

# Better Outcomes through Linked Data: Local Authority Homelessness and Rough Sleeping Data Linking

# Contents

Contents .....	2
Foreword.....	<b>Error! Bookmark not defined.</b>
Executive Summary.....	3
1. Introduction .....	8
2. Datasets and Methodology .....	10
3. Results .....	12
4. Conclusion and Recommendations .....	20
5. Further Information .....	<b>Error! Bookmark not defined.</b>
Annex.....	23

## Executive Summary

At present, both statutory homelessness and rough sleeping data are collected by local authorities and submitted to MHCLG. However, the former contains individual case level detail whilst the latter does not, making it difficult to understand the overlaps between the two and develop a deeper understanding of individuals that are experiencing homelessness and rough sleeping.

The Homelessness Pilot Team, as part of the [Better Outcomes through Linked Data \(BOLD\) programme](#), worked with Cheshire West and Chester Council (CWAC), to link these two datasets and analyse the results. The research explored user journeys between 1 January 2023 and 31 December 2023 in this local authority.. These findings are not representative of all local authorities and relate only to Cheshire West and Chester Council.

The analysis found that of the 437 individuals recorded as sleeping rough by outreach services in CWAC, 370 people (85%) were also found in the homelessness dataset.

- Individuals recorded as sleeping rough were likely to be male, with a male to female ratio of 82:18, consistent with the 2023 [national data for those rough sleeping](#) at 82:15.
- 57% of these individuals were UK nationals, lower than the national average of 62%, although 16% did not state their nationality (versus 11% nationwide).
- 288 people (78%) of the 370 found in both datasets had approached Housing Options (the CWAC team who provide housing advice and process homelessness applications) before being seen recorded as sleeping rough in 2023.

The 288 individuals who made homelessness applications were far more likely to be seen sleeping rough than those who did not. Housing Options determined that most of these individuals were either: not in priority need, or; ‘intentionally homeless’. Consequently, most individuals sleeping rough were not entitled to ongoing main duty homelessness support.

Even those determined to be in priority need by Housing Options – or, those entitled to a main duty – were recorded by outreach services sleeping rough on average four times more than those who did not contact Housing Options. Whilst those recorded as not having a priority need were ten times more likely to be sleeping rough, and those recorded as intentionally homeless were sixteen times more likely. This reflects the difficulty facing support service teams in finding suitable accommodation for those in precarious housing situations.

The 82 individuals found only in the rough sleeping dataset included a high number of individuals with support needs, who were more likely to avoid interacting with Housing Options (unless their support need was due to an offending history which incurs an automatic duty to refer. This aligns with the high proportion of people sleeping rough upon leaving prison or a bail hostel in the dataset.)

Ultimately this research provides evidence of the value of linking homelessness and rough sleeping datasets. By doing so, it could potentially enable targeted future commissioning of accommodation and support services, thereby increasing prevention. It is however worth noting the limitations (see full report) of this project, and the resource-heavy nature means it may not be suitable for all local authorities, nor should it be a replacement for central data linking projects.

## Background

The [2017 Homelessness Reduction Act](#) (HRA) reformed England's homelessness legislation by placing duties on local authorities (LAs) to intervene at earlier stages to prevent homelessness in their areas. MHCLG compiles [statutory homelessness statistics](#) from Homelessness Case Level Information Collection (H-CLIC) data provided by local authorities in England to report on activities undertaken under statutory homelessness duties.

Rough sleeping data, while collected by some Local Authorities at an individual level, is summarised and only the total numbers are sent to the department. This means that while some demographic analysis can be conducted at a departmental level, it is not informative.

Preliminary analysis of the nationwide data submitted to MHCLG showed discrepancies between the number of people recorded in local authorities' rough sleeping datasets and H-CLIC as sleeping rough. Full overlaps can never be expected, as not everyone sleeping rough will have applied for housing support in their local councils and not all individuals presenting as homeless will not end up sleeping rough. However initial investigations and anecdotal evidence from outreach workers and local government representatives suggested that this was not the full picture.

The BOLD project aims to fill this gap and help to develop our understanding of the links between homelessness and rough sleeping. Considering the above, MHCLG needed a partner local authority. The BOLD programme provided a grant to enable a Cheshire West and Chester Council to link their homelessness and rough sleeping data and analyse the resulting dataset.

This research was therefore commissioned with the following objectives:

- To explore user journeys through statutory homelessness and rough sleeping and identify system gaps in preventing rough sleeping.
- To identify whether particular cohorts of people sleeping rough are systematically included or excluded from homelessness data.
- To highlight risk factors within the homelessness data that are rough sleeping indicators, potentially informing prevention services provision.
- To empower the local authority to demonstrate responsible data stewardship and maximise efficiency of service delivery.

## Summary of Key Findings

A summary of the analytical findings by theme are provided below. Further details of the results are provided in Section 3 of this report. The definitions of key terminology can be found at the end of the Executive Summary.

### Overlaps between rough sleeping and homelessness data 1 January 2023 to 31 December 2023

- 4,860 households (3,514 (72%) single people) approached Housing Options for housing advice and homelessness support and were subsequently recorded in the homelessness dataset.
  - Of these, 2,000 households (1,286 (64%) single people) made a homelessness application.
- 437 people were found on daily outreach patrols and therefore recorded in the rough sleeping dataset.
  - The vast majority (370 people or 85%) of people seen sleeping rough made a homelessness application at some point during 2023.
- 67 people (15%) were found only within the rough sleeping dataset in 2023.

All of these individuals were single people (i.e. not including households with children).

