

Chapter 1

Data Linkage, Information Governance, and Data Sources

1.1 Introduction

This chapter outlines the process of obtaining administrative data suitable to answer the thesis research questions. A brief description of data linkage research and the associated advantages and disadvantages of this approach are first discussed. The following section provides a detailed description of the strict information governance protocols required including: the infrastructure used, the approvals process, and the legal framework enabling the research to take place. Finally, a description of the data sources used in the research is provided.

1.2 Data Linkage

Record linkage refers to a merging that brings together information from two or more sources of data with the object of consolidating facts concerning an individual or an event that are not available in any separate record.

[OECD, 2006]

Administrative data is data that is generated when individuals use a service of some description. Often in research terms, and exclusively in this thesis, administrative data

refers to data generated by the use of *public* services (Pavis and Morris, 2015; Mazzali and Duca, 2015). This data can describe the provision of a specific service or how it was administered by the provider (Pavis and Morris, 2015; Mazzali and Duca, 2015). As the above definition outlines, record linkage involves joining data about individuals from two or more administrative databases together (Fleming, Kirby and Penny, 2012; Harron, 2016) and is being increasingly used in social science research (Atherton *et al.*, 2015; Bell *et al.*, 2016).

Using administrative data for research purposes has a number of advantages and disadvantages. The data is not collected for research purposes and as such may lack specific information relevant to a researcher's line of inquiry (Mazzali and Duca, 2015). This also reduces the ability of a study to adjust for all potential confounding variables, decreasing the ability to make causal inferences from the data (Mazzali and Duca, 2015). There is the potential for ambiguity about the coding of variables in a database and what each code represents (Mazzali and Duca, 2015; Atherton *et al.*, 2015; Walesby, Harrison and Russ, 2017) which means specialist knowledge of the database and collection methods are required (Mazzali and Duca, 2015). Administrative databases also have the potential to contain data of questionable quality and high levels of missing data (Walraven and Austin, 2012; Hashimoto *et al.*, 2014; Harron *et al.*, 2017). Data can be missing for the same reasons as seen in other forms of research but, in addition, individuals may also be missing due to failure to interact with a service or because insufficient information was available to accurately match records during the data-linkage process (Harron *et al.*, 2017).

Advantages of administrative databases are that they enable large, often population sized, sample sizes because they are generated from service use (Mazzali and Duca, 2015; Pavis and Morris, 2015; Walesby, Harrison and Russ, 2017). This characteristic also reduces the potential for sampling bias (Mazzali and Duca, 2015). Well maintained administrative data can offer information over long periods of time including very recent data (Pavis and Morris, 2015). This can make inferences from research findings more robust with excellent levels of external validity without the extra cost traditional observational studies might incur (Mazzali and Duca, 2015; Harron *et al.*, 2017). The potential to link administrative databases from a number of sources is a significant advantage and offers insights into how services interact (Mazzali and Duca, 2015; Atherton *et al.*, 2015; Walesby, Harrison and Russ, 2017).

There are two main methods of linking data from disparate sources; deterministic matching and probabilistic matching (Fleming, Kirby and Penny, 2012; Harron, 2016; Doidge and Harron, 2018). Where differing datasets possess common unique identifiers, deterministic matching simply links data using this identifier. Probability matching methodology can be employed in the absence of a common unique identifier (Fleming, Kirby and Penny, 2012; Harron, 2016; Doidge and Harron, 2018). Using this method, a probability of two records

matching correctly is calculated based on how well the records match based on a set of common partial identifiers such as name, date-of-birth, and postcode (Fleming, Kirby and Penny, 2012; Harron, 2016; Doidge and Harron, 2018). An important consideration when using probabilistic linkage is making an assessment of false-positive match rates (Fleming, Kirby and Penny, 2012; Harron, 2016; Doidge and Harron, 2018). There are three main strategies to assist with this assessment; measuring error using “gold-standard” data (such as a validated external datasets), sensitivity analyses (comparing results across differing linkage parameters), and comparing linked and unlinked data according to characteristics (such as sociodemographic subgroups) (Harron, 2016).

Scotland is home to some of the best administrative databases in the world (Pavis and Morris, 2015). This is particularly due to the high-quality of health datasets that have been collected and maintained for over 40 years (Fleming, Kirby and Penny, 2012; Pavis and Morris, 2015). Whilst linkage of differing health datasets has become common over this period, new cross-sectoral linkages are beginning to emerge such as health and educational data (Wood *et al.*, 2013), and health and social care data (Witham *et al.*, 2015). These cross-sectoral linkages are providing new insights that have the potential to have lasting impact on policy and provision of services (Pavis and Morris, 2015; Atherton *et al.*, 2015).

