



Ministry
of Justice



Delivering Better
Outcomes by Linking Data,
evidence and experiences

Linking Police and Sexual Violence Support Service Victim Data - Annexes

Paul Allen, Eliz-Leyla Mani and Tom Toor-Cliff

Ministry of Justice

Ministry of Justice Analytical Series

2026

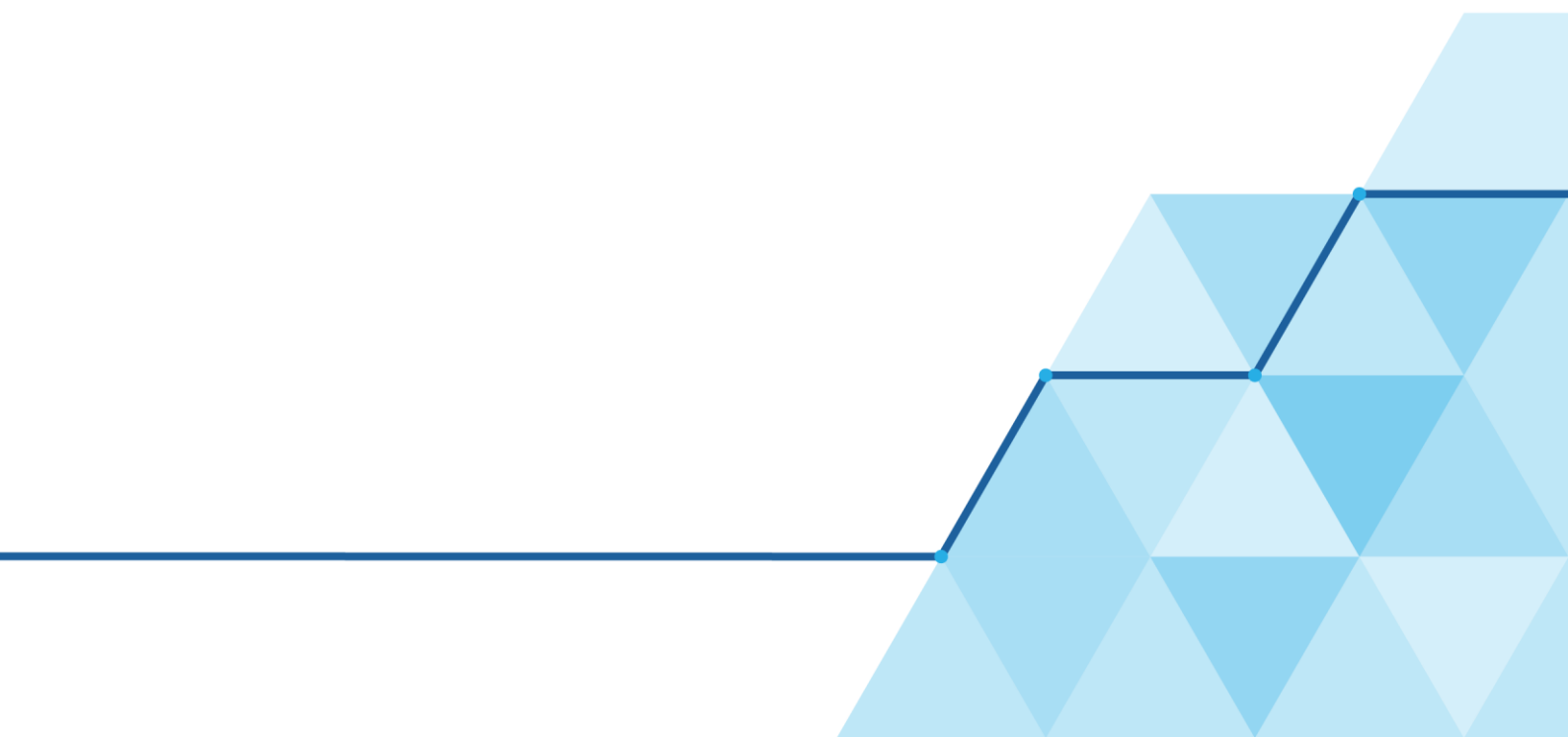
[Please don't delete the section break beneath this text.]



