BOLD
Better Outcomes through Linked Data
Gwell canlyniadau drwy ddata cysylltiol

GIG CYMRU NHS WALES | Iechyd Cyhoeddus Cymru Public Health Wales

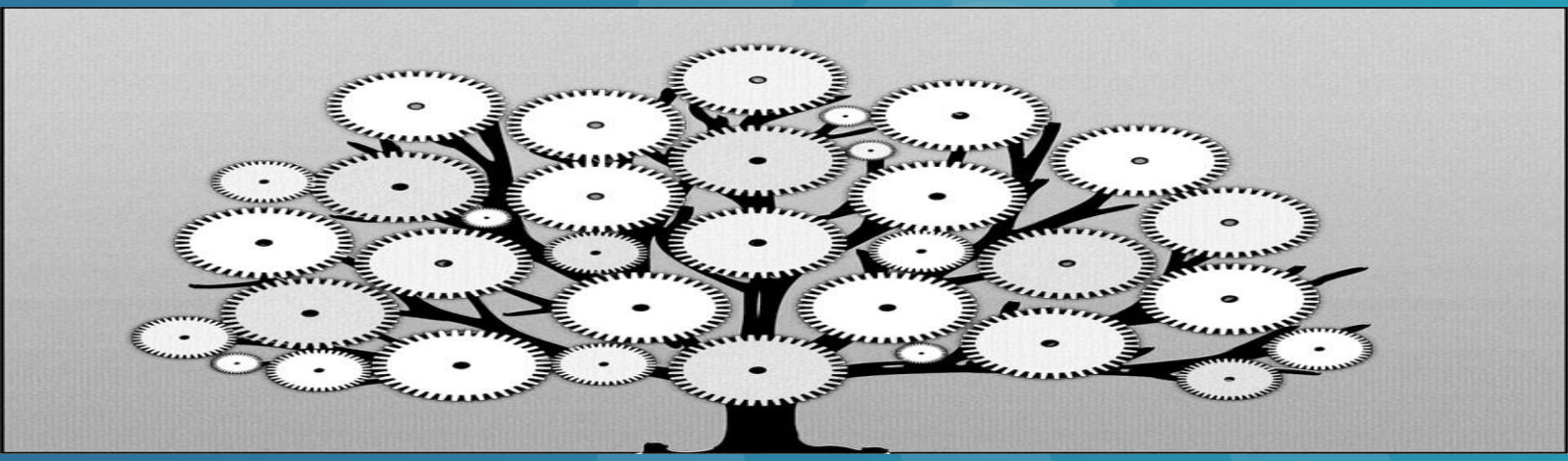
YDG CYMRU | ADR WALES

SAIL DATABANK

Llywodraeth Cymru Welsh Government

DATA EXPLAINED: Welsh National Database for Substance Misuse (WNDSM) data quality audit and considerations

Data Report



Authors
Ryan Phillips[1], Ian Farr[2], Columbus Ohaeri[1], Hywel T Evans[2], Grace Bailey[2], Josh Dixon[3], Josie Smith[1]

Affiliations
[1] Public Health Wales, No 2 Capital Quarter, Cardiff, United Kingdom
[2] Population Data Science, Swansea University School of Medicine, Swansea
[3] Welsh Government, Cathays Park, Cardiff

Acknowledgements
This report is a joint publication between Public Health Wales, ADR Wales, the Secure Anonymised Information Linkage (SAIL) Databank and Welsh Government as part of the Better Outcomes through Linked Data (BOLD) programme. BOLD is a HM Treasury funded, cross-governmental programme (2021-2025) designed 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.

Table of Contents

Tables

Background

This Data Explained summarises the content and potential limitations to users of the Welsh National Database for Substance Misuse (WNDSM) dataset. This is the primary dataset relating to individuals referred to, assessed and/or provided treatment within specialist substance misuse services in Wales.

This paper aims to guide researchers who are interested in using the WNDSM dataset to investigate the role or impact of those engaged with substance misuse treatment services in Wales. This Data Explained has been produced by researchers working together within Public Health Wales, ADR Wales and SAIL team within Swansea University.