1.3 Information Governance

1.3.1 Introduction

Confidentiality of data subjects is an important consideration in any data linkage project. The benefits of administrative data linkage, outlined in section 1.2, are dependent on research being conducted in a legally and ethically competent fashion. Whilst full anonymisation would be an effective way to protect data subjects confidentiality, it is almost impossible to achieve this with individual-level data suitable for research purposes (Harron *et al.*, 2017).

As an alternative, a process involving robust approvals review, researcher training with associated responsibilities and sanctions, and safe haven settings are used to preserve data subject confidentiality (Harron *et al.*, 2017). This section outlines how this process was applied for the purposes of the data linkage completed in this thesis. Firstly, the various organisations that provide the infrastructure that enabled the data linkage to take place are briefly described. An overview of the various approvals and ethical panels is then provided, followed by the legal framework which enabled data processing to take place with a brief description of how confidentiality is maintained during the linkage process.

1.3.2 Infrastructure

The key test for an acronym is to ask whether it helps or hurts communication.

Elon Musk

1.3.2.1 Scottish Informatics and Linkage Collaboration

The Scottish Informatics and Linkage Collaboration (SILC) is an umbrella term for a number of support services that are available to individuals wishing to conduct research using linked administrative data (SILC, 2017c). Services include computing resources (provided by the University of Edinburgh), research and project coordination advice (provided by the electronic Data Research and Innovation Service (eDRIS)), and an indexing service (provided by the National Records of Scotland (NRS)) (SILC, 2017c). SILC currently has three partner institutions; the Administrative Data Research Centres (ADRC), the Farr Institute, and the Urban Big Data Centre (UBDC) (SILC, 2017c).

1.3.2.2 Urban Big Data Centre

Funding for this PhD was provided by the Scottish Government and the Economic & Social Research Council (ESRC). The bid for funding was won by UBDC which is based within the University of Glasgow. UBDC as an organisation is also funded by the ESRC and brings together data scientists and social scientists with research interests relevant to urban living such as; housing, transport, migration, and health (UBDC, 2017b). UBDC has six partner universities; Edinburgh, Bristol, Cambridge, Reading, Sheffield, and Illinois-Chicago.

The linkage project described in this thesis was completed with the assistance of UBDC's controlled data service. This service helps researchers to access personal data that exists in administrative databases (UBDC, 2017a). In addition to a vigorous approval process, access to data is tightly controlled via safe haven IT architecture which monitors use of data and output of analyses to ensure individual anonymity is maintained (UBDC, 2017a). UBDC arranges access to the safe haven environment through liaison with eDRIS, provided by the Information Services Division (ISD) of NHS National Services Scotland (NSS) under the auspices of SILC. A more detailed description of the safe haven is given in section 1.3.4.2.

1.3.2.3 electronic Data and Research Innovation Service

ISD is a subdivision of NHS NSS (ISD, 2010b). NSS is a national NHS board in its own right and works with the other NHS boards, particularly the 14 geographic health boards, to provide centralised services such as; procurement, legal support, IT, and public health intelligence (NSS, n.d.). As a division of NSS, ISD provides, among other things, support for the latter two of these services (ISD, 2010b). This includes administering the large number of databases containing information on health service use in Scotland varying from maternity & births, to cancer services (ISD, 2010b). ISD held databases used in this thesis, the Prescribing Information System and Urgent Care Data Mart, are described more fully in sections 1.4.4 and 1.4.5.

eDRIS is part of ISD and provides services under SILC (SILC, 2017b). It is detailed specifically with assisting research using health administrative datasets. Researchers using the eDRIS service have a named research assistant who provides advice on; data sources, study design, the information governance approvals system, access to the safe haven environment, and review of analysis outputs to ensure disclosive information cannot be inferred (ISD, 2010b).

1.3.2.4 National Records of Scotland

NRS collects and maintains information about the people of Scotland including births, deaths, and marriages (NRS, 2017). In addition to producing annual reports and population estimates, NRS provides the indexing service under SILC which enables anonymous linking of administrative databases as a Trusted Third Party. This is made possible using the NHS Central Register (NHSCR) which is held by NRS (NRS, 2018a) and described more fully in section ?? .