Of the 252 people that made a homelessness application the number of people that went on to make a second homelessness application during the 12-month period was less than 10, but many people repeatedly approached Housing Options for homelessness support (not necessarily a new application) during the 12 months. For example:

- 22 people were found to be in priority need and not intentionally homeless. Of these, 13 people appear in the homelessness dataset again and were seen by outreach on average another 4 times.

- 28 people were found to be in priority need but intentionally homeless. Of these 16 people appear in the homelessness dataset again and were seen by outreach on average another 16 times.
- 117 people were not found to be in priority need, and of these 70 people appear in the homelessness dataset again and were seen by outreach on average another 10 times.
- 15 people had their applications withdrawn (no decision made), of which 12 people appear in the homelessness dataset again and were seen by outreach on average another 2 times.

## Definitions

**Data linking:** The process of identifying the same record in two or more datasets, with a view to create a more valuable linked dataset.

**Duty to Refer:** Specific public authorities (such as prisons) must notify local authorities if they are aware of a person being homeless or at risk of homelessness

**Everyone In:** A government scheme launched during the pandemic to help those sleeping rough or at risk of sleeping rough.

**Fuzzy matching:** A method for linking data which involves identifying similar, but not identical elements in datasets.

**Homelessness:** The definition for statutory homelessness is broader than the definition for rough sleeping and also includes households in insecure or unsuitable housing.

**Housing Options:** The team at CWAC responsible for providing housing advice and recording homelessness applications and data.

**Prevention Duty:** This is the duty that a local authority has towards someone at risk of becoming homeless to prevent them from becoming homeless.

**Priority need:** A priority need is one of several special reasons why a local authority must provide someone with more help (such as emergency housing) if they are homeless or facing homelessness.

**Relief duty:** This is the duty a local authority has towards a person who is already homeless, to help them secure accommodation for at least six months.

**Rough sleeping:** Also referred to as “street homelessness”, this refers to sleeping, about to bed down (sitting on/in or standing next to their bedding) or actually being bedded down in the open air (such as on the streets, in tents, doorways, parks, bus shelters or encampments), or sleeping in buildings or other places not designed for habitation (such as stairwells, barns, sheds, car parks, cars, derelict boats, stations, or ‘bashes’).

**Section 21:** The notice under the Housing Act 1988 section 21, which a landlord must give to their tenant to begin the process to take possession of a property let on an assured shorthold tenancy without providing a reason for wanting to retake possession.

**Single homelessness:** Homelessness experienced by people or households with no dependent children.

**Splink:** A Python package for probabilistic record data linkage.

**Stage 1 accommodation (Rough Sleeping Pathway):** Cheshire West and Chester Council provides a wide range of homelessness support services and accommodation for people sleeping rough to support them through the Rough Sleeping Pathway.

The Rough Sleeping Pathway is a three-staged approach:

Stage 1) emergency accommodation with homelessness and support needs assessment for people rough sleeping

Stage 2) short term supported accommodation provided by a number of landlords

Stage 3) move on into a permanent home in social or private housing

**Supported accommodation:** Accommodation provided alongside support, supervision or care to help people live as independently.

# 1. Introduction

## 1.1 Better Outcomes through Linked Data (BOLD)

The BOLD programme is led by Ministry of Justice in partnership with the Ministry for Housing, Communities and Local Government (MHCLG), Department of Health and Social Care, Welsh Government and Public Health Wales. It was created to demonstrate how people with complex needs can be better supported by linking and improving the government data held on them in a safe and secure way. BOLD has initially focused on four pilot areas: reducing homelessness, substance misuse, re-offending and supporting victims of crime.

## 1.2 Statutory Homelessness

The [Homelessness Reduction Act 2017](#) reformed England's homelessness legislation by placing duties on local authorities to intervene at earlier stages to prevent homelessness in their areas. This includes providing support if a household is threatened with homelessness within 56 days, known as a prevention duty, and to provide support to households that are homeless, known as a relief duty.

The legislative changes have been reflected in an enhanced Homelessness Case Level Information Collection (H-CLIC) data specification for local authorities to follow since April 2018.

MHCLG produces statistics from data returns by local authorities to monitor levels of statutory homelessness across England, and to report on the activities carried out by local authorities to meet their statutory homelessness duties.

The data returns indicate whether a household has reported a support need of a history of repeat homelessness, based on the information and understanding of the local authority and the assessor. However, the information collected does not specify the previous number of applications made, which household members had been affected or verify this against previous occurrences in H-CLIC records. The reference numbers of the applicants' previous homelessness case may also be recorded in H-CLIC, but this is optional, and usage may vary between local authorities. These sets of information may still be useful to the authority in supporting those households but may not always be a complete reflection of any repeat homeless experiences.

## 1.3 Rough Sleeping

Rough sleeping data, while collected by local authorities at an individual level, is summarised and only the total numbers are sent to the department. This means that while some demographic analysis can be conducted at a departmental level, it is not possible for MHCLG alone to conduct any analysis.

Preliminary analysis of the nationwide rough sleeping data highlighted discrepancies between the number of people in local authorities' rough sleeping datasets and the number of people recorded in H-CLIC as sleeping rough. Full overlaps between the two datasets can never be expected, as not everyone sleeping rough will have applied for housing support in their local councils and the majority of people presenting as homeless will not end up sleeping rough. However initial investigations and anecdotal evidence from outreach workers and local government representatives suggested that this was not the full picture.

## 1.4 Homelessness and Rough Sleeping Data Linking

To understand the discrepancies and overlaps between homelessness and rough sleeping data, the MHCLG BOLD programme provided a grant to enable Cheshire West and Chester Council to link their homelessness and rough sleeping data and analyse the resulting dataset.