Ministry
of Justice



Delivering Better
Outcomes by Linking Data,
evidence and experiences



Analysis exists to improve policy making, decision taking and practice by the Ministry of Justice. It does this by providing robust, timely and relevant data and advice drawn from research and analysis undertaken by the department's analysts and by the wider research community.

Disclaimer

The views expressed are those of the authors and are not necessarily shared by the Ministry of Justice (nor do they represent Government policy).

First published 2026



© Crown copyright 2026

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit nationalarchives.gov.uk/doc/open-government-licence/version/3

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

Any enquiries regarding this publication should be sent to us at

Evidence_partnerships@justice.gov.uk

This publication is available for download at

<http://www.justice.gov.uk/publications/research-and-analysis/moj>

ISBN 978 1 911763 43 7

[Please do not delete the section break below, as this will alter the format of the previous pages.]

Contents

Annex A - Data Sharing and Linking	4
Data share	4
Structure of data used for linking	4
SWERCC	4
Essex Police	5
Linking the data	6
Direct linking	6
Tidying data followed by linking	7
Duplicate Individuals	8
Crime type matching	9
Annex B - Methodology	11
User research methodology	11
Annex C - Coverage and Completeness of Essex Police Data	13
Breakdown of Incidents	13
Missing values	14
Annex D - Breakdown of Risks, Issues and Incident Impacts	17
Annex E – Variable Groupings	18

Annex A - Data Sharing and Linking

Data share

A full Data Protection Impact Assessment (DPIA) was produced by the BOLD Victims Pilot (VP) team with input from colleagues from Essex Police to make sure that all required adherence to GDPR and the Data Protection Act, 2018 was met. This included providing justification for the lawful basis and proportionality of the data share. The document was signed by representatives from both Essex Police and BOLD.

After governance was fully approved, the data was extracted by Essex Police using Business Objects from the crime reports section of their Athena software.

Essex Police included all fields that had been agreed in the DPIA, ensuring individuals could not be identified by BOLD. The tables were then uploaded on to the MOJ Analytical Platform by Essex Police.

No free text variables were supplied, due to the potential for personally identifiable information (PII) to be present in the free text fields. Whilst deidentification was possible, the level of resource and time needed to deidentify the data was too much for this project, putting free text out of scope.

Access to this dataset was granted to three data analysts in the BOLD VP team via the GitHub platform to be analysed according to the stated purposes of the data share.

Structure of data used for linking

SWERCC

In the SWERCC dataset, the relevant data regarding police investigations into the crimes suffered by victim-survivors was stored in the Police Incidents table. This contained information that victim-survivors had told SWERCC about their police interaction including the Police Crime Reference Number (PCRN), date of report and the latest known outcome of the police investigation. This outcome did not come from the police, it contained the latest information that the victim-survivor had told SWERCC.

This table could then be linked to the Incident table in the SWERCC data using the incident ID, which was present in both tables.

The Incident table could then in turn be linked to the Cases table. This table contained information about the victim survivor, such as gender, age and ethnicity, as well as information on the victim-survivors case such as primary incident, time elapsed before seeking support and their relationship to the primary perpetrator.

The SWERCC dataset also contained information on victim-survivors who hadn't reported to the police. These records could not be linked to the police data due to lack of common information between the datasets.

Essex Police

The shared Essex Police data was in a single table. Therefore no internal linking was needed before linking to the SWERCC data.

However, there could be multiple rows of data with the same PCRN reference – a 'long format' data structure. This was mainly due to (a) multiple victims associated with a single incident; and (b) situations where a victim has multiple vulnerabilities or other specific characteristics that are not mutually exclusive. For example, if a victim associated with a particular PCRN had both their sexual orientation and political beliefs recorded as vulnerability indicators, there would be two lines in the data with the same PCRN. One line would have sexual orientation listed in the vulnerability indicator field, and the second would have political beliefs listed. All other fields would have the same information on both lines.

