

Which mental health difficulties and contextual risk factors are associated with children and young people's involvement in crime and violence?



Secondary Data Project Analysis Plan

Evaluating institution: University of Cambridge

Principal investigator(s): Robbie Duschinsky and Guy Skinner

Analysis Plan for YEF Secondary Data Analysis Projects

Project summary

Project title	Which mental health difficulties and contextual risk factors are associated with children and young people's involvement in crime and violence?
Research Team	University of Cambridge; National Children's Bureau
Principal investigator	Professor Robbie Duschinsky and Dr Guy Skinner
Analysis plan author(s)	Dr Guy Skinner, Professor Robbie Duschinsky, Dr Barry Coughlan, Dr Jamiee Mallion, Dr Matt Woolgar, Professor Darrick Jolliffe, Julia Mannes, Nicole Marshall
Overarching research question	Which mental health difficulties and contextual risk factors are the best predictors of children and young people's involvement in crime and violence?
Supporting research question(s)	For children and young people engaged in with mental health services, which combination of mental health diagnoses, sociodemographic characteristics, and risk factors best predict criminal and violent behaviour in CYP?
Dataset(s) to be used	Computer Record Interactive Search (CRIS) South London and Maudsley NHS Foundation Trust (SLaM); Hospital Episode Statistics (HES)
Population characteristics	4- to 17-year-olds accessing CAMHS services
Years data spans	2007 to 2023
Geographic coverage	England
Primary outcome(s) investigated	Violent behaviour, violent incidents, criminal justice involvement
Main method(s) to be used or tested	Logistic regression and survival analyses

Analysis plan history

Version	Date	Reason for revision
1	04/12/2025	
1.1	13/03/2026	Round 1 peer review comments.
1.0 [original]		<i>[leave blank for the original version]</i>

Any changes to the design or methods need to be discussed with the YEF. Describe in the table above any agreed changes made to the design.

Table of contents

1. About the project.....	4
1.1. Background to the project	4
1.2. Research question(s)	6
1.3. Hypotheses	9
1.4. Key concepts	12
2. About the datasets	14
2.1. Overview of datasets used.....	14
2.2 Secondary data source(s)	14
2.3. Linking datasets.....	16
2.4. Access and data protection	17
3. About the data	18
3.1. List of variables.....	18
3.2. Measurement of key concepts	35
3.3. Missing data and attrition	38
3.4. Other sources of bias.....	39
4. About the analysis	40
4.1. Overview of analytical approach	40
4.2. Approach to addressing research question(s)	43
Research questions 1 and 2: approach and methods.....	43
Research question 3: approach and methods.....	48
5. Project management	52
5.1. Risks and mitigations	52
5.2. Timeline	54
6. References.....	55

1. About the project

1.1. Background to the project

Youth involvement in crime and violence is a pressing social issue with far-reaching consequences for individuals, families, and communities. Understanding the predictors of such behaviours is crucial for developing targeted prevention and intervention strategies. One important area of focus is the role of mental health, as emerging evidence shows that psychological difficulties often intersect with—and may compound—risks for justice-system contact among young people.

In this study we use the term mental health difficulties as an umbrella term to describe a range of emotional, behavioural, and psychological conditions experienced by children and young people. These include both common mental health conditions (e.g., depression, anxiety) and behavioural disorders (e.g., conduct disorder), which are frequently treated within Child and Adolescent Mental Health Services (CAMHS). We also include neurodevelopmental conditions, such as attention-deficit/hyperactivity disorder (ADHD), which are commonly managed within CAMHS and are strongly associated with behavioural, social, and developmental outcomes relevant to this research. While ADHD is conceptually distinct from mental health conditions and is typically classified as a neurodevelopmental disorder, it is included here due to its well-established links with behavioural difficulties, service involvement, and criminal justice outcomes.

We have seen dramatic increases in rates of long-standing mental health difficulties among children and young people (CYP) in the UK. According to recent research conducted in 2023, 1 in 5 CYP have a probable mental health difficulty. Furthermore, 12.9 times more was spent on adult mental health than CYP in England in 2019/20, despite 59% of adult mental health difficulties commencing by adolescence.

It is important to note that most individuals with mental health difficulties are not dangerous. Still, mental health is sometimes a relevant and co-occurring factor among young people who become involved with the justice system. Young people who become involved with the criminal justice system tend to have particularly high levels of mental health needs when compared to CYP in the general population, and early intervention has been recognised as key to improving outcomes. Existing literature highlights a strong link between mental health and neurodevelopmental issues such as conduct disorder, ADHD, depression, substance abuse, and criminal justice involvement in CYP. Contextual risk factors, including exposure to violence, parental criminality, and peer influences, further exacerbate their vulnerability to being criminalised. However, the interplay of these factors and their predictive strength require further exploration through comprehensive data analysis.

In interpreting these findings, we adopt a perspective consistent with the Child First framework used by the Youth Justice Board in England and Wales, which emphasises understanding children who come into contact with the justice system primarily as children with needs and vulnerabilities rather than as offenders. Many of the contextual factors examined in this study—including victimisation, family adversity, and deprivation—are

outside of children's direct control and may increase vulnerability to both mental health difficulties and criminal justice involvement. Understanding these pathways may therefore support earlier identification of need and more effective prevention strategies.

This study aims to conduct a retrospective cohort analysis conducted within a CAMHS-engaged population.

This research project has four objectives:

- For CYP engaged with CAMHS, to examine the relationship between mental health difficulties and youth involvement in crime and violence.
- To explore the impact of sociodemographic characteristics, child-level educational or developmental needs (e.g. SEND) and a wide range of risk factors identified in CAMHS on the likelihood of CYP becoming involved in crime.
- To determine which combination of mental health, sociodemographic characteristics, child-level educational or developmental needs and risk factors make CYP more vulnerable to being criminalised.
- To explore time-dependent covariates, providing a clearer picture of the temporal sequence of events between mental health and violent behaviour and CJ involvement.

These findings can inform:

- Early intervention and prevention in mental health and youth offending services: Understanding which mental health difficulties and risk factors are most strongly associated with criminal behaviour can help develop targeted interventions to prevent youth from entering the CJ system.
- Resource allocation: Identifying key associations can help policymakers and service providers allocate resources more efficiently, focusing on high-risk individuals and communities.
- Policy development: Findings can guide legislative measures to support mental health services, school interventions, and family-based programs that address underlying risk factors.

In preparation for this proposal, we consulted with the National Children's Bureau (NCB) Youth Justice Research Team and held focus groups with experts-by-experience with knowledge of CAMHS and the justice system. We have also discussed our proposal with a Youth Offending Team (YOT) in Cambridgeshire. This highlighted that there is a considerable appetite from experts-by-experience, policymakers, and professionals to investigate such questions. Yet, this has not been possible to date due to a lack of appropriate data linkages and issues with statistical power due to small sample sizes.

This proposal addresses previous barriers, leveraging extensive administrative data from the Clinical Record Interactive Search (CRIS) at South London & Maudsley (SLaM) NHS Foundation Trust and the linkage with Hospital Episode Statistics (HES). These datasets permit the examination of whether and which mental health difficulties and risk factors are most strongly associated with violent behaviour, criminal justice involvement and involvement in violent incidents.

In preparation for this application, our pilot work with CRIS SLaM indicates that the dataset includes over 53,000 CYP with at least one risk assessment conducted by CAMHS. The large sample size will permit the examination of moderator effects of sociodemographic characteristics (ethnicity, gender, age, IMD, educational challenges) and risk factors identified by CAMHS (e.g., safeguarding needs, domestic violence, parent/carer mental health difficulties, parent/carer substance misuse, emotional abuse, physical abuse, sexual abuse, neglect, self-neglect, physical disability, running away, self-harm, risk-taking behaviour, lack of concern by parents/caregivers, lack of concern from other agencies).

We have also partnered with the NCB. The NCB's Youth Justice Research Team will provide guidance in an advisory capacity. The Youth Justice Research Team will provide written feedback and attend a group meeting at the beginning and end of each WP. The interpretation of the results from this research will be further informed by input at the end of each WP from experts-by-experience provided by NCB, who have experience of CAMHS and interaction with the criminal justice system. This will take the form of group meetings led by GS and an experienced facilitator from NCB.

1.2. Research question(s)

- 1) Which mental health difficulties are associated with criminal justice (CJ) involvement and violent behaviour?
 - a. What types of mental health diagnoses (e.g., conduct disorder, depression, anxiety) are most strongly associated with subsequent involvement in crime and violence among CYP (before age 18)?
 - b. How does the co-occurrence of multiple mental health difficulties influence the likelihood of CJ involvement and violent behaviour?
 - c. To what extent do sociodemographic characteristics (e.g., ethnicity, gender, age, indices of multiple deprivation (IMD)), and child-level educational or developmental needs moderate the relationship between mental health difficulties and CJ involvement?
- 2) Which risk factors in the CAMHS risk assessment are associated with CJ involvement and violence?
 - a. Which risk factors from the CAMHS risk assessment are most strongly associated with future CJ involvement and violent behaviour?
 - b. How do combinations of risk factors in CAMHS assessments interact to impact the association of CYP's involvement in crime and violence?
 - c. Are there critical thresholds or clusters of risk factors that significantly increase the likelihood of CJ involvement or violence?
- 3) What combination of mental health diagnosis, sociodemographic characteristics, child-level educational or developmental needs and risk factors captured in CAMHS are most strongly associated with CYP's violent behaviour and CJ involvement?
- 4) What role do the duration and timing of mental health difficulties play in the association between criminal justice involvement, violent incidents and violent behaviour among children and young people?

Table 1.2. How will the questions be addressed at each stage?

Question Number	Interim report	Final report
1, 2, 3	Building on a previously funded YEF Secondary Data Analysis project, we will provide a descriptive account of the prevalence of violent behaviour, criminal justice involvement, and violent incidents using the most recent sweep of data collection. Violent behaviour will be captured through the SLAM Risk Assessment, which documents violent and aggressive behaviour; the CAMHS Item Sheet, which documents antisocial behaviour in terms of fighting and violent assault; and the CAMHS Brief Risk Assessment, which documents destructive and dangerous behaviour towards others. CJ involvement will be captured through the forensic history taken in the CAMHS Assessment, by juvenile conviction records in the CAMHS Information and whether the CYP received a criminal justice intervention in their Risk Management Plan. Violent incidents will be captured from HES Accident and Emergency data. This descriptive account will include information on the cohort of CYP in terms of their mental health (ICD-10), ethnicity, gender, age, and IMD. We will also provide a descriptive account of the following risk factors: domestic violence, parent/carers mental health difficulties, parent/carers substance misuse, emotional abuse, physical abuse, sexual abuse, neglect, destructive behaviour, dangerous behaviour, self-neglect, physical disability, running away, self-harm,	Minor edits to descriptive analysis and logistic regression models following interim report feedback.

	risk-taking behaviour, lack of concern by parents/caregivers, lack of concern from other agencies, and educational challenges (Special Education Needs and Disabilities (SEND)). (IMD), and service use characteristics. To examine the influence of co-occurring mental health difficulties and cumulative risk exposure, additional models will incorporate variables capturing the number of diagnoses and cumulative risk factors. Interaction terms will be used to assess whether associations differ by key sociodemographic characteristics (e.g., gender, ethnicity, IMD) and child-level educational or developmental needs (SEND), and stratified analyses will be conducted where appropriate to explore moderation effects.	
4	Survival analyses will be piloted to ensure feasibility.	Survival analysis techniques will be employed to investigate the timing and risk of first-time involvement in crime and violence among CYPs with mental health difficulties. Cox proportional hazard models will be utilised to estimate hazard ratios, providing insights into how specific mental health diagnoses, sociodemographic characteristics, and CAMHS risk factors influence the time to criminal justice involvement. This approach will help identify critical periods for intervention and prevention, allowing policymakers to design timely support strategies. The analysis will account for censored data, ensuring accurate estimates even when follow-up periods vary across individuals. Kaplan-Meier survival curves will also be used to visually represent differences in time-to-event distributions across different

		subgroups, such as ethnicity, gender, and IMD.
--	--	--

1.3. Hypotheses

Research Question 1: Which mental health difficulties are associated with criminal justice (CJ) involvement and violent behaviour?

