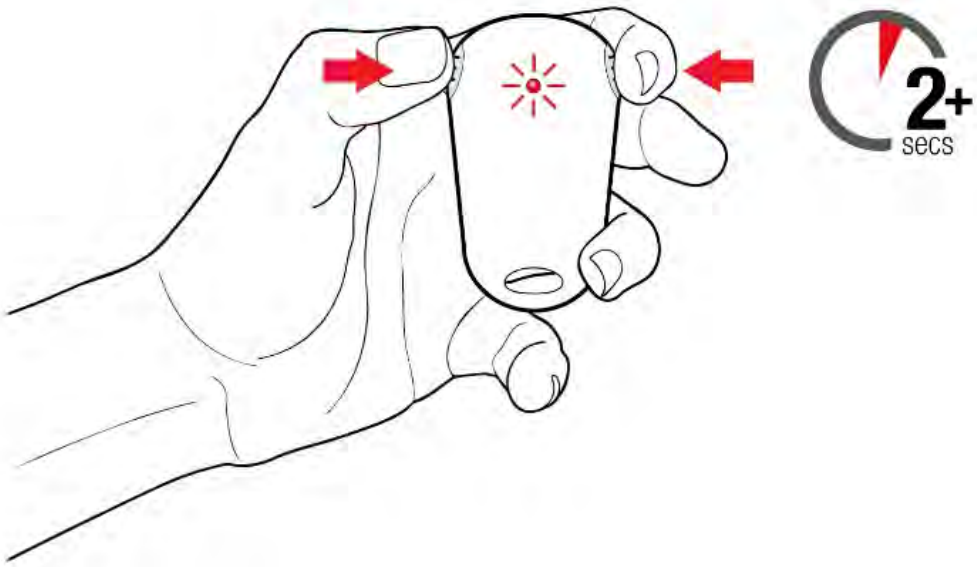
Successes:

- Randomisation is working well.
- Photograph capture of stickers is working well and allowing for quick feedback if stuck in the wrong place.
- Regular data auditing is checking that the audio alarms are working.
- Engineers for NM are feeling part of the policing family.
- Good evidential recording.
- Remote accessing of audio achieved.
- Engineers happily calling in cases of concern for well-being.
- Feedback-loop allowing for alterations to be made to the way the data-base is set-up to capture data. Greater emphasis on the use of structured text.