To address the duplication inherent to the data, individuals represented in the data were distinguished by a combination of date of birth, gender, and latest known postcode, in combination with PCRN. If any of these fields did not contain information, the individual could not be identified.¹ In total there were 19,124 PCRN with individuals that could not be identified, representing 2.6% of the total dataset.

¹ There were 10,990 PCRN with no recorded postcode, 7,908 with no recorded gender and 1,190 with no recorded date of birth.

To link the SWERCC and Essex Police datasets, common variables between both datasets had to be considered. These were;

- Police Crime Reference Number
- Date of birth.

Gender was considered. However, Essex Police only recorded binary gender,² whereas SWERCC recorded transgender information in the gender field.³

This linking assumed that there were no people with the same date of birth that shared a crime number in the data, for example, adult twins.

PCRN was not a mandatory field in the SWERCC referral form. This meant if it was not provided by the police on referral, it could be provided by the victim-survivor to SWERCC. This may not have been recorded or recalled accurately.

The Essex Police dataset defined crimes using His Majesty's Inspectorate of Constabulary and Fire & Rescue Services crime tree hierarchy.⁴ Crime Level 2 and Crime Level 3 were provided in the data share. For example, Crime Level 3 contained Rape and Other Sexual Offences. These were under a Crime Level 2 of Sexual Offences, and a Crime Level 1 of Victim-Based Crime.

Linking the data

Direct linking

In the SWERCC data that was shared, there were 3,666 unique combinations where both the PCRN and the date of birth fields were both populated. In the Essex Police data there were 730,802.

² Male and female.

³ Male, female, transgender male, transgender female, transgender and non-binary.

⁴ [About the data - His Majesty's Inspectorate of Constabulary and Fire & Rescue Services \(justiceinspectorates.gov.uk\)](https://www.justiceinspectorates.gov.uk/hmip/)

After linking these fields directly, there were 1,580 combinations of PCRN and date of birth that exist in both datasets. This included 1,570 unique PCRN and 1,426 unique SWERCC person IDs.

Tidying data followed by linking

PCRN tidying

In the SWERCC data, the PCRN appeared to be a free text field. Each entry was not necessarily in the standard PCRN format expected. For example, a field may have contained more than one PCRN, a PCRN that did not contain the correct special characters or PCRN with extra spaces between characters.

As the PCRN field in SWERCC was not in a standard format, some transformation work was carried out on that field to maximise the number of records linked.

Using pattern matching (via regular expressions) it was possible to maximise the number of PCRN that appear in the field and format them into the standard PCRN pattern.

After doing this, there were 3,073 unique PCRN and date of birth combinations in the SWERCC data. Linking this to the police data gave 1,793 combinations. This was an extra 213 above the direct linking method.

Date of birth matching

There was potential that the date of birth may not have been recorded accurately in one or both datasets.

Taking away the records that had not already been linked by the previous method described above, left 1,280 in the SWERCC data and 729,023 in the Essex Police data that had potential to be linked.

Linking was then done using PCRN, month of birth and year of birth. This allowed for linking records where the day of the month of birth had been recorded incorrectly. This gave 40 unique combinations of PCRN and date of birth that had not already been linked.

In total, of the 3,073 combinations of person ID and PCRN in SWERCC, 1,833 (60%) were linked. This was 253 more than the direct join with no manipulation and corresponded to 1,819 unique PCRN and 1,631 unique SWERCC person IDs.

The resultant table contained data from both SWERCC, such as Person ID, gender, date of birth and ethnicity as well as data from Essex Police such as gender, date of birth and postcode.

Duplicate Individuals

Following the data linking, the resultant table (containing 1,833 rows) was assessed for duplicate victims. The first step in this process was to identify rows that contained the same SWERCC person IDs but different identifiable data. The variables used initially for this process were date of birth, postcode and gender. The duplication of three of these IDs was, at least in part, due to a difference in postcode, while the date of birth remained the same. Suggesting that the SWERCC data identified this as one individual, but the Essex Police data identified them as more than one person.