The information presented in this Data Explained has been reviewed by Public Health Wales, ADR Wales and SAIL colleagues along with additional insights provided by WNDSM leads within Digital Health and Care Wales (DHCW) with expertise within this thematic area and is accepted to be Data Explained accurate at the point of publication.

This Data Explained report presents the findings of a study conducted as part of the BOLD Substance Misuse Demonstrator Pilot (Phase 1) in Wales, which focuses on data quality of the WNDSM. Further information on the BOLD programme can be found here: Better Outcomes Through Linked Data (BOLD)

Introduction

Substance misuse, the recurrent use of alcohol or illicit drugs causing physical or psychological harm[1] represent a public health challenge and reliable data is required to respond effectively. Individuals who use substances, including alcohol, licit and illicit drugs, can seek advice, support, and treatment interventions through a range of substance misuse services which aim to provide confidential and accessible support to reduce the harms caused by drugs and alcohol to individuals, families and the community. In Wales, individual-level data on those presenting for treatment are captured via monthly inputs within the Welsh National Database for Wales (WNDSM). A copy of this dataset can be accessed within the Secure Anonymised Information Linkage (SAIL) Databank (https://saildatabank.com/) , referred hereafter as Substance Misuse Dataset (SMDS).[2]

Established on 1st April 2005, the SMDS comprises routine data submitted from all substance misuse service providers in Wales receiving funding from the Welsh Government and is collated by Digital Health and Care Wales (DHCW) for reporting. This data contains sociodemographic characteristics and can capture multiple events

that may occur during a client's journey including diagnostic characteristics such as quantity and frequency of substance use, treatment interventions and outcomes measured by Treatment Outcome Profiles (TOPs). TOPs is a tool used to evaluate the effectiveness of substance misuse treatment interventions.

Substance misuse treatment data in Wales is used for planning, research and policy decision-making. As such, data quality audit and assessment are required to ensure an accurate picture.

In 2006 (and refreshed in 2016/17), the Welsh Government introduced formal reviews of the performance of substance misuse treatment services. Service activity and outcomes are monitored against the national Key Performance Indicators (KPI). KPIs are currently in place in relation to waiting times, engagement rates, self-report reduction in substance use, improved quality of life and number of clients with treatment completion, specifically:

1) "<20% did not attend post assessment" measures those who are assessed and then stop attending treatment without being classed as successfully completed
2) ">80% time between referral and treatment" defined as accessing treatment from a referral within a 20-day period
3) "Reduction in problematic substances between start and most recent review/exit TOP" (≥86.5% target)
4) "Quality of life is improved between start and most recent review/exit TOP" (≥84.2% target)
5) "Percentage of cases closed as treatment completed" (≥76.9% target).

In the last decade, the number of individuals entering substance misuse treatment has fallen, whereas the number of deaths whilst in treatment has increased across England and Wales.[3] Given this, in 2021 the UK Government published a 10-year drugs strategy.[4] As part of this, they identified the urgent need to estimate the impact that specialist substance misuse treatment has on substance-related harm. To achieve this, it was proposed that linkage of substance-misuse data could be used to enhance the availability of information.[5] This was delivered through a government pilot programme, Better Outcomes through Linked Data (BOLD), a four-year HM Treasury-funded, cross-governmental programme led by the Ministry of Justice and was designed to demonstrate how people with complex needs can be better supported by linking and improving data held in a safe and secure way, with the goal to improve life outcomes through assessment of individual's needs.[6]

The SMDS represents a key dataset in the BOLD Substance Misuse Wales research programme. The Substance Misuse Wales pilot was a collaborative project between Public Health Wales, SAIL databank, ADR Wales and the Welsh Government. Further information on the BOLD programme can be found here: Better Outcomes Through Linked Data (BOLD)

This Data Explained report presents the findings of a study conducted as part of the BOLD Substance Misuse Demonstrator Pilot (Phase 1) in Wales, which focuses on data quality issues within the Welsh National Database for Substance Misuse (WNDSM) and supports and informs all other BOLD Wales research projects.