Hypothesis 1 (H1): Externalising disorders—specifically conduct disorder, Oppositional Defiant Disorder (ODD), and substance use disorders—will be the strongest mental-health associations with CJ involvement and violent behaviour. CYP with multiple co-occurring diagnoses will be substantially more vulnerable, and sociodemographic factors will moderate these associations.

Direction/Magnitude: Externalising disorders are expected to be more strongly associated with CJ involvement and violent behaviour than internalising disorders such as anxiety and depression. Children with conduct-related difficulties are anticipated to have a markedly higher likelihood of involvement in violent behaviour or CJ pathways. Comorbidity is expected to significantly amplify risk, reflecting the compounding effect of overlapping behavioural and emotional difficulties. Sociodemographic characteristics—including ethnicity, gender, age, deprivation level (IMD), and educational challenges—are anticipated to modify these relationships, with some subgroups experiencing disproportionately elevated risk.

Supporting Evidence: Conduct disorder is strongly associated with chronic and serious offending.^{1,2} ADHD, particularly impulsive and hyperactive presentations, is associated with increased aggression and justice involvement.³ Substance misuse elevates the likelihood of violence through disinhibition and association with risky contexts.⁴ Comorbidity is associated with more severe behavioural dysregulation and poorer outcomes.^{5,6}

Potential Mechanisms: Impulsivity, emotional dysregulation, poor behavioural control, and peer deviance increase susceptibility to conflict and risk-taking. Family stress, weak parental supervision, and school exclusion create environments that reinforce antisocial behaviour. Comorbidity reflects heightened complexity and instability in home and school settings, increasing exposure to harmful contexts. Sociodemographic moderators may shape access to support, exposure to adversity, systemic bias, and responses from schools and agencies.

Specific and Testable Nature: Mental health diagnosis variables will be included in logistic regression models assessing associations with CJ involvement and violent behaviour before age 18. Comorbidity will be examined through diagnosis count variables and interaction terms. Moderation effects will be tested using interaction terms for sociodemographic variables (e.g., diagnosis × gender; diagnosis × ethnicity) and through stratified models. Model fit and effect sizes will guide the interpretation of which diagnostic profiles and sociodemographic contexts are most strongly associated with CJ involvement and violence.

Research Question 2: Which risk factors in the CAMHS risk assessment are associated with CJ involvement and violence?

Hypothesis 2 (H2): Risk factors relating to family adversity (domestic violence, parental substance misuse, parental mental illness), abuse or neglect, risk-taking behaviours, and self-harm will be the strongest antecedents of CJ involvement and violent behaviour. Multiple risk factors are expected to have cumulative or synergistic effects.

Direction/Magnitude: Children exposed to domestic violence, maltreatment, or parental dysfunction are expected to show a substantially heightened likelihood of both violent behaviour and CJ involvement. Abuse and neglect are anticipated to have a particularly strong influence, likely exceeding that of other single risk factors. CYP experiencing multiple risk factors are expected to show a rapidly increasing risk, with each additional adversity compounding the overall likelihood of harmful outcomes. Risk is expected to escalate non-linearly, such that beyond a certain threshold (e.g., several co-occurring risk factors), vulnerability increases sharply. Distinct clusters of risk (e.g., trauma-related, behavioural, or neighbourhood) may have differential but consistently elevated associations with violent or offending outcomes.

Supporting Evidence: ACEs significantly increase the likelihood of violence, offending, and CJ contact.⁷ Parental substance misuse and criminality predict behavioural difficulties and antisocial behaviour across generations,⁸ While experiences of abuse and neglect are linked to later aggression and justice involvement.⁹ Evidence further suggests that exposure to multiple adversities produces disproportionately large increases in behavioural problems, indicating risks often compound rather than operate independently.¹⁰ Theories of cumulative disadvantage suggest that early adversity can initiate trajectories in which disadvantages compound over time, leading to increasingly constrained developmental pathways.¹¹ Relatedly, stress proliferation theory proposes that exposure to one stressor increases the likelihood of subsequent stressors, producing cascades of disadvantage that intensify vulnerability.¹² These frameworks suggest that risk factors such as abuse, parental mental health difficulties, poverty, and school exclusion may not simply have additive effects but may interact in ways that amplify the likelihood of violence and criminal justice involvement. Consistent with these perspectives, clusters of risk factors have been shown to predict early-onset and persistent offending trajectories.¹³

Potential Mechanisms: Adversities can disrupt emotion regulation, producing heightened vigilance, aggression, or dissociation. Family instability reduces parental monitoring and increases exposure to criminogenic environments. Accumulated adversities overwhelm coping resources and increase engagement in risk-taking behaviours. Risk clusters reflect broader ecological patterns such as chronic deprivation, community violence, or intergenerational trauma.

Specific and Testable Nature: Logistic and survival models will identify the strongest individual CAMHS risk factors associated with CJ involvement and violence. Risk-factor counts will test for cumulative effects across adversities. Pairwise and higher-order interaction terms will test for synergistic influences.

Research Question 3: What combination of mental health diagnosis, sociodemographic characteristics and risk factors captured in CAMHS are most strongly associated with CYP's violent behaviour and CJ involvement?

Hypothesis 3 (H3): A combined model incorporating mental health diagnoses, sociodemographic factors (gender, ethnicity, age, IMD), child-level educational or developmental need (SEND) and CAMHS risk factors will provide the strongest association with violent behaviour and CJ involvement. CAMHS risk factors are expected to have the strongest overall association, enhanced further when combined with clinical and demographic information.

Direction/Magnitude: The combined model is expected to show substantially improved explanatory accuracy compared with any single-domain model. CAMHS risk factors are likely to contribute the largest share of explanatory power, with mental health diagnoses and sociodemographic characteristics providing important but more moderate contributions. Sociodemographic variables may strengthen or weaken the impact of mental health difficulties and risk factors (e.g., stronger effects in more deprived communities). Interaction effects between domains (e.g., diagnosis × IMD; risk factor × ethnicity) are expected to provide additional discriminatory value, particularly for identifying high-risk subgroups. Overall, risk will be most elevated among CYP who have combinations of clinical vulnerabilities, contextual adversities, and sociodemographic disadvantages.

Supporting Evidence: Multidomain models outperform single-domain models in predicting youth violent behaviour.¹⁴ Ecological models highlight the interaction between individual conditions and contextual challenges.¹⁵ Sociodemographic factors shape exposure to risk and access to services.^{16,17} Combined risk approaches may provide superior classification accuracy in offending research.

Potential Mechanisms: Mental health difficulties increase vulnerability; risk factors capture proximal adversities; sociodemographic structure exposure to stress and service pathways. Interactions across ecological levels (individual, family, community) may contribute to the emergence of violence and CJ involvement. Structural inequalities amplify or dampen the influence of clinical and contextual risks.

Specific and Testable Nature: Compare performance of mental-health-only, risk-only, sociodemographic-only, and combined models. Expect the combined model to demonstrate the best overall predictive performance.

Research Question 4: What role do the duration and timing of mental health difficulties play in the association between criminal justice involvement, violent incidents, and violent behaviour among children and young people?

Hypothesis 4 (H4): Earlier onset and longer duration of mental health difficulties will be associated with a higher risk of CJ involvement and violent behaviour. CYP who develop externalising difficulties at a younger age, or who experience persistent/chronic symptoms over time, will be at heightened risk of earlier and more severe CJ involvement.

Direction/Magnitude: Early-onset behavioural problems are expected to be associated with earlier and more entrenched pathways into antisocial behaviour. Longer duration of difficulties, particularly persistent externalising symptoms, is anticipated to be associated with accelerated time to first CJ contact or violent incident. CYP with chronic mental health difficulties are expected to show significantly higher hazard ratios compared to those with later-onset or shorter-duration problems.

Supporting Evidence: Early-onset externalising difficulties are associated with more persistent patterns of antisocial behaviour across development. Persistent ADHD symptoms and related neurodevelopmental differences are linked to impulsivity and behavioural regulation difficulties that can increase conflict within institutional settings such as schools.¹⁸ However, emerging research highlights that many of the risks experienced by these children arise not solely from individual behavioural traits but from institutional responses to neurodevelopmental differences. Children with ADHD and related conditions are disproportionately represented among those excluded from mainstream education and placed in alternative provision.¹⁹ These exclusionary processes may reflect limitations in mainstream educational systems' ability to support neurodivergent learners and can contribute to pathways into social marginalisation and criminalisation, sometimes described as elements of a "school-to-prison pipeline". Earlier onset of emotional or behavioural difficulties may therefore be associated with cumulative exposure to educational disruption, adverse peer environments, and other structural disadvantages, which together increase vulnerability to violence and criminal justice involvement over time.

Potential Mechanisms: Longer duration of behavioural dysregulation may reinforce maladaptive coping, conflictual relationships, and exclusion from prosocial settings (e.g., school). Early onset disrupts normative developmental pathways, reduces access to stabilising environments or stabilising environments are not provided to accommodate the needs of these children, and increases exposure to antisocial peers or high-risk contexts – potentially because they have been excluded from mainstream school environments. Chronic difficulties may also strain family functioning, increase service involvement, or reduce engagement with protective supports.

Specific and Testable Nature: Survival analysis methods (Cox models, Kaplan–Meier estimates, and interval-censored approaches where necessary) will be used to assess the influence of timing (age at onset) and duration (time from onset to event or end of follow-up) on CJ involvement and violent behaviour. Onset and duration variables will be constructed from dates of first recorded diagnosis or first presentation. Time-dependent covariates will be used where feasible. Sensitivity analyses will test the robustness of findings given uncertainty in event dates (e.g., referral vs discharge documentation). Logistic regression models will supplement survival analyses as timing-insensitive checks.

