

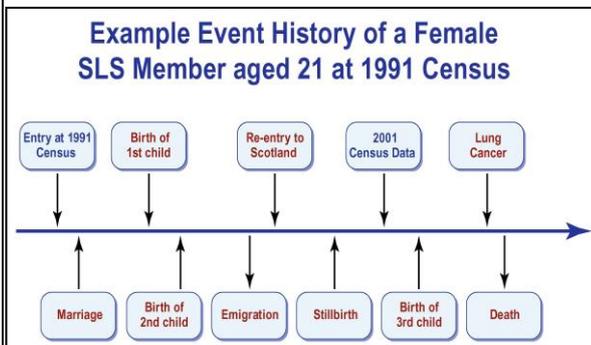
Institution: University of St Andrews



Unit of Assessment: 17 – Geography, Environmental Studies and Archaeology

Title of case study: Establishing a blueprint for administrative data based longitudinal studies in the UK

1. Summary of the impact



The Scottish Longitudinal Study (SLS) is a pioneering study, combining census, civil registration, health and education data (administrative data). It has established an approach that allows the legal and ethical use of personal, sensitive information by maintaining anonymity within the data system. This approach has become a model for the national data linkage systems that are now being established across the UK. The SLS has also enabled policy analysts to monitor key characteristics of the Scottish

population in particular health inequalities (alerting policy makers to Scotland's poor position within Europe), migration (aiding economic planning) and changing tenure patterns (informing house building decisions). Finally, the study has become fully embedded in Scotland's National Statistical agency, allowing it to produce new informative statistical series.

2. Underpinning research

Longitudinal studies have a long history in British social and epidemiological research. Most are based on surveys relying on re-interviews of the same persons over time. This can result in a high proportion of study members becoming lost to follow-up, potentially introducing important biases to the study. The Scottish Longitudinal Study (SLS) is different. It has been set up at the University of St Andrews to collect data that is either required by law (Census, birth registration, death registration, marriage registration) or is a standard administrative function within Britain (hospital admissions data). The SLS was proposed in 2001 as a key tool for the Scottish devolved administration by a team of Geographers (from the Universities of St Andrews – P. Boyle (Professor 1999-2010), R. Flowerdew (Professor 2000-2012), Dundee – A. Findlay, now at St Andrews (Professor 2011-2013) and medical researchers (from Edinburgh and Glasgow). Policy makers were convinced of its value, funding it through amongst other sources the CSO (Scotland's Chief Scientific Office) as well as ESRC and the Scottish Funding Council. The study was led by Boyle until 2009 and then by Dibben St Andrews (Reader 2004-2013).

The key underpinning research is summarised in 5 areas.

1. *Legal, ethical and governance research*

The SLS is founded on the linking of personal data for which consent cannot be practically sought between individuals. This creates a circular problem where the data used has to remain anonymous, to comply with data protection legislation, while in order to link the datasets, names and addresses have to be used. Legal and governance research (2001-2004) revealed a method based on 'firewalls' and 'Trusted Third Party' mechanisms (where the linkage is carried out by an organisation geographically separate from that managing the research dataset, this means that the research organization does not need to hold names and addresses greatly reducing the risk of disclosure) that allow linkage while also maintaining anonymity [1].

2. *Sample design and data development*

Research was carried out into suitable sampling strategies that would ensure that the sampling of birthdates across the year did not produce seasons of the year with no coverage. Considerable work was involved (2001-2004) in processing the census forms in particular, retrospectively coding the 90% of 'difficult to code' census information (e.g. occupation) that were not available electronically. Automatic systems for coding these were developed to allow cost-effective processing [1].

Impact case study (REF3b)

3. *Linking methodology*

None of the datasets that were to be linked for individuals (to produce the breadth and length of record for individuals) could be simply matched. Instead a method had to be developed that would use information such as address and date of birth to find the appropriate record for an individual (2001-2004). This method had to be sensitive to misspellings and changes to this information (i.e. people moving). We therefore developed a complex system of probabilistic and manual matching stages, all of which were implemented through a process that limited the amount of information any one organization had, to reduce the risk of information disclosure. This process was very successful, leading to final tracing and matching rates of >98% [1].

4. *Research demonstrating the utility of a census-administrative data based longitudinal study*

In order for the large investment in the setting up and running of the SLS to be made, a continuing case for the utility of such a study had to be established. Research therefore into gaps into the Scottish policy evidence base, the utility of administrative data-based research and a potential SLS methodology was undertaken and fed into the case for support for the study took place 2007 onwards after the data became available for analysis [2-4]. This led to the initial investment in the SLS by multiple funders [1] [A-C].

5. *Estimating new variables in the data*

The SLS is based on census and administrative data with variables limited to those collected in these systems. Research has therefore led to the estimation of 'synthetic measures', using a number of modeling methods, of variables of research importance 2011-13. This has included estimates of smoking propensity and income [5].

Thus, a complex system has been put in place which allows anonymous individual-level data drawn from a range of different sources to be linked and held in the SLS.