Structure of the dataset

Utilising the Secure Anonymised Information Linkage (SAIL) Databank for the study, we identified all records from the SMDS between 2010 and 2022. Data is split into events, with drug and alcohol agencies submitting the events that are applicable to the individuals. An individual will have a maximum of one client and referral event, one assessment event, and one contact end event, whilst there may be multiple treatments and/or TOPs.

It is possible that an individual will not have all of the events listed as they may drop out, e.g. if they drop out before treatment there will be no treatment event. An individual client's treatment journey can be tracked through these data tables using the ALF and SUBSTANCEMISUSEREF_ID field. This field contains an identification number specific to a particular spell of treatment and is used to following a client from referral to assessment to treatment and discharge.

In April 2014, the datasets and data definitions were updated; historic data prior to this date is available within the 'Historic' table, although the remaining six entities also cover data pre- and post-April 2014. The seven SMDS datasets are detailed below.

Through identification of an individual, each treatment episode pathway was mapped, and where multiple treatment episodes occurred for one individual, these were mapped over time.

Within SAIL, each individual is assigned a unique identifier, known as an Anonymised Linkage Field (ALF). In brief, the ALF allows linkage across several datasets. In keeping with previous literature, only individuals with a deterministic or probabilistic matching ≥0.9 were included.

The seven SMDS datasets detailed were:

Client: Information related to the client's details. This includes demographic information such as ethnicity, sex and age, as well as referral information which allows clients to be followed throughout their treatment journey.

Referrals: Referrals may operate on an 'open referral system', whereby individuals can contact services directly to obtain an initial assessment, advice and support. Alternatively, the criminal justice system, social services, family or friends, or any healthcare professional that is supporting an individual can make a referral.

Variables include key performance indicators such as waiting times, as well as demographic information such as referral age, sex and ethnicity, referral information

including referral date, source of referral and agency code

Assessment: The assessment should establish the nature and severity of substance use; explore the reason(s) for use; assess the impact of substance use on their physical, psychological and social functioning; and establish available support from family and friends

The variables collected include:

- Problem substance
- Injecting status
- Parental responsibility and number of children living in the household (if applicable)
- Previous substance use treatment
- Current accommodation need
- Current receipt of care from primary or secondary mental health services

Treatments

The treatment modalities/interventions a client receives as part of their treatment journey are documented here. There are two types of modalities: structured modalities and less-structured modalities. Other variables include modality referral date (agreed engagement with intervention date), modality start and end date, and modality exit status:

- **Planned exit** - individual has completed as set out in the treatment plan, referred to another service or moved to GP-led prescribing
- **Unplanned exit** - where treatment is withdrawn by the provider for one of the following reasons:
    - Did not attend or respond to follow-up contact
    - Moved from area
    - Individual was retained in custody/prison
    - Individual deceased
    - Individual declined treatment

Treatment Outcome Profiles (TOPs)

TOPs are completed at the start of the treatment, reviewed every twelve weeks, and completed again at the end. In Wales, everyone in receipt of treatment who are 16 and over, are required to have a TOPs. The TOPs assesses seven different modalities:

1. Inpatient detoxification
2. Specialist prescribing
3. GP prescribing
4. Psychosocial intervention
5. Structured day programme
6. Residential rehabilitation

7. ‘other’ structured intervention.

The following data is available:

- TOP interview date and treatment stage (start, review, exit or post-treatment);
- Number of days:
    - using substances (crack, opioid, etc.) and injecting (if relevant)
    - shoplifting
    - selling drugs
    - undertaking paid work
    - attending college or school
- Perceived physical, psychological health and quality of life rated on scales of 0 (poor) to 20 (good).
- Yes/No flags looking at criminal offence status, assault/violence status, urgent housing problem, risk of eviction.

Contact ended (discharge)

This dataset includes the date and reason contact ended.