## 1.5 Aims and Objectives

This research was commissioned with the following objectives:

- To explore user journeys through statutory homelessness and rough sleeping and identify system gaps in preventing rough sleeping.
- To identify whether particular cohorts of people sleeping rough are systematically included or excluded from homelessness data.
- To highlight risk factors within the homelessness data that are rough sleeping indicators, potentially informing prevention services provision.
- To empower the local authority to demonstrate responsible data stewardship and maximise efficiency of service delivery.

## 2. Datasets and Methodology

### 2.1 Homelessness Dataset

The homelessness dataset is within a system called Housing Jigsaw MRI (referred to as “MRI”) which is used by the Council’s Housing Options team to record housing advice and homelessness activities under Part 7 of the Housing Act 1996.

### 2.2 Rough Sleeping Dataset

The rough sleeping dataset is an Excel spreadsheet completed on a daily basis by the commissioned Forfutures outreach team to record information about people sleeping rough. The spreadsheet was introduced to help complete and submit the [Rough Sleeping Monthly Survey](#)<sup>1</sup>.

### 2.3 Methodology

#### 2.3.1 Data Linking

The MRI system uses two unique identifiers to identify people and cases, namely *Customer ID* and *Case ID*. In comparison the people in the rough sleeping dataset are identified by their names and date of birth.

In order to match records from both sources, exact matches were paired on first name, last name, and date of birth. This yielded a match for 230 people. Due to the data inconsistencies in names and date of birth between the two datasets, fuzzy matching (a method for linking data which involves identifying similar, but not identical elements in datasets) was then employed to improve the matching rate.

By the end of this process 370 single people out of the 437 in the rough sleeping dataset were identified as having a positive match in the MRI system. These are referred to as being recorded in both datasets.

#### 2.3.2 Local Authority Selection

To find a suitable delivery partner for this project, MHCLG created a shortlist of 20 local councils by filtering for those authorities with a suitable number of people recorded as sleeping rough in both the homelessness and rough sleeping datasets. The shortlist was assessed by experts in local government delivery and an ‘Expression of Interest’ form was

---

<sup>1</sup> Initially collected for the ‘Everyone In’ scheme during the pandemic, the Monthly Rough Sleeping Survey helps monitor local authority performance and accountability towards ending rough sleeping.

circulated. Local councils were asked to rate the quality of their data and to provide details about the variables in each dataset, along with how far back the data extended and if the data was well documented, with the final selection step involving scoring project plans for deliverability and value for money.

## 2.4 Limitations of this Study

- CWAC's IT system for recording H-CLIC changed in December 2022, with the MRI platform replacing Locata. Data from Locata was not accessible for the research; research analysis therefore covers the 12 months from 1 January 2023 to 31 December 2023. This may have implications on the findings as individuals who appeared in the homelessness data before January 2023, but not since then, would not be possible to match.
- MRI is a complex system with a vast amount of reporting fields, some of which are incorrectly populated, or left incomplete at the triage stage. This can lead to incorrect information within the dataset.
- Within both datasets there are common issues such as names being misspelt, or middle names used as forenames in one dataset but not the other. Dates of birth sometimes appeared in the USA format. Both required manual matching and amending.
- Instances of duplication through multiple case IDs in the MRI system. While there were 381 unique customer IDs in the MRI data for people who were also identified in the rough sleeping dataset, manual checks revealed that 11 of these were likely to be duplicates, as they were recorded as having the same demographics, contact details and national insurance numbers.
- The quality of the reports themselves, especially from MRI which has proven to be non-malleable in terms of generating custom reports with significant blank values leading to ambiguity.
- The inherent limitations of the data must also be considered. The hidden nature of rough sleeping, especially in certain demographics such as women, means that there will be individuals recorded rough sleeping who do not interact with outreach teams. Others may interact but do not engage or provide personal details. Similarly, for legal reasons or lack of awareness around available support, many people experiencing homelessness will not approach Housing Options. There is a risk that this might lead to lower representation of specific groups such as non-UK nationals.

## 3. Results

### 3.1 Frequency of and length of time between a person appearing in both datasets

#### 3.1.1 From Sleeping Rough to Housing Options

82 people are recorded in the rough sleeping dataset before being recorded in the Housing Options homelessness dataset (22% of all those in both datasets).

The length of time between being recorded in the rough sleeping dataset and then the homelessness dataset ranges from 1 to 322 days, with an average of 62 days. Some people (fewer than 10 over the research period) chose not to engage with any support, leading to an inflated mean. (As the research covered the year 1 January to 31st December 2023 there is a maximum possible value of 365 days.)

14 people are first recorded in the rough sleeping dataset, then in the homelessness dataset, before again being in the rough sleeping dataset.

#### 3.1.2 From Housing Options to Sleeping Rough

288 (78%) people are recorded in the homelessness dataset before being recorded in the rough sleeping dataset.

The average length of time between being recorded in the homelessness dataset and then the rough sleeping dataset is 60 days (ranging from 1 to 315 days). The maximum possible value is 365 days.

#### 3.1.3 From Housing Options to Sleeping Rough to Housing Options

154 people are first recorded in the homelessness dataset, then the rough sleeping dataset, before again recorded in the homelessness dataset (42%). This means that a new case was opened by Housing Options.

In their second (or further) homelessness applications, only 97 of the 154 people (63%) met the minimum threshold required to qualify for support.

