

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/164627/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Thomas, Ian and Mackie, Peter 2023. Using administrative data linkage to drive homelessness policy: Experiences from Wales, UK. *Cityscape* 26 (1) , pp. 117-125.

Publishers page: <https://www.huduser.gov/portal/periodicals/cityscap...>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Using administrative data linkage to drive homelessness policy: Experiences from Wales, UK

Abstract

This article profiles Administrative Data Research Wales (ADR-W) and their use of data linkage to support homelessness policy and practice in Wales, UK. Despite having a national integrated data system for nearly a decade, Wales has lacked the capacity, and more importantly the homelessness administrative data necessary to engage in linkage based research. However, the formation of ADR-W, with a remit to make better use of public sector data, has instigated a rapid shift in the use of administrative data linkage in Wales. The article introduces the ADR-W, situating it as part of a wider turn to administrative data in the UK, before providing an overview of the operation of the integrated data system used by ADR-W to conduct their research—the Secure Anonymised Information Linkage Databank. The article offers insights into some of the homelessness research conducted by ADR-W, highlighting key policy-relevant findings—including on the effectiveness of the Covid-19 response. Critical reflection is provided on some of the challenges with the current homelessness administrative data landscape in Wales, concluding with a note on its future direction.

Keywords: Wales; administrative data; data linkage; homelessness; Administrative Data Research Wales

Dr Ian Thomas¹ **

Professor Peter Mackie²

** Corresponding author

1. Administrative Data Research Wales/Cardiff University, School of Social Sciences, SPARK, Maindy Road, Cardiff, CF24 4HQ

2. Cardiff University, School of Geography and Planning, Glamorgan Building, King Edward VII Avenue, Cardiff, CF10 3WA

Introduction

Homelessness represents a violation of human rights, being the deprivation of adequate housing, privacy, and security (United Nations, 2016). Accordingly, the government of Wales ('Welsh Government') have committed to creating a nation where homelessness is rare, and, if people become homeless, to ensure it is a brief and non-recurrent experience (Welsh Government, 2021a). What makes Wales—and the UK—unique internationally is that, since 1977, there has been a legal duty on local authorities to ensure that accommodation is made available to certain 'priority need' households experiencing homelessness, i.e., families with dependent children (Fitzpatrick and Davies, 2021). Subsequent amendments to homelessness legislation in Wales, which came into force in 2015, increased legal duties to provide assistance to prevent homelessness (Mackie, Thomas, and Bibbings, 2017). Latest data suggest that in 2022/23, of an estimated 1.4million households in Wales, roughly 9,250 households were assisted to prevent their homelessness, whilst 12,540 households were assisted with homelessness (StatsWales, 2023).

Whilst policy makers and practitioners in Wales endeavour to base their decision making on up-to-date evidence, quantitative research in relation to homelessness has lagged behind other nations, particularly the US (Culhane, Fitzpatrick, and Treglia, 2020). UK homelessness research has historically leaned more heavily toward qualitative methods, whilst the US has stronger research links to psychological and health sciences, and therefore greater affinity for quantitative methods (Fitzpatrick and Christian, 2006). Where quantitative research does take place in Wales/UK, this has largely been based on small scale non-random surveys funded by charities for the purposes of political advocacy—rather than larger scale studies with greater levels of generalizability beyond particular 'at risk' sub-groups within the homeless population (Pleace and Quilgars, 2003). Apart from surveys, administrative data open up interesting opportunities to expand the repertoire of quantitative homelessness research in Wales.

Being the ‘data exhaust’ (Hand, 2018) of day-to-day processes, administrative data can provide insight into the experiences of people as they are assisted by housing and homelessness services. Furthermore, by linking together different administrative data sets, insights can be gained into the interactions of people across multiple systems, i.e., housing, health, and education (Culhane, 2016). Over the past decade, administrative data and their linkage has seen an increased interest amongst devolved governments and the research community in the UK (Elias, 2018).

A report by the Administrative Data Taskforce in 2012 recommended the formation of a UK wide network to facilitate access to and linkage of administrative data (Economic and Social Research Council, 2012), leading to the formation of the Administrative Data Research Network (ADRN) in 2014. Over time, the ADRN has transitioned into Administrative Data Research (ADR) UK, shifting its remit from assisting the research community to access administrative data, to directly engaging in policy relevant research using administrative data and its linkage (Gordon, 2020). Each of the UK nations has its own national ADR centre, which generate evidence relevant to the specific national context—this article relates to the ADR-Wales (ADR-W). The ADR centres each adopt different infrastructures for storing and linking together administrative data, with ADR-W using the Secure Anonymized Information Linkage (SAIL) Databank (Ford et al., 2009).