1.3.2.5 Health and Social Care Analysis

Health and Social Care Analysis (HSCA) is a division within the Scottish Government that provides statistic, economic, and research evidence to inform policy making in this area (Scottish-Government, 2017a). It is one of many Analytical Service Divisions (ASD) that provide analytical support, advice, and briefing to the Government. HSCA creates reports and publications in a number of key areas including; social care, care homes, and mental health and is responsible for collecting and publishing the Social Care Survey (SCS) described in section 1.4.2.

1.3.3 Approvals

As described in section 1.3.1, one of the ways in which data subject confidentiality is maintained in data linkage projects is through a rigorous and robust approvals process. Three separate approvals were required for the purposes of data linkage in this thesis.

1.3.3.1 Research Approvals Committee

Data linkage for the project was facilitated by UBDC's controlled data service. In order to utilise this service, the research proposal required approval from the UBDC Research Approvals Committee (RAC). A full list of RAC members is available online (UBDC, 2017a). This committee is independent of UBDC and approves use of funds and infrastructure in UBDC and includes a lay member of the public (UBDC, 2017a). An application to use the controlled data service is judged on its academic merit, public benefit, skill of research team, and alignment with UBDC aims before being approved (UBDC, 2017c).

The approval for the main linkage project is shown in Appendix C.

1.3.3.2 Ethics

Ethical approval for data analysis was sought and gained from the University of Glasgow College of Social Sciences Research Ethics Committee (CoSS REC). A blanket ethical approval, obtained by eDRIS from the NHS East of Scotland REC, covers research that uses NHS Health data, does not involve direct contact with data subjects, has peer-review approval, stores data in the national safe haven, and is conducted by research teams based in the UK (ISD, 2010a). The main linkage project therefore only required further approval from CoSS REC to cover the non-health related data (i.e. the Social Care Survey).

The CoSS REC approval letter for the main linkage project is shown in appendix E.

1.3.3.3 Public Benefit and Privacy Panel for Health & Social Care

In addition to RAC and College ethical approval, the main thesis project also required clearance from the Public Benefit and Privacy Panel for Health & Social Care (PBPP). This was because data from NHS sources were being used. The PBPP acts as a decision making body with delegated responsibility from NHS Scotland Chief Executive Officers and the Registrar General (Scottish-Government, n.d.). Using terms of reference and guiding principles, the panel adjudicates whether research projects using administrative data generated by the NHS in Scotland can be used for research purposes. The panel ensures that the basis for disclosing data has a clear public benefit and ensures the legal framework for accessing and processing data is sound.

The approval letter for the main thesis project is shown in Appendix D, a fuller description data processing and the legal basis for this is presented in section 1.3.4.

1.3.4 Data processing

1.3.4.1 Legal framework

The permissions and linkage of data for this project were completed in advance of the European Union (EU) General Data Protection Regulation (GDPR) coming into effect in May 2018. The information governance was informed by antecedant laws including the Data Protection Act (DPA) (1998). However, as the study period was known to overlap with the implementaion of GDPR, all legal documentation was completed to ensure compliance with the incoming regulation.

Data sharing and processing can be completed without consent of data subjects as long certain criteria, explicitly named in legislation, are met (Bell *et al.*, 2016). For the purposes of this thesis fair processing of data was completed, without consent, in accordance with three legislative paragraphs:

- Schedule 2:(6) of the DPA.
 1. The processing is necessary for the purposes of legitimate interests pursued by the data controller or by the third party or parties to whom the data are disclosed, except where the processing is unwarranted in any particular case by reason of prejudice to the rights and freedoms or legitimate interests of the data subject.
- Schedule 3:(8) of the DPA (emphasis added)
 1. The processing is necessary for medical purposes and is undertaken by
 - (a) a health professional, or
 - (b) a person who in the circumstances owes a duty of confidentiality which is equivalent to that which would arise if that person were a health professional.
 2. In this paragraph “medical purposes” includes the purposes of preventative medicine, medical diagnosis, *medical research*, the provision of care and treatment *and the management of healthcare services*
- Paragraph 9 of the Data Protection (Processing of Personal Data Order 2000(SI 2000 No.417)).

The project has clear and substantial public interest in the information it will provide to inform the delivery of public services. The data processing is necessary to enable this research to take place. The project does not support measures of analysis with respect to any individual. Finally, the project will not cause any substantial damage or distress to any individual.

Lawful processing of data for the purposes of the project is in accordance with a further two legal acts:-

- Social Work (Scotland) Act 1968.