## 3.2 Other Individual Journeys

33 people are recorded in the homelessness dataset and the rough sleeping dataset at the same time.

67 people are recorded in the rough sleeping dataset but are not recorded in the homelessness dataset.

11 people are recorded in the rough sleeping dataset before they are recorded in the homelessness dataset but are not recorded in the latter as having a history of rough sleeping in the support needs section of MRI.

Separate to the 370 people recorded in both datasets, a further 120 people who approached Housing Options for assistance stated they were sleeping rough but were not recorded in the rough sleeping dataset during the 12 months.

The demographics of this cohort are:

- Rounded to the nearest 10:  
Male: 90 (75%), Female 30 (25%),
- UK nationals 44 (37%), EEA National fewer than 10, Non-EEA or UK National fewer than 10, blank or not stated 74 (62%)
- 18-30 years old 43 (36%), 31-40 years old 40 (33%), 41-50 years old 20 (17%), 51+ (collapsed categories of 51-60 and 60+) 17 (14%)
- When assessed, fewer than 10 people were owed a prevention duty and 40 people were owed a relief duty (33%)

### 3.2.1 People Sleeping Rough with Available Accommodation

At the time of the research every new person sleeping rough was made an offer of emergency accommodation under *Everyone In*. A minority of people sleeping rough had accommodation available to them but were still recorded sleeping rough by outreach and appear in the rough sleeping dataset.

Out of the 437 people recorded in the rough sleeping dataset, 30 (7%) had accommodation available to them at the time of being seen sleeping rough, such as supported housing or their own tenancy.

### 3.2.2 Local Connection to the Borough

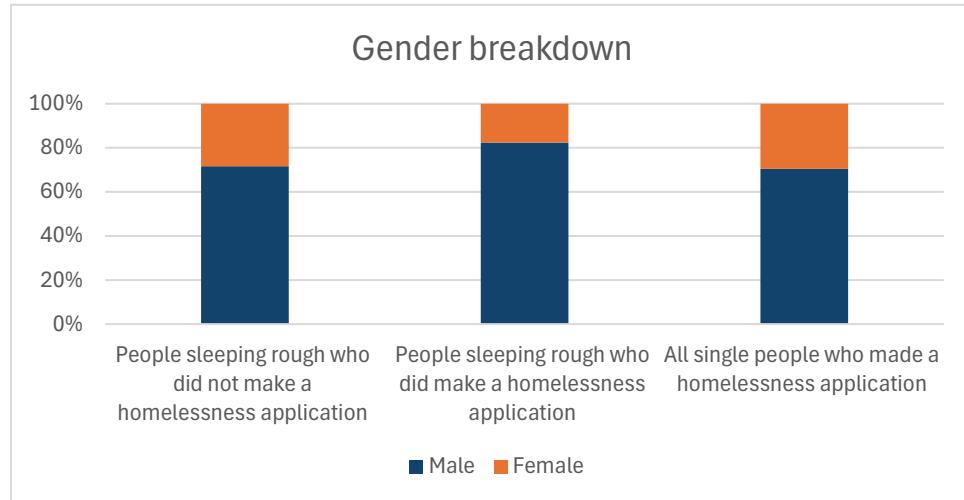
The rough sleeping dataset records whether people are from Cheshire West and Chester or another area. Of the 437 people in the rough sleeping dataset 375 (86%) are from the area. 12 people (3%) are from Wrexham and a small number are from Flintshire. 40 people (9%) are from 23 other areas and fewer than 10 people are recorded as unknown.

Of the 67 people who were not recorded in the homelessness dataset, 25 people (37%) are from other areas and do not have a local connection with Cheshire West and Chester.

## 3.3 Differences Between the Cohorts

### 3.3.1 Gender

Figure 3.1



The above graph shows the gender breakdown of those who made a homelessness application. *\*While there were people recorded as transgender in both datasets, these figures are lower than 10 in any given cohort and are therefore not included in the graph.*

70% of single people who made a homelessness application were male, and 82% of those who both slept rough and made a homelessness application were male. When compared to the general population of CWAC (roughly a 50-50 split<sup>2</sup>), this demonstrates a marked over-representation of males in both the single homeless and the rough sleeping population.