Data linkage infrastructure in Wales: The SAIL Databank

Initially piloted in 2006, the SAIL Databank acts as an ‘integrated data system’ (Zanti et al., 2022) for Wales, storing de-identified yet linkable individual-level data, and facilitating access to these data for research purposes. When data are ingested into SAIL they undergo de-identification, whereby personal data are replaced with an identification number that is unique

to each person in Wales. The de-identification process involves matching to a population spine created when people register their addresses with general practitioners/family doctors in Wales (Lyons et al., 2009). Matching can be achieved deterministically, through the use of national healthcare numbers or exact matching on name, date of birth, gender, and postcode; alternatively, matching is achieved probabilistically using combinations of name, date of birth, gender, and postcode. Once matched to the population address spine, the person's national healthcare number is extracted and forms the basis of their unique identifier. It is this unique identifier, also known as the Anonymised Linkage Field (ALF), that is used to link information about the same person across data held by SAIL. As an added layer of privacy protection, de-identification is undertaken by a 'trusted third party' so that SAIL do not have sight of personal data along with the 'clinical information' relating to people's service interactions.

The original basis for the SAIL Databank was the study of population health. SAIL therefore hold a comprehensive array of healthcare information for the population of Wales, covering primary care/family medicine to hospitalisations. Over-time, data sources have diversified, and SAIL currently hold a range of other data sets including substance misuse treatment service data, education outcomes (from schools to university), Census records, and criminal and civil courts records. Researchers can apply to the SAIL Information Governance Review Panel to use these data, with projects being assessed on their ability to generate new knowledge of scientific and/or practical value, i.e., research that has 'public' benefit. As a condition of accessing data in SAIL, researchers must undergo training in information governance to be considered a 'safe researcher'. Once approvals have been given and conditions of access met, data are accessed by the research team within a secure virtual environment. To ensure outputs are 'safe', any analysis being requested out of the virtual environment undergoes disclosure control checks to ensure individuals cannot be identified or inferred from outputs.

In addition to enabling relatively streamlined access to de-identified data already held in the SAIL Databank, researchers are also able to have data they possess brought into the SAIL environment. The ability to bring data into SAIL can enable, among other things, the use of data linkage to obtain routinely collected data for participants within housing and homelessness interventions. For example, as part of the ‘PHaCT’ randomised control trial of a critical time intervention with prison leavers facing homelessness (Lewsey, 2023), personal data related to trial participants are being imported into SAIL to allow the trial team to extract their health records—with the aim of comparing health outcomes between trial arms. Outside of this novel potential use of SAIL to conduct evaluative research, the SAIL infrastructure enables linkage between homelessness and other data sources. However, limited availability of individual/case-level linkable homelessness data in Wales has proved a challenge for ADR-W.

The state of homelessness administrative data (linkage) in Wales

Despite legal obligations placed on local authorities to help prevent homelessness, and assist households experiencing homelessness, there is no mandate to collect case-level data relating to households. This contrasts to requirements placed on Continuum of Care in the US to collect ‘universal data elements’ on people/households assisted by services in receipt of Department of Housing and Urban Development funding. However, local authorities in Wales do submit aggregate information to Welsh government for monitoring purposes, and as such, collect similar data on individual cases to complete these ‘aggregate returns’. But, without a top-down mandate to collect the same data in the same format, local authorities have designed divergent data collections. The statutory homelessness data landscape in Wales is therefore fractured and inconsistent.

Without a centralised case-level national data collection for Wales, access to statutory homelessness data is only possible through data negotiations with each of the 22 local authority housing services in Wales. However, in 2018 the ADRN obtained data from a single local authority housing service, which has gone on to form the basis for ADR-W's pioneering programme of research on homelessness. Several studies using these data have centred on the theme of severe and multiple disadvantage, exploring the interactions of people experiencing homelessness with other public services, including health, substance misuse, and the police (Browne-Gott, 2019; Thomas, 2021). A second research theme has been educational experiences of children in homeless families. This programme of work proved particularly challenging as the available local authority homelessness data only contain personal information for heads of households, not their children. Outputs from this work focused on attainment (Welsh Government, 2020) and absenteeism (Welsh Government, 2021b), and found that becoming homeless or at risk of homelessness was associated with a 7% increase in total half-day sessions absent from school.

From initial conversations, to data being deposited in SAIL, acquiring this single local authorities homelessness data took roughly 2-years to complete. A large part of the delay in data acquisition was due to uncertainty within the local authority housing team of legal obligations when sharing data, combined with the necessary bureaucratic steps to enable sharing of personal data—for example, undertaking impact assessments, creating legal documents, and finding appropriate people to authorise and take ownership of the process. Dealing with bureaucracy is a not insignificant aspect of any data share, which in this instance placed additional burden on an already overstretched housing service, contributing to delays.