8 Research

1. The Secretary of State may conduct or assist other persons in conducting research into any matter connected with his functions or the functions of local authorities in relation to social welfare, and with the activities of voluntary organisations connected with those functions.
2. Any local authority may conduct or assist other persons in conducting research into any matter connected with their functions in relation to social welfare.
3. The Secretary of State and any local authority may make financial assistance available in connection with any research which they may conduct or which they may assist other persons in conducting under the provisions of this section.

- National Health Service (Scotland) Act 1978

47 Education and research facilities. (2)Without prejudice to the general powers and duties conferred or imposed on the Secretary of State under the Scottish Board of Health Act 1919, the Secretary of State may conduct, or assist by grants or otherwise any person to conduct, research into any matters relating to the causation, prevention, diagnosis or treatment of illness, or into such other matters relating to the health service as he thinks fit."

Information governance for the project was also informed by the “Guiding Principles for Data Linkage” report produced by the Scottish Government (2012). These guidelines, themselves informed by legislation such as the DPA, highlight the importance of public interest, transparency, and privacy when conducting data linkage projects with publicly held data sets.

In order to preserve anonymity whilst linking administrative data from different agencies, a method known as “linkage using a separation of functions” is employed (Pavis and

Morris, 2015; Harron, 2016). This process involves the use of a Trusted Third Party (TTP) to process non-anonymised information in order to link more than one dataset together. The TTP receives personal information (e.g. names, addresses and dates-of-birth) from the data controllers of the administrative datasets to be used and creates index “keys” to send back to the data controllers to attach to their data (Pavis and Morris, 2015; Harron, 2016). The TTP creates a lookup table of index “keys” relevant to each dataset and sends these to a linkage agent. The linkage agent receives data from the data controllers *without* personal information and links them together using the “keys” created by the TTP and makes this available to a researcher in a secure environment (Pavis and Morris, 2015; Harron, 2016). This process means the TTP receives lots of personal information but no other information, the researcher has access to information relevant to their study but no personal information, and the data controllers share information about individuals in their datasets without compromising anonymity and without seeing data from each others databases (SILC, 2017a).

1.3.4.2 Safe Haven environment

Another integral part of ensuring the confidentiality of data subjects within large, linked administrative data is by holding such data in a safe haven environment (Harron *et al.*, 2017). As described in section 1.3.2.2, access to data for this thesis was administered via UBDC’s controlled data service and further liaison with eDRIS to enable use of the NSS National safe haven. All data shared for the purposes of the thesis was transferred by data controllers to the safe haven by secure file transfer protocol.

The safe haven enables secure data storage and access via a Virtual Private Network (VPN) connection with strict access control. This environment does not enable external access of any kind i.e internet or saving & printing facilities (ISD, 2010c). In order to retrieve output of analyses, work was submitted for statistical disclosure control which was conducted by eDRIS employees. This process ensures that data taken out of the safe haven cannot be used, either on its own or by being combined with other data, to breach the privacy of any individual (ISD, 2010c; Harron *et al.*, 2017). A full guide to statistical disclosure control is provided by Lowthian & Ritchie (2017).

1.3.4.3 Data sharing agreement

For the purposes of the main linkage project, a three-way data sharing agreement (DSA) between the University of Glasgow, NHS National Services Scotland, and Scottish Ministers (Scottish Government) was signed. This detailed the purpose of data sharing, as well as the transfer, protection, and security of data. The roles and responsibilities of each

organisation in relation to relevant data protection legislation are clearly detailed in the DSA. (include as appendix??)

1.4 Data Sources for Linkage

1.4.1 Introduction

Research conducted with administrative data requires a thorough description of databases used (Walraven and Austin, 2012). This should include a description of the purpose of the data collection and the methods employed to collect data. This enables appraisal of any potential biases that may exist within the databases (Walraven and Austin, 2012; Mazzali and Duca, 2015). There are 4 main sources of data used in the main analyses of this thesis: the Social Care Survey, the Scottish population spine and death records, the Prescribing Information System, and the Unscheduled Care Data Mart. Each of these are described in more detail below.

1.4.2 Social Care Survey

The Social Care Survey (SCS) is collected annually by HSCA for the Scottish Government to provide descriptive statistics of the amounts of social care delivered by each of Scotland’s 32 local authorities (Scottish-Government, 2017b). Results are collated and published annually by HSCA in the “Social Care Services, Scotland” report (Scottish-Government, 2017b). The SCS reports provide an overview of social care services for the public and policy makers. In addition, certain measures captured by the SCS are used in funding formulae to calculate allocation of resources to each local authority (e.g. number of people receiving home care) (Scottish-Government, 2016a).