### 3.3.2 Age

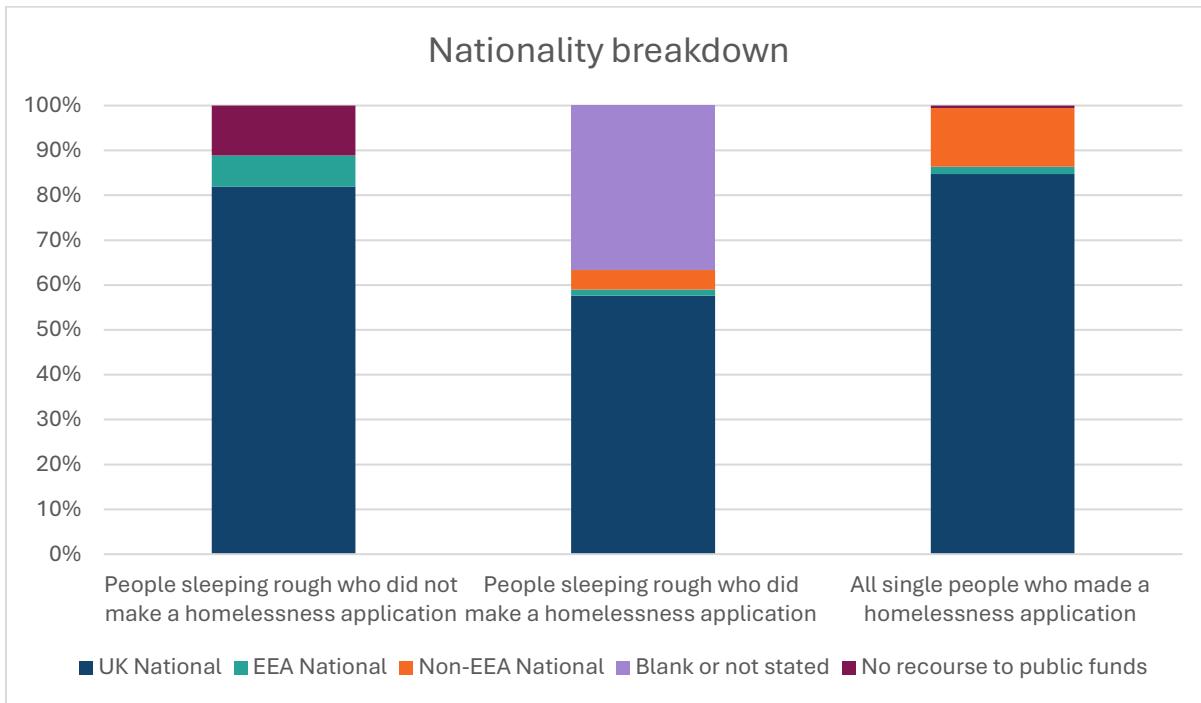
Age	People sleeping rough who did not make a	People sleeping rough who did make a homelessness application	All single people who made a homelessness application

<sup>2</sup> [Estimates of the population for England and Wales - Office for National Statistics \(ons.gov.uk\)](https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/estimatesofthepopulationforenglandandwales/sep2024) (accessed 9 September 2024)

	<b>homelessness application</b>		
<b>Totals</b>	<b>67</b>	<b>370</b>	<b>1,286</b>
16 - 17	Fewer than 10	Fewer than 10	Fewer than 10
18 - 30	21 (31%)	101 (27%)	439 (34%)
31 - 40	20 (30%)	125 (34%)	337 (26%)
41 - 50	17 (25%)	82 (22%)	233 (18%)
51 - 60	Fewer than 10	48 (13%)	159 (12%)
61+	Fewer than 10	14 (4%)	109 (8%)

The majority of people over 60 who were sleeping rough made a homelessness application, in comparison to young age groups such as for people aged between 18 and 50. Amongst the latter age group, individuals were almost equally likely to have not made a homelessness application as to have made one. It is important to note that the 61+ cohort is much smaller overall.

### 3.3.3 Nationality



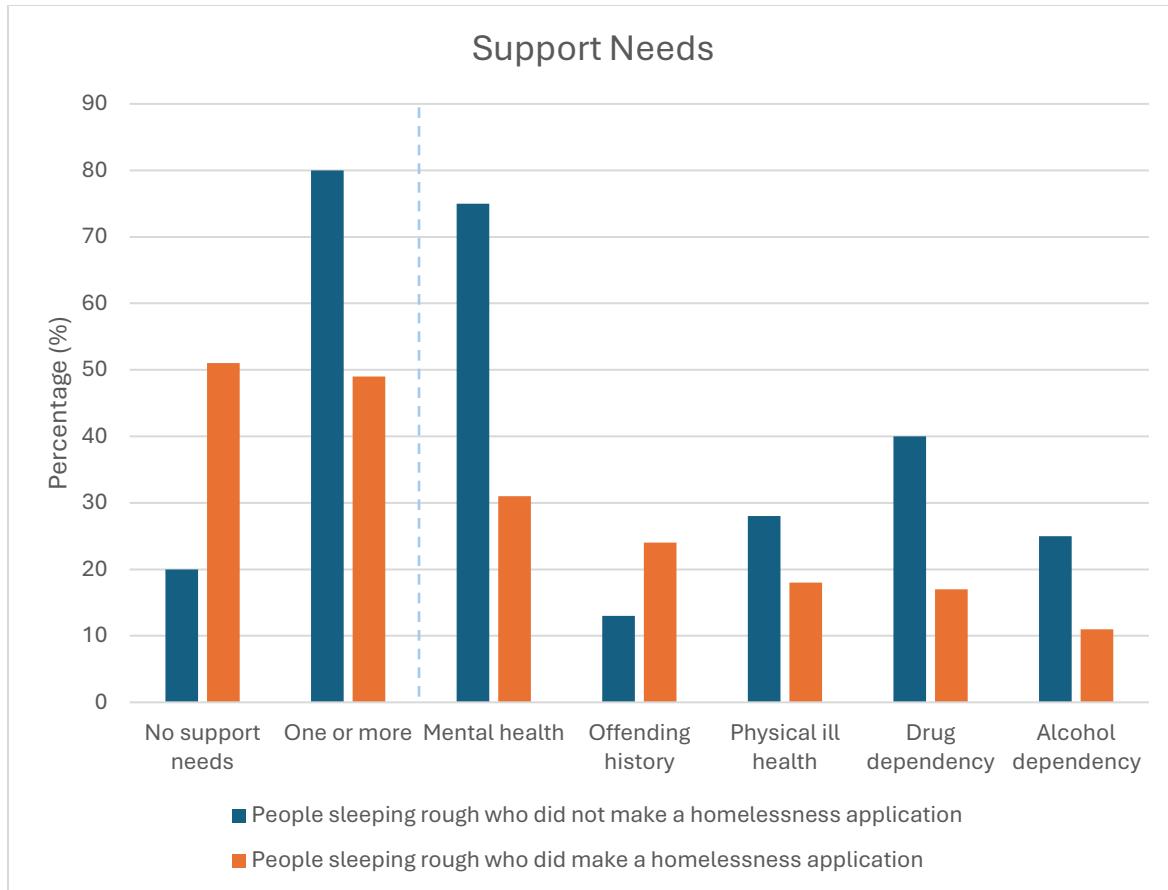
The above graph shows the nationality breakdown of those who made a homelessness application.