Since 2021, Welsh Government, in partnership with ADR-W, have been engaging in a pilot project to secure statutory homelessness data from additional local authorities in Wales. The intention of this pilot has been to acquire additional data sets to enable ADR-W to engage

in far more nuanced analyses of under-served groups within the statutory homeless population, such as minority ethnic headed households and children in families accessing housing services. But, as with the previous acquisition of local authority homelessness data by the ADRN, the pilot project has been beset by delays. Though there is a strong desire to engage in this work from local authorities, government analysts, senior policymakers in Welsh Government, and academics, progress has been slow.

Lacking up-to-date administrative data on people being assisted under the statutory homelessness system in Wales, ADR-W have used innovative methods to identify people experiencing homelessness using existing population level data collections within the SAIL Databank. Many of the health data sets held by SAIL contain indicators for homelessness as a social determinant of ill-health. For example, the 'ICD-10' system of classifying diseases used in hospitalisation data contains a code specifying homelessness. Alternatively, as part of the substance misuse data collection, service users are directly asked about their housing support needs, with the clinician provided examples of specific homelessness experiences to guide their recording of different levels of housing need. Within this measure, 'urgent housing problems' and 'housing problems' cover instances of severe homelessness, ranging from living on the streets to sleeping in different accommodation each night.

Using these codes and measures, ADR-W were able to conduct research during the Covid-19 pandemic that generated evidence of the potential protective effect of the Welsh Government crisis response—which included accommodating people experiencing homelessness in 'suitable' temporary accommodation. It was found that the prevalence of SARS-CoV-2 infection amongst people experiencing homelessness was 5.0%, compared to 6.9% among a not-homeless matched comparator (Thomas and Mackie, 2021). A similar methodology was used to demonstrate the ineffectiveness of the Welsh Government policy to prioritise people experiencing homelessness for Covid-19 vaccination. Thomas and Mackie

(2023) found incidence of Covid-19 vaccination after 350-days of follow-up was 60.4% among people experiencing homelessness, compared to 81.4% among a matched adult comparator. Despite these groundbreaking—in Wales at least—insights, the linkage and analysis of administrative data has its limitations, particularly when researching homelessness (Thomas, 2020a; Thomas and Tweed, 2021).

Limitation of ADR-W's homelessness research

Foremost, ADR-W research on people being assisted under the statutory homelessness system has had to draw on data from a single authority, and covering a relatively short period of time—people assisted between 2011 and 2017. Analysis using this data set has therefore been limited to more general questions related to the association between homelessness and people's outcomes/access to public services. An added complication has been that the statutory data covers a period of time during which Welsh homelessness legislation was being reformed, meaning that the categorisations of how households were assessed under legislation were in flux. As a result, ADR-W have been unable to look at outcomes of households assessed under the current Housing (Wales) Act in any complex multi-variate way, due to insufficient sample size and follow-up time.

Where ADR-W have resorted to using 'non-housing' administrative data to identify people experiencing homelessness, generalisability to some larger 'homeless population' may be limited. It is likely that health diagnoses codes related to homelessness are used in instances where the clinician believes homelessness was a factor in the health event—which may not be the case in all healthcare interactions, leading to an under-reporting of homelessness. Though the housing need measure within the substance misuse data set applies to all people within the data, and is therefore potentially less biased by recording behaviour, the population in this data

set is quite obviously limited to people with higher needs seeking assistance with problematic substance use. Resultantly, people experiencing homelessness identified in ‘non-housing’ data sources are potentially at the more precarious end of the homelessness spectrum due to underlying healthcare and substance misuse related issues. However, this ‘bias’ has proved useful in the Covid-19 vaccines analysis, as it demonstrated that vaccine prioritisation was not able to reach the most vulnerable homeless people in Wales.

As with all data linkage research, ‘missed matches’ can be problematic and a source of bias if matched and un-matched people have differing characteristics (Harron et al., 2017). In the case of ADR-W research, missed matches occur either where a record could not be de-identified and assigned a unique identifier, or where records are excluded due to low matching quality, i.e., where unique identifiers are assigned probabilistically and there is uncertainty that it is the correct identifier. From ADR-W’s experience of using different data sources, healthcare data in SAIL have demonstrated far lower rates of missed matches (~5% missed) when compared to the local authority homelessness data set (~25% missed). The ‘data maturity’ of the organisations collecting data, combined with their different information needs, may contribute to these differences in match rates.