Removing the postcode criteria, it was found that three SWERCC person IDs were duplicated based on a difference in gender or date of birth. On further inspection, the date of birth for each SWERCC person ID remained constant while the gender recorded by Essex Police was the varying factor. The gender of the individuals with these person IDs was also recorded by SWERCC, two of which were consistently recorded as female, and one consistently recorded as transgender male. This reinforces the fact that it is crucial to record whether an individual identifies as transgender, and it would be advantageous for Essex Police to incorporate this capability into their data systems.

The second step in assessing duplicates was to determine whether any individuals had been assigned more than one SWERCC person ID. This assessment identified 20 unique postcode and date of birth combinations (from Essex Police data) that were linked to multiple SWERCC person IDs. Given the unlikelihood that two different individuals with the same date of birth would live at the same postcode and be referred to SWERCC as victims of a similar offence, we can conclude that it is likely that 20 individuals were erroneously assigned multiple SWERCC IDs.

Crime type matching

There was a small percentage of cases in the joined table for which the crime type recorded by Essex Police was not a sexual offence. Of the 1,833 cases, 2.89% (53 cases) had an Essex Police level 2 crime type other than “SEXUAL OFFENCES”. The breakdown of the level 2 crime types for these cases can be seen in the Table below.

Table 1 - The number of linked cases where the Essex Police Level 2 Crime Type did not match a sexual offence.

Level 2 Crime Type	Count	Percentage
Violence against the person	43	81%
Damage and arson offences	3	5.7%
Public order offences	3	5.7%
Misc crimes against society	2	3.8%
Theft offences	2	3.8%

The majority of these cases (81%, 43 cases) have an Essex Police level 2 crime type of “Violence against the person”, with the remaining 19% (10 cases) fairly evenly divided between “Damage and arson offences”, “Public order offences”, “Misc crimes against society” and “Theft offences”.

For 50 of these 53 cases (94%), it was possible to estimate whether or not a sexual offence for them was recorded via a different PCRN (by using date of birth and postcode to identify and individual victim). Essex Police contained a record to show that for 66% (33) of these cases, the individual was a victim of a sexual offence, but this record had a different PCRN than the one referenced by SWERCC. This showed it is unlikely Essex

Police are incorrectly referring very few, if any,⁵ individuals that are not victims or sexual offences to SWERCC.

Table 2 - Breakdown of linked cases in SWERCC data

SWERCC Database	Count
Cases with information entered in the PCRN field	3,998
Cases with retrievable PCRN	3,073
Essex Police related PCRN	2,430
Linked data	1,833
Potential duplicates removed	1,819

Base = 3,998 SWERCC cases

⁵ As the linking between the SWERCC and Essex Police datasets was carried out using only the PCRN and victim date of birth, there were limitations when identifying unique victim-survivors between different PCRN in the Essex Police data. Postcode was therefore used as a second identifying factor, meaning that two cases in the joined dataset could not be found in the Essex Police dataset for this particular analysis, slightly restricting the conclusions that could be drawn.

Annex B - Methodology

User research methodology

The user research findings included in this report form part of a wider service design programme taking a design-led ‘test and learn’ approach to designing interventions to better meet victims’ support needs. Design-led research methods are:

- Often small scale which means engaging a small sample of people to conduct qualitative research, with a focus on depth of insight over breadth.
- Focused on obtaining actionable insights to inform tangible ideas or solutions to identified problems.
- Not designed to be representative of the views or experiences of every user group.

The purpose of including the findings from the user research in this report was to bring to life the lived experience behind the data, by contextualising and adding richness to the quantitative insights.

The user research methodology involved the following activities:

- **Three Frontline Staff Workshops** (a mix of ISVAs and counsellors from SWERCC) to understand the experiences of frontline staff and their reflections on how victim-survivors engaged with services.
- **Two Blueprinting Workshops** with operational staff to map out the ‘frontstage’ and ‘backstage’⁶ of service delivery, with the aim of understanding end-to-end service provision and to identify bottlenecks and opportunities for increased efficiency and/or effectiveness.

