This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

Data production and data analysis for the project on neighbourhood crime and mental health: Expected results 6.2 and 6.3

Gergő Baranyi, ESR 6, The University of Edinburgh

1.0 Introduction

Since the early 2000s, an increasing amount of research is focusing on the question whether residential environment is associated with the mental health and wellbeing of individuals living within the same areas, and if yes, how exposure to advantaged and disadvantaged neighbourhoods might lead to different health outcomes. Systematic reviews point out that the majority of findings are still based on cross-sectional studies, which preclude from a more causal interpretation of the relationship.[1-3] Longitudinal and life-course research is needed, in order to understand the underlying mechanism and the relevance of the findings for different age groups. Furthermore, as neighbourhood is overwhelmingly described in research with its structural features such as socioeconomic deprivation, residential instability or residential segregation, research should focus on the possible ways leading from neighbourhood structural disadvantage to depression.[4]

One of the mediating pathways between neighbourhood poverty and mental health problems can lead through the increased level of local crime in disadvantaged areas.[4] Becoming a victim or witnessing violent acts might directly cause mental health conditions such as major depression, anxiety disorders and post-traumatic stress disorder, in particular for women.[5] On the other hand, increased local neighbourhood crime can elevate the fear of being victimise (fear of crime), which leads to avoidance behaviour, unhealthy stress-reduction and chronic psychological distress, all known as potential risk factors of common mental disorders.[6]

The aim of this project is to investigate the complex relationship between neighbourhood poverty, crime and different mental health outcomes, by using a longitudinal population sample linked with objective information on neighbourhood exposures and mental health outcomes. Furthermore, modelling changes of crime in the past decade and differentiating trajectories might help to understand how long-term exposure and changing environment increases the risk of psychiatric problems.

The following report provides a summary on the different data sources used for this project, a description how they have been linked together and gives information on the main statistical analyses used. As the dataset contains particularly sensitive information about individuals, which is a subject of thorough final data clearance, neither preliminary nor final results will be presented in this report, but hold back until final publication.

This project is part of a larger investigation on economic recession and health outcomes (https://sls.lscs.ac.uk/projects/view/2015_015/).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

2.0 Data sources

This project makes use of four different sources of information:

(1) The Scottish Longitudinal Study, which provides individual level information on sociodemographic characteristics and some general health indicators of its member. This is the core sample and base for all data linkage;

(2) The Scottish Index of Multiple Deprivation with aggregated area based measure of police recorded crime and income deprivation;

(3) Prescription Information System from NHS Scotland with data on prescribed medication for mental health problems; and

(4) NHS GP registration dataset providing history of residential postcodes since 2000.

2.1 Scottish Longitudinal Study

The Scottish Longitudinal Study (SLS) is a 5.3% nationally representative sample of the Scottish population linking different administrative and statistical data sources. The sample captures around 270,000 individuals since 1991, using 20 semi-random birthdays as inclusion criteria.[7,8] The datasets consist of information on census variables (1991, 2001 and 2011 Census), vital events (birth, marriage, death, migration), education (school attendance and qualification), geographical and ecological information (e.g. Carstairs deprivation index); and on weather and pollution (e.g. fine particle exposure). SLS is a dynamic dataset, which has been continuously updated and linked with other data sources since established.[9] Data management practices and security measures ensure the exceptional high quality of SLS, while maintaining confidentiality and privacy of the study participants (see 4.0). As a nationally representative longitudinal dataset, SLS provide valuable information on the Scottish population for researchers and policy makers.

2.2 Scottish Index of Multiple Deprivation

The Scottish Index of Multiple Deprivation (SIMD) is the official tool of the Scottish Government, providing publicly available information on different types of deprivation on data zone level (500 - 1000 household residents). SIMD 2012 includes seven domains: Employment; Income; Health; Education, Skills and Training; Geographic Access to Services; Crime; and Housing. The Income Domain captures the number of people claiming for relevant benefits (Adults and children in Income Support or Income-based Employment Support Allowance households; Adults in Guarantee Pension Credit Households; Adults and children in Job Seekers Allowance households; Adults and children in Tax Credit Families). The Crime Domain consists several categories of police recorded crimes or offences, grouped into five major indicators (Crimes of Violence, Sexual Offences, Domestic housebreaking, Vandalism, Drugs Offences, Common Assault).[10] Number of individuals with social benefits and counts of police reported crime have been divided with the respective total population of the data zone, and standardised by computing ranks from the most deprived (1) to the least deprived areas (6,505) for each indicator.[10] SIMD were first released in 2004; this study makes use of the 2006, 2009 and 2012 SIMD releases. Before analysis, five equal groups with the same number of neighbourhoods were computed for both deprivation scores. Figure 1 shows the crime deprivation across Scotland based on the 2012 SIMD release (which uses police reported crime from 2010 and 2011).

