

# Secondary Data Analysis of Youth Diversion in London



## Secondary Data Project Analysis Plan

Evaluating institution: Behavioural Insights Team & the Metropolitan Police Force

Principal investigator(s): Edward Flahavan, Tim Hardy, Neeraj Rahal

---

## Analysis Plan for YEF Secondary Data Analysis Projects

### Project summary

Project title	Secondary Data Analysis of Youth Diversion in London
Research Team	Behavioural Insights Team (BIT) and the Metropolitan Police Force (MPS or 'the Met')
Principal investigator	Ed Flahavan
Analysis plan author(s)	Ed Flahavan, Neeraj Rahal, Tim Hardy
Overarching research question	What is the impact of diversion on youth offending?
Supporting research question(s)	<ol style="list-style-type: none"><li>1. Can police data and Youth Justice Services (YJS) data be linked to improve understanding and monitoring of diversion?</li><li>2. What are the characteristics of the Children and Young People (CYPs) who are diverted, and how does this compare to CYPs who aren't diverted?</li><li>3. Is there variation across London boroughs in the use of different types of diversion and who is diverted?</li><li>4. What factors predict a CYP not admitting to an offence?</li><li>5. What is the relationship between diversion and re-offending?</li><li>6. What types of diversion are most effective at reducing re-offending?</li></ol>

Dataset(s) to be used	<p>Datasets will be drawn from data held in systems by MPS and YJS, including:</p> <p>MPS systems:</p> <ul style="list-style-type: none"> <li>● Crime Report Information System (CRIS)</li> <li>● Case Overview and Preparation Application (COPA)</li> <li>● National Strategy for Police Information Systems custody system (NSPIS)</li> <li>● Missing Persons and Related Linked Indices (MERLIN)</li> <li>● Stops</li> </ul> <p>YJS systems<sup>1</sup>:</p> <ul style="list-style-type: none"> <li>● Core Plus</li> <li>● Career Vision</li> <li>● Child View</li> </ul>
Population characteristics	10 - 17-year-olds when they are first arrested, or recorded as being suspected of committing a crime
Years data spans	2015 - 2022
Geographic coverage	Greater London, England
Primary outcome(s) investigated	<ul style="list-style-type: none"> <li>● Violent offending (as measured in official police records)</li> <li>● Total offending (total number of arrests, or charges for any type of offence, within the period of this study)</li> </ul>
Main method(s) to be used or tested	Descriptive analysis, regression analysis and quasi-experimental designs (instrumental variables, difference-in-differences, matching, pre-post - exact method to be used TBD at interim reporting stage)

### Analysis plan history

Version	Date	Reason for revision
1.X [ <i>latest</i> ]		

<sup>1</sup> Some YJSs may use different systems to store data regarding the CYPs they work with, and the outcomes of youth diversion programmes. Therefore data may be collected from additional IT systems used by various YJS teams.

1.1		
1.0 [ <i>original</i> ]	June 2023	

## Table of contents

<b>1. About the project</b> .....	<b>5</b>
1.1. Background to the project.....	5
1.2. Research question(s) .....	6
1.3. Hypotheses .....	10
1.4. Key concepts.....	14
<b>2. About the datasets</b> .....	<b>18</b>
2.1. Overview of datasets used .....	18
2.2. Secondary data source(s) .....	19
2.3. Primary data collection.....	21
2.4. Linking datasets .....	21
2.5. Access and data protection .....	23
<b>3. About the data</b> .....	<b>24</b>
3.1. List of variables .....	24
3.2. Measurement of key concepts .....	30
3.3. Missing data and attrition .....	32
3.4. Other sources of bias.....	33
<b>4. About the analysis</b> .....	<b>34</b>
4.1. Overview of analytical approach .....	34
4.2. Approach to addressing research question(s).....	37
<b>5. Project management</b> .....	<b>55</b>
5.1. Risks and mitigations .....	55
5.2. Timeline .....	57
<b>6. References</b> .....	<b>59</b>

## 1. About the project

### 1.1. Background to the project

Against the backdrop of increased pressure on courts leading to delays in hearing and prosecuting cases,<sup>2</sup> there is renewed interest in the potential for youth diversion to keep young offenders out of the Criminal Justice System (CJS), reduce re-offending, and improve life outcomes.<sup>3</sup> Youth diversion broadly refers to schemes or activities which do not result in a criminal record, avoid escalation into the formal CJS, and provide an alternative outcome for children or young people (CYPs) who have been linked to an offence.<sup>4</sup>

Youth diversion has shown promising evidence in reducing re-offending.<sup>5 6</sup> Moreover, when young people who are involved in criminal activities progress through the formal CJS (e.g., court proceedings or incarceration), they are more likely to commit additional offences. For example, an international meta-analysis of 29 studies involving over 7,300 CYPs, found that formal processing through the CJS increased the prevalence, incidence and severity of future offences.<sup>7</sup>

However, there is limited research into the effects of youth diversion in the UK. Whilst there have been a number of studies which investigate the effects of youth diversion, the majority of studies are from the US, so the applicability to the UK context may not be conclusive.<sup>8</sup>

---

<sup>2</sup> National Audit Office (2021). Reducing backlog in the criminal courts. London: Crown Copyright.

<sup>3</sup> Crest (2022). Making the criminal justice system work better: how to improve out-of-court disposals and diversion scheme. London. Retrieved from: [https://64e09bbc-abdd-42c6-90a8-58992ce46e59.usrfiles.com/ugd/64e09b\\_7df3289b60fe45aaa442e5320d67cd98.pdf](https://64e09bbc-abdd-42c6-90a8-58992ce46e59.usrfiles.com/ugd/64e09b_7df3289b60fe45aaa442e5320d67cd98.pdf)

<sup>4</sup> Youth Justice Board (2021). Definitions for prevention and diversion. Retrieved from: [https://yjresourcehub.uk/images/YJB/Definitions\\_for\\_Prevention\\_and\\_Diversion\\_YJB\\_2021.pdf](https://yjresourcehub.uk/images/YJB/Definitions_for_Prevention_and_Diversion_YJB_2021.pdf)

<sup>5</sup> Gaffney, H., Farrington, D. P., & White, H. (2021). Pre-Court Diversion. Toolkit Technical Report. London: Youth Endowment Fund.

<sup>6</sup> YEF Toolkit. Retrieved from: <https://youthendowmentfund.org.uk/toolkit/>

<sup>7</sup> Petrosino A, Turpin-Petrosino C, Guckenberg, S. (2010). Formal System Processing of Juveniles: Effects on Delinquency. Campbell Systematic Reviews, 6: 1-88

<sup>8</sup> Gaffney et al.(2021)

There have been only two UK published studies on the effect of diversion, both of which were conducted over 10 years ago.<sup>9</sup>

Moreover, there is limited information about the provision and outcomes of youth diversion within the UK. It wasn't until April 2020 that the Youth Justice Board (YJB) began collecting aggregate data on the total number of CYPs receiving certain types of youth diversions from Youth Justice Services (YJS).<sup>10</sup> Currently, it's unclear what the number and profiles of CYPs being diverted are, and the effects of being diverted on outcomes like recidivism.

This project is co-delivered with the UK's largest police force, the Metropolitan Police Service (MPS or 'the Met'). It will involve obtaining data on CYPs held both by the Met, and by YJS teams across London. By gathering and linking police data with data held by YJS teams, this can offer unique insights into questions regarding youth diversion, which existing analysis has yet not answered.

Analysis of this data will help to address the research questions outlined below.

## 1.2. Research question(s)

The overall aim of the project is to understand which kinds of CYPs are diverted from the CJS in London, and what impact this diversion has on arrests and violent re-offending.

Specifically, this project will seek to answer the following research questions:

1. Can police data and YJS data be linked to improve understanding and monitoring of diversion? (RQ1)
2. What are the characteristics of the CYPs who are diverted, and how does this compare to CYPs who aren't diverted? (RQ2)
3. Is there variation across London boroughs in the use of different types of diversion and in who is diverted? (RQ3)
4. What factors predict a CYP not admitting to an offence? (RQ4)<sup>11</sup>

---

<sup>9</sup> Wilson, D. B., Brennan, I., & Olaghere, A. (2018). Police-initiated diversion for youth to prevent future delinquent behavior: A systematic review. *Campbell Systematic Reviews*, 14(1), 1-88.

<sup>10</sup> Youth Justice Board (2023), Prevention and Diversion Project – final report. Retrieved from: [https://yjresourcehub.uk/images/YJB/Prevention\\_and\\_Diversion\\_Project\\_Final\\_Report\\_YJB\\_Feb\\_2023.pdf](https://yjresourcehub.uk/images/YJB/Prevention_and_Diversion_Project_Final_Report_YJB_Feb_2023.pdf)

<sup>11</sup> Currently, it is unclear how researchers may be able to answer this research question. Based on conversations with stakeholders in the Met and YJS, there doesn't appear to be a consistent method to

5. What is the relationship between diversion and re-offending? (RQ5)
6. What types of diversion are most effective at reducing re-offending? (RQ6)

The research questions are listed in the order in which this project aims to answer them. In particular, RQ1 is a critical milestone for undertaking additional analyses to answer the remaining research questions, within the intended scope and approach of this project.

Given the complexity involved in obtaining and linking multiple sets of data, as well as potential issues with the type and quality of data held (both within the Met, and from different YJS teams), we note that it may not be possible to answer RQ1 – or that we find the answer to RQ1 is that data from the YJS teams and the Met cannot be linked.

In this case, this project will focus exclusively on utilising the data held by the Met, to fully address the remaining research questions possible. Given this project is being conducted in collaboration with the Met, obtaining data held by the Met is considered a feasible outcome. This will enable secondary data analysis to be undertaken without the use of data from YJS teams. It is acknowledged that reliance solely on Met data may limit the level of detail and depth of analysis to fully address the remaining research questions. Whilst we would anticipate being able to address all the remaining research questions, the lack of YJS data is likely to limit our ability to fully address research questions 3 and 6.

**Table 1.2. How will the questions be addressed at each stage?<sup>12</sup>**

Question Number	Interim report	Final report
1	Description of the feasibility of merging data from the Met and YJS teams, as well as a description of how internal Met data (from different systems) has been	Description of whether data sets from the Met and YJS were able to be linked to inform this research project. This will also include a description of which sources of data were obtained to undertake analysis for this project.

---

identify CYPs on the basis of whether they admitted to an offence (regardless of whether they were diverted or not). However, this may change as we gain greater clarity and sight of the data.

<sup>12</sup> Responses to how research questions will be answered are based on current understanding of data held by the Met and select systems used by YJS teams (Core Plus, Child View). Researchers have not yet had sight of the data held by the Met or YJS teams, limiting the level of detail able to be provided regarding how research questions may be answered in either the interim or final report.

	collated into a single dataset for analysis. <sup>13</sup>	Identification of challenges and barriers to the data collection process and ease of cleaning and merging data.
2	<p>Initial descriptive analysis of characteristics of CYPs who are diverted (including those who are diverted more than once).</p> <p>Initial descriptive analysis of characteristics of CYPs who have never been diverted, within the same time period.</p> <p>These characteristics are likely to include sex, self-identified ethnicity, age, number of previous arrests, reasons for arrest, type and severity of previous offences.<sup>14</sup></p>	<p>Descriptive analysis of characteristics of CYPs who are diverted (including those who are diverted more than once), as well as CYPs who have never been diverted.</p> <p>These characteristics are likely to include sex, self-identified ethnicity, age, number of previous arrests, reasons for arrest, type and severity of previous offences.<sup>15</sup></p> <p>For CYPs who are diverted this may also include the following additional characteristics: mental health conditions, religion, immigration status, care/care leaver status, number of previous arrests, number of previous convictions, number of previous diversions, number of successfully completed diversion activities.<sup>16</sup></p> <p>Additional analysis on whether differences between CYPs who are diverted, and CYPs who have never</p>

<sup>13</sup> It is noted that at the time of the interim report, data may not have been obtained from any YJS teams (or only from a limited number of YJS teams). Regardless, the interim report will outline which data sets have been obtained, and the extent to which these have been merged into a single dataset.