⁶ Frontstage and backstage describe what customers do and do not see. Customers interact with frontstage parts like counter staff, but not with backstage areas like support staff, systems, or partners working behind the scenes.

- **Sixteen Journey Mapping Sessions** with victims of rape, with a mix of experiences of victim support services and the CJS. Journey mapping is the process of visually depicting the steps, interactions, and emotions a user experiences across a service – this was used to understand victim-survivors' experiences of support services (SWERCC) and the CJS, and to identify met and unmet needs.
- Recruitment was split into two stages, or 'sprints'. All participants were required to have had a primary incident of rape, be aged over 18, and identify as female. In addition:
 - Sprint 1 participants were required to have reported to the police, with a mix of referral source⁷ and type of engagement⁸ with the CJS.
 - Sprint 2 participants were required to have had a service closure (with a mix of planned and unplanned closures)⁹ and a mix of whether they waited longer than two years to engage with support services or not.

A discussion was also held with Essex PCC and a representative from Essex Police Witness Care Unit to better understand the wider service ecosystem in which SWERCC services operate.

⁷ Referral to support services, split into those made by police and referrals not made by the police (for example, where the referral is made by a healthcare or social service provider).

⁸ In particular an effort was made to include with a 'No Further Action' outcome and those that decided to drop out of the CJS.

⁹ A SWERCC case ends in a 'planned closure' when a victim feels that they no longer need support and agree this with their support service workers. If a victim does not complete any service, the case will have an 'unexpected closure'.

Annex C - Coverage and Completeness of Essex Police Data

Breakdown of Incidents

The table below shows the breakdown by year of report in this data. This shows a relatively even spread of incidents each year between 2018 and 2023.

Table 3 - The number of unique incidents reported to Essex Police per year

Year of Report	Number of incidents	Proportion (%) of total incidents
2018	114,335	16%
2019	127,786	17%
2020	120,076	16%
2021	125,808	17%
2022	127,060	17%
2023	115,736	16%
2024	1	0%

Base = 730,802 unique incidents. Essex Police records.

The breakdown by crime type of these incidents is shown in the table below.

Table 4 - The number of unique incidents reported to Essex Police by crime type

Crime Type	Number of Unique Incidents	Percentage
Violence against the person	369,165	51%
Theft offences	172,203	24%
Public order offences	77,117	11%
Damage and arson offences	64,522	9%
Sexual offences	28,835	4%
Miscellaneous crimes against society	9,067	1%
Robbery	8,107	1%
Possess weapon offences	1,771	0%
Drug offences	13	0%
N/A	2	0%
Total	730,802	

Base = 730,802 unique incidents. Essex Police records.

Missing values

One of the core data quality dimensions in the Government Data Quality Framework¹⁰ is completeness of the variables. Specifically, “the degree to which records are present”. For a dataset to be deemed of good quality, a high completion rate for variables is important. Missing data can lead to analysis not being representative, skewed or contain biases.

The figure below shows the rate that each of the variables that were not used for victim identification (PCRN date of birth and gender) were missing.¹¹ This revealed considerable variation in the completeness of these fields. The only field that was well populated was “location district” (2% missing for sexual offences). Other fields were less complete (14-65% missing). The lack of data in the occupation field, for example, (60% missing for sexual offences) meant that no meaningful analysis could be conducted using that field. Analysis using ethnic appearance, with 19% missing for sexual offences would have to be interpreted with caution.¹² In contrast, SWERCC variables were more complete with the missing rate for common variables ranging between 6% and 8%¹³.

