The Promises and Pitfalls of Administrative Data Linkage for Tackling Homelessness

Ian Thomas and Emily Tweed

Cardiff University, Wales
University of Glasgow, Scotland

Introduction

To meet the challenges of tackling homelessness in a changing world, we need to understand its extent, causes, and consequences, and the impacts of efforts to prevent and alleviate it. Robust evidence from a diverse range of sources is therefore required to inform policy making and service delivery. One approach to meeting this need is the use of administrative data and data linkage. Administrative data refers to information generated as a by-product of the day-to-day activities of services and organisations (Hand, 2018). Examples of administrative data are records of interactions with housing services, healthcare diagnoses, and benefits received. Although not collected for the purposes of research or evaluation, these data can be enormously useful for studying homelessness, particularly when linked together over time and/or with other data sources (Culhane, 2016). Data linkage involves the joining together of information, usually at an individual level, either using personal details (e.g., name, date of birth, address), or through a unique identifier (e.g., health or social security number).

The use of administrative data linkage in tackling homelessness differs internationally and is largely a product of the availability and quality of homelessness data, alongside the presence of data linkage infrastructures, both technical and legal, where data can be processed, linked, and made available for research use. For example, Denmark has collected administrative data on people using homeless shelters since 1999 and has a well-developed national data linkage infrastructure, where administrative data are collected on the basis that they will be routinely linked. Combined, these two factors have enabled natural experiments and population-level data linkage analysis of patterns of shelter use (Benjaminsen and Andrade, 2015), family backgrounds of shelter users (Benjaminsen, 2016a), factors that increase the risk of shelter use (Benjaminsen, 2016b), and psychiatric disorders

and mortality amongst shelter users (Nielsen et al., 2011). In this article we draw on published research and use case studies of linkages undertaken by one of the authors to reflect on the promises and pitfalls of administrative data linkage to understand and tackle homelessness.

The Promises of Administrative Data Linkage

Administrative data have several key features that make them an invaluable source of information on homelessness. They represent people's interactions with organisations and services ('systems') in the 'real world', often over extensive periods of time (depending on how long a service has been running). They often contain larger samples of people experiencing homelessness than possible through surveys, sometimes covering entire populations of people accessing services. However, administrative data are usually limited to information needed by an organisation to undertake their work and are not collected for the purposes of research, which may limit the extent and quality of information available.

Data linkage can enhance individual administrative data sources by contributing additional information on people's characteristics and experiences: for instance, access to substance use services (see case study 1) or educational history (see case study 3) among people experiencing homelessness. This can provide a greater understanding of the population affected by or at risk of homelessness, as well as the impact of interventions. For instance, Benjaminsen (2018) linked data from Housing First records to those on shelter use nationally to assess the coverage rate of Housing First and evaluate its scaling-up.


Case study 1: Understanding access to substance use services amongst people experiencing homelessness (Thomas, 2021)

What was found
1 741 people living in the City and County of Swansea were assessed by substance use services during the two-year period between April 2014 and March 2016. A total of 4 079 heads of households were accessing services for housing related issues during the same time period. Twenty three percent of people accessing substance use services were also accessing the housing team; whilst 10% of heads of households known to the housing team had accessed substance use services.

Figure 1: Overlap in service user populations
By type of services accessed
April 2014 to March 2016, City and County of Swansea

1 341   400   3 679

☐ Housing team only
☐ Substance use services only
☐ Both housing team and substance use services
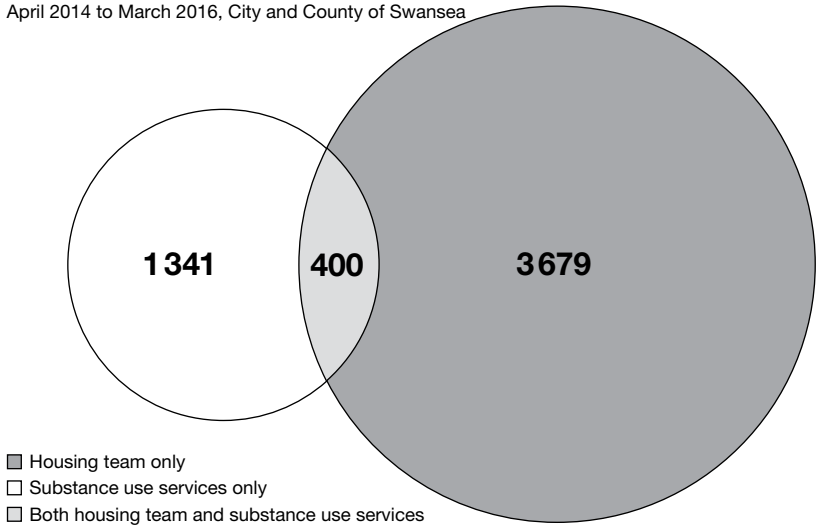
How data linkage was used

People experiencing homelessness were identified using data from the City and County of Swansea's housing team—a 'statutory' homelessness service run by the Local Authority. As the homelessness data available did not provide a reliable measure of drug and alcohol use, data linkage was used to the Substance Misuse Data Set (SMDS) Wales, a data set relating to people accessing state funded substance use services.