1.4. Key concepts

Table 1.4 Definitions of key concepts

Terms	Definition used
Child and Adolescent Mental Health Services (CAMHS)	CAMHS is the name for the NHS services that assess and treat young people with emotional, behavioural, or mental health difficulties.
Violent incidents	Violent incidents are any form of assault recorded during an admission to hospital or Accident and Emergency (A&E). This will include assaults by bodily force, sharp object, blunt object, firearm, chemical or other means. In this study, this variable is intended to capture involvement in violent contexts that result in hospital contact, rather than victimisation per se. To reduce the likelihood that the measure reflects victim status alone, children and young people with explicit indicators of victimisation in other data sources were excluded from this outcome definition. The resulting measure therefore represents instances where CYP were recorded as being involved in violence-related events requiring hospital treatment, without assuming that they were necessarily victims or perpetrators.
Violent behaviour	Violent behaviour will be conceptualised through recorded violent and aggressive behaviour, antisocial behaviour in terms of fighting and violent assault, and destructive and dangerous behaviour towards others.
Criminal justice involvement	Criminal justice involvement will be defined as CYP having a conviction record and whether they have received a criminal justice intervention in their Risk Management Plan.
Mental health	We will be using clinician reported variables from the ICD-10 and self-, parent-, and teacher-reported variables in the SDQ to conceptualise mental health. This encompasses a broad spectrum of mental health difficulties.
CAMHS Risk factors	This will include measures on the following: domestic violence, parent/carer mental health difficulties, parent/carer substance misuse, emotional abuse, physical abuse, sexual abuse, neglect, self-neglect, physical disability, running away, self-harm, risk taking behaviour, lack of concern by parents/caregivers, lack of concern from other agencies, lack of concern by CYP.
Educational Characteristics	Educational characteristics are defined as CYP having SEND or SALT. We will also include school attendance and fixed-term exclusions. Special Educational Needs and Disabilities (SEND) will be captured as a binary indicator recorded by clinicians within the CAMHS dataset, reflecting whether a child or young person has identified special educational needs or disabilities. This variable does not capture the full set of SEND categories used within the educational system (e.g.

	the 13 primary need categories defined in the SEND Code of Practice), but instead represents a clinician-recorded indicator of whether a CYP has recognised educational or developmental support needs. In our analysis SEND will represent any CYP who are noted by clinicians to have had any of the following conditions: Autism, Dyslexia, Dyspraxia, ADHD, Cerebral Palsy, learning difficulties, profound and multiple learning difficulties, sensory impairment, Social, Emotional, Mental Health (SEMH), Hearing or visual impairment, Down's syndrome, and Epilepsy. SALT stands for speech and language therapy. In our analysis SALT will represent any CYP who have speech and/or language difficulties and those who have received speech and language therapy.
Socioeconomic status	Socioeconomic status (SES) is a way to describe a person's or family's economic and social standing in relation to others. In this analysis we use Indices of Multiple Deprivation (IMD) to define groups of SES.
Safeguarding needs	In this analysis, safeguarding needs of a CYP will be conceptualised by the involvement of social care services, whether the CYP is on the child protection register, and whether the CYP is a looked-after child.

2. About the datasets

2.1. Overview of datasets used

Data will be extracted for CYP attending CAMHS in South London and the Maudsley NHS Foundation Trust (SLaM) between 2007-2023. SLaM is one of the largest mental health providers in Europe, with a catchment area of over 1.3 million people and covers four London boroughs: Croydon, Lambeth, Southwark, and Lewisham²⁰. Data will be extracted using the Clinical Record Interactive Search (CRIS).

Data will also be extracted from the Hospital Episode Statistics (HES). HES is a curated data product containing details about admissions, outpatient appointments and historical accident and emergency attendances at NHS hospitals in England. We will match and extract all data on violent incidents for CYP with at least one risk assessment in CRIS SLaM CAMHS between 2007-2023.

2.2 Secondary data source(s)

Table 2.2a Dataset Description - Clinical Record Interactive Search (CRIS) at South London & Maudsley (SLaM) NHS Foundation Trust

Name of dataset	Clinical Record Interactive Search (CRIS) at South London & Maudsley (SLaM) NHS Foundation Trust
Data owner(s)	Maudsley Biomedical Research Centre (BRC) Clinical Records.
Type of data	Administrative dataset.

Availability of data	CRIS SLaM has been approved for secondary data analysis by the University of Oxford (Ref: 08/H0606/71+5). The CRIS oversight committee have approved this project (Ref: 24-062)
Team member(s) who will have access	Robbie Duschinsky, Guy Skinner, Barry Coughlan, Matt Woolgar.
Population/geographic coverage or sampling frame	CYP referred to South London and Maudsley NHS Foundation Trust.
Years covered or survey waves	2007 to 2023.
Exclusion criteria	All 10–17-year-olds referred to CAMHS between 2007 to 2023 with at least one risk assessment.
Expected population/sample size (following exclusion criteria)	53,344 CYP with at least one risk assessment conducted by CAMHS.
Documentation	Information on the CRIS Dataset - https://www.maudsleybrc.nihr.ac.uk/facilities/clinical-record-interactive-search-cris/

Table 2.2b Dataset Description - Hospital Episode Statistics (HES)

Name of dataset	Hospital Episode Statistics (HES)
Data owner(s)	NHS England.
Type of data	Administrative dataset linked to CRIS SLaM.
Availability of data	CRIS SLaM has been approved for secondary data analysis by the University of Oxford (Ref: 08/H0606/71+5). The CRIS oversight committee have approved this project (Ref: 24-062)
Team member(s) who will have access	Robbie Duschinsky, Guy Skinner, Barry Coughlan, Matt Woolgar
Population/geographic coverage or sampling frame	CYP referred to South London and Maudsley NHS Foundation Trust CAMHS with at least one risk assessment.
Years covered or survey waves	2007 to 2023.
Exclusion criteria	All 10–17-year-olds who can be matched with those referred to CAMHS between 2007 to 2023 in CRIS SLaM with at least one risk assessment.
Expected population/sample size (following exclusion criteria)	Expected matching with 48,009 CYP with at least one risk assessment conducted by CAMHS in the CRIS SLaM

	dataset. This is based on a 90% matching rate in previous work.
Documentation	Information on the HES dataset - https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics HES Data Dictionary - https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary

2.2. Primary data collection

No primary data will be collected.

2.3. Linking datasets

The Maudsley Biomedical Research Centre have set up a Clinical Data Linkage Service (CDLS), which enables data to be linked from CRIS with other clinical courses through a secure trusted research environment (TRE). TRE's have been set up in the NHS to ensure that confidential patient information can be linked in a way that guarantees the legal and ethical rights of patients.

The CDLS service has linked CRIS with HES data. This linkage has a data controller and a data processor. A data controller is the individual or organisation or organisation who controls and is responsible for how data is kept and used. South London and Maudsley (SLaM) NHS Foundation Trust is the data controller for CRIS. In linking CRIS to HES, SLaM is also the data processor. NHS England (formally NHS Digital) are the data controller for the HES. HES contain details of all hospital admissions, outpatient appointments and A&E attendances at NHS hospitals in England. Further details of the data held within HES can be found on the NHS England [website](#).

In order to link CRIS data with HES data, CRIS identifiers are compiled by CDLS and transferred to the Health and Social Care Information Centre (HSCIC) using an NHS-approved secure file transfer protocol. HSCIC then adds the CRIS ID to all HES records that match CRIS records and extracts all other HES records for patients within the four catchment boroughs served by SLaM (the control group). HSCIC destroys patient identifiers leaving only the CRIS ID and HES extract ID. As with other linked data sets, the CRIS-HES data are transferred back to CDLS to be held and provided to researchers in a fully anonymised format. Please see Figure 1 for a flow diagram of this process.

This linkage is legally permissible under the Health Research Authority's support under Section 251 of the NHS Act 2006. This support enables temporary use of identifiers in a secure environment for accurate linkage.

All applications to use CRIS-HES data must be approved by the CRIS Oversight Committee.

In September 2021, the CDLS was audited by NHS Digital to ensure that the terms and conditions set for this linkage were being abided by and that data were being kept safe and secure.

Based on the audit, NHS Digital identified several areas of good practice and concluded that the overall risk of a breach of the terms and conditions with respect to compliance, duty of care, confidentiality or integrity was low.

The CRIS-HES data linkage has been used successfully by many researchers with a high degree of matching and data quality.²¹

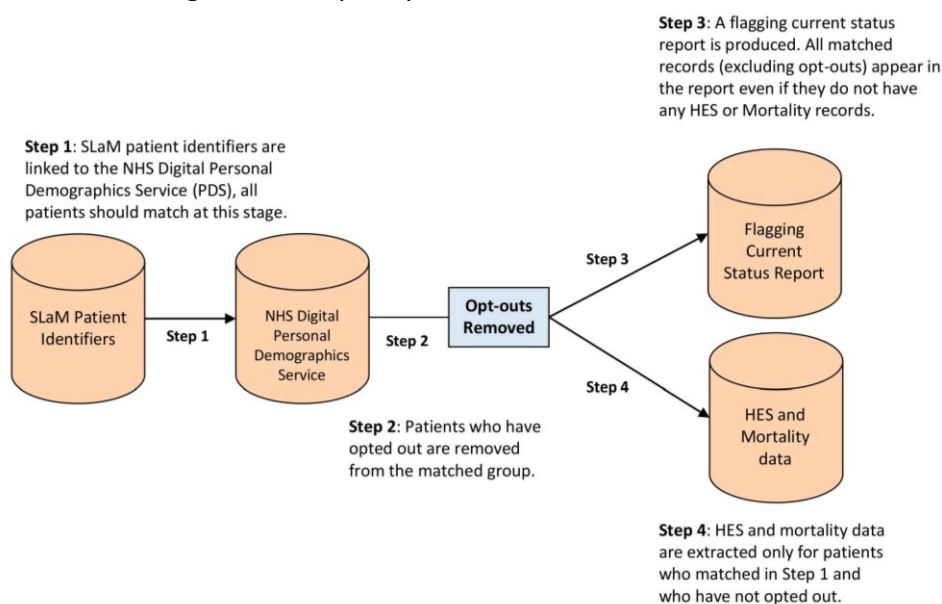


Figure 1. Flow diagram of the linkage process between SLaM records and HES data. Cited from Jewell et al. (2020).²²

2.4. Access and data protection

In developing this project, we have collaborated with Cambridge University's Research Governance Office. All research in this proposal has been developed and will be governed by the guidelines set out in the UK Policy Framework for Health and Social Care Research.

Access to CRIS SLaM and HES is available without cost to YEF through Dr Matt Woolgar. Confidence in the feasibility of this proposal is aided by this pre-existing access to data from CRIS SLaM, and the feasibility work already conducted. We have scoped variables, investigated data quality, and prepared a dataset which is cleaned and ready for the proposed analysis.

We recognize that concerns around the usability and quality of CAMHS data, such as variability in data completeness and consistency, are well-documented. The study by Coughlan et al. (2024)²³ highlights the risk rates and profiles of young people at intake to CAMHS in South London, identifying common patterns in data quality and coverage within CRIS SLaM. Drawing on their methodological approach, we have taken steps to ensure our dataset is robust. This includes detailed scoping of key variables—such as referral dates,

intake outcomes, and clinical measures—along with an evaluation of their completeness and reliability. Moreover, their findings on latent class profiles and risk categorization have informed our variable selection and approach to handling missing data. By leveraging these insights and conducting rigorous feasibility checks, we have addressed the common challenges associated with CAMHS data and are confident in the quality and usability of the dataset for our proposed analysis.

CRIS SLaM has been approved for secondary data analysis by the University of Oxford (Ref: 08/H0606/71+5). This project has been approved by the CRIS oversight committee (Ref: 24-062). Two experienced research assistants have been recruited to ensure the linkage between HES and CRIS is established within WP1. Both research assistants have two years' experience working with the CRIS dataset. The CRIS oversight committee has also approved the linkage between CRIS SLaM and HES data (Ref: 24-062).

CRIS gleans anonymised data from electronic health records. To analyse anonymised CRIS data, authorised researchers submit a project proposal to the CRIS oversight committee at King's College, London giving full details of their study. This project proposal is reviewed by a panel including researchers and experts-by-experience. Searches are audited regularly, and every 3-6 months researchers submit a report outlining their activities. The research team will use a Virtual Private Network (VPN) to access and work with the data unless on site in SLaM. Following the CRIS security model, all data for this proposed project will be analysed and stored within the SLaM firewall. Within the firewall, data will be stored on a T:Drive for researchers authorised to use CRIS and all files will be password protected. In line with the CRIS security model, to avoid threats to patient anonymity, cell sizes of n=9 or fewer will not be presented. In line with current regulations, linked data will be accessed from a secure Office for National Statistics research location.