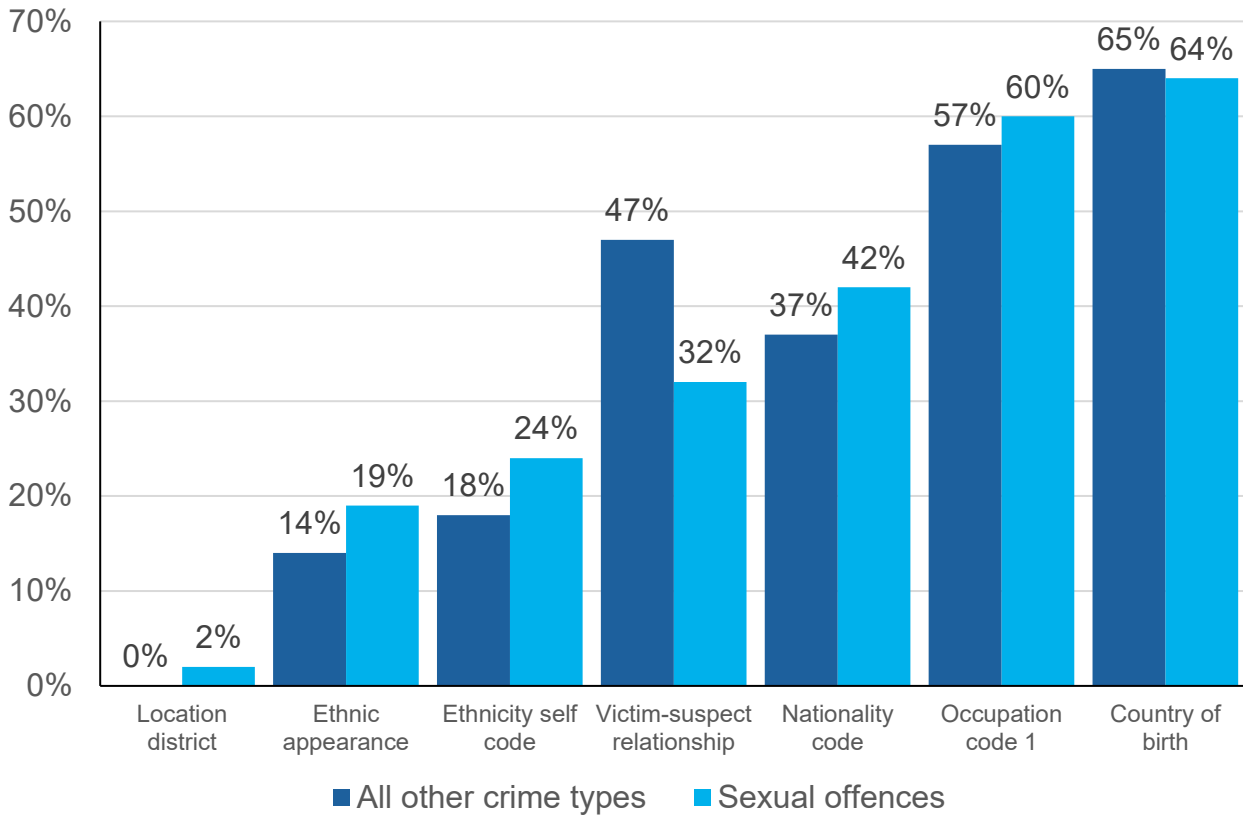
¹⁰ The Government Data Quality Framework: found at <https://www.gov.uk/government/publications/the-government-data-quality-framework/the-government-data-quality-framework#Data-Quality-Principles>

¹¹ Entered as “Unknown”, “Unspecified”, “Not Stated” or left blank.

¹² Analysis in this report used the field “Ethnicity Self Code” for police data and a self-defined “Ethnicity1” field for SWERCC.

¹³ Ethnicity had 8% missing, nationality had 6% of entries missing, primary perpetrator had 6% of entries missing. A full breakdown of the missing values in the SWERCC dataset can be found here: [Adult victim-survivor's journey through a sexual violence support service in Essex - GOV.UK](#).

Figure 1 - Rate of missing values (% of total identifiable victims) in the Essex Police data, by crime type



Base = 730,802 total identifiable victims.

The variation in completeness between sexual offences and all other crime types for the variable “victim-suspect relationship” may indicate that this field was considered an important piece of information for sexual offences, but less so for other crimes. This highlighted that there are external factors that can affect the quality of data collection.