<sup>14</sup> Should data not be obtained (or be available) from YJS teams at time of interim report, this will rely on data held by the Met.

Based on our current understanding of data held by the Met, the characteristics of CYPs listed above should be available to researchers.

<sup>15</sup> It is currently unclear whether data by the Met, or YJS teams, capture 'severity of offences' within a separate data field in their systems. However, we may be able to determine the severity of offences, based on matching the offences with external metrics of crime harm or severity (e.g., the Cambridge Crime Harm index)

<sup>16</sup> Based on current understanding of data held by YJS teams, these characteristics should be available to researchers.

However, inclusion of all factors listed above is contingent on the outcome of RQ1, and the quality of data provided from YJS teams.



		been diverted are statistically significant, across each of these key characteristics.
3	Initial descriptive analysis of the number of CYPs provided with a type of diversion, split by Basic Command Units (BCUs) over time.	<p>Additional analysis of descriptive analysis (completed at time of interim report), based on any additional datasets received (i.e. from additional YJS teams).</p> <p>This will include descriptive analysis of disposal outcomes, across different BCUs, split by ethnicity, age, number of previous offences, type and severity of previous offences.</p> <p>We will also seek to include a descriptive analysis of CYPs who receive the 'Triage' type of diversion, across different BCUs split by the additional following characteristics: religion, immigration status, care/care leaver status.<sup>17</sup></p>
4	Identification of feasibility to answer research question <sup>18</sup> , as well as testing appropriate statistical method to undertake	Undertaking statistical method identified in interim report to identify factors predictive of admitting guilt in CYPs. <sup>20</sup>

---

<sup>17</sup> This descriptive analysis of additional characteristics is limited to CYPs who receive the 'triage' type of diversion, as these CYPs are referred to YJS teams, who we understand to hold information on these additional characteristics (i.e. immigration, religion, care/leaver status). Therefore it is unclear if, or how, these characteristics could be compared to CYPs who have not been diverted.

<sup>18</sup> We will be assessing the feasibility of this research question in the interim report, as it is unclear at this stage whether we will be able to feasibly extract data on whether CYPs have, or have not, admitted to an offence. We have been informed that this information is not captured within a specific data field in the Met's systems but is noted as free text within crime reports (which would require manual review of reports, for this data to be extracted for analysis). However, we have also been informed that admission of guilt may be captured in one system as a separate data field, though the quality and completeness of this data is unknown at this stage.

<sup>20</sup> This is contingent on determining whether such an analysis is feasible, at the interim report stage of the project.

	prediction modelling, based on data available. <sup>19</sup>	
5 & 6	<p>Assessment of feasibility for different quasi-experimental design (QED) methods to undertake analysis. Based on our understanding of the data available, the most likely options are instrumental variables design, difference-in-differences, matching, or pre-post design approaches.<sup>21</sup></p> <p>We will also perform regression analysis (which should not be interpreted causally) to investigate the relationship between being diverted and re-offending.</p>	Should feasible QED designs and appropriate statistical methods be identified at the interim stage of this project, this will be implemented as outlined shortly following the interim stage.

### 1.3. Hypotheses

#### Research Question 1

We do not have a hypothesis for this research question, given this research question is not linked to specific prior research or theories.

#### Research Question 2

We hypothesise that male CYPs from White backgrounds are more likely to be diverted.

There is limited information available about the characteristics of CYPs who are diverted in the UK. One study in the UK identifies the characteristics of CYPs ( $n = 1,027$ ) in an evaluation

---

<sup>19</sup> We anticipate using a logistic regression as the statistical method to answer this research question, which will include individual-level co-variables (e.g., sex, sex, self-identified ethnicity, age, number of previous arrests, type of offence, location of crime (BCU), etc.). However, as noted in the footnote above, this is dependent on being able to identify at an individual level in the data available, which CYPs did (or did not) admit to their offences.

<sup>21</sup> It is acknowledged that once we have sight of the data held by neither of these strategies may be tenable. In this case we will outline any additional options which may be able to inform a response to this research question – or whether none exist given the data available.

of a pilot diversion scheme in six sites across the UK.<sup>22</sup> Overall, researchers identified that CYPs involved in the diversion scheme were on average 14.7 years old, predominantly male (71%), and White British or Northern European (67.5%). Whilst this included pilot sites in two London boroughs, it is unclear how representative this is of existing characteristics of CYPs in London who are diverted from the youth CJS. This is particularly since London's population is more diverse, compared to populations served by other police forces in the UK.<sup>23</sup> Nevertheless, this suggests that male CYPs from White backgrounds may be more likely to be diverted.

Additionally, CYPs from Black and Asian backgrounds have been found to be significantly more likely to plead not guilty at court, compared to White CYPs.<sup>24</sup> This suggests that individuals from Black and ethnic minority backgrounds could also be less likely to admit to an offence when arrested and/or charged with an offence. However, we note this is based on individuals pleading guilty *at court*, and the same trends may not necessarily hold for individuals admitting their offences to police. Regardless, given that CYPs have to admit to offending to be diverted,<sup>25 26</sup> it is plausible that CYPs from BAME backgrounds may be less likely to meet this criterion (compared to White CYPs), and thus not be diverted from the CJS.

### Research Question 3

We do not have preferred hypotheses for this research question.

Although there is some limited information suggesting variation in the implementation of youth diversion practices across London, this evidence relies on anecdotal and incomplete survey data.<sup>27</sup> Moreover, this does not provide data indicating factors which may be linked

---

<sup>22</sup> Haines, A, Goldson, B, Haycox, A, Houten, R, McGuire, J, Nathan, T, Perkins, E, Richards, E and Whittington, R (2012) Evaluation of the Youth Justice Liaison and Diversion (YJLD) Pilot Scheme. Retrieved from: [https://e-space.mmu.ac.uk/621343/1/dh\\_133007.pdf](https://e-space.mmu.ac.uk/621343/1/dh_133007.pdf)

<sup>23</sup> Her Majesty's Inspectorate of Probation (2019). Annual report: inspection of youth offending services (2018-2019)

<sup>24</sup> Uhrig, N. (2016). Black, Asian and Minority Ethnic disproportionality in the Criminal Justice System in England and Wales. Ministry of Justice: Crown Copyright.

<sup>25</sup> Refer to Table 1.4 below, for additional details on which types of diversion require an admission of guilt.

<sup>26</sup> Ely, C., Fadda-Archibald, F., Jolaoso, B., Kilgallon, A., Robin-D'Cruz, C., Waters, R., and Whitehead, S. (2019). Understanding Youth Diversion in London: Evidence and practice briefing. London: Centre for Justice Innovation.

<sup>27</sup> Ibid.

to whether CYPs admit to an offence. As a result, the current state of diversionary activities across London, including differences of the types of diversion, remains unclear.

#### **Research Question 4**

We hypothesise that having Black and Ethnic Minority (BAME) backgrounds may predict CYPs not admitting to an offence.

The Lammy review<sup>28</sup> highlighted that individuals from BAME backgrounds were found to be more likely than White individuals to plead not guilty at court, due to a lack of trust with the CJS. Additional research with practitioners in the youth justice field has also highlighted that receiving poor legal advice, and mistrust of the CJS were seen as contributing to CYPs from BAME backgrounds being less likely to admit to an offence or provide 'no comment' at police interviews.<sup>29</sup>

#### **Research Question 5**

We hypothesise there is a significant negative relationship between any type of diversion and re-offending rates, compared to CYPs who are not diverted out of the youth CJS. We further hypothesise that diversion has a smaller impact on reducing re-offending rates for CYPs from Black or other ethnic minority backgrounds, compared to White CYPs.

The first hypothesis is based on meta-analytic studies of youth diversion evaluations which have found that any type of pre-court diversion is associated with a reduction in re-offending rates compared to CYPs who are not diverted.<sup>30 31</sup> However, one meta-analysis found a non-significant negative relationship between youth diversion, and the prevalence of re-offending (i.e. the re-offending rate) as well as the average severity of offences. This suggests that diversion might not have a strong impact on whether CYPs reoffend.

---

<sup>28</sup> Lammy, D. (2017). *The Lammy Review: An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System*. London.

<sup>29</sup> Ofori, A., Jolaoso, B., Robin-D'Cruz, C., & Whitehead, S. (2021). *Equal diversion? Racial disproportionality in youth diversion*. London: Centre for Justice Innovation.

<sup>30</sup> Wilson, D. B., Brennan, I., & Olaghere, A. (2018). Police-initiated diversion for youth to prevent future delinquent behavior: A systematic review. *Campbell Systematic Reviews*, 14(1), 1-88

<sup>31</sup> Wilson, H. A., & Hoge, R. D. (2013). The effect of youth diversion programs on recidivism: A meta-analytic review. *Criminal justice and behavior*, 40(5), 497-518

Although the same study found a significant negative effect of youth diversion on the incidence of re-offending (the average number of offences), suggesting that diversion reduces the frequency at which CYPs re-offend.<sup>32</sup> Nonetheless, a review of the evidence of pre-court diversion by the YEF found that pre-court diversion activities reduce reoffending by around 13%.<sup>33</sup> Although both the review from YEF and these meta-analyses were primarily based on studies in the US, so it is unclear how applicable they may be to the UK.

Regarding the second hypothesis, one meta-analysis of youth diversion evaluations found the effect sizes for re-offending in studies of majority black participants provided with pre-court diversion outcomes, was not statistically significant. Conversely, in studies with majority white participants who were diverted, there were significant reductions in re-offending.<sup>34</sup> This may indicate racial or ethnic background as a factor which impacts CYPs experience and outcomes with the diversion process.

#### **Research Question 6**

We do not have a preferred hypothesis for this research question.

There is mixed evidence on which types of diversion (if any) are associated with a greater effect on reducing re-offending. One meta-analysis reported that types of diversion which involve providing some kind of support service or intervention, had a greater impact on reducing re-offending, compared to types of diversion which didn't involve additional services.<sup>35</sup> However, another meta-analysis found no significant differences between diversion activities which involved providing support service or restorative justice practices, compared to diversion types which didn't involve any additional support for CYPs.<sup>36</sup> Given the suggested variation in practice within London,<sup>37</sup> it is unclear whether different types of diversion are associated with greater effects on reducing re-offending.

---

<sup>32</sup> Petrosino et al. (2010)

<sup>33</sup> YEF Toolkit: pre-court diversion. Retrieved from: <https://youthendowmentfund.org.uk/toolkit/pre-court-diversion/>

<sup>34</sup> Wilson & Hoge (2013).