All researchers accessing data in CRIS have received honorary contracts from SLaM and undergone research governance training.

3. About the data

3.1. List of variables

Table 3.1: Variable definitions

Variable abbreviation	Variable definition	Variable source	Derivation or specification
X85	Assault by drugs, medications and biological substances Incl.: homicidal	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.

	poisoning by (any): biological substance, drug, medicament		
X86	Assault corrosive substance Excl.: corrosive gas	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X87	Assault by pesticides Incl.: wood preservatives Exclude.: plant food and fertilizers	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X88	Assault by gases and vapours	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X89	Assault by other specified chemicals and noxious substances Incl.: homicidal poisoning	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X90	Assault by unspecified chemical	HES – Admitted Patient	Possible values: Yes/No. Captured closest to referral

	or noxious substance Incl.: homicidal poisoning NOS	Care and Accident and Emergency.	and closest to discharge.
X91	Assault by hanging, strangulation and suffocation	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X92	Assault by drowning and submersion	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X93	Assault by handgun discharge	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X94	Assault by rifle, shotgun and larger firearm discharge	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X95	Assault by other and unspecified firearm discharge	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.

X96	Assault by explosive material Excl.: incendiary device	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X97	Assault by smoke, fire and flames Incl.: arson, cigarettes, incendiary device	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X98	Assault by steam, hot vapours and hot objects	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
X99	Assault by sharp object	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Y00	Assault by blunt object	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Y01	Assault by pushing from high place	HES – Admitted Patient Care and Accident and	Possible values: Yes/No. Captured closest to referral and closest to discharge.

		Emergency.	
Y02	Assault by pushing or placing victim before moving object	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Y03	Assault by crashing of motor vehicle	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Y04	Assault by bodily force Incl.: unarmed brawl or fight Excl.: assault by: strangulation, submersion, use of weapon; assault by bodily force	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Y05	Sexual assault by bodily force Incl.: rape (attempted), sodomy (attempted)	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Y08	Assault by other	HES – Admitted	Possible values: Yes/No. Captured

	specified means	d Patient Care and Accident and Emergency.	closest to referral and closest to discharge.
Y09	Assault by unspecified means Incl.: assassination (attempt), homicide (attempt), manslaughter (nonaccidental), murder (attempted)	HES – Admitted Patient Care and Accident and Emergency.	Possible values: Yes/No. Captured closest to referral and closest to discharge.
Social_care_involvement	Does the CYP have social care involvement	CRIS SLaM – CAMHS Full Assessment	Binary = 1 where any of 'Present', 'Past, Present', 'Past' Captured closest to referral
Child_protection_register	Is the CYP on the child protection register	CRIS SLaM – CAMHS Full Assessment	binary = 1 where [Child_Protection_Register_ID] = 'yes' during index referral
Looked_after_child	Is the CYP a Looked After Child	CRIS SLaM – CAMHS Full Assessment	binary = 1 where [Looked_After_Child_ID] = 'yes' during index referral
Special_Education_Needs_ID	Does the CYP have special education needs?	CRIS SLaM – CAMHS Information	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium"

			in spell/ episode of care THEN = TRUE Captured closest to referral
Exclusion_From_School_ID	Has the CYP been excluded from school?	CRIS SLaM – CAMHS Information	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
Parent_Carer_Mental_Health_ID	Does the CYP's parent/carer have mental health difficulties?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
Parent_Carer_Substance_Misuse_ID	Does the CYP's parent/carer have substance misuse difficulties?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
From_Others_Emotional_Abuse_ID	Is the CYP at risk of emotional abuse from others?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
From_Others_Physical_Abuse_ID	Is the CYP at risk of physical abuse from others?	CRIS SLaM – CAMHS Brief Risk	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium"

		Assessment	in spell/ episode of care THEN = TRUE Captured closest to referral
From_Others_Sexual_Abuse_ID	Is the CYP at risk of sexual abuse from others?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
From_Others_Neglect_Lack_Supervision_ID	Is the CYP at risk of neglect/lack of supervision?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
Situational_Domestic_Violence_ID	Is the CYP at risk of witnessing domestic violence?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE
To_Others_Violence_Towards_Others_ID	Has the CYP been violent towards others?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral and closest to discharge
To_Others_Destructive_Behaviour_ID	Has the CYP displayed destructive behaviour	CRIS SLaM – CAMHS Brief Risk	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium"

	to wards others?	Assessm ent	in spell/ episode of care THEN = TRUE Captured closest to referral and closest to discharge
To_Others_Offending_ID	Has the CYP committed a criminal offence involving another person?	CRIS SLaM – CAMHS Brief Risk Assessm ent	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral and closest to discharge
To_Others_Dangerous_Behaviour_ID	Has the CYP displayed dangerous behaviour towards others?	CRIS SLaM – CAMHS Brief Risk Assessm ent	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral and closest to discharge
To_Others_Physical_Abuse_ID	Has the CYP been physically abusive towards others?	CRIS SLaM – CAMHS Brief Risk Assessm ent	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral and closest to discharge
To_Others_Sexual_Abuse_ID	Has the CYP been sexually abusive towards others?	CRIS SLaM – CAMHS Brief Risk Assessm ent	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE

			Captured closest to referral and closest to discharge
To_Self_Physical_Disability_ID	Does the CYP have a physical disability?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
To_Self_Learning_Disability_ID	Does the CYP have a learning disability?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
To_Others_Exclusion_From_School_ID	Has the CYP been excluded from school?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
To_Self_Not_Attending_School_ID	Has the CYP not attended school?	CRIS SLaM – CAMHS Brief Risk Assessment	Possible values: High, Low, Medium, None, Not Known. IF "High", "Medium" in spell/ episode of care THEN = TRUE Captured closest to referral
History_Of_Violence_ID	Does the CYP have a history of violence?	CRIS SLaM – Full Risk Assessment	Possible values: Don't Know, No, Yes IF "Yes" in episode of care/ spell THEN = "TRUE"

			Captured closest to referral and closest to discharge
Antisocial_Behaviour_ID	Has the CYP displayed antisocial behaviour?	CRIS SLaM – Full Risk Assessment	Possible values: Don't Know, No, Yes IF "Yes" in episode of care/ spell THEN = "TRUE" Captured closest to referral and closest to discharge
Incidents_Of_Violence_ID	Has the CYP been involved in incidents of violence?	CRIS SLaM – Full Risk Assessment	Possible values: Don't Know, No, Yes IF "Yes" in episode of care/ spell THEN = "TRUE" Captured closest to referral and closest to discharge
Communicating_Needs_ID	Does the CYP have difficulty communicating their needs?	CRIS SLaM – Full Risk Assessment	Possible values: Don't Know, No, Yes IF "Yes" in episode of care/ spell THEN = "TRUE" Captured closest to referral
Learning_Disability_ID	Does the CYP have a learning disability?	CRIS SLaM – Full Risk Assessment	Possible values: Don't Know, No, Yes IF "Yes" in episode of care/ spell THEN = "TRUE" Captured closest to referral
Violence_Community_ID	Has the CYP committed violence in the community?	CRIS SLaM – Full Risk Assessment	Possible values: Don't Know, No, Yes IF "Yes" in episode of care/ spell THEN = "TRUE" Captured closest to referral and closest to discharge

Planned_Intervention_ID_YP_Criminal_Justice_Interventions	Does the CYP have a planned criminal justice intervention?	CRIS SLaM – Risk Management Plan	Binary =1 where YP Criminal Justice Interventions "yes" during index/episode of care Captured closest to referral and closest to discharge
First SDQ Emotional_Symptoms_Score	Emotional symptoms score	CRIS SLaM – SDQ Parent report, self-report, teacher report	Captured closest to referral
First SDQ Conduct_Problems_Score	Conduct problems score	CRIS SLaM – SDQ Parent report, self-report, teacher report	Captured closest to referral
First SDQ Hyperactivity_Score	Hyperactivity score	CRIS SLaM – SDQ Parent report, self-report, teacher report	Captured closest to referral
First SDQ Peer_Problems_Score	Peer problems score	CRIS SLaM – SDQ Parent report, self-report, teacher report	Captured closest to referral
First SDQ Prosocial_Score	Prosocial score	CRIS SLaM –	Captured closest to referral

		SDQ Parent report, self- report, teacher report	
First SDQ Total_Difficulties_Score	Total difficulties score	CRIS SLaM – SDQ Parent report, self- report, teacher report	Captured closest to referral
First SDQ Impact_Score	Impact score	CRIS SLaM – SDQ Parent report, self- report, teacher report	Captured closest to referral
T1 ICD10 Diagnosis_Date	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 closest to referral
T1 ICD10 Primary Diag (Axis 1a)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 closest to referral
T1 ICD10 Secondary Diag 1 (Axis 1b)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 closest to referral
T1 ICD10 Secondary Diag 2 (Axis 2)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 closest to referral
T2 ICD10 Primary Diagnosis_Date	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to first Event Completion_Point_I D "During Therapy"

T2 ICD10 Primary Diag (Axis 1a)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to first Event Completion_Point_I D "During Therapy"
T2 ICD10 Secondary Diag 1 (Axis 1b)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to first Event Completion_Point_I D "During Therapy"
T2 ICD10 Secondary Diag 2 (Axis 2)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to first Event Completion_Point_I D "During Therapy"
T3 ICD10 Diagnosis_Date	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to discharge
T3 ICD10 Primary Diag (Axis 1a)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to discharge
T3 ICD10 Secondary Diag 1 (Axis 1b)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to discharge
T3 ICD10 Secondary Diag 2 (Axis 2)	ICD10 diagnosis	CRIS SLaM – Diagnosi s	ICD10 diagnosis closest to discharge
Counts inpatient stay (episode)	Number of inpatient stay	CRIS SLaM – Inpatien t Episode	Counts of inpatient episodes (in episode of care)
Length of inpatient stay	Length of inpatient stay	CRIS SLaM – Inpatien t Episode	Average of inpatient stay in weeks (within episode of care); "Admission_Date" to "Discharge_Date"
Event_Type_ID Direct contact episode	Number of contacts with CAMHS	CRIS SLaM – Inpatien t episode	Counts of "Direct contact" Event_Type_ID (within episode of care)
Event_Type_ID Missed appointment (DNA) episode	Number of missed	CRIS SLaM –	Counts of " Missed appointment (DNA)" Event_Type_ID

	appointments	Inpatient episode	(within episode of care)
Antisocial_Behaviour_Regular_Defiance_ID	Does the CYP display antisocial behaviour in the form of regular defiance?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral and closest to discharge
Antisocial_Behaviour_Stealing_ID	Does the CYP display antisocial behaviour in the form of stealing?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral and closest to discharge
Antisocial_Behaviour_Damage_ID	Does the CYP display antisocial behaviour in the form of causing damage to property?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral and closest to discharge
Antisocial_Behaviour_Truancy_ID	Does the CYP display antisocial behaviour in the form of truancy?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral and closest to discharge