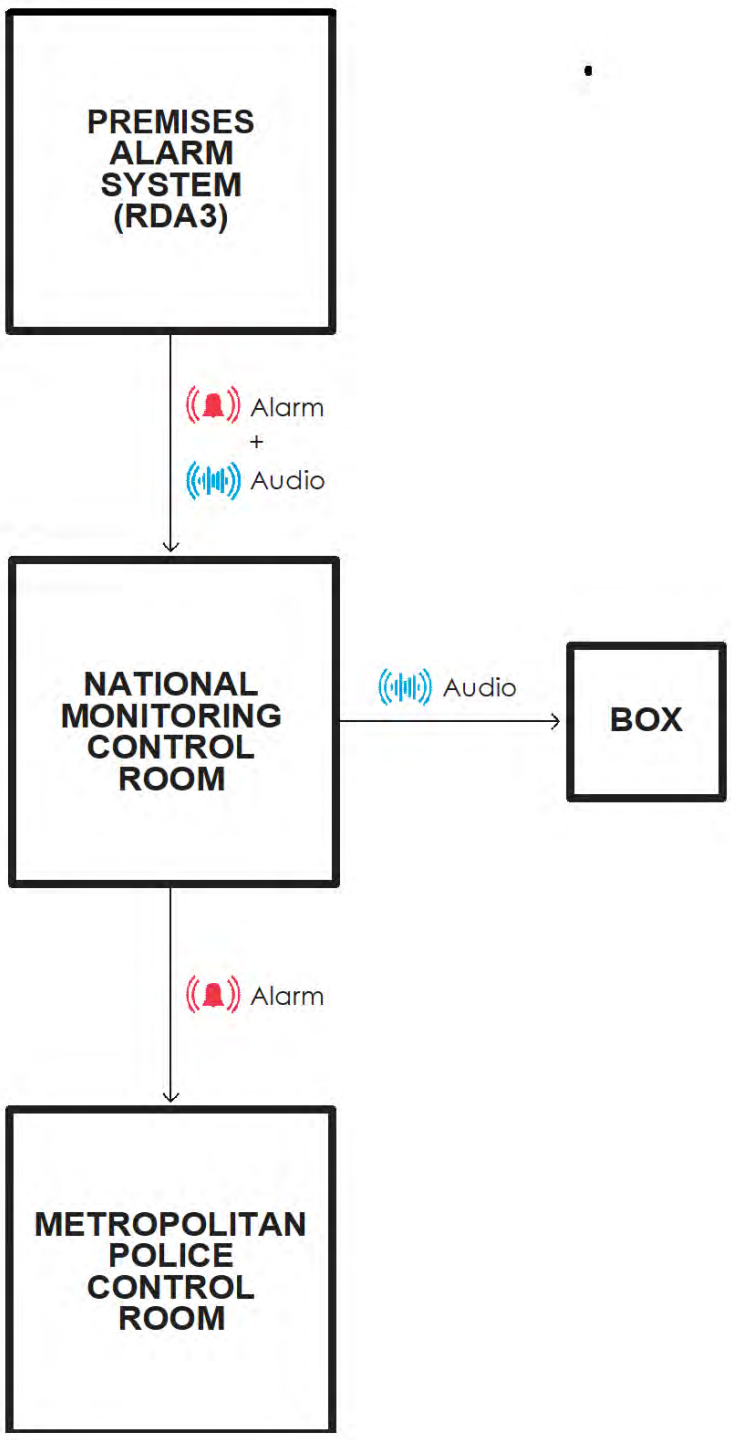
Appendix B: RDA2 Alarm

Appendix C: RDA3 Alarm

Appendix D: Alarm Activation Fob



Appendix E: High Level RDA3 Process Map



Appendix F: Audio Alarm Consent Form

CONSENT FORM

SYSTEM ID	

I consent to the installation of this audio panic alarm in this address.

I confirm that I have been given instructions about how the audio panic alarm works. I understand how to activate it. I understand that the device will only capture audio once the alarm is activated sending National Monitoring and the Metropolitan Police Service the 5 minutes of audio recorded before alarm was activated and then continuing to send audio until the recording is switched off because the police have arrived or it has been decided the recording can stop.

Whilst the audio panic alarm has been tested to be reliable, I understand that the audio panic alarm is not a guaranteed service. I will where possible try to contact the police as well.

I understand that is important that anyone entering this address is aware of the audio panic alarm. Notices have been put up at all the entrances to your address. If the notices are removed, people might say that the audio panic alarm was recording them without their knowledge and this could stop the police being able to use any evidence it records. I consent to these notices being used.

I am aware of the Metropolitan Police Service's Privacy Notice, their policy for Protecting Special Category and Criminal Convictions Policy and have also been given information as to how National Monitoring and the Metropolitan Police Service work together to provide me with an audio panic alarm and then process any data recorded by it.

I understand that the Metropolitan Police Service is the data controller and processes my personal data based on it being necessary for its law enforcement purposes.

To let the police perform their law enforcement role, I consent to any recording being listened to by the police and National Monitoring and understand any recording can be used or played in court.