<sup>35</sup> Petrosino et al. (2010)

<sup>36</sup> Wilson & Hoge (2013)

<sup>37</sup> Ely et al. (2019)

## 1.4. Key concepts

**Table 1.4 Definitions of key concepts**

Terms	Definition used
Youth diversion	<p>Definitions of youth diversion vary and are not applied consistently, leading to different conceptualisations of the activities and practices which are considered diversionary.<sup>38</sup></p> <p>A common element across definitions for youth diversion refers to providing CYPs with an alternative pathway from the formal or traditional CJS.<sup>3940</sup> However, there appear to be different interpretations on how much involvement with the CJS constitutes 'formal' processing, given that diversion can occur at different points in the CJS (e.g. pre-charge, post-charge, or in court).<sup>41 42</sup></p> <p>For this project, we conceptualise youth diversion as: providing CYPs alternatives to being charged, or prosecuted, with criminal offences. This definition excludes Out-Of-Court-Disposals (OOCs) provided by courts to CYPs. This is because once an individual is charged with and/or prosecuted for an offence, they may be considered to have been substantively involved in the formal CJS (regardless of the outcome of any prosecution).</p> <p>Alternatives to being charged or prosecuted can include both informal, non-statutory and formal, statutory Out-Of-Court-Disposals</p>

---

<sup>38</sup> Youth Justice Board (2023)

<sup>39</sup> <https://yjlc.uk/resources/legal-terms-z/diversion>

<sup>40</sup> Centre for Justice Innovation (2021). Youth justice - Frequently Asked Questions Getting the terminology right. Retrieved from: [https://justiceinnovation.org/sites/default/files/media/document/2021/youth\\_justice\\_faq\\_second\\_edition.pdf](https://justiceinnovation.org/sites/default/files/media/document/2021/youth_justice_faq_second_edition.pdf)

<sup>41</sup> Centre for Justice Innovation (2022). Effective point-of-arrest diversion for children and young people. Suffolk: Clinks.

<sup>42</sup> Gaffney, H., Farrington, D. P., & White, H. (2021). Pre-Court Diversion. Toolkit Technical Report. London: Youth Endowment Fund.

	<p>(OOCs)<sup>43</sup>. Statutory OOCs are more likely to result in an individual having a criminal record, with the offence also being disclosable on future Disclosure and Barring Service (DBS) checks. This may present barriers to employment or engagement in other pro-social organisations or community-based activities, as well as potentially helping to internalise the notion of having a criminal record for CYPs which may reinforce identifying with labels such as ‘criminal’, ‘offender’ or ‘delinquent’.</p> <p>Given this distinction, we have opted to specify two types of diversions for this project, as noted below:</p> <ol style="list-style-type: none"> <li>1) <b>Non-statutory diversion:</b> where CYPs with a linked offence receive an alternative outcome which <i>avoids a criminal record, and escalation into the formal youth justice system</i>. This may involve being referred to specific diversionary schemes or having the YJS delivering support services (including intervention programmes) which may or may not be voluntary. Such diversions include: <ol style="list-style-type: none"> <li>a) <b>Community resolution:</b> a diversionary police outcome that can only be used when children have accepted responsibility for an offence. It is an outcome commonly delivered, but not limited to, using restorative approaches.<sup>44</sup> CYPs are required to accept responsibility for the offence (a lower standard than required for admitting guilt for an offence), in order to be issued with a community resolution.<sup>45</sup> Receipt of a community resolution is recorded on MPS systems, but not recorded on the PNC and does not form part of a formal criminal record.<sup>46</sup></li> </ol> </li> </ol>
--	--

<sup>43</sup> Excluding OOCs offered by courts to CYPs.

<sup>44</sup> Youth Justice Board (2021)

<sup>45</sup> National Police Chief’s Council (2022). Community Resolutions (CR) Guidance 2022. Retrieved from: <https://library.college.police.uk/docs/NPCC/Community-resolution-guidance-2022.pdf>

<sup>46</sup> HM Inspectorate of Probation and HM Inspectorate of Constabulary and Fire & Rescue Services (2018). Out-of-court disposal work in youth offending teams. Retrieved from: <https://www.justiceinspectores.gov.uk/cji/wp-content/uploads/sites/2/2018/03/Out-of-court-disposal-work-in-youth-offending-teams-reportb.pdf>

	<p>b) <b>Triage:</b> a suite of voluntary interventions ranging from light-touch sessions, to structured programmes (e.g. drug and alcohol counselling, mentoring programmes, restorative justice interventions) typically provided by YJS, though can involve third sector organisations. Interventions aim to provide holistic support to children (and their parents or carers), as an alternative to being formally processed through the justice system.<sup>47</sup> Based on conversations with the Met, we understand CYPs are required to admit to an offence, in order to be referred to YJS for triage.</p> <p>c) <b>No Further Action:</b> An outcome used when the police decide not to pursue an offence for various reasons, including determining it is not in the public interest to proceed with prosecution, or where diversionary, educational or intervention activity has taken place or been offered, and it is not in the public interest to take any further action.<sup>48</sup></p> <p>2) <b>Statutory diversion:</b> where CYPs with a linked offence receive an alternative outcome which <i>minimises their involvement with the formal youth justice system</i>. Such outcomes include:</p> <p>a) <b>Youth Caution:</b> A formal caution provided by police as an alternative to charging a CYPs with an offence. The CYP must admit to the offence, and police must have enough evidence to prove an offence was committed. Following a Youth Caution, the CYP is referred to the YJS which may offer an assessment and additional services or intervention programmes. A Youth Caution forms part of a CYPs criminal record and can be disclosed to employers in some circumstances.<sup>49</sup></p>
--	--

---

<sup>47</sup> Centre for Justice Innovation (2021). Youth justice - Frequently Asked Questions Getting the terminology right. Retrieved from: [https://justiceinnovation.org/sites/default/files/media/document/2021/youth\\_justice\\_faq\\_second\\_edition.pdf](https://justiceinnovation.org/sites/default/files/media/document/2021/youth_justice_faq_second_edition.pdf)

<sup>48</sup> Youth Justice Board (2021)

<sup>49</sup> Youth Justice Board (2013). Youth Cautions Guidance for Police and Youth Offending Teams. Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/354050/yjb-youth-cautions-police-YOTs.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/354050/yjb-youth-cautions-police-YOTs.pdf)



	<p>b) <b>Youth Conditional Caution:</b> A formal caution provided by police, with one or more conditions attached. A CYP is referred to YJS for assessment, which recommends conditions attached to the conditional caution. If a CYP does not adhere to these conditions, they could be prosecuted for the original offence. The CYP must admit to the offence and consent to conditions linked with the conditional caution.<sup>50</sup></p> <p>Based on current understanding of data held by the Met and YJS, we believe these two types of diversion can be identified in the data and would be able to help determine whether there are any differences in outcomes for CYPs, based on the kinds of pre-court diversions. However, as researchers have not had sight of data (by either the Met or YJS), the definitions outlined above may be subject to revision, to better align with the data available for analysis.</p>
Formal processing	We define formal processing through the CJS as any further involvement a CYP has with the criminal justice system. This includes being charged and prosecuted with an offence. A CYP is considered to have been formally processed, irrespective of the outcome of the prosecution (e.g. conviction, acquittal, OOC etc.)
Offending	For the purposes of this study, offending refers to an arrest or charge relating to any type of offence, as measured by police recorded crime data.
Violent offending	For the purposes of this study, violent offending is defined as an arrest or charge, relating to an offence which results in, or is intended to, cause someone's death or physical harm (excluding sexual harm). <sup>51</sup>

<sup>50</sup> Ministry of Justice (2013). Code of Practice for Youth Conditional Cautions. Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/243443/9780108512179.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/243443/9780108512179.pdf)

<sup>51</sup> The definition of violent offending is informed by the definition of a violent offence as noted within section 31 of the *Criminal Justice Act 1991*.

	We anticipate this includes offences relating to violence against the person and robbery, based on police recorded crime data.
Re-offending / re-arrest	<p>We anticipate using re-arrest data for CYPs, as a measure of re-offending (i.e. commission of multiple crimes by the same individual). This is due to our understanding of data held by the Met and YJS teams, which does not consistently capture individuals' conviction history.</p> <p>However, we anticipate being able to identify the arrest records of those CYPs within our sample, based on our understanding of the data the Met holds on CYPs.</p> <p>Following gaining sight of crime data held by the Met, and/or YJS teams during our data preparation phase, we may revisit the feasibility of relying on arrest/re-arrest data as a measure of CYPs offending.</p>

## 2. About the datasets

### 2.1. Overview of datasets used

Datasets held by the Met include the following:

- Crime Reporting Information System (CRIS): captures reports on criminal incidents. The crime report is specific to the incident (not the individual). Therefore, multiple individuals may be captured within a single CRIS report.
- National Strategy for Police Information System (NSPIS): captures information about an individual when they are brought into custody, and holds data related to the investigation of a crime for an individual. NSPIS includes information around the arrest, information around the investigation outcome (e.g. triage, community resolution, youth caution, youth conditional caution).
- Case Overview and Preparation Application (COPA): captures details that would be required to prepare the case for court, to be sent to the CPS. COPA contains a detailed description of the police interview. Admission of guilt must be recorded.
- Missing Persons and Related Linked Indices (MERLIN): captures safeguarding reports (e.g. checks and concerns raised regarding potential issues with mental health, police protection, pre-birth, child sexual exploitation, county lines etc.) as well as information related to missing persons.

- Stops: captures details of everyone stopped and searched by the Met. Includes outcomes for individuals who were stopped and searched.

These systems were used across all BCUs in the Met, although since October 2022 these have been migrated to a single new system within the Met (ConnectPlus).

YJS services hold data on CYPs which they work with on different systems, including:

- Career Vision
- Child View
- Core Plus

However, all YJS teams are required to provide quarterly information to the YJB on the CYPs they work with, which we understand to include the following types of data:

- Demographic information: such as name, address, ethnicity, religion, immigration status
- Offence information: PNC number, arrest date (for each offence), offence description
- Intervention information: intervention programme offered to CYPs, type of programme, start and end date of programme offered.

## 2.2. Secondary data source(s)

**Table 2.2a MPS data – Data held across CRIS, NSPIS, COPA, MERLIN & Stops**

Name of dataset	MPS data
Data owner(s)	Metropolitan Police Service
Type of data	Administrative data, cross-sectional and longitudinal data
Availability of data	Privately available to a BIT researcher with MPS vetting, and members of the MPS SIU team partnering on this project.
Team member(s) who will have access	MPS team members: <ul style="list-style-type: none"> <li>● Lewis Kelly</li> <li>● Katie Harper</li> </ul> BIT Team members <ul style="list-style-type: none"> <li>● Tim Hardy</li> <li>● Ed Flahavan</li> </ul>

Population/geographic coverage or sampling frame	CYPs aged 10 - 17 years old when they are first arrested, or recorded as being suspected of committing a crime, across all London boroughs.
Years covered or survey waves	January 2015 – September 2022
Exclusion criteria	Data to be limited to 10-17-year-olds who were first arrested or recorded as being suspected of committing an offence between 2015-2022.
Expected population/sample size (following exclusion criteria)	Unclear at this point in time due to not having sight of the data, however, this will be refined during the data preparation phase of the project.
Documentation	Not publicly available.

**Table 2.2b YJS data – Data held across Core Plus, Career Vision, Childview**

Name of dataset	YJS data
Data owner(s)	YJS teams across London boroughs
Type of data	Administrative data, cross-sectional and longitudinal
Availability of data	Data will be provided securely to the Met. Therefore, this will be privately available to a BIT researcher with MPS vetting, and members of the MPS SIU team partnering on this project.
Team member(s) who will have access	MPS team members: <ul style="list-style-type: none"> <li>● Lewis Kelly</li> <li>● Katie Harper</li> </ul> BIT Team members <ul style="list-style-type: none"> <li>● Tim Hardy</li> <li>● Ed Flahavan</li> </ul>
Population/geographic coverage or sampling frame	CYPs aged 10-17 years old when they are first referred to YJS teams, across all London boroughs.
Years covered or survey waves	January 2015 – September 2022

Exclusion criteria	Data to be limited to 10-17-year-olds who were first arrested or recorded as being suspected of committing an offence between 2015-2022.
Expected population/sample size (following exclusion criteria)	Unclear at this point in time due to not having sight of the data, however, this will be refined during the data preparation phase of the project.
Documentation	Not publicly available.