Though local authorities in Wales are developing a culture of data use, there remain gaps and weaknesses in their data (Audit Wales, 2018). This compares to the national health service in Wales, where data are core to its functioning, leading to higher quality data. Furthermore, national healthcare numbers are collected when people access healthcare related settings, which then allows deterministic matching to the population spine used when de-identifying data. Without healthcare numbers, local authority homelessness data were de-identified using probabilistic methods, which was, we would suggest, impacted by poorer quality collection of personal information. As data linkage becomes ‘mainstreamed’ as part of the data processing and evidence landscape in Wales, we would hope that the data maturity of

local authorities and housing and homelessness support services improves, reducing missed match rates.

Future directions

Regardless of the challenges faced, ADR-W are slowly developing a portfolio of research that demonstrates the potential use of data linkage to support evidence driven homelessness policy and practice. Something that the ADR-W team have consistently argued for, for over a decade, is the need for a national individual/case-level data collection related to statutory homelessness in Wales (Mackie, Thomas, and Hodgson, 2012; Thomas, 2020b). Putting this in context, Wales are now the outlier of the devolved UK nations, with Northern Ireland, Scotland, and England, all having individual level homelessness collections. However, there are some promising developments in this area. The Ending Homelessness Action Plan for Wales makes a commitment to improving data (Welsh Government, 2021a), with Welsh Government engaging ADR-W to scope out a new data collection system (Thomas, 2020b), drawing inspiration from other UK approaches. Rather than being just about better measurement of homelessness and its impacts, the key opportunity that such a system would enable is the ability to evaluate interventions to establish what works in ending and more importantly preventing homelessness in Wales.

Acknowledgements

This article reports on findings from research that has made use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who made anonymised data available for research via the SAIL Databank. Responsibility for the interpretation of the information supplied by SAIL is the authors' alone. Linked data sets cannot be made available to researchers due to data sharing agreements limiting their access to the research team. However, the individual data sets that were used in this research can be accessed via application to the SAIL Databank. This analysis was undertaken as part of the Administrative Data Research Centre Wales, funded by the Economic and Social Research Council (ES/W012227/1).

References

- Audit Wales. 2018. "The maturity of local government in use of data." Cardiff: Auditor General for Wales.
- Browne-Gott, Hannah. 2019. "Using administrative data to understand the service interactions of people experiencing homelessness," *International Journal of Population Data Science* 4(3):1-1. <https://doi.org/10.23889/ijpds.v4i3.1333>
- Culhane, Dennis. 2016. "The Potential of Linked Administrative Data for Advancing Homelessness Research and Policy," *European Journal of Homelessness* 10(3):109-126
- Culhane, Dennis, Suzanne Fitzpatrick, and Dan Treglia. 2020. "Contrasting Traditions in Homelessness Research between the UK and US." In *Using Evidence to End Homelessness*, edited by Lígia Teixeira and James Cartwright. Bristol: Policy Press.

- Elias, Peter. 2018. "The UK Administrative Data Research Network: Its Genesis, Progress, and Future," *The ANNALS of the American Academy of Political and Social Science* 675(1):184-201. <https://doi.org/10.1177/0002716217741279>
- Economic and Social Research Council. 2012. "The UK Administrative Data Research Network: Improving Access for Research and Policy," Swindon: UK Economic and Social Research Council
- Ford, David V., Kerina H. Jones, Jean-Philippe Verplancke, Ronan A. Lyons, Gareth John, Ginevra Brown, Caroline J. Brooks, Simon Thompson, Owen Bodger, Tony Couch, and Ken Leake. 2009. "The SAIL Databank: building a national architecture for e-health research and evaluation," *BMC Health Service Research* 9. <https://doi.org/10.1186/1472-6963-9-157>
- Fitzpatrick, Suzanne and Julie Christian. 2006. "Comparing Homelessness Research in the US and Britain," *International Journal of Housing Policy* 6(3):313-333. <https://doi.org/10.1080/14616710600973151>
- Fitzpatrick, Suzanne and Liz Davies. 2021. "The 'ideal' homelessness law: balancing 'rights centred' and 'professional-centred' social policy," *Journal of Social Welfare and Family Law* 43(2):175-197. <https://doi.org/10.1080/09649069.2021.1917712>
- Gordon, Emma. 2020. "Administrative Data Research UK," *Patterns* 1(1):1-2. <https://doi.org/10.1016%2Fj.patter.2020.100010>
- Hand, David J. 2018. "Statistical Challenges of Administrative and Transaction Data," *Journal of the Royal Statistical Society Series A: Statistics in Society* 181(3):555–605. <https://doi.org/10.1111/rssa.12315>