Name:_____

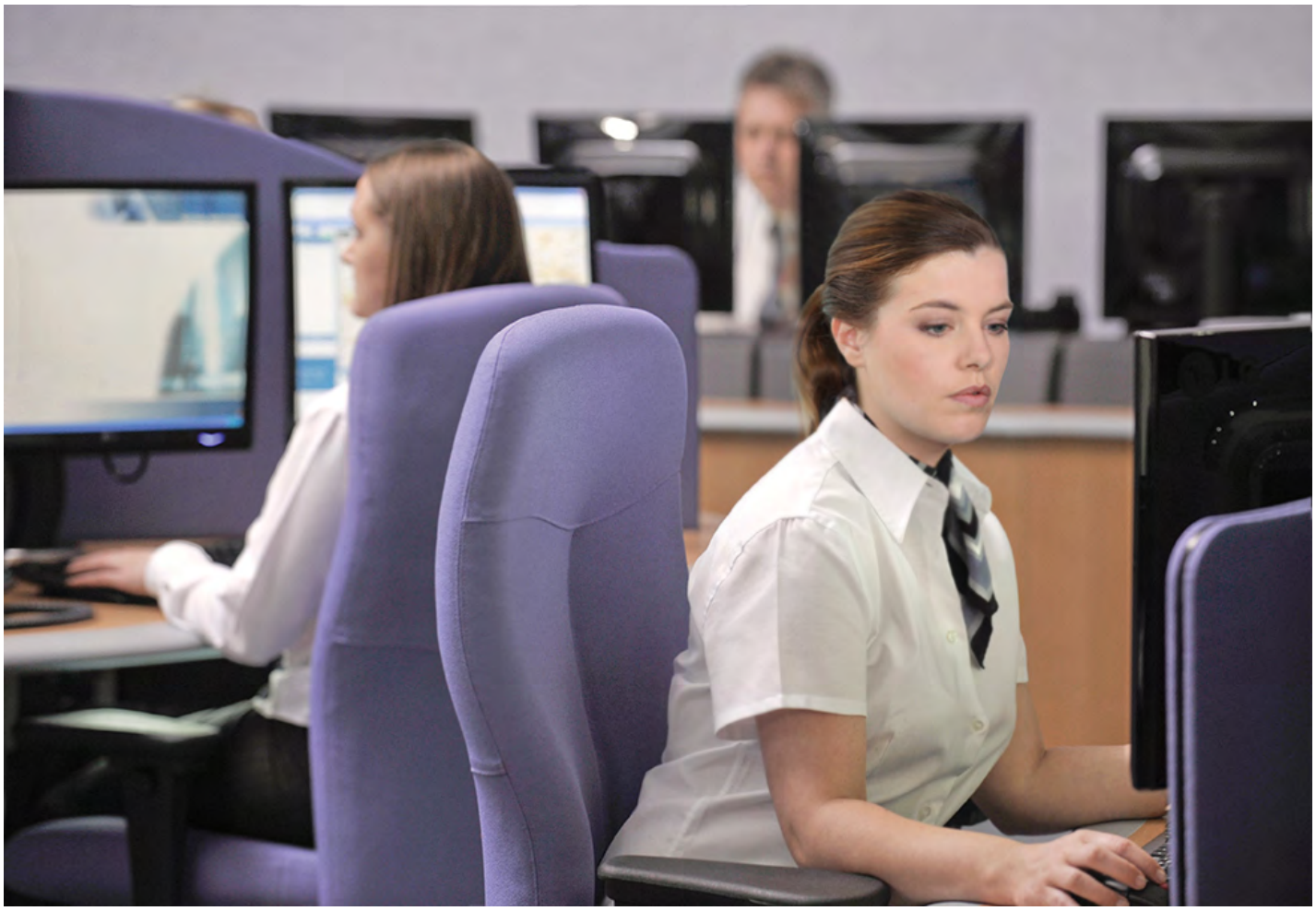
Signed:_____ Date: _____

If you no longer wish to have this device and/or display the warning notices you need to inform your local safeguarding unit via 101.