Historic data

This table includes older data from prior to April 2014 (before implementation of the updated definitions now used in SMDS). As a result, different variables from the time period are available. This includes information about substance use, including time using,

Data limitations

Study population

For the purposes of analysis reported here, we used a cohort to describe the overall individual-level SMDS data. A total of 299,058 referrals involving 103,203 individuals were identified in the SMDS data between 01/01/2010 and 01/09/2022.

The following is a breakdown of the major data quality problems experienced while undertaking analysis with the substance misuse dataset:

Data quality issue 1: Treatment episodes data present but exit reason indicating no treatment[1]

There are a total of 12 possible outcomes for a treatment episode, these are recorded in the SMDS_CONTACT_ENDED table and are:

1. **Treatment completed – problematic substance free**: The client no longer requires a treatment intervention and is judged by the case worker as not

[1] The WNDSM dataset was amended in 2014, so the exit reason would not have been available for records prior to April 2014.

using any of the clients reported problematic substances.

2. **Treatment completed:** The client has reached their treatment goal(s) as agreed at commencement of treatment
3. **Treatment withdrawn by provider**: The treatment provider has withdrawn treatment provision from the client. This item could be used, for example, in cases where the client has seriously breached a contract leading to their discharge. It should not be used if a client has simply 'dropped out'
4. **Referred to another service**: A client has finished treatment at this provider agency but still requires treatment and the individual has been referred to another substance misuse provider for this. This code should only be used if an appropriate referral path is available.
5. **Did not attend or respond to follow-up contact**: The treatment provider has lost contact with client for 8 weeks or more without a planned discharge and attempts to re-engage the client have not been successful.
6. **Moved from area**: Client has moved from the geographical area in which they are receiving treatment and not referred to another service.
7. **Prison / Retained in Custody**: The client is no longer in contact with the treatment provider as they are in prison or another secure setting.
8. **Deceased**: During their time in contact with the treatment provider the client has died.
9. **Inappropriate Referral**: Client has been contacted following referral and states that they are not ready to engage in treatment or when assessed there is no substance misuse treatment need for client.
10. **Client unaware of referral**: A third-party referral has been received and when client has been contacted they were unaware of the referral.
11. **Treatment commencement declined by the client**: The treatment provider has received a referral and has undertaken an initial face-to-face assessment with the client, after which the client has chosen not to commence a recommended treatment intervention.
12. **Moved to GP Led Prescribing (Primary Care)**: The client is receiving GP Led Prescribing (Primary Care) and is no longer involved with the substance misuse treatment agency.

Contact end reasons 9, 10 and 11 should be used for circumstances where no treatment was started, however, there are 7,744 treatment episodes where treatment data exists, representing 0.5% of all treatment episodes within the study period. There was clear evidence of large geographic variation in miscoding in relation to exit reason as shown in Table 1. It is not clear from the dataset why these

geographic variations in miscoding exit reasons exist.

Table 1: Number and percent of 'inappropriate exit reason' with treatment data recorded by health board.

Health Board	Miscoded recorded exit reason (n)	Percent of HB treatment total (%)
Betsi Cadwaladr University Health Board (BCUHB)	2,270	8.5
Abertawe Bro Morgannwg University Health Board (ABMUHB)	1,735	8
Cardiff & Vale University Health Board (CVUHB)	789	6.5
Aneurin Bevan University Health Board (ABUHB)	1,653	5.7
Powys Teaching Health Board (PTHB)	268	5.3
No health board provided	31	4.2
Other - Outside Wales	195	3.1
Cwm Taf University Health Board (CTMUHB)	450	2.5
Hywel Dda University Health Board (HDUHB)	353	2

Treatment Modalities

It is common for treatment episodes to contain multiple treatment modalities, which can range from less structured methods such as brief intervention and harm reduction to more structured modalities such as inpatient treatment as well as pharmacological interventions including the prescribing of opioid substitute treatment. It is therefore possible for an episode of inappropriate exit reason to contain multiple treatment modalities that may have been recorded incorrectly.