### 2.3. Primary data collection

No primary data will be collected for this project.

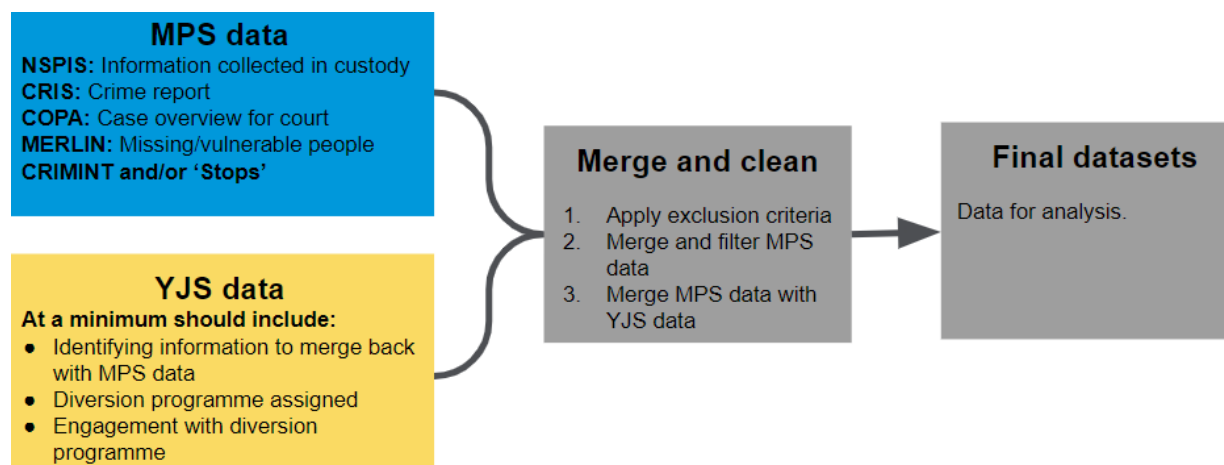
### 2.4. Linking datasets

#### Project dataflow

1. Data will be extracted from across different internal Met systems. Should data not be able to be obtained from YJS, data from the Met will be used to conduct the analysis for this project.
2. Data will be obtained from selected YJS teams to determine the feasibility of merging data from YJS teams with the Met's data. We anticipate being able to approach two YJS offices (Tower Hamlet and Croydon) where we have existing relationships with stakeholders in the YJS, which should help gather YJS data in a timely manner.
3. If initial YJS data can be merged with Met data, then we will seek to obtain a second tranche of data from YJS offices across all London boroughs. We hope to leverage relationships with stakeholders in the London Office of Technology and Innovation (LOTI) and the Information Governance for London working group, to be able to engage with YJS at a pan-London level, rather than having to approach individual YJS teams in each borough. However, we acknowledge that, even if we are able to engage YJS teams at a pan-London level, we may be provided with data from different YJS teams at different timescales.

A summary diagram showing the ideal dataflow for this project is provided in Figure 1 below.

Figure : Summary diagram of dataflow for project



### Joining databases

To collate data from internal Met systems we will be relying on matching data using one of two reference criteria, depending on the completeness of data available.

#### Option 1: PNC ID

Based on our understanding in conversations with the Met and YJS, PNC identification numbers represent one option to match unique individuals across different Met systems. Additionally, we understand YJS teams are also required to record the PNC ID numbers of CYPs they work with. Therefore, we should be able to link YJS data with MET data using PNC ID number.

However, there are limitations to this approach. Based on our understanding, not all individuals in Met systems or YJS systems will have a PNC ID number – especially if CYPs are diverted, or have previously been diverted, prior to being arrested. Additionally, we understand that PNC IDs were not regularly recorded in the Met’s systems (particularly for historical records older than three years) as PNC ID numbers could not be accessed by most Met officers until relatively recently.

Therefore, merging data using PNC ID in both Met and YJS systems may not capture all relevant individuals. Although we have been informed anecdotally by contacts in YJS teams, that almost all CYPs YJS teams interact with, will have PNC ID numbers. However, it is unclear how representative this is for all YJS offices across London.

#### Option 2: Custom identifier

As an alternative approach, we may use a unique identifier to merge datasets. Based on conversations with stakeholders in the Met police, we were informed that the Met collates

and merges data for its own internal analysis using a unique identifier. This identifier is a string of characters which is made up of an individual's forename, surname and date of birth (DD/MM/YYYY). Given that relevant data fields (i.e. name and date of birth) are captured in all Met and YJS systems, this may provide a greater coverage of individuals in merging datasets.

However, this approach does have a (minor) limitation as it risks merging data from separate individuals, if these individuals have the same name and date of birth. Nevertheless, this may present the most feasible option, as it would facilitate merging across all datasets from Met and YJS systems.

Moreover, whilst we have not applied this approach to joining data from Met or YJS systems, being aware that the Met itself uses a similar approach to conduct its own internal analysis, helps establish the feasibility of utilising this approach to merging the data.

### **Thresholds and incomplete matching**

We anticipate that we may be able to determine the success of linking databases (in either collating internal Met data, or through merging Met data with YJS data), by establishing an appropriate threshold for data completeness and integrity. This will be established during the data preparation phase of the project, once a BIT researcher will have sight of Met data and/or YJS data.

To address incomplete matching, we may utilise fuzzy matching to identify and link similar or closely related records, based on the identifier used to merge the data. For example, if using an individual's forename, surname and date of birth (DD/MM/YYYY) as an identifier, we may match data based on similar names, rather than exact names, in case of errors in how the individual's name may have been recorded in different systems.

Depending on the difficulty involved in the merging process, we may also employ data standardisation methods such as normalising formats, removing inconsistencies, and applying standard rules to facilitate accurate merging of data across datasets.

### **2.5. Access and data protection**

This project seeks to partner with the Met, to access data held by the Met's systems, and to leverage existing Data Sharing Agreements (DSAs) between the Met and YJS teams across London.

All data from YJSs will be transferred securely to the Met. Existing data sharing agreements identify Part 3 (law enforcement) of the Data Protection Act 2018 as legal basis for processing

data held by the Met and provided by YJS. Given this research project will use data to analyse the effects of youth diversion in reducing the likelihood of young people re-offending, this project is considered to fall within the remit of existing data sharing agreements.

Data gathered from MPS and YJS systems will include personal data (e.g. names, date of birth addresses) and special category data (e.g. racial or ethnic background) of CYPs. We will also aim to include personal or identifiable data of officers involved in either charging or diverting CYPs (refer to section 4.2 for additional details on how this data will be analysed).

To access these kinds of data held by the Met, a Data Protection Impact Assessment (DPIA) has been drafted collaboratively between the Met and BIT, to outline how data will be appropriately accessed, analysed and stored appropriately during this project. This notes that data extracted from the Met's systems will be collated and stored securely on the MPS AWARE system. This will also store data received from YJS. This data will not be transferred or taken out of the MPS AWARE system.

A BIT analyst will have access to this data to undertake merging and analysis (once MPS vetting has been processed and authorised). Selected individuals from the Met's Strategic Insight Unit (SIU) will also have access to the data.

The Met's DPIA for this project has also identified this is a "one-off" strategic project so specific Code of Practice on the Management of Police Information (MoPI) retention periods do not apply. Once the data has been analysed, the report written and peer reviewed and signed off, the data will only be retained for a specific length of time for follow-up requests by senior management and/or relevant agencies being informed by the project findings. We estimate that the data will be retained for up to two years after it has been extracted and starts to be analysed by BIT researchers.

In addition to gathering and storing data from MPS and YJSs systems, this data will be restructured into a single dataset for analysis. This dataset will be pseudonymised to protect individuals' identity. This will likely include replacing forenames and surnames with unique identifiers.

### **3. About the data**

#### **3.1. List of variables**

#### **Table 3.1: Variable definitions<sup>52</sup>**

---

<sup>52</sup> We note that the variables listed in Table 3.1 are extracted from several different datasets and some variables may involve multiple observations for the same individual (e.g. if someone had multiple arrests).



Variable abbreviation	Variable definition	Variable source	Derivation or specification
Sex	Categorical variable capturing a CYP's sex	MPS (NSPIS, CRIS, COPA, MERLIN, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project.
Ethnicity	Categorical variable capturing a CYP's ethnicity (6 categories used by MPS, 16 categories used by YJS)	MPS (NSPIS, CRIS, COPA, MERLIN, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project.
Age	Age, in years	MPS (NSPIS, COPA, MERLIN, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project. Some MPS systems will only list an individual's date of birth, rather than their age. In these datasets, an individual's age will be calculated based on their date of birth, for each individual cases.
Religion	Categorical variable capturing a CYP's religion	YJS data	We will determine the feasibility of how this variable will be specified during the

			data preparation phase of the project.
Immigration status	Categorical variable capturing a CYP's religion	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project.
Care/leaver status	Categorical variable capturing a CYP's religion	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project.
Residential address	Categorical variable which indicates an individual's residential address	MPS (NSPIS, CRIS, COPA, MERLIN, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project. We anticipate that address data will be extracted at the postcode level, for ease of data extraction (rather than the free-text fields of the full address)
Linked offence(s)	Categorical variable which indicates the specific offence(s) an individual is suspected or charged with having committed	MPS (NSPIS, CRIS, COPA, MERLIN, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Linked category of arrest	Categorical variable which identifies the	MPS (NSPIS, CRIS, COPA, MERLIN,	We will determine the feasibility of how

	type of offence(s), for which an individual is suspected or charged with having committed	Stops) and YJS data	this variable will be specified during the data preparation phase of the project
Date of arrest	Date of each arrest during observation period (2015-2022)	MPS (NSPIS, CRIS, COPA, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Number of arrests before observation period	Number of times an individual has been arrested before the observation period (i.e. prior to 2015)	MPS (NSPIS, CRIS, COPA, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Location of crime	Categorical variable which indicates the address at which a crime was committed	MPS (NSPIS, CRIS, COPA, Stops) data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project. We anticipate that address data will be extracted at the postcode level, for ease of data extraction (rather than the free-text fields of the full address)
CR	Categorical variable identifying the Community Resolution disposal outcome for an individual suspected	MPS (NSPIS, CRIS, Stops) data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project

	of committing an offence.		We anticipate this will rely on Disposal outcome or Elimination from investigation data fields.
Triage	Categorical variable identifying the Triage outcome for an individual suspected of committing an offence.	MPS (NSPIS, CRIS, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
NFA	Categorical variable identifying the No Further Action outcome for an individual suspected of committing an offence.	MPS (NSPIS, CRIS, Stops) data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
YC	Categorical variable identifying the Youth Caution outcome for an individual suspected of committing an offence	MPS (NSPIS, CRIS, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
YCC	Categorical variable identifying the Youth Conditional Caution outcome for an individual suspected of committing an offence	MPS (NSPIS, CRIS, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Arrest	Categorical variable specifying whether an	MPS (NSPIS, CRIS, Stops) and YJS data	We will determine the feasibility of how this variable will be

	individual has been arrested for an offence		specified during the data preparation phase of the project
Charge	Categorical variable specifying whether an individual has been charged with an offence	MPS (NSPIS, CRIS, Stops) and YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Intervention	Categorical variable specifying the type of diversionary intervention programmes accepted by CYPs diverted to YJS	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Start date of intervention	Start date of intervention accepted by CYPs diverted to YJS	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
End date of intervention	End date of intervention accepted by CYPs diverted to YJS	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project
Intervention outcome	Categorical variable specifying whether an individual who was provided with a diversionary intervention	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project

	programmes successfully completed the intervention		
Engagement with intervention	Variable specifying the level of attendance during the intervention offered to CYPs diverted to YJS	YJS data	We will determine the feasibility of how this variable will be specified during the data preparation phase of the project