Comparisons on the effects of nationality are difficult as the nationality field was blank or 'not stated' for almost 60% of all cases in the homelessness dataset and in 36% of linked cases. While the data suggests that people sleeping rough without access to public funds are less likely to approach Housing Options, the numbers are too small to confidently draw conclusions.

A higher percentage of those who sleep rough and make a homelessness application have a nationality not stated in either system.

### 3.3.4 Support Needs

Support needs are not recorded in the same way in both datasets. For example, the rough sleeping dataset does not record learning disabilities or offending history as support needs and physical health is recorded separately.



The above graph shows individuals with support needs who made a homelessness application.

Compared to those with no support needs, those with one or more support needs appear less likely to have made a homelessness application. The only support need for which this pattern was not true was “offending history”, which could be due to people with offending histories coming to Housing Options through referrals from prison and probation services, who have a ‘Duty to Refer’ those at risk of homelessness.

The difference in whether or not people were likely to have made a homelessness application was especially pronounced for people who had mental health issues or substance or alcohol dependencies. For each of these support needs, people sleeping rough were at least twice as likely to have not made a homelessness application in the research year as to have made one.

This highlights the importance of multi-disciplinary teams in encouraging people sleeping rough to engage with formal support services.

### 3.3.5 Duty to Refer

Of the 370 people found in both datasets, 88 (24%) had been recorded under a Duty to Refer from a public body. There were also 45 cases recorded under Duty to Refer but from an agency (not public body)<sup>3</sup>.

<b>Duty to Refer</b>	<b>People who were seen sleeping who did make a homelessness application (370)</b>	<b>Total number of households who approached Housing Options (4,860)</b>
Duty to Refer from Public Body	88 (24%)	444 (9%)
Duty to Refer from an agency	45 (12%)	396 (8%)
Not referred	237 (64%)	4,020 (83%)

The table above shows that people referred from public bodies rather than agencies are more likely to be seen sleeping rough. (It is important to note that the figures provided in the right-hand column of the table include all households, whereas other analysis in this report refers only to single applicants.)

### 3.3.7 Homelessness Applications

Of the 370 people in both datasets, 252 people (68%) made a homelessness application (they met the 'reason to believe' threshold and had a 'needs and circumstance assessment' taken).

118 people (32%) did not make a homelessness application but were given housing advice.

---

<sup>3</sup> See Glossary of terms for an explanation of Duty to Refer

### 3.4 Factors Associated with Rough Sleeping

Category	Sub-category	Total number of people in the homelessness data (4,860)	Number of people in the homelessness data who went on to be found sleep rough	Rate of people in this category that went on to be found sleep rough
<b>Gender</b>	Male	2544	235	9%
	Female	2283	53	2%
	Not known	<10	<10	13%
	Transgender	25	<10	Low value suppressed
<b>Nationality</b>	UK National	1703	267	16%
	EEA National	49	13	27%
	Non-EEA National	263	<10	3%
<b>Age</b>	18-30	1797	81	5%
	31-40	1432	100	7%
	41-50	838	60	7%
	51-60	467	34	7%
	60+	276	13	5%
<b>Reasons for approaching Housing Options</b>	Rough sleeping	283	108	38%
	Asked to leave by family	731	47	6%
	Leaving prison	109	33	30%
	Friends no longer willing to accommodate	203	25	12%
	Leaving bail hostel	60	22	37%
	Sofa surfing	307	21	7%
	Bail conditions – unable to return home	56	16	29%
	Asylum Seeker			
	accommodation ended	161	12	7%
	Relationship breakdown	322	12	4%
	Eviction from supported housing	44	11	25%
	Section 21 notice <sup>4</sup>	430	10	2%

<sup>4</sup> See Glossary of terms for a definition of Section 21 notices.

The above table shows the risk of rough sleeping associated with demographic and situational characteristics, by comparing the cohort of people who approached Housing Options and were then seen sleeping rough, versus those who approached Housing Options but were not seen sleeping rough within the year. (Whilst this shows an association between these characteristics and sleeping rough, the former do not necessarily cause rough sleeping.)

Men are at higher risk of sleeping rough than women, although the risk may be even higher for those with an unknown gender. This analysis should be used with caution given the low numbers of people in this category.

In spite of there being far more UK nationals both approaching Housing Options and sleeping rough, the risk of ending up sleeping rough was almost twice as high for EEA nationals (27% versus 16%). There is a chance that this could be because of the reduced likelihood of the latter to be eligible for statutory support. Again, this should be treated with caution, given the low numbers of people in the datasets who were EEA nationals.

Age did not appear to play a large role in someone's likelihood of not sleeping rough after approaching Housing Options, with similar risk levels across all age groups.

## 4. Conclusion and Recommendations

### 4.1 Feasibility of Data Linking at a Local Authority Level

While CWAC was the data controller for both datasets, this did not mean that data was easily accessible, as demonstrated by the need to limit data to the 1-year timeframe. This was in part due to the change in software suppliers, and the limited ways in which data was able to be downloaded. In cases where other local authorities may also be using two or more software systems for their homelessness and rough sleeping records, or software systems from which data is not easily downloadable or transferable, resourcing will need to be dedicated to find solutions to this.