3. References to the research

The design of the SLS was set out in a series of working papers and these were then combined into a main summary paper [1]. This was then published in the *International Journal of Epidemiology*, the main journal where new cohort/ longitudinal studies are introduced. The international excellence, in terms of originality and significance, of the research on which the study is based, is recognised by this and the continuing funding of the study by the Economic and Social Research Council.

Research Grants for maintaining the SLS during the REF period

- A. ESRC 2011 Extending the Longitudinal Studies Centre - Scotland (LSCS) 2012-17 PI C. Dibben £1.5 million
- B. ESRC 2011 Extending the Longitudinal Studies Centre - Scotland (LSCS) 2011-12 PI C. Dibben £0.3 million
- C. ESRC 2009 Extending the Longitudinal Studies Centre - Scotland (LSCS) from 2009 to 2011. PI Paul Boyle £0.4 million.

1. Boyle, P., Feijten, P., Feng, F., Hattersley, L., Huang, Z., Nolan, J. and Raab, G. (2009) Cohort Profile: The Scottish Longitudinal Study (SLS), *International Journal of Epidemiology* 38(2):385-392. doi: 10.1093/ije/dyn087
2. Popham, F., Boyle, P., O'Reilly, D. & Leyland, A.H. (2011) Selective internal migration. Does it explain Glasgow's worsening mortality record? *Health & Place*, 17(6): 1212-1217. doi: 10.1016/j.healthplace.2011.08.004
3. Boyle, P., Feng, Z. & Raab, G. (2011) Does widowhood increase mortality risk? Comparing different causes of spousal death to test for selection effects. *Epidemiology* 22: 1-5. doi: 10.1097/EDE.0b013e3181fdcc0b
4. Popham, F., & Boyle, P.J. (2011) Is there a 'Scottish effect' for mortality? Prospective observational study of census linkage studies. *Journal of Public Health*, 33(3): 453-458. doi: 10.1093/pubmed/fdr023
5. Dibben, C. and Clemens, T. (2012) Estimating an occupational based wage in the census: a mixed model approach to generate empirical bayes estimates. Longitudinal Studies Centre Scotland, Research Working Paper 10.
<http://www.lscs.ac.uk/sls/working%20papers/WP10%20Dibben.pdf>

Impact case study (REF3b)

4. Details of the impact

The Scottish Longitudinal Study (SLS) has had impact in a number of significant areas in Scotland but also more widely across the UK

- It has changed National Records of Scotland's (NRS) statistical infrastructure – allowing new statistical series to be produced
- It is used by local, national government and NHS officials for policy analysis, impacting local and national policy decision making
- The study has trained over 100 researchers in longitudinal data analysis using administrative data
- The SLS data system has become a model for the newly emerging UK national administrative data infrastructure

Changed National Records of Scotland's statistical infrastructure. The SLS has been accepted as a Scottish National study and as such it is now co-supported and housed within the National Records of Scotland (NRS) - the National Statistical Agency since 2004. As a longitudinal study it replaces the need for expensive traditional longitudinal surveys collected through face-to-face questionnaires (often costing up to £10 million) [S1]. The recognition of the study by the Scottish equivalent of the Office of National Statistics as being part of the National statistical system is testament to the quality and reliability of the study. The SLS has changed the type of statistical series that NRS are producing. For example the General Registrars' report (2010) [S6], on new demographic findings, makes extensive use of the SLS. (Since 1855 - the General Registrars' report is annually laid before Parliament as the major statement on Scotland's population). NRS have used it to ask important questions about the nature of occupational coding (and therefore social class) on death certificates (a key statistic for government), investigating the potential exaggeration of someone's occupation status at death [S1]. The achievement of the SLS in Scotland has had influence across the UK so that the Northern Ireland Statistical Agency have argued that the SLS has provided a roadmap for a similar study in Northern Ireland - "*There is real SLS impact*" [S2].

Impacted local and national policy decision making. Since its creation, the SLS has also been used by analysts outside the academy to examine a wide range of research questions feeding into government social, health and housing policy. This has included, for example, reports and studies conducted on behalf of the Scottish Government [S10], Scottish Public Health Observatory [S11] and the NHS [S12]. To give two examples. Researchers in Glasgow City Council used it to investigate local patterns of housing tenure change and in particular the slowing of the fall in demand for social housing [S7]. These findings were incorporated into a demographic model of tenure change, which then fed into the research base for a number of key strategic policy documents including the Glasgow and the Clyde Valley Housing Needs and Demand Assessment, Glasgow and the Clyde Valley Strategic Development Plan, and Glasgow's Housing Strategy. A researcher working within Scottish Government working on 'return migration' [S9] work form key evidence for the Scottish Government report, 'Characteristics and intentions of immigrants to and emigrants from Scotland – Review of existing evidence' (Eirich, 2011), this in turn was discussed by Skills Development Scotland, Migrants' Rights Network, National Coalition of Anti-Deportation Campaigns and the Information Centre about Asylum and Refugees. It was also referenced in the UK Needs Analysis Report of the EU Portfolio of Integration projects.