### 3.2. Measurement of key concepts

**Table 3.2 Measurement of key concepts**

Concept	How the concept will be measured and encoded
Diversion	<p>Based on data held within MPS, we anticipate combining the disposal outcomes for individuals (where these include CR, NFA, Triage, YC and YCC) to group these individuals into a single category of disposal outcome, to derive an overall measure of youth diversion.</p> <p>This is likely to be coded as a categorical variable, however given BIT researchers have not yet had sight of the data, the process and method for measuring this concept will be refined during further phases in the project.</p>
Non-statutory diversion	<p>Based on data held within MPS, we anticipate combining the disposal outcomes for individuals (where these include CR Triage, NFA outcomes) to group these individuals into a single category of disposal outcome, to derive non-statutory diversion.</p> <p>This is likely to be coded as a categorical variable, however given BIT researchers have not yet had sight of the data, the process and method for measuring this concept will be refined during further phases in the project.</p>
Statutory diversion	<p>Based on data held within MPS, we anticipate combining the disposal outcomes for individuals (where these include YC and YCC) to group these individuals into a single category of disposal outcome, to derive non-statutory diversion.</p>

	<p>This is likely to be coded as a categorical variable, however given BIT researchers have not yet had sight of the data, the way of measuring this concept will be refined during further phases in the project.</p>
Offending	<p>We understand that data held in Met systems will be able to consistently identify the arrest and charge history of CYPs within London.</p> <p>Therefore we anticipate measuring offending based on whether a CYP has been arrested or charged, for any offence, based on information help by the Met on individual CYPs. This is likely to be coded as a categorical variable, however given BIT researchers have not yet had sight of the data, the process and method for measuring this concept will be refined during further phases in the project.</p>
Violent offending	<p>We anticipate using data from the Met’s system to identify the offence that a CYP has been arrested for, or charged with, and determine whether this offence is considered a violent offence.<sup>53</sup> We anticipate that arrests or charges will relate to the offence categories of violence against the person, and burglary offences.</p> <p>This is likely to be coded as a categorical variable, however given BIT researchers have not yet had sight of the data, process and method for measuring this concept will be refined during further phases in the project.</p>
Re-offending / re-arrest	<p>We anticipate this will involve deriving this from the offending variable for each individual CYP, to measure the number of arrests during the time period this project will seek to cover. This will also involve linking the date of these arrests, to specify arrests which occurred following previous diversions, or formal processing outcomes.</p> <p>However, given BIT researchers have not yet had sight of the data, the way of measuring this concept will need to be refined during further phases in the project.</p>

---

<sup>53</sup> We anticipate using the Home Office’s Crime Recording Rules (specifically the Counting rules for notifiable offences and notifiable reported incidents), to determine which offences relate to the categories of violence against the person and robberies.

### 3.3. Missing data and attrition

We have a preliminary impression of the likely quality of the data we will receive from the Met and YJS systems based on conversations with individuals in the Met and select YJS offices (Tower Hamlets and Croydon). We anticipate our understanding of data quality issues will evolve during the data preparation phase of the project, once BIT researchers have gained access to the Met and YJS data.

Based on our conversations with stakeholders in the Met, we anticipate some issues with the quality of data for certain data fields across systems in the Met, as noted below:

- Residential address: this is provided as free text in most systems and the level of detail can vary, with some individuals having multiple addresses listed. The Stops database in particular relies solely on the voluntary disclosure of a suspect's home address (which may or may not be accurate).
- Age: in some systems (particularly CRIS) an age range (e.g. 17-25) might be listed, if officers aren't able to identify a person's actual age. We also understand that the mobile app interface for Stops which enables officers to capture data immediately, has a known issue (which the Met is seeking to resolve) whereby age data isn't integrated into the reporting data for the Stops system.
- Admission of guilt: this may be captured within the free text of interview records, but is not recorded as a distinct data field, which would limit our ability to feasibly extract this data to use in our analysis.
- Offending history: information on previous arrests is available (primarily recorded on NSPIS), but this is only for individuals who have been arrested within London. If individuals were arrested for additional offences in other parts of the country, this would not be captured in MPS systems.
- Location of crime: the address recorded can vary, especially if the exact location of a crime is unknown (e.g. if an individual has a phone stolen from them whilst they're on a tube journey). This can sometimes result in the location of a crime being recorded at the victim's home address.
- Referral to YJS: there isn't a specific data field which captures when an individual has been referred to YJS services. Relevant disposal outcomes (e.g. Youth Triage) can be used as a proxy for when individuals have been referred to YJS for support/intervention. However, we have been informed that in practice, officers may also record this outcome (i.e. Triage) when disposing of cases using Community Resolutions, and vice versa.
- Court outcomes: information on the disposal of cases which progress to prosecution and are resolved in court (e.g. convictions), is inconsistently captured in Met systems.



We also anticipate some issues with the quality of data for certain data fields from YJS systems, as noted below:

- Type of intervention/support delivered to youth who have been diverted to YJS (i.e. 'Triage'): not all YJS record specific types of intervention programmes offered to youth who have been diverted to YJS.

Based on our current understanding, we will aim to assess and calculate the number of missing data fields across datasets provided by the Met and/or YJS. This may provide an initial understanding of the level of missing data, across relevant data fields in this project, to help inform our approach to carrying out relevant analyses for this project.

Should missing data/data quality impact the ability of this research project to address research questions in a valid manner, we will highlight this in further updates to this analysis plan, outlining further steps we may be able to take to mitigate against missing data or data quality issues.

Our assumption is that we will add a separate category for variables which have missing data when we conduct the analysis. However, depending on the types of data which are missing, we may instead use strategies such as data imputation<sup>54</sup>, using subsets of data (with complete information) to perform analyses, or conducting sensitivity analyses to determine the appropriateness of conducting analysis involving datasets with missing data.

### 3.4. Other sources of bias

We acknowledge that data held by the MPS and YJSs may be impacted by biases in how data is collected and/or how processes are implemented – and the potential impact of having CYPs from ethnic minority backgrounds being overrepresented. For instance, individuals from ethnic minority groups account for a disproportionate share of individuals involved at different stages of the CJS, including arrest, prosecution, conviction and imprisonment.<sup>55</sup> Based on stop and search figures in England and Wales from between April 2020 and March 2021, individuals from black backgrounds were about 7 times more likely to be stopped and

---

<sup>54</sup> For example, if ethnicity and age were missing for some CYPs, we could impute these variables using chained equations, with other CYP characteristics, diversion type and re-offending outcomes as predictors. Depending on how long the imputations take and what proportion of the sample has missing data, we may choose to perform a single stochastic imputation or perform multiple imputation. The latter entails imputing several datasets, so descriptive statistics / regression estimates would be averaged across datasets using Rubin's rules.

<sup>55</sup> Lammy, D. (2017). *The Lammy Review: An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System*. London.

searched by police, compared to individuals from white backgrounds.<sup>56</sup> Practices of over-policing populations of ethnic minority backgrounds has been suggested as a factor leading to the over-representation of individuals from ethnic minority backgrounds in the CJS.<sup>57</sup> We will not be able to accurately measure within data obtained for this project, whether over-representation of ethnic minorities is a result of over-policing these populations, biases of officers or other factors. This is not the primary purpose of this study, we will consider these factors when undertaking our analysis.

We will seek to minimise the potential impact of any biases in how crime data has been collected, prior to undertaking analysis of the data (especially where this relates to exploring ethnic disproportionality in youth diversion). Whilst we recognise that we may not be able to account for biases in how crime data has been collected, we will strive to ensure that the data quality available for our analysis, is not biased towards different ethnic groups (as this could lead to analysing data which may disproportionately include individuals from certain ethnic backgrounds). To mitigate against this, we will split the data by ethnicity prior to any analysis, to assess data completeness and quality of crime data by ethnicity. This will help ensure that any differences (or lack of differences) are not the result of differences in data quality. If necessary, we will speak to MPS data analysts or even frontline officers to understand any discrepancies we find in the data.

## 4. About the analysis

### 4.1. Overview of analytical approach

Analysis of these datasets will include exploratory data analysis (EDA) and descriptive analysis, to investigate the spatial, demographic and temporal distribution of CYPs arrested. This will also seek to identify the frequency of arrests and diversions of CYPs, including how this is impacted by different factors such as location/BCU of arrest, CYPs' ethnicity, and other trends in youth offending. EDA methods may include visualisations (e.g., Sankey plots) and clustering analyses.

Regression models will be used to assess the predictive value of CYP characteristics (e.g. demographics, location, offence types) related to being diverted. This will also look at

---

<sup>56</sup> Home Office (2022). *Police powers and procedures: Stop and search and arrests, England and Wales, year ending 31 March 2021 second edition*. Retrieved from: <https://www.gov.uk/government/statistics/police-powers-and-procedures-stop-and-search-and-arrests-england-and-wales-year-ending-31-march-2021>

<sup>57</sup> Shankley, W., & Williams, P. (2020). Minority ethnic groups, policing and the criminal justice system in Britain. In Byrne, B., Alexander, C., Khan, O., Nazroo, J., & Shankley, W. (Eds.), *Ethnicity, Race and Inequality in the UK: State of the Nation* (p. 51-71). Bristol: Policy press.

variability across different BCUs, as well as to examine whether diversion affects the probability of re-offending.

Based on assessing the feasibility of undertaking analysis using quasi-experimental design (QED). Such QED analysis may include difference-in-differences and matching approaches, to explore which types of diversion are most effective for preventing re-offending.

**Table 4.1 Overview of analysis methods to address research questions<sup>58</sup>**

<i>Research question</i>	<i>Analyses informing interim report</i>	<i>Analyses informing final report</i>
1. Can police data and YJS data be linked to improve understanding and monitoring of diversion?	N/A  No analysis will be carried out, given the nature of this research question focuses on feasibility of data linkage.	N/A  No analysis will be carried out, given the nature of this research question focuses on feasibility of data linkage.
2. What are the characteristics of the CYPs who are diverted?	Exploratory data analysis of diversion types across individual-level characteristics of CYPs in London, as well as individual-level characteristics of CYPs who have never been diverted.	Summary statistics and descriptive analysis of characteristics of CYPs in London who are diverted, as well as CYPs in London who have not been diverted.  Statistics analysis of differences of individual-level characteristics between CYPs who have been diverted, compared to CYPs who have never been diverted.  This will also include data visualisations and plots of key findings.