To compare individuals accessing homelessness and substance use services within the same geographic region, records in the SMDS were limited to individual's resident in the City and County of Swansea between April 2014 and March 2016. If a person had multiple assessments during this period, they were only counted once. The data were linked at the person level, thereby giving an indication of the scale of overlap between housing and substance use services.

Administrative data linkage becomes especially powerful when combining data from different policy areas and services to explore cross-sectoral solutions. Case study 2 provides an example of how record linkage has helped develop our understanding of the health needs of the homeless population in Scotland. Similar techniques can also be used to evaluate interventions: in Canada, survey data collected

as part of the At Home/Chez Soi homelessness intervention is being linked to administrative health data in order to evaluate the impacts on healthcare utilisation (Wiens et al., 2020).

Case study 2: Understanding the health needs of people experiencing homelessness in Scotland (Waugh et al., 2018)

What was found
The study identified 435 853 individuals assessed by local authorities as homeless during the period 2001-2016, a coverage of 76% of homelessness assessments undertaken during that time. This indicated that at least 8% of the Scottish population had been homeless at some point during the study period. Of these, 119 786 (27%) had been homeless on multiple occasions.

People who had experienced homelessness were more likely to have used healthcare services than their non-homeless peers, with Accident and Emergency departments attendances being 1.8 times higher among the ever-homeless cohort compared to controls living in the most deprived areas of Scotland; mental health admissions were 4.9 times higher. Interactions with health services also peaked just before the date of first homelessness assessment, but had been increasing relative to controls for some years before that point. This suggests crucial opportunities for early intervention in the healthcare sector to prevent or mitigate episodes of homelessness.

The results were featured in the Scottish Chief Medical Officer's report and have influenced the work programme of the Scottish Government policy team, particularly in relation to the needs of people experiencing severe and multiple disadvantage.

How data linkage was used
This project linked national administrative data on people applying to Local Authorities for homelessness support between 2001 and 2016 with health datasets on hospital admissions, interactions with drug treatment services, medication dispensing, and deaths. This linkage represented a major scaling-up of a previous demonstration project undertaken in a single Local Authority (Fife), and an opportunity to understand this relationship at a whole-population level rather than through small-scale studies.

Each individual in the 'ever homeless' cohort was matched on age and sex to two controls without experience of homelessness: one person from the 20% most deprived areas in Scotland and one person from the 20% least deprived areas in Scotland.

Data linkage can therefore help push for greater joined up working, by evidencing the impacts this can have in preventing future homelessness or reducing costs in other policy areas (Downie, 2018). In case study 3, data linkage between housing team data and education data enabled analysis of the association between homelessness and absenteeism from school. This same data linkage could equally be used to explore the impacts of school-based homelessness prevention efforts on the retention of pupils in school—such as Upstream Cymru (End Youth Homelessness Cymru, 2020).


Case study 3: Exploring the impacts of homelessness on pupil absenteeism (Thomas and Mackie, 2020)

What was found

Persistent absenteeism, as defined as missing more than six weeks of school, was higher amongst 'homeless' compared to 'non-homeless' pupils, even after taking into account characteristics such as gender and socioeconomic deprivation. On average, homeless pupils missed five days more of school than their non-homeless peers (Thomas and Mackie, 2021).

Figure 2: Persistent absenteeism
Percentage pupils absent for more than 6 weeks of school
By academic year and pupil housing status

☐ Not homeless   ■ Homeless

	2012/13	2013/14	2014/15	2015/16
Not homeless	7%	5%	3%	3%
Homeless	12%	13%	15%	10%
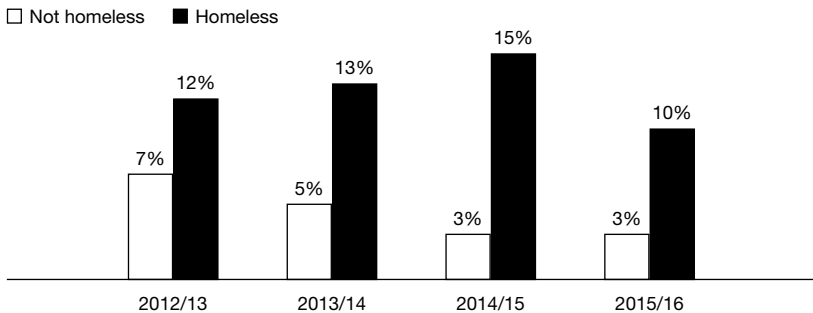
How data linkage was used