Antisocial_Behaviour_Fighting_ID	Does the CYP display antisocial behaviour in the form of fighting?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral and closest to discharge
Antisocial_Behaviour_Violent_Assault_ID	Does the CYP display antisocial behaviour in the form of violent assaults?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral and closest to discharge
Speech_and_Language_Poor_or_Unusual_Communication_ID	Does the CYP have poor speech and language or unusual communication?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral
Speech_And_Language_Disorder_Of_Articulation_ID	Does the CYP have a speech and language disorder in the form of difficulty articulating themselves?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral
Speech_And_Language_Delay_In_Language_ID	Does the CYP have a speech and language	CRIS SLaM – CAMHS	Possible values: Dubious, Not Present, Present, Unknown

	disorder in the form of a delay in language?	Item Sheet	IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral
Speech_And_Language_Selective_Mutism_ID	Does the CYP have a speech and language disorder in the form of mutism?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral
Speech_And_Language_Abnormal_Language_ID	Does the CYP have a speech and language disorder in the form of abnormal language?	CRIS SLaM – CAMHS Item Sheet	Possible values: Dubious, Not Present, Present, Unknown IF "Present" in episode of care/spell THEN = TRUE Completed closest to referral
School_Attendance_Score_ID	Extent to which school attendance is a problem	CRIS SLaM – Health of the Nation Outcome Scales for Children and Adolescents (HoNOS CA)	Possible values: 0 = no problem 1= minor problem requiring no action 2= mild problem but definitely present 3= moderately severe problem 4= severe to very severe problem IF "2", "3", "4" THEN= PROBLEM
Counts inpatient stay (episode)	Number of inpatient stays	Inpatient episode	Counts of inpatient episodes (in episode of care)
Length of inpatient stay	Average length of	Inpatient episode	Average of inpatient stay in weeks (within episode of care);

	inpatient stay		"Admission_Date" to "Discharge_Date"
Event_Type_ID Direct contact episode	Number of contacts with CAMHS	Inpatient episode	Counts of "Direct contact" Event_Type_ID (within episode of care)
Event_Type_ID Missed appointment (DNA) episode	Number of missed appointments	Inpatient episode	Counts of " Missed appointment (DNA)" Event_Type_ID (within episode of care)

3.2. Measurement of key concepts

Table 3.2 Measurement of key concepts

Concept	How the concept will be measured and encoded
Violent Incidents	Violent incidents will be defined in accordance with the International Classification of Diseases 10th Edition (ICD-10) diagnosis codes recorded during an admission to categorise admissions as being due to violence. This will include all types of assaults, including bodily force, sharp object, blunt object, firearm, chemical, or other means. Violence was also defined by previously validated codes, which would trigger consideration of violence by carers, peers, or strangers.²⁴ This concept will use HES Admitted Patient Care and Accident and Emergency ICE-10 codes X85-X99, Y00-Y05, Y08-09. Violent incidents will be encoded as: Yes/No. If a CYP has any record within X85-X99, Y00-Y05, Y08-09 they will be encoded as 'Yes' to having a violent incident. Violent incidents are likely to be rare events. However, we hope that our large sample size will provide sufficiently powered analyses. As noted above, we initially plan to investigate different types of violent incident reported within the HES data. If events are too rare within each type of violent incident, we will look to combine specific types of violent incidents into a single variable, maximising statistical power of this outcome. In this study, this variable is intended to capture involvement in violent contexts that result in hospital contact, rather than victimisation per se. To reduce the likelihood that the measure reflects victim status alone, children and young people with explicit indicators of victimisation in other data sources were excluded from this outcome definition. The resulting measure therefore represents instances where CYP were recorded as

	being involved in violence-related events requiring hospital treatment, without assuming that they were necessarily victims or perpetrators.
Violent behaviour	We will be using clinician reported variables to form a violent behaviour outcome. Violent behaviour will be conceptualised through the SLAM Risk Assessment, which documents violent and aggressive behaviour, the CAMHS Item Sheet, which documents antisocial behaviour in terms of fighting and violent assault, and through the CAMHS Brief Risk Assessment, which documents destructive and dangerous behaviour towards others. Violent behaviour will be encoded as: Yes/No. If a CYP has any record indicating violent and aggressive behaviour, antisocial behaviour, and/or destructive and dangerous behaviour towards others, they will be encoded as 'Yes' to having displayed violent behaviour. Depending on the prevalence of variables reported, this concept may be further split into more detailed types of violent behaviour (e.g. violent assault, antisocial behaviour). We will report in the interim report our final decision on this concept. Those involved in violent incidents (victim and potentially perpetrator) could be different from those involved in violent behaviour (capturing more on the perpetrator role). We are ensuring that all possible data has been extracted through these discussions, searching the respective data dictionaries and using a Natural Language Processing App in CRIS to identify further incidents of violence and potential distinctions between perpetration and being a victim of violence.
Criminal justice involvement	Criminal justice involvement will be conceptualised through the forensic history taken in the CAMHS Assessment, juvenile conviction records in the CAMHS Information, and whether the CYP received a criminal justice intervention in their Risk Management Plan. Criminal justice involvement will be encoded as: Yes/No. If a CYP has any record indicating they have a juvenile conviction and/or having received a criminal justice intervention, they will be encoded as 'Yes' to having criminal justice involvement.

	Depending on the prevalence of variables reported, this concept may be further split into more detailed types of criminal justice behaviour (e.g. juvenile conviction; criminal justice intervention).
Mental health	We will be using clinician reported variables from the ICD-10 and self-, parent- and teacher-reported variables in the SDQ to conceptualise mental health. Analyses will be conceptualised as diagnoses of mental health conditions and a broader assessment of self-/other-reported mental health. Mental health will be encoded as having a referrals for one of the following (CYP may have referrals for multiple mental health difficulties): (Suspected) First episode psychosis; Ongoing or recurrent psychosis; Bipolar disorder; Depression; Anxiety; Obsessive-compulsive disorder; Phobias; Organic brain disorder; Drug and alcohol difficulties; Post-traumatic stress disorder; Eating disorders; Perinatal mental health issues; Personality disorders; Self-harm behaviours; Conduct disorders; Neurodevelopmental conditions; In crisis; Relationship difficulties; Gender discomfort issues; Attachment difficulties; Self-care issues; Adjustment to health issues; Gambling disorder; Community perinatal mental health; Partner assessment; Behaviours that challenge due to a learning disability. This concept may be organised into broader categories to allow better matching (e.g. ICD-10 Axis 1 and Axis 2). We will report in the interim report our final decision on this concept. Mental health will also be encoded by parent-, self- and teacher-reported SDQ total difficulties scores. Scores range from 0-40, with higher scores indicating more abnormal behaviours. This concept may be organised into broader categories to allow better matching (e.g. internalising score, externalising score). We will report in the interim report our final decision on this concept.
Risk factors	We will be using clinician reported variables to form a cumulative risk variables. This will include measures on the following: domestic violence, parent/carer mental health difficulties, parent/carer substance misuse, emotional abuse, physical abuse, sexual abuse, neglect, self-neglect, physical disability, running away, self-harm, risk taking behaviour, lack of concern by parents/caregivers, lack of concern from other agencies, lack of concern by CYP. Each risk factor will be scored dichotomously (Present; Not Present).

	Cumulative risk factors will be encoded as an ordinal variable, where a higher score indicates a greater level of cumulative risk. The maximum score for this concept will be 15.
Educational Characteristics	Educational characteristics are defined as CYP having SEND or SALT. We will also include school attendance and fixed-term exclusions. SEND will be encoded dichotomously (Present; Not Present). SALT will be encoded dichotomously (Present; Not Present). Fixed-term exclusions will be encoded dichotomously (Present; Not Present). School attendance will be encoded dichotomously (Problem; Not a problem).
Safeguarding needs	In this analysis, safeguarding needs of a CYP will be conceptualised by the involvement of social care services, whether the CYP is on the child protection register, and whether the CYP is a looked-after child. Involvement of social care services, whether the CYP is on the child protection register, and whether the CYP is a looked-after child will each be scored dichotomously (Present; Not Present). The concept of safeguarding needs will be scored dichotomously (Yes; No). If a CYP has any of these three variables reported as present, they will be encoded as having a safeguarding need. Depending on the prevalence of variables reported, this concept may be further split into more detailed types of safeguarding need (e.g. involvement of social care services, whether the CYP is on the child protection register, and whether the CYP is a looked-after child) or scored on an ordinal scale ranging from 0 (no safeguarding need) – 3 (involvement of social care services, on the child protection register, and a looked-after child all present). We will report in the interim report our final decision on this concept.

3.3. Missing data and attrition

In preparation for this application, our pilot work with CRIS SLAM indicates that the dataset includes 53,344 CYP with at least one risk assessment conducted by CAMHS. The reporting of variables including ethnicity, geographic location, gender, age, socioeconomic status, cumulative risk factors, and educational characteristics are well reported and have been explored in other research.²⁵

We are currently interrogating patterns in the missing data to assess whether certain groups of CYP are more or less likely to have incomplete violent behaviour and criminal justice involvement variables. We have reviewed the data dictionaries for both CRIS SLAM and HES

with data managers at the Maudsley Biomedical Research Centre to ensure all possible information is being captured in extraction. We have also requested data on violent behaviour, violent incident, and criminal justice involvement to be extracted closest to referral, closest to discharge and whether a variable has 'ever' been recorded. This will maximise our likelihood of retrieving all available violent behaviour and criminal justice involvement. This process has also provided us with a chance to conceptualise relevant variables (e.g. violent behaviour being defined from three different assessments conducted in CAMHS).

Missing key variables could lead to unmeasured confounding or omitted variable bias, reducing the validity of causal inferences.

Missing data will be imputed using multivariate imputation of chained equations (MICE). Sensitivity checks will be conducted to test the assumption of whether variables are Missing at Random (MAR) or Missing Not at Random (MNAR).

3.4. Other sources of bias

There are CYP with more than one admission to CAMHS. Therefore, we will treat each referral and period of service use as a single 'episode' and present analyses on the first episode of care for a CYP in order that our findings might speak to the dilemmas for clinicians and mental health services in understanding and planning for CYP's needs at intake.

To address potential period effects, we will incorporate the year of referral as a covariate in our models to account for changes in CAMHS policies, resource allocation, or societal factors over time. These analyses provide insights into temporal trends and mitigate potential confounding by period effects. While our analyses will account for period effects to the extent possible, we acknowledge that unmeasured temporal factors (e.g., evolving diagnostic criteria or external societal influences) may still influence the results.

The categories derived from administrative data may have limited reliability and validity because it is not possible to know whether clinicians are using these data fields in the same way and for the same reasons each time. Various assessor (e.g., professional background and experience), informant (e.g., relationship to young person), and service-related (e.g., limited resources) factors may influence how different clinicians perceive and assess variables proposed in this analysis. We also do not know, other than for the SDQ, what different sources of evidence (e.g., CYP disclosure, parents, social workers etc) clinicians used in their assessment. Violence is not always recognised at an admission, or consistently recorded, and therefore this study's estimates of prevalence of violence are likely to be underestimates.

We recognise that the relationship between severity of mental health difficulties and criminal justice involvement or violent incidents may not be linear. For CYP who have been admitted to inpatient care, who likely have the most severe mental health difficulties, violent incidents and criminal justice involvement may be lower. This is because they would not have the opportunity to engage with the criminal justice system when staying in hospital. We will therefore explore possible non-linear relationships between severity and outcomes.

CYP with input from services beyond CAMHS (e.g. social care involvement) may be reported as engaging in more violent behaviour due to the increased opportunity and monitoring of behaviour compared to those who do not have additional service input.

Equality, diversity, and inclusion have been central to the design of this research proposal. Professor Duschinsky has served as Special Advisor to the Wellcome Trust for EDI. We have undertaken a formal Equality Impact Assessment to ensure equality, diversity, cohesion, and integration have been considered thoroughly in this proposal.