<sup>58</sup> Details of analysis methods and approach are provided in Section 4.2

<p>3. Is there variation across London boroughs in the use of diversion and in who is diverted?</p>	<p>Initial descriptive analysis of the number of CYPs diverted across each BCU, and how this varies between 2015-2022.</p>	<p>Descriptive analysis for the number of CYPs who Triaged to YJS teams across London boroughs have been, and how this varies between 2015-2022.</p>
<p>4. What factors predict a CYP not admitting to an offence?</p>	<p>N/A</p> <p>No analysis will be carried out for the interim report. The interim report will identify the feasibility of statistical methods to answer this research question</p>	<p>We anticipate running a logistic regression analysis, where the outcome is CYPs not admitting to an offence.</p> <p>We will include individual-level characteristics of interest as covariates (e.g. sex, self-identified ethnicity, age, number of previous arrests, type of offence, location of crime (BCU), year, season (spring, summer, autumn or winter)</p>
<p>5. What is the relationship between diversion and re-offending?</p> <p>&amp;</p> <p>6. What types of diversion are most effective at reducing re-offending?</p>	<p>Regression analysis to investigate the relationship between being diverted the different types of diversion, and re-offending.</p> <p>The interim report will identify the feasibility of conducting additional QED analysis methods for the final report.</p>	<p>Based on the feasibility of methods identified in the interim report, we anticipate utilising one the following types of analyses:</p> <ul style="list-style-type: none"> <li>● Instrumental variables design (within- and between-BCU comparison)</li> <li>● Matching (within- and between-BCU comparison)</li> <li>● Matching (between-BCU comparison)</li> </ul>

		<ul style="list-style-type: none"> <li>• Difference-in-differences (between-BCU comparison)</li> <li>• Pre-post (within-BCU comparison)</li> </ul>
--	--	--

#### 4.2. Approach to addressing research question(s)

##### Research question 1: approach and methods

<i>Research question</i>	Can police data and YJS data be linked to improve understanding and monitoring of diversion?
<i>Hypothesis, if relevant</i>	N/A
<i>What will you be able to say by the interim report</i>	We anticipate being able to state whether it is feasible or not to merge MPS data and YJS data.
<i>Descriptive analysis, if relevant</i>	N/A
<i>Models, specifications and statistical techniques used, if relevant</i>	N/A
<i>Estimating equation, if relevant</i>	N/A
<i>What does the approach need to succeed (constraints/assumptions)?</i>	N/A
<i>Uncertainty and inference</i>	N/A
<i>Robustness checks</i>	N/A

<i>Subgroup you intend to study</i>	N/A
<i>Changes to the analysis</i>	N/A

### Note on research questions 2-6: approach and methods

Due to current limitations in being able to access and gain a detailed understanding of the type of data available, its quality and completeness (either from the Met or YJS), we have not yet been able to outline a detailed approach and method to answering the project’s research questions (beyond the methods already outlined within this plan).

We have provided below our anticipated approach and methods to answering these research questions. We will develop and note a fulsome analytical approach to answering the project’s research questions, following gaining sight of data held within MPS systems. We anticipate this will provide us with necessary information to develop our approach to the analysis for most of the remaining research questions.

### Research question 2: approach and methods

<i>Research question</i>	What are the characteristics of the CYPs who are diverted?
<i>Hypothesis, if relevant</i>	We hypothesise that male CYPs from White backgrounds are more likely to be diverted
<i>What we will be able to say by the interim report</i>	<p>The interim report will include detailed exploratory analysis of diversion types for CYPs in London, as well as CYPs who have never been diverted.</p> <p>We will be able to say how different individual-level characteristics are associated with different diversion types, as well as different individual-level characteristics for CYPs who have not been diverted. These characteristics may include:</p> <ul style="list-style-type: none"> <li>• Sex</li> </ul>

	<ul style="list-style-type: none"> <li>● Self-identified ethnicity</li> <li>● Age</li> <li>● Number of previous arrests</li> <li>● Reason for arrest</li> </ul>
<p><i>Descriptive analysis, if relevant</i></p>	<p>Since this is a novel dataset, we intend to do an extensive exploratory analysis. This will include:</p> <p>Summary statistics for all diverted, and non-diverted individuals by:</p> <ul style="list-style-type: none"> <li>● Year in the dataset</li> <li>● Sex</li> <li>● Self-identified ethnicity</li> <li>● Age</li> <li>● Number of previous arrests</li> <li>● Reason for arrest</li> </ul> <p>Diversions types for CYPs split by:</p> <ul style="list-style-type: none"> <li>● Year in the dataset</li> <li>● Sex</li> <li>● Self-identified ethnicity</li> <li>● Age</li> <li>● Number of previous arrests</li> <li>● Reason for arrest</li> </ul> <p>We will generate data visualisations to communicate the findings. Where we find interesting trends, we will plot them (e.g. if the main reasons for arrest for CYPs have changed over time).</p> <p>Since our dataset includes the Covid-19 pandemic, this exploratory analysis will also allow us to understand and visualise the impact of the pandemic on diversion.</p> <p>We will also examine whether there are significant differences in each of the characteristics above between CYPs who are diverted and CYPs who have never been diverted using two-sample tests:</p>

	<ul style="list-style-type: none"> <li>• Year in the dataset: t-test</li> <li>• Sex: chi-squared test</li> <li>• Self-identified ethnicity: chi-squared test</li> <li>• Age: t-test</li> <li>• Number of previous arrests: t-test</li> </ul> <p>Reason for arrest: chi-squared test</p>
<i>Models, specifications and statistical techniques used, if relevant</i>	N/A
<i>Estimating equation, if relevant</i>	N/A
<i>What does the approach need to succeed (constraints/assumptions)?</i>	Data on characteristics from MPS, diversion data from YJSs
<i>Uncertainty and inference</i>	N/A
<i>Robustness checks</i>	N/A
<i>Subgroup you intend to study</i>	CYPs who are diverted
<i>Changes to the analysis</i>	If MPS data cannot be merged with YJS data, we will seek to rely solely on data held by MPS.

### Research question 3: approach and methods

<i>Research question</i>	Is there variation across London boroughs in the use of diversion and in who is diverted?
--------------------------	---



<i>Hypothesis, if relevant</i>	We do not have a specific hypothesis for this research question.
<i>What we will be able to say by the interim report</i>	The interim report will include an initial descriptive analysis of the number of CYPs diverted for each Basic Command Unit (BCU), and how this varies over time (in the period covered by the data)
<i>Descriptive analysis, if relevant</i>	<p>We will calculate headline summary statistics (e.g. the mean and standard deviation per year) on the number of CYPs provided with a type of diversion across the Met's 12 BCUs. We will also provide summary statistics of CYPs diverted in each BCU split by:</p> <ul style="list-style-type: none"> <li>● Year in the dataset</li> <li>● Sex</li> <li>● Self-identified ethnicity</li> <li>● Age</li> <li>● Number of previous arrests</li> <li>● Reason for arrest</li> </ul> <p>In addition, we will calculate the mean number of CYPs diverted by BCU, split by year and the type of diversion itself.</p> <p>After obtaining YJS data, we will also provide summary statistics for CYPs who Triaged to YJS across each borough have been (for which we have data available), split by:</p> <ul style="list-style-type: none"> <li>● Year in the dataset</li> <li>● Sex</li> <li>● Self-identified ethnicity</li> <li>● Age</li> <li>● Number of previous arrests</li> <li>● Number of previous convictions</li> <li>● Previous type of diversion</li> <li>● Initial engagement with YJS</li> </ul>

	<ul style="list-style-type: none"> <li>• Whether the CYP successfully completed their previous diversion activity</li> </ul> <p>We will generate data visualisations to communicate these findings (e.g. how the number of CYPs diverted changes over time for all BCUs combined, or all how the mean number per year differs across BCUs).</p>
<i>Models, specifications and statistical techniques used, if relevant</i>	N/A
<i>Estimating equation, if relevant</i>	N/A
<i>What does the approach need to succeed (constraints/assumptions)?</i>	Data on location of crime (BCU) from MPS, diversion data from YJSs
<i>Uncertainty and inference</i>	N/A
<i>Robustness checks</i>	N/A
<i>Subgroup you intend to study</i>	CYPs who are diverted
<i>Changes to the analysis</i>	If MPS data cannot be merged with YJS data, we will seek to rely solely on data held by MPS.

#### Research question 4: approach and methods

<i>Research question</i>	What factors predict a CYP not admitting to an offence?
--------------------------	---

<i>Hypothesis, if relevant</i>	We hypothesise that CYPs having Black and Ethnic Minority (BAME) backgrounds may predict not admitting to an offence.
<i>What we will be able to say by the interim report</i>	<p>We will identify whether it is feasible to answer this research question in the final report from the data available.<sup>59</sup></p> <p>We will also identify based on the data available what statistical method is most appropriate to answer this question in the final report.</p>
<i>Descriptive analysis, if relevant</i>	N/A
<i>Models, specifications and statistical techniques used, if relevant</i>	<p>The methodology will depend on our findings in the interim report, but it is likely that our approach will be regression analysis. Specifically, we plan to run a logistic regression where the outcome is not admitting to an offence and the covariates are all observed individual characteristics of interest (e.g. sex, self-identified ethnicity, age, number of previous arrests, type of offence, location of crime (BCU), year, season (spring, summer, autumn or winter)).</p> <p>In both cases, the sample would be offences committed by CYPs for whom we have data on whether they admitted / did not admit to an offence.</p>
<i>Estimating equation, if relevant</i>	<p>We would estimate a logistic specification such as:</p> $Y_i \sim \text{Bernoulli}(1, p_i) ; \text{logit}(p_i) = \alpha + \beta_X X_i$ <p>where <math>Y_i</math> is an indicator for a CYP not admitting to an offence in a given scenario (index scenario by <math>i</math>), <math>p_i</math> is the</p>

<sup>59</sup> Based on our current understanding there does not appear to be a method to consistently capture whether CYPs admitted to an offence or not, or in a format which is feasibly able to be identified for this analysis. This may impact our ability to answer this research question. However, this may change as we gain greater clarity and sight of the data.

	probability of this, and $X_i$ is a vector of covariates. Standard errors would be clustered by CYP.
<i>What does the approach need to succeed (constraints/assumptions)?</i>	<p>This is observational data analysis, so results are not causal.</p> <p>We require data on the relevant characteristics and outcomes from MPS.</p>
<i>Uncertainty and inference</i>	<p>We will present p-values for each variable in the regression. We will not adjust these p-values for multiple comparisons since this is not causal analysis, but we will highlight in reports that there is a high chance that at least some statistically significant correlations are spurious.</p> <p>We will also convert all estimated effects from the logistic models into effects in percentage points at the overall mean rate of not admitting an offence in the sample.</p>
<i>Robustness checks</i>	We will use a linear probability model as a robustness check.
<i>Subgroup you intend to study</i>	Offences committed by CYPs for whom we have data on whether they admitted / did not admit to an offence
<i>Changes to the analysis</i>	<p>If MPS data cannot be merged with YJS data, we will seek to rely solely on data held by MPS.</p> <p>If we are unable to identify whether individuals did not admit to an offence (i.e. maintained their innocence) within MPS data, we will seek to identify a suitable proxy variable which may indicate whether an individual maintained their innocence, based on the data available.</p> <p>However, if no such proxy variable is identified, we may not be able to undertake an appropriate analysis to answer this research question.</p>

## Research question 5: approach and methods

<i>Research question</i>	What is the relationship between diversion and re-offending?
<i>Hypothesis, if relevant</i>	<p>We hypothesise that there is a significant negative relationship between any type of diversion and re-offending rates, compared to CYPs who are not diverted out of the youth CJS.</p> <p>We also hypothesise that diversion has a smaller impact on reducing re-offending rates for CYPs from Black or other ethnic minority backgrounds, compared to White CYPs.</p>
<i>What we will be able to say by the interim report</i>	<p>In the interim report, we will present an assessment of feasibility for different quasi-experimental methods (e.g. instrumental variables, matching, difference-in-differences, pre-post) that could be used to undertake the analysis.</p> <p>We will also perform regression analysis (which should not be interpreted causally) to investigate the relationship between being diverted and re-offending.</p>
<i>Descriptive analysis, if relevant</i>	N/A
<i>Models, specifications and statistical techniques used, if relevant</i>	As above, we are not sure on the exact quasi-experimental method we will use at this stage, but we expect it to be one of the following <sup>60</sup> :

<sup>60</sup> Our proposal also suggested that we could use a “regression discontinuity design” (RDD), where we would compare re-offending rates either (i) in neighbouring BCUs, some of which have a much higher rate of youth diversion (as a proportion of offenders) than others, or (ii) in a short window of time before/after youth diversion was introduced or dramatically increased in a BCU or set of BCUs. However, based on our current understanding of the data, we do not think that an RDD is feasible here. In both cases, the sample size will have to be very small for the method to be valid. Additionally, in case (i), it is very difficult to construct an accurate measure for the distance from the location of the crime to the border between BCUs. This would be the “running variable”; it is best practice to use some polynomial of this variable in estimation.