The procedure for linking data involved a non-trivial degree of manual data cleaning, which is likely to be the case in other local authorities if there are not strict guidelines in place with regards to formatting of dates and names. Some of these local authorities might not have the necessary resources, especially if the research is not limited to a 1-year timeframe as in this project

## 4.2 Recommendations for Local Authorities Conducting Linking Projects

This project has highlighted various areas of improvement for CWAC's data collections and have given CWAC insights into risk factors and journeys of people experiencing homelessness and rough sleeping in their area. It could be worthwhile other authorities replicating this.

On data stewardship and data structures in general:

- Creating a unique identifier (such as a client ID) for each person
- Establishing processes early on for searching people when recording new interactions, to prevent duplicates (e.g. always checking dates of birth)
- Establishing guidelines early on with regard to formats of fields such as names and dates
- Building in validation checks into software systems (e.g. only allowing dates to be entered in a certain format; not allowing the user to proceed without completing necessary fields)
- Establishing a quality assurance process
  - Including running regular deduplication processes
- Keeping an up-to-date data dictionary
- Requesting the software provider:
  - Pre-populate fields when adding a new record for the same person – this will minimise the need for repetition on the part of the person entering data, therefore resulting in a lower risk of typing errors and a lower risk of duplication
  - Allow direct downloads of data into a common format such as CSV
  - Accommodate data transfers or downloads when switching provider
- Ensuring there is a process to account for any gaps during transition periods when switching provider

- If there are separate homelessness and rough sleeping teams, communicating all proposed changes in data fields to one another
- Keeping data for as long as it has the potential to be useful, while adhering to data protection rules. For example, it may be beneficial to retain records for several years if it can help to identify when people are returning to sleeping rough after a long period of no contact.
- Better recording practices like keeping a singular universal file in the Cloud/SharePoint accessible to all users and adding dropdown and data validation features in the majority of the columns would reduce the inconsistencies in the data
- Choosing a suitable linking method depending on the completeness of the data and the project aims. Several probability thresholds for probabilistic matching may need to be tested
- Recognising the limitations of the linking, including that the analysis will be limited to contacts that occurred within a local area (i.e., it might not provide insight into people who have previously slept rough in other parts of the country).

# Annexes

## Annex A: Full List of Research Questions

### Exploring user journeys

- What journeys do users take through the homelessness relief system?
- What is the frequency of a person first appearing in the homelessness dataset and then in a rough sleeping dataset?
- What is the frequency of a person first appearing in the rough sleeping dataset and then in the homelessness dataset?
- What is the frequency of a person first appearing in the homelessness dataset, then in the rough sleeping dataset, then again in the homelessness dataset?
- What is the frequency of a person first appearing in the rough sleeping dataset, then in the homelessness dataset, then in the rough sleeping dataset again?
- What is the frequency of people appearing in the homelessness dataset and the rough sleeping dataset at the same time?
- What is the average length of time between each appearance in either dataset in each of the instances above?
- What is the frequency of people appearing multiple times in either dataset?
- How are all of the above affected by demographics?
- What are the outcomes for those recorded as “Lost contact” in the homelessness dataset?
- What is the frequency of people who are shown as “lost contact” in the homelessness dataset later appearing in the rough sleeping dataset?
- How many people are on the streets in spite of having been allocated accommodation?
- What is the frequency of people recorded in the homelessness dataset as having been placed into accommodation then appearing in the rough sleeping dataset?
- How is the above affected by demographics?
- How many people are on the streets in spite of being eligible for relief duties?
- How many people appear in the rough sleeping dataset who have access to public funds?
- How is the above affected by demographics?
- What gaps are there that could be addressed to prevent people from ending up on the streets?

- What are the risk factors for someone appearing initially in the homelessness dataset (not recorded as sleeping rough) and later appearing in the rough sleeping dataset? This could be based on (but not exclusive to):
  - Date of first application (I.e. does how quickly the case was dealt with have an impact on the likelihood of ending up sleeping rough?)
  - Gender
  - Type of applicant (e.g. single, adult with dependent children)
  - Number of children
  - Nationality (Reasons for eligibility for assistance and Nationality)
  - Ethnicity
  - Sexual orientation
  - Age
  - Employment status
  - Benefits towards housing costs
  - Accommodation at time of application
  - Last settled accommodation
  - Main reason for loss of settled home
  - Reason for loss of assured hold tenancy
  - Reason for loss of social rented tenancy
  - Reason for loss of supported housing
  - Referral agency
  - Support needs
    - Total number of support needs
    - For each support need
    - Whether engaged in each type of assistance for support needs

#### Coherence between systems

- How many people who are sleeping rough are captured in the rough sleeping dataset but not in the homelessness dataset, or vice-versa?
- How many people appear in both datasets at the same time but are not recorded in the homelessness dataset as sleeping rough on approach?
- How many people appear in the rough sleeping dataset before they appear in the homelessness dataset but are not recorded in the homelessness dataset as having a history of rough sleeping?
- How many people appear in the homelessness dataset as rough sleeping on approach but do not appear in the rough sleeping dataset?