- Harron, Katie, Chris Dibben, James Boyd, Anders Hjern, Mahmoud Azimae, Mauricio L Barreto, and Harvey Goldstein. 2017. "Challenges in administrative data linkage for research," *Big Data & Society* 4(2):1-12. <https://doi.org/10.1177/2053951717745678>
- Lewsey, Jim 2023. "Preventing homelessness, improving health for people leaving prison." <https://www.isrctn.com/ISRCTN46969988>.
- Lyons, Ronan A., Kerina H. Jones, Gareth John, Caroline J. Brooks, Jean-Philippe Verplancke, David V. Ford, Ginevra Brown, and Ken Leake. 2009. "The SAIL databank: linking multiple health and social care datasets," *BMC Medical Informatics and Decision Making* 9. <https://doi.org/10.1186/1472-6947-9-3>
- Mackie, Peter, Ian Thomas, and Kate Hodgson. 2012. "Impact analysis of homelessness legislation in Wales: A report to inform the review of homelessness legislation in Wales." Cardiff: Welsh Assembly Government
- Mackie, Peter, Ian Thomas, and Jennie Bibbings. 2017. "Homelessness Prevention: Reflecting on a Year of Pioneering Welsh Legislation in Practice," *European Journal of Homelessness* 11(1):81-107.
- Pleace, Nicholas and Deborah Quilgars. 2003. "Led rather than leading? Research on homelessness in Britain," *Journal of Community & Applied Social Psychology* 13:187-196
- StatsWales. 2023. "Households for which assistance has been provided by outcome and household type." <https://statswales.gov.wales/Catalogue/Housing/Homelessness/Statutory-Homelessness-Prevention-and-Relief>

- Thomas, Ian. 2020a. "Analysing Administrative Data." In *SAGE Research Methods Foundations*, edited by Paul Atkinson, Sara Delamont, Alexandru Cernat, Joseph W. Sakshaug, and Richard A. Williams. <https://doi.org/10.4135/9781526421036871241>
- Thomas, Ian. 2020b. "Feasibility study into a new case level homelessness data system for Wales." <https://housingevidence.ac.uk/publications/feasibility-study-into-a-new-case-level-homelessness-data-system-for-wales/>
- Thomas, Ian. 2021. "Severe and multiple disadvantage." Swansea: Swansea University. https://www.adruk.org/fileadmin/uploads/adruk/Documents/Data_Insights_Severe_and_multiple_disadvantage_March_2021_.pdf
- Thomas, Ian, and Peter Mackie. 2021. "A population level study of SARS-CoV-2 prevalence amongst people experiencing homelessness in Wales, UK," *IJPDS Special Issue: Population Data Science for COVID-19* 5(4):1-12. <https://doi.org/10.23889/ijpds.v5i4.1695>
- Thomas, Ian, and Peter Mackie. 2023. "Assessing the coverage and timeliness of coronavirus vaccination among people experiencing homelessness in Wales, UK: a population-level data-linkage study," *BMC Public Health* 23:1-8. <https://doi.org/10.1186/s12889-023-16432-x>
- Thomas, Ian, and Emily Tweed. 2022. "The Promises and Pitfalls of Administrative Data Linkage for Tackling Homelessness," *European Journal of Homelessness* 15(3):177-188.
- United Nations. 2016. "Report of the Special Rapporteur on adequate housing as a component of the right to an adequate standard of living, and on the right to non-discrimination in this context." <https://www.ohchr.org/en/hr-bodies/hrc/regular-sessions/session31/list-reports>

Welsh Government. 2020. “Exploratory analysis of education outcomes of children and young people living in homeless households.” <https://www.gov.wales/exploratory-analysis-education-outcomes-children-and-young-people-living-homeless-households>

Welsh Government. 2021a. “Ending Homelessness in Wales: A high level action plan 2021-2026.” <https://www.gov.wales/ending-homelessness-wales-high-level-action-plan-2021-2026>

Welsh Government. 2021b. “Absence from school amongst children living in homeless households.” <https://www.gov.wales/absence-school-amongst-children-living-homeless-households-html#:~:text=Absenteeism%20varied%20with%20academic%20year,total%20sessions%20absent%20from%20school.>

Zanti, Sharon, Emily Berkowitz, Matthew Katz, Amy Hawn Nelson, T.C. Burnett, Dennis Culhane, and Yixi Zhou. 2022. “Leveraging integrated data for program evaluation: Recommendations from the field,” *Evaluation and Program Planning* 95. <https://doi.org/10.1016/j.evalprogplan.2022.102093>