The higher completion rate of variables in the SWERCC data showed that there can be different standards of data collection depending on factors such as differing ways of working and priorities. For example, SWERCC are focused on victims, whereas the police focus on the crime, the perpetrator and the victim.

Missing data values may be recorded elsewhere in the Essex Police data, such as other databases or free-text fields including victim impact statements. However, relying on this

information from free text fields makes analysis more resource intensive and prone to inaccuracies.

A high quality dataset can allow for more in-depth and robust analysis. Analysing data on fields with a high number of missing values may lead to inaccurate conclusions. While linking datasets can go some way to improving the quality of analysis, accurate linking and the derivation of unique insights relies on well-completed variables.

Annex D - Breakdown of Risks, Issues and Incident Impacts

This annex sets out the breakdown of selected risks, issues and impacts faced by victims who had been referred to SWERCC.

Table 5 - Comparison of mental health and domestic abuse risks faced by victims of CSA

Time Elapsed Before Seeking Support	Mental Health Risk	Domestic Abuse Risk
Less than a week	33%	17%
Less than a month	20%	NA
Less than 6 months	45%	27%
Less than 2 years	50%	12%
More than 2 years	34%	3%

Base – 4,436. All first time cases in the SWERCC data

Table 6 - Level of isolation as an incident impact faced by victims of all primary incidents

Time Elapsed Before Seeking Support	Loneliness, Isolation, or Relationship Loss
Less than a week	14%
Less than a month	17%
Less than 6 months	25%
Less than a year	28%
Less than 2 years	24%
More than 2 years	30%

Base – 4,436. All first time cases in the SWERCC data

Table 7 - Level of suicidal thoughts or attempts reported by rape victims

Time Elapsed Before Seeking Support	Suicidal Thoughts or Attempts
Less than a week	16%
Less than a month	21%
Less than 6 months	14%
Less than a year	20%
Less than 2 years	22%
More than 2 years	25%

Base – 4,436. All first time cases in the SWERCC data

Annex E – Variable Groupings

The tables below show the groupings used to simplify the analysis by reducing the number of entries for each variable.

Table 8 - Groupings for Essex Police variable "Victim-suspect relationship"

Victim-Suspect Relationship	Victim-Suspect Relationship Group
Appropriate adult	Known individual
Associate	Known individual
Business Associate	Known individual
Colleague	Known individual
Customer	Known individual
Doctor/Medical	Known individual
Employee	Known individual
Employer	Known individual
Friend	Known individual
Guardian	Known individual
Landlord	Known individual
Neighbour	Known individual
Overseen by	Known individual
Patient	Known individual
Responsible for	Known individual
Teacher	Known individual
Tenant	Known individual
Co-habitee	Non-relative living with
Lodger	Non-relative living with
Partner	Partner
Spouse	Partner
Relationship type refused	Relationship unknown
Relationship unknown	Relationship unknown
Aunt	Relative
Brother	Relative
Brother in law	Relative
Carer	Relative
Cousin	Relative
Daughter	Relative
Daughter in law	Relative
Dependent	Relative
Ex foster parent	Relative
Father	Relative

Father in law	Relative
Foster parent	Relative
Grandchild	Relative
Grandparent	Relative
Great grandchild	Relative
Great grandparent	Relative
Mother	Relative
Mother in law	Relative
Nephew	Relative
Niece	Relative
Relative	Relative
Relative other	Relative
Sister	Relative
Sister in law	Relative
Son	Relative
Son in law	Relative
Stepdaughter	Relative
Stepparent	Relative
Stepson	Relative
Uncle	Relative
Described person	Unknown individual
Stranger	Unknown individual

Table 9 - Groupings for SWERCC variable "Primary Perpetrator"

Primary Perpetrator	Primary Perpetrator Grouped
Ex boyfriend	Ex-Partner
Ex girlfriend	Ex-Partner
Ex partner	Ex-Partner
Ex partner/Husband	Ex-Partner
Ex partner/Wife	Ex-Partner
Group - all known	Group of Perpetrators
Group - all related	Group of Perpetrators
Group - all unknown	Group of Perpetrators
Group - some known	Group of Perpetrators
Group - some related	Group of Perpetrators
Acquaintance	Known Individual
Army worker	Known Individual
Babysitter	Known Individual
Boyfriends Father	Known Individual
Boyfriends brother	Known Individual
Boyfriends friend	Known Individual
Boyfriends uncle	Known Individual