The data sets used included the City and County of Swansea statutory housing team's data, attendance and exclusions data, the Pupil Level Annual School's census (PLASC), which records information on every pupil attending school in Wales, and the Welsh Demographic Service (WDS), a list of people registered with a GP as living in Wales. The education data sets (i.e., attendance, exclusions, and PLASC) were combined to form an education panel covering all pupils aged roughly 5 to 16 years old in state funded schools in Swansea for the academic years 2012/13 to 2015/16—roughly 27 000 pupils per year.

A complicating factor in this linkage analysis was that the housing team data related only to the main 'adult' head of household who had applied to the housing team. Children and young people living with the adult applicant were therefore identified by linking to the WDS, which contains a unique residential identifier. A flag was then included in an education panel data set for whether a pupil lived in a residence accessing the housing team—taken to be an indication of homelessness/housing instability.

Though administrative data can be used to undertake research and evaluative work, it can also help answer one of the more basic questions when tackling homelessness: how many people are experiencing homelessness? 'Triangulating' data from different administrative sources by linking them together can help make up for any potential bias from using a single data source. In Scotland, for example, combining data on people applying to councils for statutory homelessness support (the HL1 dataset) with those accessing the preventative pathway known as Housing Options (PREVENT1), can ensure a broader spectrum of those experiencing homelessness and housing insecurity are represented in research.

Moreover, linkage to non-housing administrative data can help identify people not approaching housing services who are nonetheless homeless and who may fall under a much broader definition of 'hidden homeless' (Pleace and Hermans, 2020). In case study 4, linkage between data from housing services, primary healthcare, and substance use services helped identify a group of people who were experiencing homelessness but were not accessing statutory services.

Case study 4: Improving estimates of homelessness using linkage to non-housing data (Thomas and Mackie, 2021)

What was found
There were 4 290 people experiencing homelessness attending the City and County of Swansea housing team during the April 2013 to March 2015 period. There were 426 people who had attended either their GP or accessed a substance use programme during the same period and were recorded as experiencing homelessness. The combination of SMDS and GP data with the Swansea housing team data increased estimates of homelessness in Swansea by roughly 5%, or 211 unique individuals.

Figure 3: Number of people experiencing homelessness
By type of services accessed
April 2013 to March 2015, City and County of Swansea

211   215   4075

☐ Housing team only
▨ Health & substance use services only
☐ Multiple service types
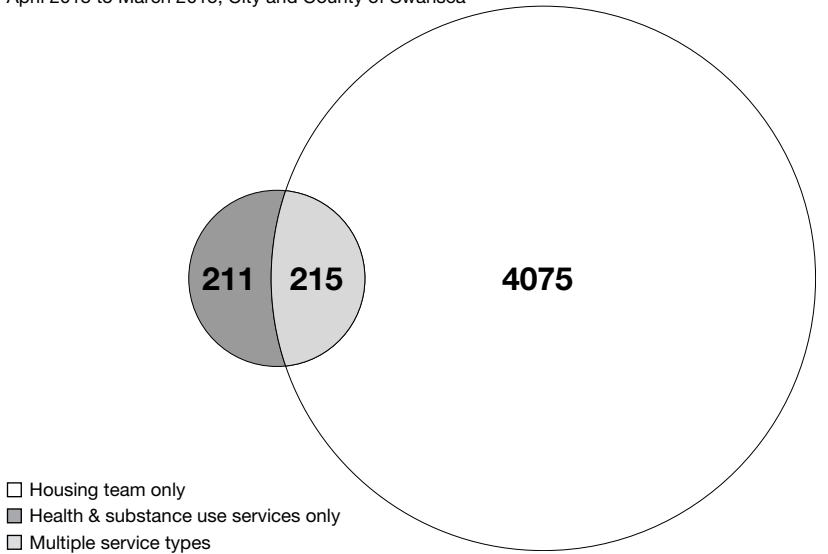
How data linkage was used

This analysis used data from the Substance Misuse Dataset (SMDS) Wales, General Practice (GP) data, and data from the Swansea Local Authority housing options team. The SMDS includes information on a person's 'accommodation need', covering a range of homeless situations, including sofa surfing, rough sleeping, staying in hostels, Bed and Breakfasts, or squatting. The GP data uses specific codes where a diagnosis is linked to homelessness. As with the SMDS, the homelessness codes can relate to general 'homelessness', or to specific forms of homelessness, such as sleeping on the streets or in shelters. The GP data and SMDS were combined to create a flag for whether a person had been recorded as experiencing homelessness/housing related issues outside of a housing setting.

