Using Data Linkage to Improve Estimates of Homelessness

Authors: Dr Ian Thomas (ThomasIR@Cardiff.ac.uk)  
Dr Peter Mackie (mackiep@cardiff.ac.uk)

Date: February 2021

This data insight outlines exploratory analysis using data linkage between local authority data and health data in order to improve estimates of the scale of homelessness at the local authority level.

What we did

This analysis used data from the Substance Misuse Data Set (SMDS) Wales, General Practice (GP) data, and data from the Swansea local authority housing options team.

Given the different definitions of 'homeless' used across each data set, our analysis relates to homelessness in its broadest sense. For example, homelessness covers sofa surfing with friends or family, staying in Bed and Breakfasts, alongside a complete lack of accommodation.

SMDS data are generated by public and third sector organisations when assessing people entering onto substance use programmes. 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.

Our analysis limits the substance misuse and general practitioner data to those people who were recorded as being 'homeless' according to the different data set's definition. Data were also limited to those recorded as living in Swansea to produce estimates for a single area.

In their strategy for preventing and ending homelessness in Wales, Welsh Government outline that a better understanding of the scale of homelessness is a key question to address when tackling this social issue. Current sources of information on homelessness in Wales include aggregate data on people approaching local authority housing teams and the annual (aggregated) count of rough sleepers.

In this insight report we consider whether data held by public services can be used to improve our estimates of homelessness. Furthermore, by using data linkage, we can generate counts of unique individuals experiencing homelessness, in contrast to the current aggregate estimates which may overlap to an unknown extent.
Both the SMDS and GP data were combined to generate an indicator for whether a person had been seen in either data set during the period. Combining SDMS and GP data was necessary because small numbers within the data otherwise risked disclosure of individuals. We 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.

**What we found**

There were 4,075 homeless individuals who attended the Swansea housing options team during the April 2013 to March 2015 period. 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.

Though there was some overlap in people experiencing homelessness in both the combined health data and Swansea housing team data, over half of those who appeared at health services who were homeless had not also been seen by statutory housing services. This is not to suggest that at some point outside of the April 2013 to March 2015 observation period, these people have not already engaged with the housing team or will go on to.

It should be noted that of the 6,105 cases/applications of support in the Swansea housing team data opened in the April 2013 to March 2015 period, 4,845 cases/applications could be linked to a person. There are therefore 1,260 cases unaccounted for (21%), and which could include people in the SMDS/GP data.

---

5 Percentages do not sum to 100 due to rounding
Why it matters

Within year estimates of homelessness based solely on statutory data may lead to an underestimate of vulnerable groups. By linking to non-homelessness service data, we have been able to identify people who are homeless, but, who for various reasons do not or have not yet sought assistance from statutory services.

A 5% increase in estimates of people experiencing homelessness may seem a trivial improvement. However, this analysis drew on only two additional data sets, amongst many that could potentially be drawn on, e.g. projects that interact with homeless populations being funded through Supporting People. Of particular benefit is linkage to ‘non-homelessness’ data, which might reveal homelessness amongst a wider population group, e.g. GP and hospital data.

References

1 https://gov.wales/homelessness-strategy

2 https://gov.wales/homelessness-statistics

3 https://gov.wales/national-rough-sleeper-count


5 Individuals/households can make multiple applications to the authority during the time-period of interest; the number of cases is greater than the number of unique individuals. Data cleaning also reduced the number of cases between raw data upload from Swansea, and the final data set used in this analysis.
ADR Wales brings together specialist teams, data science experts, and statisticians as part of the Economic and Social Research Council (part of UK Research and Innovation) funded ADR UK. Our team is made up of specialists in their field from Swansea University Medical School, the Wales Institute of Social and Economic Research, Data and Methods (WISERD) at Cardiff University and the SAIL Databank at Swansea University with statisticians, economists and social researchers from Welsh Government. Together ADR Wales develops new evidence which supports the Welsh Government’s national strategy, Prosperity for All to improve the lives of people in Wales.

For further information please contact Cathrine.E.Richards@Swansea.ac.uk