Brothers friend	Known Individual
Caretaker	Known Individual
Careworker	Known Individual
Caseworker	Known Individual
Celebrity	Known Individual
Client	Known Individual
Doctor	Known Individual
Employer	Known Individual
Fathers friend	Known Individual
Friend	Known Individual
Friend of family	Known Individual
Friends boyfriend	Known Individual
Friends brother	Known Individual
Friends father	Known Individual
Friends grandfather	Known Individual
Friends uncle	Known Individual
Grandmothers boyfriend	Known Individual
Health professional	Known Individual
Housemate	Known Individual
Inmate	Known Individual
Internet acquaintance	Known Individual
Landlord	Known Individual
Lodger	Known Individual
Member of faith group	Known Individual
Mothers boyfriend	Known Individual
Mothers ex-boyfriend	Known Individual
Mothers friend	Known Individual
Mothers girlfriend	Known Individual
Neighbour	Known Individual
Nurse	Known Individual
Patient	Known Individual
Person in position of trust	Known Individual
Pimp	Known Individual
Police officer	Known Individual
Priest	Known Individual
Prison officer	Known Individual
Professional	Known Individual
Punter	Known Individual
Pupil at school	Known Individual
Sisters boyfriend	Known Individual
Sisters friend	Known Individual
Social worker	Known Individual
Substance supplier	Known Individual
Teacher	Known Individual
Work colleague	Known Individual

Youth worker	Known Individual
NA	NA
Not disclosed	Not disclosed
Boyfriend	Partner
Girlfriend	Partner
Partner	Partner
Partner/Husband	Partner
Partner/Wife	Partner
Adopted brother	Relative
Adoptive father	Relative
Adoptive mother	Relative
Aunt	Relative
Brother	Relative
Brother-in-law	Relative
Carer	Relative
Cousin	Relative
Ex Brother-in-law	Relative
Father	Relative
Father-in-law	Relative
Female relative	Relative
Foster Parents relative	Relative
Foster brother	Relative
Foster carer	Relative
Foster parent	Relative
Grandfather	Relative
Grandmother	Relative
Half brother	Relative
Mother	Relative
Other family member	Relative
Other male relative	Relative
Sister	Relative
Son	Relative
Son-in-law	Relative
Step brother	Relative
Step father	Relative
Step grandfather	Relative
Step grandmother	Relative
Step mother	Relative
Step sister	Relative
Step son	Relative
Uncle	Relative
Gang Member	Unknown Individual
Not known - incapacitated	Unknown Individual
Sex trafficker	Unknown Individual
Stranger 1	Unknown Individual

Stranger 2	Unknown Individual
------------	--------------------

Table 10 - Groupings used for SWERCC ethnicity variable

Ethnicity Self Code	Ethnicity Self Code Group
White British	White
Any Other White Background	White
White Irish	White
Any Other Black Background	Any other ethnic group
Any Other Asian Background	Any other ethnic group
Indian	Any other ethnic group
Caribbean	Any other ethnic group
Pakistani	Any other ethnic group
Any Other Ethnic Group	Any other ethnic group
Bangladeshi	Any other ethnic group
Chinese	Any other ethnic group
Gypsy or Irish Traveller	Any other ethnic group
Arab	Any other ethnic group
White And Black Caribbean	Any other ethnic group
Any Other Mixed Background	Any other ethnic group
White And Asian	Any other ethnic group
White And Black African	Any other ethnic group
African	Any other ethnic group
Any Other Ethnic Group	Any other ethnic group

[To add further appendices to the report, add a heading (with the style 'Heading - appendix reference') and give it the appropriate letter, add a second heading (with the style 'Heading - appendix name'), then type the text of the appendix.]