Main analysis explored the overlap of the combined 'homeless-in-health-data' indicator with people approaching the Swansea housing options team at any point during the period from April 2013 to March 2015.

Conclusion: The Pitfalls of Administrative Data Linkage

Despite the potential promise of homelessness research using administrative data linkage, it is important to acknowledge several important challenges. For someone to be visible in administrative data, they must have some form of contact with the systems which generate that data (Pleace and Hermans, 2020). As a result, those who are not engaged with services – or on a sporadic basis – may be missed. However, research suggests that most people who are experiencing homelessness are in contact with some sort of service at some point (Pleace and Hermans, 2020). Maximising the period of time for which data are available to ensure inclusion of those engaging on a sporadic or transient basis (Benjaminsen and Andrade, 2015), and triangulating across datasets from multiple sources, as in our third case study, can therefore help fill in the gaps of people's homelessness pathways.

Where data on personal identifiers (such as name, date of birth, or health or social security numbers) are incompletely or incorrectly recorded, individuals may be missed from the linkage process, potentially skewing results. For example, 25% of records from the City and County of Swansea housing team data used in the three research case studies could not be assigned to a unique person and were not linkable, compared to less than 5% from the SMDS, where personal details are a requirement of the collection. We would therefore advocate systems collect personal identifiers where feasible and legal for them to do so, and to maintain the quality of this data, to maximise the potential for linkage research. However, we recognise that some homelessness services, particularly low threshold services such as 'soup-runs' and other meal programmes, may have transitory interactions with people using their services, during which it is not appropriate or feasible to collect personal details. There may therefore always be some services and interactions with people experiencing homelessness that are outside the realms of feasible administrative data collection, and therefore data linkage research.

A fundamental aspect of administrative data are their origin from organisational processes; they therefore reflect these processes and the ways in which they are socially constructed by policy, practice, and human behaviour (Gomm, 2004). For example, statutory homelessness data collections across the UK reflect housing legislation, and how it is enacted by local authorities and 'street level bureaucrats', more so than they do an objective reflection of the state of 'homelessness' in the UK. Their use and interpretation require an understanding of the process by which they are created, how processes and data have changed over time, and who might be missed as a result. For instance, correct interpretation of linkage projects using statutory homelessness data in Scotland requires the knowledge that not everyone applies for such support on every occasion they are homeless; that eligibility for support has changed over time (for instance, with the abolition of the 'priority need'

test in 2012); and that recent policy developments on prevention, rapid rehousing, and Housing First may have changed the characteristics of people requiring such support (see Waugh et al., 2018). Close collaboration with data providers and people with lived experience of these systems can help researchers understand the value and limitations of these data in the context of the real-world processes that created them.

A final consideration in data linkage is the unique ethical and legal challenges of using these data. The size and historic nature of many administrative datasets means that gaining individual consent to use these is rarely practical. Use of administrative data therefore tends to occur under other legal provisions, such as 'public benefit'. Evidence suggests that public attitudes towards administrative data use and linkage under these provisions are generally positive, though not unconditional (Aitken et al., 2016). Fewer studies have specifically examined this question among people with experience of homelessness; a workshop in London found positive views towards the use of administrative data for health research, including the linkage of sensitive health and social data (Luchenski et al., 2017), but there is a need for more detailed research and engagement on this topic.

Public acceptance of administrative data linkage is higher where extensive safeguards are in place to ensure data are analysed safely, securely, and with the smallest possible risk of inadvertently revealing people's identities (Kispter, 2019). However, creating these 'safe settings' for administrative data research and linkage requires extensive technical infrastructure, as well as appropriate legal and governance structures—see Harron et al. (2017) for more details. Where infrastructures exist, navigating them can be a lengthy and bureaucratic process, especially when combining data from multiple providers. To avoid the challenges described here and realise the potential of administrative data use and linkage, future priorities in the UK and internationally should include the routine integration of datasets across the complex system of sectors and services relevant to homelessness. This ensures any homelessness data that is collected for linkage research can be re-used by others where data providers agree in order to maximise the future impact and value of these resources. There should also be a focus on using administrative data to evaluate 'what works', including the impacts of large-scale policy and service changes; and there should be close collaboration among data providers, researchers, and those with lived experience of homelessness to ensure a rich understanding of the data within its real-world context.