	<ul style="list-style-type: none"> <li>● <b>Instrumental variables design (within- and between-BCU comparison):</b> This design uses variation in the propensity to offer diversion to a CYP (instead of charging them for an offence) among police officers to estimate the impact of being diverted on re-offending. Specifically, our understanding is that community resolution and no further action diversion types are offered by the arresting officer<sup>61</sup>. If allocation of CYPs to arresting officers is as if random (e.g. based on which officers are patrolling on the specific day, and the schedule for patrols changes over time), then the arresting officer is a valid instrument for diversion - it affects treatment status without affecting re-offending outcomes. This design is known in the literature as a “judge leniency” design.</li> <li>● <b>Matching (within- and between-BCU comparison):</b> This is a cross-sectional comparison between a group of CYPs who are diverted and a group of CYPs who are not, after re-weighting the latter so its observable characteristics are more similar to the former. Specifically, we think the best approach would compare the re-offending rates of CYPs who are diverted and CYPs who are accused of committing a low level offence but do not admit to it (so can’t be diverted)<sup>62</sup>. We propose to use entropy balancing as our matching method, and our matching variables would include sex, self-identified ethnicity, age, number of previous arrests, type of offence, year and season (spring, summer, autumn or winter).</li> </ul>
--	--

---

<sup>61</sup> We understand that the other types of diversion are likely to be offered by a single officer embedded in the YJS team for that BCU. In that case, the variation in the propensity to offer diversion is across different BCUs and is less likely to be as if random, so this approach would be more likely to produce biased estimates.

<sup>62</sup> If data on whether someone admitted to an offence is not (almost) complete - specifically, if it is not available in at least 90% of cases - our comparison group will just consist of non-diverted CYPs (whether they admitted to an offence, did not admit to it or have missing data) and we may use whether someone admitted to an offence as a matching variable (with a “missing” category).

- **Matching (between-BCU comparison):** We could also perform a local comparison between BCUs, some of which may have a much higher rate of youth diversion (as a proportion of youth offenders) than others.

We would have to dilute estimates by the difference in take-up rates of youth diversion among offenders. We could potentially just use certain pairs of neighbouring BCUs whose diversion rates contrast most heavily in this analysis - this might improve the validity of the analysis but reduce the sample size.

- **Difference-in-differences (between-BCU comparison):** If we find that a certain BCU (or set of BCUs) introduced youth diversion methods later (or initially at a very low frequency) compared to others, we can compare the pre-post differences in re-offending rates for the two groups to estimate the effect of youth diversion (and then dilute this estimate by the difference in pre-post differences in take-up rates of youth diversion among offenders in the “treatment group”). We could also perform this comparison for certain pairs of neighbouring BCUs rather than the full sample.

- **Pre-post (within-BCU comparison):** This would involve comparing re-offending rates in a certain BCU before and after youth diversion was introduced (or dramatically and suddenly increased). Again, we would dilute estimates by the pre-post difference in take-up rates of youth diversion.

In the interim analysis, we will perform regression analysis to investigate the relationship between being diverted and re-offending. This is not a quasi-experimental method and so results should not be interpreted causally. The sample will be all CYPs in the dataset.

*Estimating equation, if relevant*

For all the methods above our outcome would be reoffending within 12 months of the date of the initial offence. Let an individual be indexed by  $i$  and  $Y_i$  be a binary variable for whether  $i$  reoffended within 12 months. Our sample will consist of CYPs' first offences (and whether they re-offended within 12 months of them) only.<sup>63</sup>

Then the estimating equations would take the following forms:

- **Instrumental variables design (within- and between-BCU comparison):** We would use two-stage least squares (2SLS) estimation in this case.

**First stage:**

$$T_i = \delta + \theta Z_i + \gamma X_i + u_i$$

**Second stage:**

$$y_i = \alpha + \beta_T \hat{T}_i + \beta_X X_i + e_i$$

Here:

- $y_i$  is the outcome for individual  $i$
- $T_i$  is a binary variable for being diverted
- $\hat{T}_i$  is the predicted value of  $T_i$  from the first stage, implicitly, that is:

$$\hat{T}_i = \hat{\delta} + \hat{\theta} Z_i + \hat{\gamma} X_i$$

- $\hat{\delta}$ ,  $\hat{\theta}$  and  $\hat{\gamma}$  are the estimated values of  $\delta$ ,  $\theta$  and  $\gamma$  respectively
- $Z_i$  is the relative leniency for the officer responsible for offering diversion to  $i$  (or charging them with an offence) - their mean diversion rate minus the mean diversion rate among all other officers in their BCU
- $X_i$  is a set of covariates

We would use heteroskedasticity-robust standard

---

<sup>63</sup> We may also include a secondary outcome measure of reoffending, within 6 months of the date of the initial offence, to capture more short-term impacts of re-offending.



errors (note that 2SLS requires us to use linear models even though the outcome is binary).

- **Matching (within- and between-BCU comparison):**

$$Y_i \sim \text{Bernoulli}(1, p_i); \text{logit}(p_i) = \alpha + \beta_T T_i + \beta_X X_i$$

Here  $p_i$  is the probability that  $Y_i = 1$ ,  $T_i$  is a binary variable for being diverted and  $X_i$  is a vector of covariates.

- **Matching (between-BCU comparison):**

$$Y_i \sim \text{Bernoulli}(1, p_i); \text{logit}(p_i) = \alpha + \beta_T T_i + \beta_X X_i$$

Here  $p_i$  is the probability that  $Y_i = 1$ ,  $T_i$  is a binary variable for being in the “treatment group” (i.e. BCUs with (the group with higher rates of youth diversion) and  $X_i$  is a vector of covariates.

- **Difference-in-differences (between-BCU comparison):**

$$Y_{it} \sim \text{Bernoulli}(1, p_{it});$$

$$\text{logit}(p_{it}) = \alpha + \beta_T T_i + \beta_P \text{Quarter}_t + \dots$$

$$\beta_{DiD} \text{Treat}_{it} + \beta_X X_{it}$$

Here  $p_{it}$  is the probability that  $Y_{it} = 1$ ,  $T_i$  is a binary variable for being in the “treatment group” (i.e. BCUs with (the group with higher rates of youth diversion),  $\text{Quarter}_t$  is a categorical variable for year and quarter,  $\text{Treat}_{it}$  is a binary variable indicating that  $i$  is in the “treatment group” after youth diversion was introduced / dramatically increased and  $X_{it}$  is a vector of covariates. The coefficient of interest is  $\beta_{DiD}$ . We would cluster standard errors by BCU.

- **Pre-post (within-BCU comparison):**

	$Y_{it} \sim \text{Bernoulli}(1, p_{it}) ;$ $\text{logit}(p_{it}) = \alpha + \beta_T \text{Post}_t + \beta_X X_{it}$ <p>Here <math>p_{it}</math> is the probability that <math>Y_{it} = 1</math>, <math>\text{Post}_t</math> is a binary variable for the offence being at a time after youth diversion was introduced and/or dramatically increased in the BCU and <math>X_{it}</math> is a vector of covariates.</p> <p>To investigate whether diversion has different impacts on reducing re-offending rates for ethnic minority CYPs compared to White CYPs, we will add interaction terms between treatment variable(s) and a categorical variable for a CYP being an ethnic minority / White / having missing ethnicity. In the difference-in-differences case, we will add interactions for both the <math>T_i</math> and <math>\text{Treat}_{it}</math> variables. In the instrumental variables case, we will use officer fixed effects and their interaction with the categorical ethnicity variable to instrument both <math>T_i</math> and the interaction between <math>T_i</math> and ethnicity.</p> <p>For the interim report's regression analysis, we will estimate the same regression as the matching (within- and between-BCU comparisons) analysis:</p> $Y_i \sim \text{Bernoulli}(1, p_i) ; \text{logit}(p_i) = \alpha + \beta_T T_i + \beta_X X_i$ <p>Here <math>p_i</math> is the probability that <math>Y_i = 1</math>, <math>T_i</math> is a binary variable for being diverted and <math>X_i</math> is a vector of covariates. We would cluster standard errors by CYP.</p>
<p><i>What does the approach need to succeed (constraints/assumptions)?</i></p>	<p>The assumptions differ for each method:</p> <ul style="list-style-type: none"> <li> <b>Instrumental variables:</b> The first key assumption is instrument exogeneity - that the arresting officer is uncorrelated with any other unobservable determinants of the outcome. For example the officer responsible for diverting / charging the CYP should only affect their probability of re-offending </li> </ul>

through this choice (this is sometimes called the exclusion restriction), and there are no factors which affect both the arresting officer and probability of re-offending (this is sometimes called the independence assumption). The second key assumption is instrument relevance - there is variation in diversion rates between officers (conditional on BCU and the observable variables used in the first stage).

- **Matching:** The main assumption is conditional independence - that we observe all factors that are correlated with both treatment status and the outcome. For example, if the comparison group has a higher underlying offence rate (in the absence of any diversion) than the treatment group and this is not perfectly accounted for by observable variables (e.g. number of previous arrests), we will overstate the effectiveness of diversion. We think it is likely that this assumption will be violated to some extent, since diversion has eligibility criteria that are correlated with propensity to reoffend.
- **Difference-in-differences:** The main assumption is parallel trends - that the underlying trends in the re-offending rate (when diversion rates are equal) are identical in the treatment and comparison groups, after accounting for observable variables. We cannot test this assumption directly, but will examine whether there are similar trends in pre-intervention outcomes between groups.
- **Pre-post:** The main assumption is that there is no time trend in the reoffending rate within the BCU, after accounting for observable individual-level variables.

All methods also assume there are no spillovers between the treatment and comparison groups.

Out of the approaches above, we believe that the instrumental variables design will provide the most

	<p>convincing results, but only if certain conditions are met. Firstly, there must be sufficient variation in diversion rates between officers in a BCU to avoid the weak instrument problem. Secondly, on top of the data on the relevant characteristics and outcomes from MPS and YJSs required by the other results, the design requires data on which police officer (or officer number) is associated with offering diversion or charging the CYP.</p>
<p><i>Uncertainty and inference</i></p>	<p>For each method, we will designate one specification as our “primary specification” and use this to report headline results. We will note whether our results are consistent across robustness checks.</p> <p>When we have fewer than 30 clusters (which will happen when we are clustering at the BCU level), we will consider using randomisation inference to get p-values, since the standard p-values tend to be too small when the number of clusters is small<sup>64</sup>.</p>
<p><i>Robustness checks</i></p>	<p>Our first robustness check will use re-offending within 6 months as the outcome. This will require excluding all observations where the first offence is committed less than 6 months before the end of the period covered by the data, so less weight will be placed on more recent data points.</p> <p>For all methods other than instrumental variables (whose estimation uses linear models), our second robustness check will replace the logistic model with a poisson model (using reoffending within 12 months as the outcome). We will use heteroskedasticity-robust standard errors in cases where we are not already using cluster-robust standard errors in the main specification.</p> <p>If we use a matching method which involves comparing within- and between-BCUs and there is good data on</p>

<sup>64</sup> MacKinnon et al. (2022). Cluster-Robust Inference: A Guide to Empirical Practice. Retrieved from: <https://arxiv.org/pdf/2205.03285.pdf>.

	whether someone admits to an offence, we will use an alternative comparison group: young people who commit a low-level offence and admit to it, but have a prior criminal record (and so aren't diverted).
<i>Subgroup you intend to study</i>	This depends on the method chosen (see above).
<i>Changes to the analysis</i>	If MPS data cannot be merged with YJS data, we will seek to rely solely on data held by MPS.