Training researchers. In addition to the impacts of managing for government a major data base, and producing research that impacts policy in the areas of health, education and employment, a third impact has been in training people in longitudinal data analysis and in supporting non-academic research use of the SLS. St Andrews researchers have been pro-active in organising training for those outside the academy wishing to access the SLS. Since 2008, 14 training events have been organised by the SLS team in Edinburgh, Belfast, Glasgow, London, Stirling, and St Andrews. In total 125 non-academic users have been trained including 4 people from local authorities, 6 from health boards, 110 from various sectors of the Scottish Government, 5 from Charities or private consulting firms. Given the relatively small community of quantitative social scientists in Scotland, this represents a good proportion of potential users. As a result of this training, 9 longitudinal research projects have been launched by non-academics in the fields of health inequalities, migration and employment.

Model for the newly emerging UK national administrative data infrastructure. The SLS has become a path-breaking model that allows the linkage, holding, and analysis of highly personal data within appropriately strict legal and ethical constraints. For example, the Scottish Government use it as an exemplar of good practice in their development of national Data Sharing and Linkage Service, “*The SLS has been of absolute fundamental importance to the development of the new National Data Sharing and Linking Service*” [S3]. A senior member of Scottish Government argues “*It is fair to say that the Scottish Longitudinal Study was a vital element in Scotland being able to make practical progress in this area quickly. This is because it had a solution that had been developed, tested and trusted. In particular, our thinking about governance of privacy and ethics issues is derived from that used for the SLS. The same is true for the processes of indexing and linking datasets themselves. These practical considerations would have taken much more time to think through if the SLS hadn’t been around, risking frustration from Ministers and loss of momentum*” [S4].

The SLS has become a very important model for other parts of the UK that are seeking to produce similar studies. The Administrative Data Taskforce (ADT) (making recommendations to David Willets, the Minister of State for Universities and Science, and BIS over the future of UK wide research infrastructure) has used the design of the SLS as a model for future UK-wide research centres) [S5]. The ADT argued that future “*ADRC [Administrative Data Research Centres] could build on best practice from the experience of the ...Scottish Longitudinal Study (SLS)*” (p.5 [S10]) and have a “*data linkage process ... similar to that used by the Scottish Longitudinal Study (SLS), where personal identifying information is not held in the ADRC, but is matched through a third party service, such as the National Health Service Central Register*” (p.6). One senior adviser to the ESRC comments that “*The design, direction and future ambitions of the Scottish Longitudinal Study, together with its impressive achievements to date, have provided the prototype for the bold step forward that is now being taken by the Economic and Social Research Council, the statistical authorities of the UK and government departments. There is no doubt in my mind that if we did not have this valuable experience and example to draw on, we would have been much less likely to have attracted the capital funding gained to establish the Administrative Data Research Network.*” [S5].

5. Sources to corroborate the impact

Archived communication or agreed referee corroborating the use of SLS for data development within the respective governmental organisations.

[S1] Head of Department, National Record Office for Scotland, Scottish Government.

[S2] Head of Department, Northern Ireland Statistical Agency.

[S3] Head of the Scottish Government’s Data Sharing and Linkage Service.

[S4] Senior government official, Scottish Government.

[S5] Senior academic adviser to the UK Economic and Social Research Council.

Reports/ papers

[S6] Scotland’s Population 2010: The Registrar General’s Annual Review of Demographic Trends 156th Edition. <http://www.gro-scotland.gov.uk/files2/stats/annual-review-2010/j176746-00.htm>

[S7] Jan Freeke “Housing tenure change 1991-2001 in Scotland, Glasgow Conurbation & Glasgow City”. Glasgow City Council. <http://calls.ac.uk/wp-content/uploads/CALLS-Impact-Case-Study-1-%E2%80%93-Jan-Freeke.pdf>

[S8] McCollum, D. (2011) *The Demographic and Socio-Economic Profile of Return Migrants and Long-Term In-Migrants in Scotland: Evidence from the Scottish Longitudinal Study*. Scottish Government Social Research Report.

<http://www.scotland.gov.uk/Resource/Doc/341539/0113589.pdf>

[S9] The UK Administrative Data Research Network: Improving Access for Research and Policy Report from the Administrative Data Taskforce – December 2012

http://www.esrc.ac.uk/images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf

[S10] Kirsty Corbett & Alan Winetrobe “Once a NEET, Always a NEET” Scottish Government.

http://sls.lscs.ac.uk/projects/view/2008_007/

[S11] Diane Stockton “The determinants of self-assessed health in Scottish adults” Scottish Public Health Observatory. http://sls.lscs.ac.uk/projects/view/2009_006/

[S12] Katharine Sharpe “Area-based versus individual measures of socioeconomic background – How do they compare in predicting cancer incidence?” NHS Scotland, Information Services Division. http://sls.lscs.ac.uk/projects/view/2009_005/