Should you wish to discuss how your data is processed please call 03303336619

Appendix G: A6 Audio Alarm Warning Sticker

Appendix H: NM Control Room



BIBLIOGRAPHY

- Ariel, B., Farrar, W.A. and Sutherland, A. (2015) 'The Effect of Police Body Worn Cameras on Use of Force and Citizens' Complaints Against the Police: A Randomized Controlled Trial', *Journal of Quantitative Criminology*, 31: 509-535.
- Bachman, R.D. and Schutt, R.K., (2017) *The Practice of Research in Criminology and Criminal Justice,* Thousand Oaks, California: Sage Publishing.
- Barnham, L., Barnes, G.B. and Sherman, L.W. (2017) 'Targeting Escalation of Intimate Partner Violence: Evidence from 52,000 Offenders', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 116-142.
- Bland, M. and Ariel, B. (2015) 'Targeting escalation in reported domestic abuse: evidence from 36,000 callouts', *Criminal Justice Review,* 25 (1): 30-53.
- Bridger, E., Strang, H., Parkinson, J. and Sherman, L.W. (2017) 'Intimate Partner Homicide in England and Wales 2011-2013: Pathways to Prediction form Multi-agency Domestic Homicide Reviews', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 93-104.
- Brown, G.E., Paige, J.A. and Godin, J-G.J. (2000) 'Chemically mediated predator inspection behaviour in the absence of predator visual cues by a characin fish', *Animal Behaviour,* 60(3): 315-321.
- Buck, A.J., Hakim, S. and Rengert, G.F. (1993) 'Burglar Alarm and The Choice Behaviour of Burglars: A Suburban Phenomenon', *Journal of Criminal Justice*, 21(5): 497-507.
- Button, I.M.D., Angel, C. and Sherman, L.W. (2017) 'Predicting Domestic Homicide and Serious Violence in Leicestershire with Intelligence Records of Suicidal Ideation

or Self-Harm Warnings: A Retrospective Analysis', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 105- 115.

- Chalkey, R. and Strang, H. (2017) 'Predicting Domestic Homicide and Serious Violence in Dorset: a Replication of Thornton's Thames Valley Analysis', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 81-92.
- Cohen, H.H. and Sloan, G.D. (2016) 'The science behind codes and standards for safe pedestrian walkways: lighting and visual cues', *Applied Ergonomics*, 52: 112-119.
- Cohen, L.E., and Felson, M. (1979) 'Social change and crime rate trends: a routine activity approach', *American Sociological Review*, 44(4): 588-608.
- Dowling, C., Morgan, A., Boyd, C., and Voce, I. (2018) 'Policing domestic violence: A review of the evidence', *AIC reports. Research Report series*, xii.
- Drover, P. and Ariel, B. (2015) 'Leading an Experiment in Police Body-Worn Video Cameras', *International Criminal Justice Review*, 25(1): 80-97.
- Dunford, F. W. (1990) 'System-initiated warrants for suspects of misdemeanour domestic assault: A pilot study', *Justice Quarterly,* 7: 631-653.
- Farrell, G. and Pease, K. (1993) *Once Bitten, Twice Bitten: Repeat Victimisation and its Implications for Crime Prevention*, London: Home Office, Police Research Group, Crime Prevention Unit Series Paper No. 46.
- Farrell, G., Clarke, K. and Pease, K. (1993) 'Arming the Toothless Tiger: Family Protection Orders, Court Injunctions and Domestic Violence', *Justice of the Peace*, February 6: 88-90.
- Fixsen, D.L, Naoom, S.F, Blasé, K.A, Friedman, R.M. and Wallace, F. (2005) (eds) *Implementation Research: A synthesis of the literature*, Tampa: University of South Florida.