A central motivation to conduct this research is that justice-involved CYP been underrepresented in research on secondary mental health care. In attending to potential inequalities and disproportionality, it is imperative to consider differences between groups in each element of my proposed research. Information about participant personal and contextual characteristics will be available from CAMHS data. We will examine socioeconomic factors, cumulative risk factors, ethnicity, gender, as well as a variety of other personal or contextual variables.

Ethical issues around the representation of justice-involved CYP in our dissemination and engagement activities will be handled carefully. Strategies to address this will include attention to strengths as well as difficulties, and co-development of materials and representations with experts-by-experience. The NCB have specialists in handling the ethical issues of representation of CYP needs in their communications team, who will provide support.

However, as others have described (e.g., Perera et al., 2016¹³), compared to national averages, SLaM has a considerably higher representation of patients from black minority groups and lower proportion of patients from Asian minority groups. Therefore, it would be beneficial to test these models in other parts of the UK.

4. About the analysis

4.1. Overview of analytical approach

Building on a previously funded YEF Secondary Data Analysis project, we will provide a descriptive account of the prevalence of violent behaviour, criminal justice involvement, and violent incidents using the most recent sweep of data collection. Violent behaviour will be captured through the SLaM Risk Assessment, which documents violent and aggressive behaviour; the CAMHS Item Sheet, which documents antisocial behaviour in terms of fighting and violent assault; and the CAMHS Brief Risk Assessment, which documents destructive and dangerous behaviour towards others. CJ involvement will be captured through the forensic history taken in the CAMHS Assessment, by juvenile conviction records in the CAMHS Information and whether the CYP received a criminal justice intervention in their Risk Management Plan. Violent incidents will be captured from HES Accident and Emergency data.

This descriptive account will include information on the cohort of CYP in terms of their mental health (ICD-10), ethnicity, gender, age, and IMD. We will also provide a descriptive account of the following risk factors: domestic violence, parent/carers mental health difficulties,

parent/carer substance misuse, emotional abuse, physical abuse, sexual abuse, neglect, destructive behaviour, dangerous behaviour, self-neglect, physical disability, running away, self-harm, risk-taking behaviour, lack of concern by parents/caregivers, lack of concern from other agencies, and educational challenges (Special Education Needs and Disabilities (SEND)).

This research will employ a retrospective cohort study design, analysing longitudinal administrative data from CRIS SLaM and HES.

To address issues of temporality, exposure variables (mental health diagnoses, sociodemographic characteristics, and CAMHS risk factors) will be anchored to the initial CAMHS referral using values recorded closest to the referral date. These variables therefore represent the clinical and contextual profile of children at the point of entry to CAMHS. While some risk indicators may refer to historical experiences (for example abuse or neglect), anchoring exposures to the intake assessment ensures that the analyses examine associations between characteristics identified at CAMHS entry and subsequent outcomes during follow-up where applicable.

Outcomes in this study are recorded both closest to referral (T1) and closest to discharge (T2). Analyses using T1 outcomes will be interpreted as cross-sectional explanatory models, examining patterns of co-occurring vulnerability identified at the point of CAMHS entry. Analyses using T2 outcomes will examine whether characteristics measured at referral are associated with later criminal justice involvement, violent incidents, or violent behaviour recorded during follow-up. This distinction allows us to separate patterns of vulnerability at intake from associations with subsequent outcomes during CAMHS involvement.

Given the potential conceptual overlap between SEND and some CAMHS diagnoses, models including both variables will be assessed for collinearity. Where substantial overlap is observed, alternative model specifications will be explored in sensitivity analyses.

Survival analysis techniques will be employed to investigate the timing and risk of first-time involvement in crime and violence among CYPs with mental health difficulties. Cox proportional hazard models will be utilised to estimate hazard ratios, providing insights into how specific mental health diagnoses, sociodemographic characteristics, and CAMHS risk factors influence the time to criminal justice involvement. This approach will help identify critical periods for intervention and prevention, allowing policymakers to design timely support strategies. The analysis will account for censored data, ensuring accurate estimates even when follow-up periods vary across individuals. Kaplan-Meier survival curves will also be used to visually represent differences in time-to-event distributions across different subgroups, such as ethnicity, gender, and IMD. For these analyses, children with evidence of criminal justice involvement or violent behaviour recorded at referral will be excluded so that survival models estimate the time to first recorded event following CAMHS entry.

Logistic regression models will then be employed to determine the likelihood of violent behaviour, violent incidents and criminal justice involvement based on the presence of specific mental health diagnoses, sociodemographic characteristics and CAMHS risk factors. These models will allow us to estimate adjusted odds ratios and assess the relative

contribution of individual risk factors while adjusting for potential confounders such as demographic characteristics and prior mental health service usage. Additional logistic regression models will be used to examine the cumulative effects of co-occurring mental health diagnoses and CAMHS risk factors by incorporating variables representing the number of diagnoses and cumulative risk exposure. Finally, stratified regression models and interaction terms will be used to explore whether gender and ethnicity moderate the relationship between mental health, CAMHS risk factors, violent behaviour and criminal justice involvement.

Survival analysis and regression models provide robust tools for identifying associations and potential causal pathways between mental health difficulties, demographic characteristics, CAMHS risk factors, and involvement in crime and violence. While observational data cannot establish causality definitively, these methods allow for statistical control of confounding variables, improving the validity of the findings.

Survival analysis enhances causal inference by accounting for time-dependent covariates and censoring, providing a clearer picture of the temporal sequence of events. Modelling the time to CJ involvement allows for a better understanding of the potential impact of interventions and the identification of high-risk periods where intervention might be most effective. Additionally, stratified survival analysis can help assess the effects of moderating variables, offering insights into the differential impact of risk factors across subgroups. Logistic regression models will then enable the estimation of adjusted odds ratios, which provide insights into the strength and direction of associations between risk factors and outcomes, controlling for known confounders.

Despite these strengths, causal inference is limited by potential residual confounding and unmeasured variables. To further strengthen causal interpretations, sensitivity analyses will be conducted to assess the robustness of findings under different assumptions, and propensity score methods will be employed to balance observed covariates between comparison groups.

Race equity is a central consideration in our research, as disparities in mental health, criminal justice involvement, and access to services disproportionately affect racially minoritized CYP. Our study will take a proactive approach to ensure that race equity is embedded in both the research design and the interpretation of findings. Professor Duschinsky has served as Special Advisor to the Wellcome Trust for EDI. We have undertaken a formal Equality Impact Assessment to ensure equality, diversity, cohesion, and integration have been considered thoroughly in this proposal.

Examining racial disparities in mental health and CJ involvement: We will systematically assess whether and how mental health diagnoses, CAMHS risk factors, violent behaviour and CJ involvement differ by ethnicity. Using the CRIS dataset, we will:

- Disaggregate findings by ethnicity to identify disparities in mental health diagnoses, sociodemographic characteristics and CAMHS risk factors.

- Examine whether certain ethnic groups experience multiple mental health diagnoses that may contribute to violent behaviour and CJ involvement.
- Examine whether certain ethnic groups experience greater cumulative risk that may contribute to violent behaviour and CJ involvement.

Addressing structural inequities in data interpretation: We recognize that structural racism and systemic bias may shape patterns observed in our data. To mitigate these biases, we will:

- Use statistical techniques such as stratified regression models and interaction effects to explore whether race moderates the relationship between mental health, risk factors, and CJ involvement.
- Contextualize findings within existing evidence on racial disparities, mental health and the CJ research.
- Engage a diverse group of experts-by-experience with lived experience of both mental health services and the CJ system to interpret findings and ensure that analyses do not reinforce harmful stereotypes.

Policy and practice implications: Findings from this study will inform race-conscious interventions by:

- Highlighting where targeted interventions may be necessary to reduce disparities in mental health treatment, risk exposure, and youth justice outcomes.
- Providing evidence to policymakers on the need for culturally responsive mental health services and equitable justice policies.

By embedding race equity considerations into each stage of our research, we aim to contribute to a more just and effective system for supporting CYP at risk of violent behaviour and CJ involvement.

The selected methodology provides a comprehensive and feasible approach given the available data, ensuring accurate, actionable insights that can be translated into policy and practice.

4.2. Approach to addressing research question(s)

Research questions 1, 2, 3: approach and methods

Research question	1. Which mental health difficulties are associated with criminal justice (CJ) involvement and violent behaviour? 1a. What types of mental health diagnoses (e.g., conduct disorder, depression, anxiety) are most strongly associated with subsequent involvement in crime and violence among children and young people?
--------------------------	--

	1b. How does the co-occurrence of multiple mental health difficulties influence the likelihood of CJ involvement and violent behaviour? 1c. To what extent do sociodemographic characteristics (e.g., ethnicity, gender, age, indices of multiple deprivation (IMD), educational challenges) moderate the relationship between mental health difficulties and CJ involvement? 2. Which risk factors in the CAMHS risk assessment are associated with CJ involvement and violence? 2a. Which risk factors from the CAMHS risk assessment are most strongly associated with future CJ involvement and violent behaviour? 2b. How do combinations of risk factors in CAMHS assessments interact to impact the association of CYP's involvement in crime and violence? 2c. Are there critical thresholds or clusters of risk factors that significantly increase the likelihood of CJ involvement or violence? 3. What combination of mental health diagnosis, sociodemographic characteristics and risk factors captured in CAMHS are most strongly associated with CYP's violent behaviour and CJ involvement?
<i>Hypothesis, if relevant</i>	H1: Externalising disorders—specifically conduct disorder, ODD, ADHD, and substance use disorders—will be the strongest mental-health associations with CJ involvement and violent behaviour. CYP with multiple co-occurring diagnoses will show elevated likelihoods, and sociodemographic factors will moderate these associations. H2: Risk factors relating to family adversity (domestic violence, parental substance misuse, parental mental illness), abuse or neglect, risk-taking behaviours, and self-harm will be the strongest antecedents of CJ involvement and violent behaviour. Multiple risk factors are expected to have cumulative or synergistic effects. H3: A combined model incorporating mental health diagnoses, sociodemographic factors (gender, ethnicity, age, IMD, SEND), and CAMHS risk factors will provide the most

	robust associations with violent behaviour and CJ involvement.
<i>What will you be able to say by the interim report</i>	In the interim report, we will enable us to report comprehensive descriptive statistics on the prevalence of violent behaviour, violent A&E incidents and criminal justice involvement, alongside distributions of mental health diagnoses, sociodemographic characteristics and CAMHS risk factors, each with associated 95% confidence intervals. Using multivariable logistic regression, we will report adjusted odds ratios, 95% confidence intervals and p-values to identify which mental health difficulties and CAMHS risk factors are associated with violent behaviour and criminal justice involvement, including the relative contribution of specific diagnoses, cumulative or co-occurring mental health difficulties, and individual or combined risk factors. Moderation analyses will allow us to report whether these associations differ by gender, ethnicity, age, deprivation or SEND status, supported by interaction terms. We will also report whether particular combinations or thresholds of risk factors sharply increase the likelihood of violence or criminal justice involvement.
<i>Descriptive analysis, if relevant</i>	Building on a previously funded YEF Secondary Data Analysis project, we will provide a descriptive account of the prevalence of violent behaviour, criminal justice involvement, and violent incidents using the most recent sweep of data collection. Violent behaviour will be captured through the SLaM Risk Assessment, which documents violent and aggressive behaviour; the CAMHS Item Sheet, which documents antisocial behaviour in terms of fighting and violent assault; and the CAMHS Brief Risk Assessment, which documents destructive and dangerous behaviour towards others. CJ involvement will be captured through the forensic history taken in the CAMHS Assessment, by juvenile conviction records in the CAMHS Information and whether the CYP received a criminal justice intervention in their Risk Management Plan. Violent incidents will be captured from HES Accident and Emergency data. This descriptive account will include information on the cohort of CYP in terms of their mental health (ICD-10), ethnicity, gender, age, and IMD. We will also provide a descriptive account of the following risk factors: domestic violence, parent/carer mental health difficulties, parent/carer substance misuse, emotional abuse, physical abuse, sexual abuse, neglect, destructive behaviour, dangerous behaviour, self-neglect, physical disability,