Table 2: Number and percent of inappropriate exit reason episodes with treatment data recorded by number of modalities.

Number of Treatment Modalities	Count (n)	Percent (%)	Percent (All Categories) (%)
1	4,495	7.3	58
2	1,386	5.2	17.9
3	1,536	4.3	19.8
4+	327	2.6	4.2

There were at least 13,183 individual treatment modalities assigned to individuals with an incorrect exit reason recorded (Table 2).

The majority (89.6%) of all inappropriate exit reasons were recorded amongst individuals receiving psychological treatment interventions (Table 3). It is worth noting, however, that 10.4% of all inappropriate exit reasons recorded indicate the prescribing of controlled substance such as methadone or buprenorphine, either with or without psychological therapy alongside. These pharmacological treatments account for 3-4% of all pharmacological treatments recorded in the dataset.

Table 3: Number and percent of 'inappropriate exit reason' with treatment data recorded by modality class

Modality Class	Count (n)	Percent (%)
Pharmacological	23	3.4
Psychological	6936	6.2
Psychopharmacological	785	3.3

Data quality issue 2: Same day or next day re-referrals where previous treatment is 'completed'.

A contact end reason of "Treatment completed – problematic substance free" or "Treatment completed" is assigned where an individual is deemed to have successfully completed their treatment regime. The contact end definitions indicate that a client would no longer need treatment, and we would not expect to see re-referrals into treatment occurring on the same or next day (as might be the case if a treatment is unsuccessful and the individual is being referred to another specialist service).

If a client does require further treatment intervention at exit from treatment, treatment providers should record an appropriate reason for contact ending for these

clients e.g. “referred to another service”.

In total, SMDS contained 7,210 same or next day re-referrals in the study period. Of these 1,018 (14.1%) follow a ‘treatment completed’ episode, and 496 (6.9%) follow a ‘treatment completed – drug free’ episode. In addition, there were 72 recorded re-referrals following a treatment episode with a contact end reason of ‘deceased’ indicating the individual had died before treatment completion (Table 4).

Table 4: Number and percent of treatment episodes ending where another treatment episode is recorded as commencing the same or next day by treatment exit reason

Exit reason of previous treatment episode	Count (n)	Percent of all same/next day re-referrals (%)
Referred to another service	4,370	60.6
Treatment completed	1,018	14.1
Did not attend or respond to follow-up contact	603	8.4
Treatment completed – problematic substance free	496	6.9
Prison	285	4
Inappropriate referral/Client unaware of referral	254	3.5
Deceased	72	1
Treatment withdrawn by provider	57	0.8
Treatment commencement declined by client	40	0.6
Unknown	15	0.2

Data quality issue 3: Successful completion of treatment when no treatment intervention is recorded.

Successful completion is defined within SMDS using the contact end reasons (provided by the service provider and treatment exit):

- ‘Treatment completed – problematic substance’ - “The client no longer requires a treatment intervention and is judged by the case worker as not using any of the clients reported problematic substances”, and
- ‘Treatment completed’ - “The client has reached their goal(s) as agreed at commencement of treatment”

These definitions indicate that a client has undertaken a treatment regime containing one or more treatment modalities and successfully completed all the required parts.

Where treatment has been undertaken (whether successful or not) the modalities undertaken should be recorded. It should not be possible for a client to have an exit reason of successful completion but not have any treatment information attached.