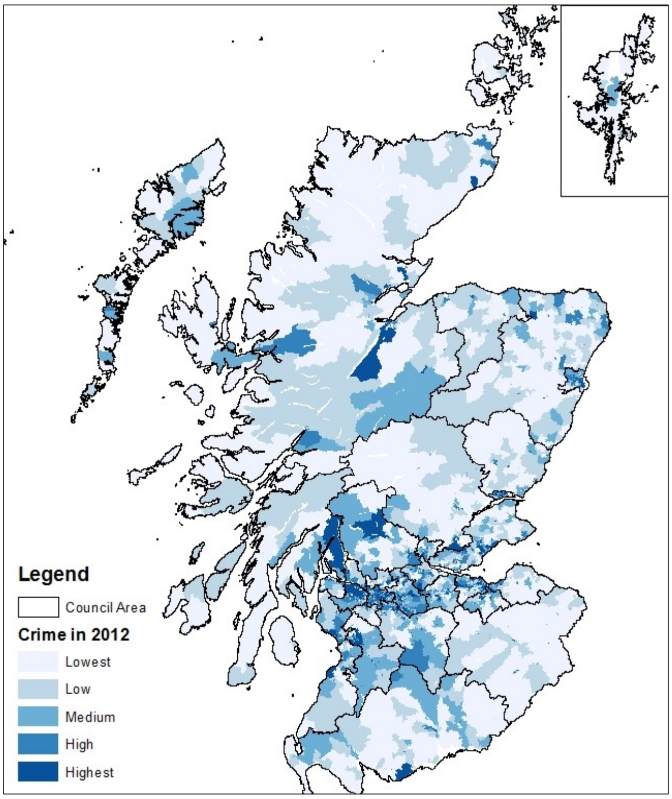

Figure 1: Reported crime in Scotland in 2010/2011 per data zone

2.3 Prescription Information System

The Information Services Division is part of the NHS National Services Scotland and hold high quality data on health service use. One of their database is the Prescription Information System (PIS), which provides prescriptions prescribed, dispensed and reimbursed in the primary care in Scotland.[11] As the NHS Scotland is universally utilized across the country, it covers nearly the entire population. For this project, data on three main categories of medicines, used for the treatment of mental health problems has been extracted between 2009 and 2015: Hypnotics & Anxiolytics (British National Formulary [BNF] 4.1), Antipsychotics and related drugs (BNF 4.2) and Antidepressants (BNF 4.3).[12] Before 2009, data quality is not considered as adequate. As mental health prescriptions are a proxy of mental health problems – antidepressants might be also prescribed for chronic pain or myalgic encephalomyelitis in various dosage –, the Information Services Division has recently developed a

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

text-mining algorithm to extract free-text dose instructions with natural language processing, in order to better identify the purpose of medication.[11]

2.4 NHS GP registration database

Lastly, this project makes use of the NHS GP postcode dataset, containing information on the postcode of residence and date of each NHS GP registration in Scotland since 2000. Although the core SLS dataset holds data on residential postcode, these are only available for the years when census information was collected (1991, 2001, 2011). This dataset provides continuous history on residential mobility, leading to a better precision of changes in time and space.

3.0 Data Linkage

3.1 The core SLS sample

The Scottish Longitudinal Study Development & Support Unit (SLS-DSU) at the University of Edinburgh is responsible for data management, maintenance of SLS but also for generating subsets of data for researchers. Tracing records and conducting data linkage between Censuses and other administrative data is done by a trusted third party (TTP). For the original dataset, sampled from the 1991 census, over 98% of participants could be traced by TTP in the NHSCR allowing further data linkage.[13] The longitudinal linkage between the Censuses is done with a combination of deterministic, probabilistic and manual matching: for the 2011 Census 84.4% of individuals were linked with deterministic, 11.0% probabilistic and 4.6% with manual methods.[14] As NHS GP postcode data is held by NHSCR, this linkage was also carried out by them in 2016. The resulting dataset required a great deal of cleaning as the original NHS GP postcode data had not been used for analyses previously.

3.2 SLS sample with other data sources

The residential location derived from the 2011 Census was used to link SIMD 2012 to the project's SLS sample. As in census form there were bar codes indicating the residential postcodes, it led to high accuracy avoiding any scanning issues. This postcode has been used to place SLS members to the corresponding data zones. For SIMD 2006 and 2009, residential address from the GP registration database has been utilized with the same linking method.