	running away, self-harm, risk-taking behaviour, lack of concern by parents/caregivers, lack of concern from other agencies, and educational challenges (Special Education Needs and Disabilities (SEND)).
<i>Models, specifications and statistical techniques used, if relevant</i>	For RQ1, we will use logistic regression to estimate associations between specific mental health diagnoses and: (a) violent behaviour, (b) violent incidents, and (c) CJ involvement. Adjust for sociodemographic characteristics and prior service use. Examine co-occurring diagnoses using multiple regression and cumulative burden indices. For RQ2, logistic regression models will be used to test whether individual risk factors (e.g. domestic violence, parental mental health, running away, substance misuse, abuse/neglect, risk-taking behaviour) are associated with violent behaviour, violent incidents and CJ involvement. We will model combinations of risk factors using interaction terms. We will identify clusters or thresholds using exploratory factor analyses. For RQ3, we will build hierarchical or multivariable models incorporating mental health diagnoses, sociodemographic characteristics, and CAMHS risk factors simultaneously. We will assess relative contributions (odds ratios, pseudo-R^2), and potential high-risk profiles. In cases of missing data, multiple imputation of chained equations (MICE) will be used. This will be a distinct step before analysis is conducted. The MICE algorithm offers a robust method of imputing values through iterative estimations based on conditional distributions. It accounts for dependencies between variables, by using the values of other variables to estimate the missing data. MICE will be applied in five steps: 1. Initial imputation: We will fill in all missing values in the dataset with simple placeholders, for example the mean or median of each variable or the most frequent for categorical variables. This allows us to impute missing values iteratively. 2. Reset missing values: We will then choose a variable at a time for imputation and change those placeholder values back to missing. 3. Build regression model: We will build a regression model to predict the variable from step 2, by using the other independent variables in the dataset. This

	model uses all of the other variables in the dataset as predictors to estimate the values of the missing data points. 4. Replace missing values: We will use the regression model to predict and fill in the missing values for this variable. At this point, this variable has now both observed and newly imputed values. 5. Repeat for each variable: The number of multiple imputations will be greater than or equal to the percentage of missing data ($M \geq \% \text{ missing data}$). Going through all the variables will complete one iteration or 'cycle'. For each cycle, the model updates imputation values, making the predictions more accurate. We will conduct 10 cycles, at which point the model is said to have reached convergence and the estimates won't improve any further.
<i>Estimating equation, if relevant</i>	N/A.
<i>What does the approach need to succeed (constraints/assumptions)?</i>	As many as 506 participants are needed to detect (at 80% power) an effect of $d=0.25$, and four times that are required to identify interaction effects – an obstacle to investigations of inequalities and intersectionality. However, as the number of CYP is over 53,000, we are confident we will have sufficient prevalence to be able to investigate our proposed multivariable and stratified models.
<i>Uncertainty and inference</i>	• Prevalence estimates with 95% confidence intervals • Logistic regression: odds ratios, 95% CIs, p-values • Model fit: AIC, BIC, and pseudo-R^2 • Predictive accuracy: AUC (95% CI) • Interaction effects: interaction term coefficients, 95% CIs, p-values • Marginal predicted probabilities with 95% CIs • Sensitivity analysis: comparison of AORs and CIs across models.
<i>Robustness checks</i>	We will assess model stability by testing reduced and expanded covariate sets, exploring whether the inclusion or exclusion of specific sociodemographic variables or risk factors meaningfully affects parameter estimates. We will evaluate the influence of infrequent or rare predictors by re-running models excluding low-prevalence diagnoses and risk factors to ensure that results are not driven by sparsely populated categories. Finally, if we encounter a larger number of correlated predictors (or sparsely populated categories), we will consider an exploratory penalised regression approach (e.g., elastic net; or elastic net Cox for

	time-to-event outcomes) as a robustness check alongside the prespecified models.
<i>Subgroup you intend to study</i>	We will test moderation by CYP's ethnicity, gender, age, socioeconomic status, year of referral, and educational characteristics using interaction terms in stratified logistic regression models. We will assess whether these characteristics modify relationships between diagnoses/risk factors and outcomes.
<i>Changes to the analysis</i>	The planned analyses are pre-specified, but we anticipate that some refinements may be required in response to the structure and quality of the data. If the prevalence of certain outcomes (e.g. specific forms of violent behaviour or CJ involvement) or predictors (e.g. rare diagnoses or risk factors) is lower than expected, we may need to collapse categories (for example, grouping related diagnoses or combining rare risk factors) or simplify models to avoid sparse-data bias and convergence problems. Similarly, if missingness is more extensive or more complex than anticipated, we may need to refine our multiple imputation strategy (e.g. revising the imputation model, excluding variables with very high missingness, or conducting additional complete-case sensitivity analyses). We may also adapt the modelling approach if exploratory work indicates that alternative specifications are more appropriate—for instance, prioritising logistic regression for binary outcomes and considering alternative functional forms or reduced interaction terms where models are unstable or over-parameterised. Any deviations from the initial plan will be driven by empirical considerations, will be minimised where possible, and will be transparently documented and justified in reporting.

Research question 1b: approach and methods

<i>Research question</i>	4. What role do the duration and timing of mental health difficulties play in the association between criminal justice involvement, violent incidents and violent behaviour among children and young people?
<i>Hypothesis, if relevant</i>	H4: Earlier onset and longer duration of mental health difficulties will be more strongly associated with CJ involvement and violent behaviour
<i>What will you be able to say by the interim report</i>	In the interim report, we should be able to report piloted survival analysis approaches are feasible.

<i>Descriptive analysis, if relevant</i>	Described above for RQ1 and RQ2.
<i>Models, specifications and statistical techniques used, if relevant</i>	To examine how the duration and timing of mental health difficulties influence the risk of CJ involvement, violent incidents and violent behaviour, we will conduct time-to-event analyses using survival modelling. The primary outcome will be the time from first CAMHS contact to the first recorded CJ involvement and first recorded violent incident or behaviour. A key limitation of the dataset is that violent behaviour, violent incidents and CJ involvement are often documented either at the point of referral or at discharge, rather than dated at the time the event occurred. To address this, documented dates will be used as approximate event times, and the analysis will treat events recorded at referral as left-censored and events recorded at discharge as interval-censored between referral and discharge. We will fit Cox proportional hazards models using these approximated event times and assess the proportional hazards assumption using Schoenfeld residuals; where violated, time-varying coefficients or stratified Cox models will be employed. Timing will also be represented through variables capturing age at first diagnosis, duration since onset of mental health difficulties and, where feasible, time-varying indicators of active risk factors. Kaplan–Meier survival curves will be used to describe unadjusted time-to-event patterns and compare subgroups defined by mental health diagnoses, gender, ethnicity, IMD and duration of difficulties. Interpretation will emphasise relative differences between groups rather than precise event timing. Cox models will estimate hazard ratios and 95% confidence intervals for the association between specific diagnoses, sociodemographic characteristics and CAMHS risk factors and the time to first recorded CJ involvement or violent incident, adjusting for confounders such as prior service use.
<i>Estimating equation, if relevant</i>	N/A.
<i>What does the approach need to succeed (constraints/assumptions)?</i>	This approach relies on several key assumptions and conditions related to the structure, quality and completeness of the administrative data. First, survival analysis requires sufficient precision in exposure timing—that is, reasonably accurate dates for first CAMHS contact, mental health diagnoses and the onset of relevant risk factors. Although outcome dates for violent behaviour, violent incidents and CJ involvement may be approximate,

the method assumes that documentation dates are reasonable proxies for event occurrence and that left-censoring and interval-censoring procedures adequately capture uncertainty. For analyses involving duration or timing of mental health difficulties, we assume that diagnosis dates and follow-up periods are reliably recorded and that temporal ordering (mental health difficulties → outcomes) is broadly correct.

A further constraint relates to the number of CYP who experience new-onset criminal justice involvement, violent behaviour or violent incidents between T1 and T2. If only a small proportion of CYP are free of these outcomes at T1 and subsequently experience an event during follow-up, the statistical power for estimating time-to-event models will be limited. In this scenario, three adaptations will be made to ensure RQ4 can still be meaningfully addressed.

The approach further assumes that missing data are manageable and that the use of multiple imputation (MICE) is appropriate—that missingness can be explained by observed covariates (MAR assumption) and that key variables are not missing at levels that prevent meaningful modelling. Sufficient sample size and event counts are needed to support Cox models, subgroup analyses (e.g., by gender, ethnicity, IMD) and sensitivity analyses; low-prevalence diagnoses or risk factors may need to be combined to avoid sparse data bias.

Survival models assume either that the proportional hazards assumption holds or that any violations can be addressed through time-varying effects or stratification. The analytic strategy also depends on accurate linkage between SLaM and HES datasets to identify violent incidents, and on stable identifiers to track children and young people longitudinally. Conceptually, the approach assumes that documented risk factors and diagnoses reasonably represent underlying clinical and social conditions and that recording practices do not differ systematically across key subgroups in ways that could bias results.

Finally, the success of the approach requires transparent documentation and careful interpretation. Given the uncertainty in exact event timing, results will focus on relative hazards and differences between groups rather than precise estimates of when events occurred. Clinical and

	contextual expertise will be essential for interpreting these temporal patterns and ensuring that findings are both plausible and policy-relevant.
<i>Uncertainty and inference</i>	• Kaplan–Meier survival probabilities with 95% CIs, median time-to-event, and log-rank tests. • Cox model hazard ratios, 95% CIs, p-values, and tests of proportional hazards. • Model fit indices (likelihood ratio tests; AIC/BIC for interval-censored models). • Time-varying coefficients or stratified results where PH assumptions are violated. • Sensitivity analyses comparing HRs and CIs under alternative event-timing assumptions. • Logistic regression AORs, 95% CIs, p-values, and pseudo-R² as timing-insensitive comparisons. • Predicted probabilities and marginal effects with 95% CIs for key subgroups
<i>Robustness checks</i>	Given uncertainty in event timing, extensive sensitivity analyses will be conducted. These will include alternative assumptions about the timing of events (e.g., treating discharge-recorded events as occurring at discharge, at referral or at the midpoint between referral and discharge), as well as secondary analyses restricting outcomes to those with more reliable timing (e.g., HES-recorded violent incidents). We will also estimate parallel logistic regression models (ever vs. never CJ involvement/violence) to provide a timing-insensitive comparison. Additional sensitivity checks will explore model stability under different covariate specifications and exclusion of low-prevalence predictors. Together, these survival and cross-sectional analyses will allow us to characterise how the onset, persistence and timing of mental health difficulties relate to criminal justice involvement, violent incidents and violent behaviour while transparently addressing uncertainty in event dates and ensuring that conclusions are robust across modelling assumptions.
<i>Subgroup you intend to study</i>	By CYP's ethnicity, gender, age, socioeconomic status, year of referral, cumulative risk factors, and educational characteristics
<i>Changes to the analysis</i>	Adaptations to the planned survival analysis may be required in response to data quality and timing constraints. If precise event dates for violent behaviour, violent incidents or criminal justice involvement are unavailable or highly clustered at referral or discharge, we may simplify the time-to-event framework by using approximate event times (e.g.,

at discharge or at the midpoint between referral and discharge) and prioritise a smaller set of survival models supported by sensitivity analyses. If interval-censored models prove unstable or infeasible, we may instead rely on standard Cox models with clearly documented timing assumptions, supplemented by discrete-time or logistic regression models (ever vs. never) as timing-insensitive robustness checks. If the proportional hazards assumption is substantially violated for key predictors, we may reduce the number of time-varying effects modelled or move to stratified analyses for a limited set of core variables (e.g., gender or broad diagnostic groups). If missing data or poor recording of onset/duration make fine-grained timing variables unreliable, we may replace detailed duration measures with simpler indicators (e.g., early vs. late onset, shorter vs. longer time in CAMHS) and focus interpretation on relative differences between groups rather than precise timing.