- How many people appear in the homelessness dataset as having a history of rough sleeping but do not appear in the rough sleeping dataset?
- How many people appear in the homelessness dataset as having slept rough more than 1 year ago (Current accommodation = No fixed abode – Not rough sleeping on approach but has slept rough at least once in the last year) but do not appear in the rough sleeping dataset?
- How many people appear in the homelessness dataset as having slept rough more than 1 year ago (Current accommodation = No fixed abode – Not rough sleeping on approach but has slept rough at least once in the last year) but appear in the rough sleeping dataset more recently?
- For the above, is there evidence that they have connections with another local authority?
- For those who appear more than once in the homelessness dataset or the rough sleeping dataset, how often are they recorded in the homelessness dataset as having a history of repeat homelessness?
- What is the frequency of matches/non-matches (in either direction) by sex/gender?
- How are the above affected by nationality?
- What is the frequency of matches/non-matches (in either direction) by nationality?
- For each referral agency in the homelessness dataset, what is the frequency of people who are recorded as accessing that service in the rough sleeping dataset but not in the homelessness dataset?
- For each referral agency with a Duty to Refer, how many people are recorded as accessing that service in the rough sleeping dataset but either do not appear in the homelessness dataset or appear in the homelessness dataset but have not been referred by that agency?
- Are people reporting more support needs in one system than in the other?
- For all matched subjects, what are the differences in reported support needs?
- On average, which system reports more support needs per person?
- Which support needs are most likely to not be reported in either system?
- How many people who are recorded on the rough sleeping dataset as eligible for statutory duties but have not contacted Housing Options?
- What is the frequency of people who have access to public funds (nationality / immigration status) but do not appear in the homelessness dataset?
- For the above, is there evidence that they have connections with another local authority?
- For the above, how is this affected by demographics?

## Data improvement

- How can categories be broken down to capture information suitably detailed for strategic purposes?
- What proportion of those recorded as having entered “supported housing or hostel” in the homelessness dataset fall into each of those categories?
- For everyone who is recorded in the homelessness dataset as having entered “supported housing or hostel”, what is the frequency that fall into supporting housing, hostels, and each of any additional accommodation options in the rough sleeping dataset?
- What detailed information is being captured under “other” for variables in the homelessness data collection? E.g., reason for loss of accommodation
- For everyone who is recorded in the homelessness dataset as X, Y, Z as “Other”, what is the frequency of people who fall into all other breakdowns in the rough sleeping dataset?
- What other changes to the homelessness collection would be beneficial?
- Any other breakdowns (i.e. not just those captured in the homelessness dataset as “other”, but all those which are more detailed in the rough sleeping dataset)

## Annex B: Data Cleansing

To start the research and analysis the data was cleaned and formatted. This took a number of weeks due to the complexities of the reporting fields within MRI and the number of decisions that are generated per case when all the reporting fields are completed (of which there are over 100).

The two datasets are very different. MRI data comprises 3 CSV files namely Needs and Circumstances, Case Details and TA (Temporary Accommodation) Reason while the outreach spreadsheet has one master list.

The majority of the data cleansing steps were applied to the outreach spreadsheet since it is a manually maintained report and has a high number of inconsistencies.

The main issues identified were incorrect names and naming conventions, inconsistent date of birth formats (switching between UK and US date formats) and leading and trailing white spaces accompanied with spelling mistakes which would make it harder to check for duplicates on certain columns.

Initial data cleansing was also applied to a subset of the data from both outreach and Needs and Circumstances (MRI) datasets to include first name, last name, date of birth and a unique identifier to enable the linking of the datasets

## Annex C: Data Protection

People found rough sleeping are asked verbally by the outreach team for consent to share their personal information in order to refer them to Housing Options. Some may have provided written consent if they are already known to services.

MRI includes a Declaration and Consents section within the Triage stage which asks for a signature and date to enable the Council to protect and use personal information. This includes data sharing consents such as requesting medical information and consent to use equality and diversity information.

The linked dataset produced on this project was not shared beyond the research team and the guidance from the CWAC's Data Protection team was followed to ensure compliance with data protection legislation.

## Annex D: Supporting Tables

Gender	People sleeping rough who did not make a homelessness application	People sleeping rough who did make a homelessness application	All single people who made a homelessness application
<b>Totals</b>	<b>67</b>	<b>370</b>	<b>1,286</b>
Male	48 (72%)	303 (82%)	898 (70%)
Female	19 (28%)	65 (18%)	377 (29%)
Transgender	-	<1%	-

Nationality <sup>5</sup>	People sleeping rough who did not make a homelessness application	People sleeping rough who did make a homelessness application	All single people who made a homelessness application
<b>Totals</b>	<b>67</b>	<b>370</b>	<b>1,286</b>
UK National	59 (88%)	212 (57%)	1,087 (85%)
EEA National	Suppressed due to low numbers	Suppressed due to low numbers	21 (2%)
Non-EEA National	Suppressed due to low numbers	16 (4%)	168 (13%)
Blank or not stated	Suppressed due to low numbers	135 (36%)	Suppressed due to low numbers
No recourse public funds	Suppressed due to low numbers	Suppressed due to low numbers	Suppressed due to low numbers

---

<sup>5</sup> It is possible for people to meet the criteria for more than one of the nationality breakdowns therefore totals may be higher than the number of people in each group.

Support needs	People sleeping rough who did not make a homelessness application (67)	People sleeping rough who did make a homelessness application (370)
No support needs	13 (20%)	189 (51%)
One or more	54 (80%)	181 (49%)
Mental health	50 (75%)	115 (31%)
Offending history	<10 (<15%)	90 (24%)
Physical ill health	19 (28%)	66 (18%)
Substance dependency	27 (40%)	62 (17%)
Alcohol dependency	<20 (<30%) <sup>6</sup>	39 (11%)
Learning disability	Not recorded in the rough sleeping dataset	15 (4%)

<sup>6</sup> Additional numbers were suppressed in this table to minimise the risk of disclosure.