A look-up (produced by NHSCR that matches the SLS number to the NHS identifying number) for the total SLS sample is held by the electronic Data Research and Innovation Service (eDRIS) at NHS Scotland. Look-ups remain in place for a longer time in order to make quick data extraction for future projects possible. Time-to-time updates are needed to include new SLS sample members due to new entries at census, immigration and births.

Figure 2 summarize how the different datasets has been linked together and the timely distribution of the data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

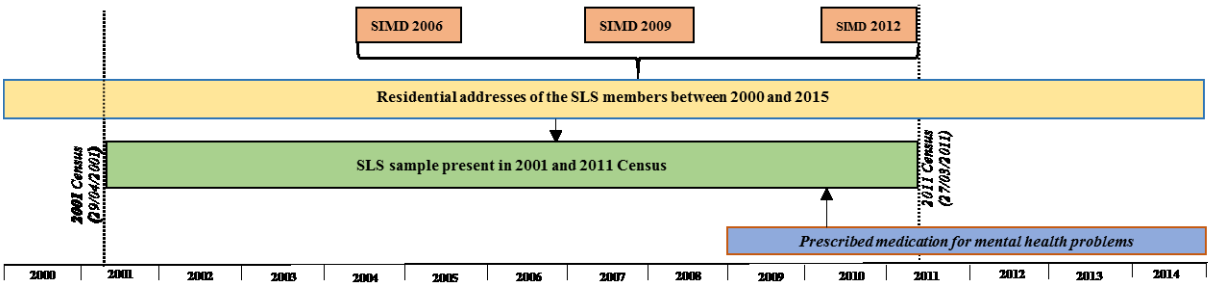

Figure 2: Model of data linkage

4.0 Data management and data security

As the linked SLS sample contains sensitive information about individuals (and for this project about their mental health status), several security measures have been introduced when the original dataset was set up, to ensure privacy and maintain confidentiality of participants.

(1) SLS is an anonymised individual-level dataset, names are not available in the dataset. Concerning the 20 birthdays, which provides the basis for the individual data sampling, only a limited number of SLS staff responsible for data management (data administrators) are aware of these dates.[9] Personal identifiers and the full postcodes (used for data linkage) are available only for SLS staff.

(2) The Steering Committee oversees the maintenance, creation and use of SLS and every project has to be approved by the SLS Research Board. Projects aiming to identify individuals are completely prohibited, similarly, researchers are not allowed to attempt to identify anyone in the dataset.[7]

(3) Before gaining access to SLS, all users have to become an Approved Researcher, which requires participation on the Safe Users of Research data Environment (SURE) training.

(4) Access to the data is strongly controlled and supervised; no raw microdata files are publicly available for users. The dataset sits on a password-protected stand-alone network, accessible in only two rooms in Edinburgh.[9] Once a project has been approved, a subset of data will be generated, containing all relevant variables requested in advance.[7] Analysis can be conducted either (a) in person at the safe setting (National Records of Scotland, Ladywell House, Edinburgh) or (b) in remote access by providing SLS staff with syntax files.

(5) Strict rules apply in the environment where the data is analysed by researchers (safe settings). Visitors are required to be escorted by SLS staff in the building and have to wear passes all the time. In the safe setting the use of electronic deceives (e.g. mobile phone, laptop) is forbidden and only a notebook provided by SLS staff can be used. The notebook is securely stored in the SLS-DSU.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

(6) Preliminary and final results have to be cleared by the SLS team before releasing them to the researchers. In order to protect confidentiality, no outputs containing possible disclosive information can be released (e.g. cell counts under 10).

5.0 Data analysis

5.1 Survival models

One of the most commonly used statistical approach to analyse longitudinal data is the survival or event history analysis, which concentrates on the time until an event of interest occurs. Unlike the name denotes, the binary outcome does not have to be failure or death. Survival times start at a fixed point, such as birth or beginning of the study, and end either at the onset of the event or at the end of the study. As the outcome often does not occur until after the end of the study period, cases without event are referred as right censored. The survival function can be calculated with the Kaplan-Meier estimator, which provides the probability of individual survival until time *t*, where $0 \leq S(t) \leq 1$.[15] For this project, the survival time until the first new medication has been calculated and tested whether the probability of a new medication differs between neighbourhoods with higher and lower crime deprivation. As comparing survival functions (e.g. with log-rank test) has several limitations while accounting for individual level covariates, a suitable analytic approach is the Cox proportional hazards regression models. It uses regression models to calculate the hazard ratio and enables to accommodate various continuous or categorical covariates.[15]