A further adaptation may be necessary if only a small number of CYP are free of CJ involvement, violent behaviour or violent incidents at T1 but experience new-onset events between T1 and T2. In this case, three contingency approaches will be used. First (1a), survival models may be simplified by limiting predictors to core diagnostic and demographic variables and removing interaction terms or granular covariates; these models will be treated as exploratory, with emphasis on effect direction rather than precise hazard estimation. Second (1b), if modelling is underpowered, RQ4 will be addressed through descriptive timing analyses, including summarising age at first recorded CJ/violent event and duration between first mental health difficulty and any outcome, with comparisons across key subgroups. Third (1c), logistic regression models predicting any CJ involvement or violent behaviour by the end of follow-up will incorporate timing-related predictors (e.g., early vs late onset; shorter vs longer duration), providing a pragmatic alternative analysis that does not depend on large numbers of new-onset events. All such modifications will be empirically driven, minimised where possible, and transparently reported, with clear documentation of assumptions and limitations arising from data constraints.

5. Project management

5.1. Risks and mitigations

Table 5.1 Risks and mitigations

Number	Risk	Likelihood (Low/Medium/ High)	Mitigation
1	Errors in linking CRIS SLaM and HES dataset (e.g. non-matching identifiers) could result in loss of critical records, reducing sample size and validity.	Low	Conduct checks on linked data to identify inconsistencies and assess whether linkage bias exists (e.g. if specific groups are less likely to be successfully linked).
2	Measurement and coding errors. Inconsistent or inaccurate coding of key variables in administrative datasets could lead to misclassification.	Low	We will review coding practices used in CRIS SLaM and HES datasets and consult clinical and administrative data experts to understand any inconsistencies. We will perform data cleaning and recoding to standardize variables before analysis.
3	Missing or incomplete data for key variables could impact the ability to draw robust conclusions.	Medium	Performing a comprehensive data audit during initial project phase to assess the extent of missingness. Use multiple imputation of chained equations for missing data to account for potential biases. Engaging with data managers and clinicians to explore whether additional variables or proxy variables can address gaps.
4	Low event frequency and power addressing RQ4	Medium	The viability of the planned survival analyses depends on there being a sufficient number of CYP who are free of CJ involvement, violent behaviour and violent incidents at T1 but experience a first event during follow-up. If the number of new-onset cases between T1 and T2 is very small, survival models will be simplified (e.g. reduced sets of predictors, no interaction terms) and treated as exploratory, with an emphasis

			on effect direction rather than precise estimates. In this scenario, RQ1b will be addressed primarily through descriptive analyses of age at first event and duration from first recorded mental health difficulty to any CJ or violent outcome, and through supplementary logistic regression models predicting any CJ/violent involvement by the end of follow-up. Findings will be interpreted cautiously, with explicit acknowledgement of limited power for detecting timing effects.
--	--	--	--

5.2. Timeline

Table 5.2 Timeline

Date	Activity	Staff responsible/leading
11/2025 – 01/2026	Host a meeting with NCB's Youth Justice Research Advisory Group. Provide a descriptive account of the prevalence of violent behaviour, offending/criminal behaviour, prolific offending, and violent incidents at the point of referral to CAMHS in CYP within CRIS SLAM. Provide a description of the cohort of CYP in terms of their mental health, ethnicity, gender, age, IMD, risk factors, and educational challenges. Submit full research protocol.	Robbie Duschinsky and Guy Skinner
02/2026-08-2026	Host a meeting with NCB's Youth Justice Research Advisory Group. Based on the descriptive accounts developed, use logistic regression models and survival analyses to investigate research questions for WP1. Submit the Interim Report once logistic regression and survival analysis models have been piloted.	Robbie Duschinsky and Guy Skinner
09/2026-04/2027	Host a meeting with NCB Youth Justice Research Advisory Group. Use logistic regression models and survival analyses to investigate research questions for WP2 and WP3. Begin writing the final report,	Robbie Duschinsky and Guy Skinner

	co-authored by Dr Jamiee Maillion, NCB Youth Justice Lead and Darrick Jolliffe, Head of Department, Security and Crime Science, University College London.	
--	--	--

6. References

- ¹ Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., Poulton, R., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *American Journal of Psychiatry*, 165(5), 603–611. Doi: 10.1176/appi.ajp.2007.07040614
- ² Moffitt, T. E. (2018). Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, 2, 177–186. Doi: 10.1038/s41562-018-0307-9
- ³ Sibley, M. H., Evans, S. W., & Serpell, Z. N. (2011). Legal outcomes of adolescents with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, 50(3), 242–251. Doi: 10.1016/j.jaac.2010.12.003
- ⁴ Bennett, T., Holloway, K., & Farrington, D. P. (2008). The statistical association between drug misuse and crime: A meta-analysis. *Aggression and Violent Behavior*, 13(2), 107–118. Doi: 10.1016/j.avb.2008.02.001
- ⁵ Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry*, 40(1), 57–87. Doi: 10.1111/1469-7610.00424
- ⁶ Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2008). Young adult outcome of hyperactive children: Adaptive functioning in major life activities. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(5), 540–547. Doi: 10.1097/CHI.0b013e318167753e
- ⁷ Hughes, K., Bellis, M. A., Hardcastle, K. A., Sethi, D., Butchart, A., Mikton, C., Jones, L., & Dunne, M. P. (2017). The effect of multiple adverse childhood experiences on health: A systematic review and meta-analysis. *The Lancet Public Health*, 2(8), e356–e366. Doi: 10.1016/S2468-2667(17)30118-4
- ⁸ Thornberry, T. P., Freeman-Gallant, A., Lizotte, A. J., Krohn, M. D., & Smith, C. A. (2003). Linked lives: The intergenerational transmission of antisocial behavior. *Journal of Abnormal Child Psychology*, 31(2), 171–184. Doi: 10.1023/A:1022584923641
- ⁹ Widom, C. S., & Maxfield, M. G. (2001). *An update on the “cycle of violence”*. National Institute of Justice. <https://www.ojp.gov/pdffiles1/nij/184894.pdf>

-
- ¹⁰ Evans, G. W., Li, D., & Whipple, S. S. (2013). Cumulative risk and child development. *Psychological Bulletin*, 139(6), 1342–1396. Doi: 10.1037/a0031808
- ¹¹ Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, 58(6), S327–S337. <https://doi.org/10.1093/geronb/58.6.S327>
- ¹² Pearlin, L. I., Schieman, S., Fazio, E. M., & Meersman, S. C. (2005). Stress, health, and the life course: Some conceptual perspectives. *Journal of Health and Social Behavior*, 46(2), 205–219. <https://doi.org/10.1177/002214650504600206>
- ¹³ Finkelhor, D., Shattuck, A., Turner, H., & Hamby, S. (2015). A revised inventory of adverse childhood experiences. *Child Abuse & Neglect*, 48, 13–21. Doi: 10.1016/j.chiabu.2015.07.011
- ¹⁴ Mulvey, E. P., Schubert, C. A., & Chassin, L. (2010). Substance use and delinquent behavior among serious adolescent offenders. *Development and Psychopathology*, 22(2), 453–475. Doi: 10.1017/S0954579410000178
- ¹⁵ Bronfenbrenner, U. (1994). Ecological models of human development. In T. Husén & T. N. Postlethwaite (Eds.), *International encyclopedia of education* (2nd ed., Vol. 3, pp. 1643–1647). Pergamon Press.
- ¹⁶ Sampson, R. J., & Wilson, W. J. (1995). Toward a theory of race, crime, and urban inequality. In J. Hagan & R. D. Peterson (Eds.), *Crime and inequality* (pp. 37–56). Stanford University Press.
- ¹⁷ 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*. UK Government. <https://www.gov.uk/government/publications/lammy-review-final-report>
- ¹⁸ Young, S., Moss, D., Sedgwick, O., Fridman, M., & Hodgkins, P. (2015). A meta-analysis of the prevalence of attention deficit hyperactivity disorder in incarcerated populations. *Psychological Medicine*, 45(2), 247–258. <https://doi.org/10.1017/S0033291714000762>
- ¹⁹ Hughes, N., Williams, H., Chitsabesan, P., Davies, R., & Mounce, L. (2017). Nobody made the connection: The prevalence of neurodisability in young people who offend. *London: Office of the Children's Commissioner*. <https://doi.org/10.13140/RG.2.2.27131.26407>
- ²⁰ Perera, G., Broadbent, M., Callard, F., Chang, C.-K., Downs, J., Dutta, R., et al. (2016). Cohort profile of the South London and Maudsley NHS foundation trust biomedical research Centre (SLaM BRC) case register: Current status and recent enhancement of an electronic mental health record-derived data resource. *BMJ Open*, 6(3), e008721. Doi: 10.1136/bmjopen-2015-008721

-
- ²¹ Davis, K. A. S., Bashford, O., Jewell, A., Shetty, H., Stewart, R. J., Sudlow, C. L. M., et al. (2018). Using data linkage to electronic patient records to assess the validity of selected mental health diagnoses in English Hospital Episode Statistics (HES). *PLoS One*, 13(3), e0195002. Doi: 10.1371/journal.pone.0195002
- ²² Jewell, A., Broadbent, M., Hayes, R. D., Gilbert, R., Stewart, R., Downs, J., et al. (2020). Impact of matching error on linked mortality outcome in a data linkage of secondary mental health data with Hospital Episode Statistics (HES) and mortality records in South East London: a cross-sectional study. *BMJ Open*, 10, e035884. Doi: 10.1136/bmjopen-2019-035884
- ²³ Coughlan, B., Woolgar, M., Hood, R., Hutchinson, D., Bywaters, P., Duschinsky, D., et al. (2024). Risk rates and profiles at intake in child and adolescent mental health services: A cohort and latent class analyses of 21,688 young people in South London. *JCPP Advances*, 4(3), e12246. Doi: 10.1002/jcv2.12246
- ²⁴ Herbert, A., Gilbert, R., Gonzalez-Izquierdo, A., & Li, L. (2015). Violence, self-harm and drug or alcohol misuse in adolescents admitted to hospitals in England for injury: A retrospective cohort study. *BMJ Open*, 5, e006079. Doi: 10.1136/bmjopen-2014-006079
- ²⁵ Coughlan, B., Woolgar, M., Hood, R., Hutchinson, D., Denford, E., Hillier, A., et al. (2024). Risk rates and profiles at intake in child and adolescent mental health services: A cohort and latent class analyses of 21,688 young people in South London. *JCPP Advances*, 4(3), e12246. Doi: 10.1002/jcv2.12246