There are 95,332 referrals recorded within SMDS with a contact end reason indicating successful completion. Of those 18,344 (19%) have no recorded treatment data (no treatment start date, treatment modality type or treatment end date), and 6,493 (6.8%) had no recorded assessment data. Again, there was substantial geographic variation recorded, as well as with substance type, as shown in Table 5 and Table 6 respectively:

Table 5: Number and percent of successful treatment episodes with no treatment data by health board

Health Board	'Successfully completed treatment' with no treatment data (n)	Percent of HB completed treatments (%)
CVUHB	3,354	39.2
BCUHB	5,457	25.1
ABMUHB	1,592	18.1
HDUHB	2,471	15.2
CTMUHB	2,098	14.7
ABUHB	2,763	14.2
PTHB	55	2

Table 6: Number and percent of recorded 'successful treatment' episodes with no treatment data, by substance type

Substance category	Treatment completion with no treatment data	Percent (%)
No record	3,968	61.1
Alcohol	10,919	20
Heroin, Methadone, Other Opiates	1,837	12.3
Amphetamines and Ecstasy	253	11.3
Other drugs	212	8.5
Cannabis	827	8.2
Cocaine and Crack Cocaine	296	6.5

Data Quality issue 4: Treatment episodes with inaccurate time frames

Each treatment episode recorded within SMDS has an episode length defined as the time between the date of referral and the contact end date. It is during this time frame that assessments and treatment take place and any TOPs are completed. This time frame can differ greatly by substance type and individual with a potential range from a single day to many years.

In total 259,140 treatment episodes with a recorded start and end date were included in this study. Of these, 5,716 (2.3%) were seen to have time frames that are inaccurate. Entries were defined as having an inaccurate time frame where the contact end date is before the referral date. There was substantial geographic variation in inaccurate time frame recording as shown in Table 7.

Table 7: Number and percent of treatment episodes by health board that had inaccurate time frames

Health Board	Count (n)	Percent of HB episodes (%)
ABMUHB	782	2.6
ABUHB	468	1.1
BCUHB	1150	1.6
CVUHB	2246	6
CTMUHB	575	2.1
HDUHB	313	1
PTUHB	70	0.9

Data Quality Issue 5: Missing Ethnicity Data

There were 298,845 referrals recorded in the study period, and of these 37,371 (12.5%) were missing ethnicity data (Table 8). Fields to record ethnicity are present both the referral and assessment data tables. For the purpose of this analysis a missing ethnicity entry was defined as having no input or input that indicates data was not recorded for example 'Not stated' or 'Unknown'

Of those referrals recorded, 195,044 reached an assessments stage where ethnicity should be recorded if not collected at the referral stage. There were 18,648 assessments where no ethnicity data was recorded, representing 9.6% of all assessments. There was substantial geographic variation in recording of ethnicity, as shown in Table 9.

Table 8: Number of referrals and assessments by recorded ethnicity

Ethnicity Group	Recorded at referral (n)	Recorded at assessment (n)
Asian	1,615	1,042
Black	1,459	904
Mixed	2,068	1,379
Other	466	260
White	255,866	172,811
Missing	**37,371**	**18,648**

Table 9: Number and percent of referral and assessment contact where no ethnicity recorded, by health board

Health Board	Count (n)	Percent of HB contacts (%)
ABMUHB	4363	17.7
ABUHB	2267	6.7
BCUHB	1760	3.5
CVUHB	6793	27
CTMUHB	599	2.5
HDUHB	1840	7.7
PTHB	46	0.8

Implications for research using these data

The provision of high-quality, comprehensive and effective substance misuse treatment services is challenging and costly, and substantial and ongoing investment is made at a national level. Accurate and timely data detailing the patient care pathway and evidence of outcomes is essential to understanding best practice and enables tailoring of policy and service provision to optimise positive outcomes and efficiency.

Key performance indicators (KPIs), implemented by the Welsh Government, provide

base line and standardized measures for activity and performance across comparable services nationally. Robust data collection and reporting are required in order to establish a reliable evidence base in relation to KPIs over time.

Alongside this, research utilising the SMDS may lead to, or be affected by, bias or erroneous conclusions if the researchers are unaware of the data quality issues existing within the dataset. This data quality audit will provide an important guide to future researchers of the identified issues to be accounted for and will act as a legacy product for SAIL users.