5.2 Multilevel models

Another type of analyses, which has been extensively used in the last years, are multilevel models. This statistical approach account for the fact that observations are often nested into a higher hierarchical structure (such as patients in hospitals) and subjects within the same cluster might be similar or response to an exposure similarly.[16] Using the terms of this project, two random individuals from the same neighbourhood might have similar mental health state than two random persons from two very different areas, even if we account for individual differences. Multilevel models, in our case multilevel logistic models, can partition the total individual variance into variation related to higher-level units, such as neighbourhoods, and remaining individual level variance.[16] A more advanced statistical method analysing the effect of crime on mental health would be combining survival analysis with multilevel models (multilevel survival analysis), which uses the length of time until event to calculate the probability of survival, but also able to account for the nested structure of the data by modifying the baseline hazard funcion.[17]

5.3 Data visualisation

Finally, maps with GIS techniques have been created in order to show spatial distribution of crime deprivation in Scotland (e.g. see Figure 1) and indicate how the crime pattern has been changed between SIMD 2006 and 2012.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676060

6.0 References

1. Barnett A, Zhang CJP, Johnston JM, Cerin E. Relationships between the neighborhood environment and depression in older adults: a systematic review and meta-analysis. *Int Psychogeriatr* 2017: 1-24.
2. Kim D. Blues from the neighborhood? Neighborhood characteristics and depression. *Epidemiol Rev* 2008; 30: 101-17.
3. Gong Y, Palmer S, Gallacher J, Marsden T, Fone D. A systematic review of the relationship between objective measurements of the urban environment and psychological distress. *Environ Int* 2016; 96: 48-57.
4. Joshi S, Mooney SJ, Rundle AG, Quinn JW, Beard JR, Cerda M. Pathways from neighborhood poverty to depression among older adults. *Health & Place* 2017; 43: 138-43.
5. Freeman K, Smith N. Understanding the relationship between crime victimisation and mental health: A longitudinal analysis of population data. Sydney, Australia: NSW Bureau of Crime Statistics and Research, 2014.
6. Lorenc T, Clayton S, Neary D, et al. Crime, fear of crime, environment, and mental health and wellbeing: mapping review of theories and causal pathways. *Health Place* 2012; 18(4): 757-65.
7. Boyle PJ, Feijten P, Feng Z, et al. Cohort Profile: the Scottish Longitudinal Study (SLS). *Int J Epidemiol* 2009; 38(2): 385-92.
8. Hattersley L, Boyle PJ. *The Scottish Longitudinal Study - A technical guide to the creation, quality and linkage of the 2001 Census SLS sample, LSCS Technical Working Paper 3.0*. St Andrews: Longitudinal Studies Centre—Scotland, 2008.
9. Hattersley L, Boyle PJ. *The Scottish Longitudinal Study - An Introduction, LSCS Technical Working Paper 1.0*. St Andrews: Longitudinal Studies Centre—Scotland, 2007.
10. Scottish Government. *Scottish Index of Multiple Deprivation - SIMD16 Technical Notes*, Scotland, Edinburgh, 2016.
11. McTaggart S, Nangle C, Caldwell J, Alvarez-Madrazo S, Colhoun H, Bennie M. Use of text-mining methods to improve efficiency in the calculation of drug exposure to support pharmacoepidemiology studies. *International Journal of Epidemiology* 2018; 47(2): 617-24.
12. Information Services Division. *Medicines used in Mental Health - Years 2005/06 to 2015/16.*Scotland: NHS National Services Scotland, 2016.
13. Hattersley L, Raab G, Boyle PJ. *The Scottish Longitudinal Study - Tracing rates and sample quality for the 1991 Census SLS sample*, *LSCS Technical Working Paper 2.0*. St Andrews: Longitudinal Studies Centre—Scotland, 2007.
14. Carsley, S. *Introduction and the 2011 Census data linkage [presentation]*, SLS 2011 Census launch, Edinburgh, Scotland, 2014.
15. Johnson LL. Chapter 26 - An Introduction to Survival Analysis. In: Gallin JI, Ognibene FP, Johnson LL, eds. Principles and Practice of Clinical Research (Fourth Edition). Boston: Academic Press; 2018: 373-81.
16. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. *Stat Med* 2017; 36(20): 3257-77.
17. Austin PC. A Tutorial on Multilevel Survival Analysis: Methods, Models and Applications. *Int Stat Rev* 2017; 85(2): 185-203.