Acknowledgements

> **References**

Aitken, M., de St. Jorre, J., Pagliari, C., Jepson, R., and Cunningham-Burley, S. (2016) Public Responses to the Sharing and Linkage of Health Data for Research Purposes: A Systematic Review and Thematic Synthesis of Qualitative Studies, *BMC Medical Ethics* 17(73) pp.1-24.

Benjaminsen, L. (2016a) The Variation in Family Background Amongst Young Homeless Shelter Users in Denmark, *Journal of Youth Studies* 19(1) pp.55-73.

Benjaminsen, L. (2016b) Homelessness in a Scandinavian Welfare State: The Risk of Shelter Use in the Danish Adult Population, *Urban Studies* 53(10) pp.2041-2063.

Benjaminsen, L. (2018) Housing First in Denmark: An Analysis of the Coverage Rate among Homeless People and Types of Shelter Users, *Social Inclusion* 6(3) pp.327-336.

Benjaminsen, L. and Andrade, S.B. (2015) Testing a Typology of Homelessness Across Welfare Regimes: Shelter Use in Denmark and the USA, *Housing Studies* 30(6) pp.858-876.

Culhane, D. (2016) The Potential of Linked Administrative Data for Advancing Homelessness Research and Policy, *European Journal of Homelessness* 10(3) pp.109-126.

Downie, M. (2018) *Everybody In: How to End Homelessness in Great* Britain (London: Crisis).

End Youth Homelessness Cymru (2020) *Heading Upstream – Working Earlier to Prevent Youth Homelessness* [Webpage] Available at: https://www.endyouthhomelessness.cymru/voices/upstream-cymru-update [Accessed 23 June 2021].

Gomm, R. (2004) *Social Research Methodology: A Critical Introduction* (Basingstoke, New York: Palgrave Macmillan).

Hand, D.J. (2018) Statistical Challenges of Administrative and Transaction Data, *Journal of the Royal Statistical Society: Series A* 181 pp.555-605.

Harron, K., Dibben, C., Boyd, J., Hjen, A., Azimaee, M., Barreto, M.L. and Goldstein, H. (2017) Challenges in Administrative Data Linkage for Research, *Big Data & Society* pp.1-12.

Kispeter, E. (2019) *Public Support for Accessing and Linking Data About People from Various Sources: Literature Review* (Warwick Institute for Employment Research, Warwick: The University of Warwick).

Luchenski, S., Clint, S., Aldridge, R., Hayward, A., Maguire, N., Story, A., and Hewett, N. (2017) Involving People with Lived Experience of Homelessness in Electronic Health Records Research, *International Journal of Population Data Science* 1(1) p.295.

Nielsen, S.F., Hjorthoj, C.R., Erlangsen, A., and Nordentoft, M. (2011) Psychiatric Disorders and Mortality among People in Homeless Shelters in Denmark: A Nationwide Register-Based Cohort Study, *The Lancet* 377 pp.2205-2214.

Pleace, N. and Hermans, K. (2020) Counting All Homelessness in Europe: The Case for Ending Separate Enumeration of 'Hidden Homelessness', *European Journal of Homelessness* 14(3) pp.35-62.

Thomas, I. (2021) *Severe and Multiple Disadvantage*. [Online] Available at: https://www.adruk.org/news-publications/news-blogs/the-overlap-of-substance-misuse-and-housing-issues-375/ [Accessed 6 July 2021].

Thomas, I. and Mackie, P. (2020) *Exploratory Analysis of Education Outcomes of Children and Young People Living in Homeless Households*. [Online] Available at: https://gov.wales/exploratory-analysis-education-outcomes-children-and-young-people-living-homeless-households-html [Accessed 6 July 2021].

Thomas, I. and Mackie, P. (2021) *Using Data Linkage to Improve Estimates of Homelessness*. [Online] Available at: https://www.adruk.org/news-publications/news-blogs/the-power-of-case-level-data-and-data-linkage-in-homelessness-research-340/ [Accessed 6 July 2021].

Waugh, A., Clarke, A., Knowles, J., and Rowley, D. (2018) Health and Homelessness in Scotland. [Online] Available at: https://www.gov.scot/publications/health-homelessness-scotland/ [Accessed 9 June 2021].

Wiens, K., Rosella, L.C., Kurdyak, P., and Hwang, S.W. (2020) Patterns and Predictors of High-Cost Users of the Health System: A Data Linkage Protocol to Combine a Cohort and Randomised Controlled Trial of Adults with a History of Homelessness, *BMJ Open* 10(e039966) pp.1-10.