- Gaines, L. and Bichler, G. (2007) *Police Response to Burglar Alarms Study*, San Bernardino County, Report to San Bernardino County Police Chiefs and Sheriff's Association.
- Goosey, J., Sherman, L.W. and Neyroud, P. (2017) 'Integrated case management of repeated intimate partner violence: a randomized, controlled trial', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 174-189.
- Grove, L.E., Farrell, G., Farrington, D.P. and Johnson, S.D. (2012) *Preventing Repeat Victimization: A Systematic Review*, Stockholm: Bra, National Council for Crime Prevention.
- Guy, S. (2012) 'Prescription warning labels need to attract more attention', S*pringer Healthcare News*, 521(1): 1-5.
- Gwillim, J. (2018) 'Implementing a Randomised Controlled Trial of Domestic Violence Protection Notices (DVPNs): A Protocol and Enrolment Phase Study', Unpublished in MSt. thesis, University of Cambridge.
- Hagan, F.E. (2006) *Research Methods in Criminal Justice and Criminology,* New York: Pearson Education.
- Hagan, F.E. (2014) *Research Methods in Criminal Justice and Criminology.* 9th ed*.,* New York: Pearson Education.
- Hakim, S., Rengert, G.F. and Shachmurove, Y. (1995) 'Burglar and Fire Alarms: Costs and Benefits to the Locality', *The American Journal of Economics and Sociology*, 54(2): 145-161.
- Jones, R.B. and Nowell, N.W. (1973) 'The effects of familiar visual and olfactory cues on the aggressive behaviour of mice', *Physiology and Behaviour*, 10: 221-223.

- Kirby, S. (2013) *Effective Policing Implementation in Theory and Practice*, London: Palgrave MacMillan.
- Knutsson, J. (1984) *Operation Identification? A Way to Prevent Burglaries?* Stockholm, Sweden: National Council for Crime Prevention
- Kyvsgaard, B. and Sorensen, D.W.M. (2020) 'Do stickers indicating the use of forensic property marking prevent burglary? Results from a randomized controlled trial', *Journal of Experimental Criminology,* 17: 287-303.
- Linton, B. and Ariel, B. (2020) 'Random Assignment with a Smile: How to Love "TheRandomiser"', *Cambridge Journal of Evidence-Based Policing*, 4(3-4): 233-237.
- London Assembly (2020) Population and Household Projections. Retrieved 23rd July 2021 from https://data.london.gov.uk/demography/population-and-household-projections/
- London Councils (2018) London's Population. Retrieved 24th May 2021 from https://directory.londoncouncils.gov.uk/demographics/population/
- Lum, C., Koper, C., Wilson, D.B., Stoltz, M., Goodier, M., Eggins, E., Higginson, A. and Mazerolle, L. (2020) 'Body-worn cameras' effects on police officers and citizen behaviour: A systematic review', *Campbell Systematic Reviews*. 2020: 16: e1112
- Mackinnon, D.P. (1995) 'Review of the effects of the alcohol warning label', *Alcohol, Cocaine and Accidents*, 7: 131-161.
- Maxfield, M.G. and Babbie, E.R., (2015) *Research Methods for Criminal Justice and Criminology,* 8th ed*.,* Boston, MA: Cengage Learning.
- Mayor's Office for Policing and Crime (MOPAC) (2015) MOPAC Annual Report 2014-15. Retrieved 24th of May 2021 from https://www.london.gov.uk/what-we-

do/mayors-office-policing-and-crime-mopac/our-priorities/police-and-crime-plan-2013-2016/mopac-annual-report

- MPS Internal Report (2020) *Operation Athena II Performance Pack June 2020.*
- Natarajan, M. (2016) 'Police response to domestic violence: A case study of TecSOS mobile phone use in the London Metropolitan Police Service', *Policing: A Journal of Policy and Practice*, 10(4): 378-390.
- Nevo, I., Fitzpatrick, M., Thomas, R.E., Gluck, P.A., Lenchus, J.D., Arheart, K.L. and Birnbach, D.J. (2010) 'The efficacy of visual cues to improve hand hygiene compliance in Simulation in Healthcare', *The Journal of the Society for Simulation in Healthcare,* 5(6): 325-331.
- Newcomer, K.E., Hatry, H.P. and Wholey, J.S. (2015) *The Handbook of Practical Program Evaluation,* 4th Ed., New-Jersey: Wiley.
- Office of National Statistics (ONS) (2020) Population Estimates. Retrieved 31st of July 2021 from https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates
- Oliver, R., Barnaby, A., Roe, S. and Wlasny, M. (2019) *The Economic and Social Costs of Domestic Abuse: Research Report 107*, London: Home Office (2019).
- Perkins, C., Beecher, D., Aberg, D. C., Edwards, P., and Tilley, N. (2017) 'Personal security alarms for the prevention of assaults against healthcare staff', *Crime Science*, 6(1): 11.
- Radford, L., Aitken, R., Miller, P., Ellis, J., Roberts, J. and Firkic, A. (2011) Meeting the needs of children living with domestic violence in London: Research report (Refuge/NSPCC research project November 2011), London: Refuge.