A number of key data quality issues have been identified relating to inappropriate, inaccurate or non-recording of treatment modalities, key milestones (treatment start, cessation and outcomes) and demographics affecting a substantial proportion of those engaged in treatment, including:

- Evidence of treatment where exit reason indicates no treatment provided. Without revisiting each treatment episode, no conclusions can reliably be drawn as to where the error/s in recording lie.
- Evidence of recording 'successful' treatment outcome followed by same or next day re-referral. For a patient or client navigating the substance misuse treatment system, clarity and transparency of process is important, as is the continuity of care to allow for treatment for as long as it may be required. Historic KPI reports may be unreliable in reflecting treatment episodes that are successfully completed.
- Evidence of recorded 'successful completion of treatment' where no treatment has been recorded. As above, this reflects systematic errors in over-reporting treatment successfully completed and introduces bias in both KPI records and in research utilising this data.
- Service failure to routinely and consistently report ethnicity, instead using 'Ethnicity not stated' (code Z). The Equality Act 2010 includes race as a protected characteristic. Race includes ethnic and racial groups. Poor data quality related to ethnicity impacts substantially on research to assess health inequalities in the provision of substance misuse treatment and trends over time.
- Evidence of inaccurate and incomplete referral, assessment and treatment dates. There is a validation check within DHCW which highlights records with inconsistent dates. These are reported back to the services when they sign off their respective files, but the records will still be uploaded as these are not deemed critical errors. Many of the inaccurate dates may be due to services referring an individual, but the assessment date/treatment date remains as the original data from the referring agency, and the date of referral is updated. Agencies should be conducting their own assessment / treatment, so that all dates are accurate.

Recommendations

1. We recommend that routine audit and data quality improvement plans be initiated involving substance misuse service managers, planners and commissioners including the Substance Misuse Area Planning Boards to improve completeness, accuracy and consistency across Wales. Although Welsh Government and DHCW have issued guidance and undertake extensive validation checks and queries on the data, geographic variation and inaccuracies persist.
2. The introduction of mandatory fields for completion and the removal of default codes (for example code Z, 'ethnicity not stated') at assessment and treatment stage be instigated to drive data quality improvement.
3. Study designs need to consider the current limitations of the Substance Misuse datasets
4. Further research into the reasons for geographic variation in data quality and implications for treatment effectiveness, service navigation and treatment pathways in Wales.

References

1. Degenhardt, L. et al. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Psychiatry 5, 987–1012 (2018).

2. Health Data Research UK. Substance Misuse Dataset (SMDS). Dataset - Health Data Research Innovation Gateway

3. National Audit Office and The Comptroller & Auditor General. Reducing the harm from illegal drugs. at https://www.nao.org.uk/wp-content/uploads/2023/10/reducing-the-harm-from-illegal-drugs.pdf%20(2023). (2023).

4. HM Government. From harm to hope: A 10-year drugs plan to cut crime and save lives. Available at: UK 10 year drug strategy

5. Government, U. Ministry of Justice: Better Outcomes through Linked Data (BOLD) - GOV.UK. at https://www.gov.uk/government/publications/ministry-of-justice-better-outcomes-through-linked-data-bold/ministry-of-justice-better-outcomes-through-linked-data-bold#the-bold-programme.

6. Thinks, B. Improving lives through linked data: Views from groups with complex needs Substance Misuse Pilot Audience summary report (2023).

Disclaimer

Due to the sensitive nature of the data, all research applications to access data within SAIL Databank are reviewed by an independent Information Governance Review Panel (IGRP). This process includes consideration by Welsh Government, to ensure that the research is an appropriate use of the data and is for the public benefit. The IGRP includes representatives from professional and regulatory bodies, data providers and the general public.

BOLD

Better Outcomes through Linked Data

Gwell canlyniadau drwy ddata cysylltiol

Add publishing departments' URLs here. (links can be found by clicking through from Departments, agencies and public bodies - GOV.UK - GOV.UK (www.gov.uk)

Associated partners

GIG CYMRU NHS WALES | Iechyd Cyhoeddus Cymru Public Health Wales

YDG CYMRU | ADR WALES

SAIL DATABANK

Llywodraeth Cymru Welsh Government