### Research question 6: approach and methods

<i>Research question</i>	What types of diversion are most effective at reducing re-offending?
<i>Hypothesis, if relevant</i>	We do not have a specific hypothesis for this research question.
<i>What we will be able to say by the interim report</i>	<p>The same as research question 5:</p> <p>In the interim report, we will present an initial assessment of feasibility for different quasi-experimental methods (e.g. matching, difference-in-differences, regression discontinuity) that could be used to undertake the analysis.</p> <p>We will also perform regression analysis (which should not be interpreted causally) to investigate the relationship between the different types of diversion and re-offending.</p>
<i>Descriptive analysis, if relevant</i>	N/A

<i>Models, specifications and statistical techniques used, if relevant</i>	We will use the same quasi-experimental approach (e.g. instrumental variables, matching, difference-in-differences, pre-post) as in research question 5, except we will compare both (i) non-statutory diversions and (ii) statutory diversions to the comparison group.
<i>Estimating equation, if relevant</i>	<p>This will be the same as in question 5, except:</p> <ul style="list-style-type: none"> <li>• For instrumental variables and matching, we will separately compare non-statutory and statutory diversions to the comparison group<sup>65</sup></li> <li>• For difference-in-differences, pre-post and regression analysis (for the interim report), we will transform the treatment variables from binary to categorical (one category for non-statutory, one for statutory)</li> </ul>
<i>What does the approach need to succeed (constraints/assumptions)?</i>	See research question 5.
<i>Uncertainty and inference</i>	This will be the same as in research question 5, except we will correct for multiple (i.e. two) comparisons using the Benjamini-Hochberg procedure.
<i>Robustness checks</i>	This will be the same as in research question 5, unless we use a matching approach. In this case, we will also run a robustness check whether we compare the two treatment groups (non-statutory diversion and statutory diversion) to each other. We will weigh both groups rather than a single group - i.e. we are estimating the average treatment effect (ATE) of one treatment relative to the other, rather than the average treatment effect on the treated (ATT).

---

<sup>65</sup> This is because we will be re-weighting the comparison group to match the characteristics of each treatment group, so the weighted comparison groups will be different.

	As exploratory analysis, we may also break down the non-statutory and statutory categories further into the different diversion types in each category (e.g. community resolution, triage, no further action, youth caution, youth conditional caution). We will not correct for multiple comparisons in this analysis, but we will not use the estimates as headline findings.
<i>Subgroup you intend to study</i>	This depends on the method chosen (see above).
<i>Changes to the analysis</i>	If MPS data cannot be merged with YJS data, we will seek to rely solely on data held by MPS.

## 5. Project management

### 5.1. Risks and mitigations

**Table 5.1 Risks and mitigations**

Number	Risk	Likelihood (Low/Medium/ High)	Mitigation
1	Fail to gain access to data held by the Met, due to lack of agreement on data sharing arrangements.	Low	We have engaged with members of the SIU at the Met to determine the practicalities of accessing Met data, prior to commencing this project. We have identified with Met colleagues a pathway for data to be extracted and processed by the MPS SIU. A BIT researcher will be vetted by the Met and granted access to analyse the data.
2	Fail to gain access to data held by YJS teams, due to	Medium	We have engaged with the LOTI which is responsible for facilitating data sharing

	lack of agreement data sharing arrangements.		arrangements across local public agencies in London. We have received feedback from both the MPS and LOTI that an existing data sharing agreement is in place to share data between YJS teams and the MPS, and that this existing agreement falls within the scope of this project. Should data sharing arrangements not be agreed with relevant stakeholders, we will seek to complete this project relying solely on data held by the Met.
3	Delays in accessing data held by the Met or YJS teams, due to issues in progressing relevant data sharing documentation (e.g. lack of available staff, lack of prioritisation).	High	We will engage with stakeholders early and consistently throughout the project to ensure buy-into the project, to mitigate against unnecessary delays. We will seek to understand and agree appropriate timeframes for stakeholders and partners to keep the project on schedule, particularly when we are depending on partner organisations to progress and agree data sharing processes and documentation.
4	Issues with the variability in the type and quality of data available, prevents substantive analysis to answer the project's research questions	Medium	We have engaged with stakeholders both at the Met and at YJS teams to understand the kinds of data held by the Met and YJS, and whether it



			<p>may be possible to merge and analyse this data.</p> <p>Noting that we have not had sight of the data, we believe that it should be feasible to merge data, though it is unclear exactly which data fields would be able to be linked at this point in time.</p> <p>We will aim to identify as early as possible the feasibility of being able to link and merge datasets.</p>
5	BIT unable to secure researcher with MPS vetting to start and/or complete analysis due to turnover or illness.	Low	We have submitted additional candidates to obtain MPS vetting, to ensure that data analysis is not dependent on a sole BIT researcher.
6	Staff turnover or illness delays BIT from delivering the work.	Low	BIT's team is extensive (150+ UK team members) and multiple people with skills and experience relevant to this project would be available. We will choose an appropriate replacement for any staff member who is taken ill or leaves during the project.

## 5.2. Timeline

**Table 5.2 Timeline**

Date	Activity	Staff responsible/leading
May 2023	Submit plan to MPS Research Ethics Committee & obtain ethics approval	Neeraj Rahal

May - June 2023	Finalise MPS DPIA and agree approach to data sharing for project between MPS & YJS	Neeraj Rahal
June 2023	BIT draft initial data analysis plan and share with YEF	Neeraj Rahal
June 2023	BIT research staff get MPS vetting	Ed Flahavan
June - August 2023	Collating, merging and cleaning internal MPS data	Tim Hardy / Ed Flahavan
June – July 2023	Agree data sharing agreements between MP and YJSs with the Information Governance for London Working group	Neeraj Rahal
April - August, 2023	Engage YJSs and other stakeholder to access data	Neeraj Rahal
August – October 2023	Interim analysis of MPS data and any available YJS data	Tim Hardy / Ed Flahavan
August - September 2023	Assessing feasibility of QED strategies	Tim Hardy / Ed Flahavan
October - November 2023	Drafting interim report and present interim report to YEF	Neeraj Rahal / Tim Hardy / Ed Flahavan
October – January 2024*	QED analysis	Tim Hardy / Ed Flahavan

January – February 2024*	Write-up of final report and recommendations	Neeraj Rahal / Tim Hardy / Ed Flahavan
December – January 2024**		
March 2024* February 2024**	Deliver final report to YEF	Neeraj Rahal
April 2024* March 2024**	Respond to comments from YEF and YEF appointed external peer review	Neeraj Rahal / Ed Flahavan
May 2024* April 2024**	Incorporate comments and submit final report	Neeraj Rahal

\*If QED analysis determined to be feasible

\*\*Timeframes if QED analysis not proceeded with

## 6. References

Crest (2022). *Making the criminal justice system work better: how to improve out-of-court disposals and diversion scheme*. London. Retrieved from: [https://64e09bbc-abdd-42c6-90a8-58992ce46e59.usrfiles.com/ugd/64e09b\\_7df3289b60fe45aaa442e5320d67cd98.pdf](https://64e09bbc-abdd-42c6-90a8-58992ce46e59.usrfiles.com/ugd/64e09b_7df3289b60fe45aaa442e5320d67cd98.pdf)

Centre for Justice Innovation (2021). *Youth justice - Frequently Asked Questions Getting the terminology right*. Retrieved from: [https://justiceinnovation.org/sites/default/files/media/document/2021/youth\\_justice\\_faq\\_second\\_edition.pdf](https://justiceinnovation.org/sites/default/files/media/document/2021/youth_justice_faq_second_edition.pdf)

Ely, C., Fadda-Archibald, F., Jolaoso, B., Kilgallon, A., Robin-D’Cruz, C., Waters, R., and Whitehead, S. (2019). *Understanding Youth Diversion in London: Evidence and practice briefing*. London: Centre for Justice Innovation

- Gaffney, H., Farrington, D. P., & White, H. (2021). *Pre-Court Diversion. Toolkit Technical Report*. London: Youth Endowment Fund.
- Haines, A, Goldson, B, Haycox, A, Houten, R, McGuire, J, Nathan, T, Perkins, E, Richards, E and Whittington, R (2012) *Evaluation of the Youth Justice Liaison and Diversion (YJLD) Pilot Scheme*. Retrieved from: [https://e-space.mmu.ac.uk/621343/1/dh\\_133007.pdf](https://e-space.mmu.ac.uk/621343/1/dh_133007.pdf)
- HM Inspectorate of Probation (2019). *Annual report: inspection of youth offending services (2018-2019)*. Retrieved from: <https://www.justiceinspectors.gov.uk/hmiprobation/inspections/youthannualreport/>
- HM Inspectorate of Probation and HM Inspectorate of Constabulary and Fire & Rescue Services (2018). *Out-of-court disposal work in youth offending teams*. Retrieved from: <https://www.justiceinspectors.gov.uk/cji/wp-content/uploads/sites/2/2018/03/Out-of-court-disposal-work-in-youth-offending-teams-reportb.pdf>
- Lammy, D. (2017). *The Lammy Review: An Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System*. London.
- MacKinnon, J., Nielsen, M., & Webb, M. (2022). Cluster-Robust Inference: A Guide to Empirical Practice. Retrieved from: <https://arxiv.org/pdf/2205.03285.pdf>.
- Ministry of Justice (2013). *Code of Practice for Youth Conditional Cautions*. Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/243443/9780108512179.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/243443/9780108512179.pdf)
- National Audit Office (2021). *Reducing backlog in the criminal courts*. London: Crown Copyright.
- National Police Chief's Council (2022). *Community Resolutions (CR) Guidance 2022*. Retrieved from: <https://library.college.police.uk/docs/NPCC/Community-resolution-guidance-2022.pdf>
- Ofori, A., Jolaoso, B., Robin-D'Cruz, C., & Whitehead, S. (2021). *Equal diversion? Racial disproportionality in youth diversion*. London: Centre for Justice Innovation.
- Petrosino A, Turpin-Petrosino C, Guckenberg, S. (2010). Formal System Processing of Juveniles: Effects on Delinquency. *Campbell Systematic Reviews*, 6: 1-88.
- Uhrig, N. (2016). *Black, Asian and Minority Ethnic disproportionality in the Criminal Justice System in England and Wales*. Ministry of Justice: Crown Copyright.

Wilson, D. B., Brennan, I., & Olaghere, A. (2018). Police-initiated diversion for youth to prevent future delinquent behavior: A systematic review. *Campbell Systematic Reviews*, 14(1), 1-88

Wilson, H. A., & Hoge, R. D. (2013). The effect of youth diversion programs on recidivism: A meta-analytic review. *Criminal justice and behavior*, 40(5), 497-518

Youth Justice Board (2021), *Definitions for prevention and diversion*. Retrieved from: [https://yjresourcehub.uk/images/YJB/Definitions\\_for Prevention and Diversion YJB 2021.pdf](https://yjresourcehub.uk/images/YJB/Definitions_for_Prevention_and_Diversion_YJB_2021.pdf)

Youth Justice Board (2023). *Prevention and Diversion Project – final report*. Retrieved from: [https://yjresourcehub.uk/images/YJB/Prevention and Diversion Project Final Report YJB Feb 2023.pdf](https://yjresourcehub.uk/images/YJB/Prevention_and_Diversion_Project_Final_Report_YJB_Feb_2023.pdf)

Youth Justice Board (2013). *Youth Cautions Guidance for Police and Youth Offending Teams*. Retrieved from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/354050/yjb-youth-cautions-police-YOTs.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/354050/yjb-youth-cautions-police-YOTs.pdf)