- Raphael, I. (2015) 'Cooling hot property? An assessment of the impact of property marking on residential burglary crime reduction, crime displacement or diffusion of the benefits and public confidence'. Unpublished Ph.D. thesis, University of Portsmouth.
- Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002) *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, New York: Houghton Mifflin Company.
- Sherman, L.W. (1998) 'Ideas in American Policing: Evidence Based Policing', in *Police Foundation*: Washington DC.
- Sherman, L.W. (2007) 'The power few: experimental criminology and the reduction of harm', *Journal of Experimental Criminology*, 3(4): 299-321.
- Sherman, L.W. (2013) 'Targeting, Testing and Tracking Police Services: The Rise of Evidence-Based Policing 1975-2025' *Crime and Justice*, Vol.43: 1-59.
- Sherman, L.W. (2018) 'Policing Domestic Violence 1967-2017', *Criminology and Public Policy*, 17(2): 453-465.
- Sherman, L.W. and Berk, R.A., (1984), 'The specific deterrence effects of arrest for domestic assault', *American Sociological Review*, 49:2: 261-272.
- Sherman, L.W. and Harris, H., (2015), 'Increased death rates of domestic violence victims from arresting vs. warning suspects in the Milwaukee Domestic Violence Experiment (MilDVE)', Journal of Experimental Criminology, 11: 1-20.
- Sherman, L.W. and Harris, H.M. (2013) 'Increased homicide victimization of suspects arrested for domestic assault: a 23-year follow-up of the Milwaukee Domestic Violence Experiment (MilDVE)', *Journal of Experimental Criminology*, 9(4): 491-514.

- Sherman, L.W., Gottfredson, D.C., Mackenzie, D.L., Eck, J., Reuter, P., and Bushway, S.D. (1997) 'Preventing Crime: What Works, What Doesn't, What's Promising', in National Institute of Justice: Research in Brief, U.S. Department of Justice: Office of Justice Programs, Maryland: University of Maryland.
- Sherman, L.W., Strang, H. and O'Connor, D. (2017) 'Introduction – Key Facts About Domestic Abuse: Lessons from Eight Studies', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 59-63.
- Strang, H., Sherman, L., Ariel, B., Chilton, S., Braddock, R., Rowlinson, T., Cornelius., Jarman, R. and Weinborn, C. (2017), 'Reducing the Harm of Intimate Partner Violence: Randomized Controlled Trial of the Hampshire Constabulary CARA Experiment', *Cambridge Journal of Evidence Based Policing,* 1: 160–173.
- The Randomiser (n.d.). Retrieved 1st of August 2020 from https://www.therandomiser.co.uk/admin/
- Thornton, S. (2017) 'Police Attempts to Predict Domestic Murder and Serious Assaults: Is Early Warning Possible Yet?', *Cambridge Journal of Evidence-Based Policing*, 1 (2-3): 64-80.
- Tilley, N. and Webb, J. (1994) *Burglary Reduction: Findings from Safer Cities Schemes,* London: Home Office, Police Research Group Crime Prevention Series: Paper No.51.
- Von Hirsch, A., Bottoms, A., Burney, E. and Wikstrom, P.O. (1999) *Criminal deterrence and sentence severity: An analysis of recent research*, Oxford: Hart Publishing.
- Walker, S. (2001) 'Domestic violence: Analysis of a community safety alarm system' *Child Abuse Review*, 10(3): 170-182.

- Wilcox, A. and Hirschfield, A. (2007) *A Framework for Deriving Policy Implications from Research,* Huddersfield: The University of Huddersfield.