All 32 Scottish local authorities collect information on social care as part of their management systems (Scottish-Government, 2016b). HSCA produce a data specification document outlining the information that should be returned for the social care survey and this is sent to HSCA via a secure web-based system called ProcXed. This system supports data validation checks on transfer to improve data accuracy (Scottish-Government, 2016b).

The SCS has been collected in its present form since 2013 as a combination of two previous data collections - the Home Care Census and the Self Directed Support (Direct Payments) Survey (Scottish-Government, 2016b). Individual-level data has been collected since 2010. Some questions have remained constant throughout this period but there have also been some changes in definitions and measures.

The most recent SCS (2017) collected information on on all individuals that received community alarm, telecare, self directed support (SDS) or social work/ support worker services during the previous financial year. In addition, individuals that received home care services, meals, housing support, shopping, or laundry services during a specified census week are included in the survey.

Before 2013, information on telecare and community alarm services was only collected for individuals receiving these services during the census week. Furthermore, the value for total weekly hours of housing support services was included in the value of home care services. The separation of these services acknowledges that housing support is often regarded as 24-hour-a-day-7-day-a-week service (Scottish-Government, 2016b).

For the value of total weekly hours of home care, HSCA request details on scheduled and actual hours of care delivered. Some local authorities are able to return both values, others only return one value. Where both are returned, actual hours of home care are used in official reporting (Scottish-Government, 2016b). Approximately 129,000 people received community alarms and/or telecare services, approximately 60,000 received home care services, and approximately 8,000 received SDS funding in 2016/17 (Scottish-Government, 2016b). The overlap of individuals who receive more than one of these services is unknown.

The cross-sectional nature of the survey, and in particular the census week variables, mean that the SCS only collects a sample of the entire population that receive social care in Scotland in any given financial year. It is unknown how large this sample is. It is also impossible to infer whether the values of total weekly hours of home care delivered to individuals is representative of the care they receive throughout the financial year. In order to gain a better understanding of these issues, an exploratory project, using a richer data set, was conducted as part of this thesis and is reported in chapter ??.

1.4.3 NHS Central Register and Death records

The ‘Research population spine’ is a copy of the National Health Service Central Register (NHSCR) and is controlled by NRS (NRS, 2018a). The spine contains over 9 million records and is updated quarterly. There are very strict policies which govern access to the offline secure server where the spine is stored.

The NHSCR is used operationally for the purpose of transferring GP records (NRS, 2018a). Despite the name and the fact it is used for transferring patients records, the population spine does not hold medical records themselves. It contains records of anyone who was born in Scotland, who registers with a GP in Scotland, or who dies in Scotland. Variables included are: forename, middle name, surname, date of birth, gender, postcode, and country of birth.

This resource is extremely useful for administrative data linkage projects. Data sources that don't contain a CHI number (such as the SCS) can be matched to the population spine using deterministic and probabilistic methods (see section 1.2). This then allows linkage to CHI-based data sources enabling cross-sectoral projects, such as the main analyses in this thesis, to be possible.

NRS also collates the register of deaths which includes details of every death in Scotland since 1855 (NRS, 2018b). Details on the date of death were requested for all individuals in the thesis study cohort.

1.4.4 Prescribing Information System

The Prescribing information system (PIS) contains all community prescribed medicines for every individual in Scotland from 2009 onwards. Data is collected to provide payment to community pharmacies for the medicines dispensed to the population. The data base can be linked to health sources via the Community Health Index (CHI) number. A full description of the PIS database and its applicability to research has been published by Alvarez-Madrazo et al (2016).

1.4.5 Urgent Care Data

The Urgent Care Data Mart (UCD) is a database collated by ISD for the purpose of understanding a patient journey through emergency and urgent care services (ISD, 2017). It is a linkage of routine health data from a number of sources controlled by ISD: NHS24 telephone triage service, Scottish Ambulance Service (SAS), primary care out-of-hours services (PC OOH), Accident & Emergency (A & E), acute emergency inpatient admissions (both general and mental health), and deaths. Data is available from 2011 with the exception of PC OOH data which is available from 2014. CHI numbers are available on all records.

Continuous Urgent Care Pathways (CUPs) are calculated that join together records from each of these sources that occur within 24 hrs of each other (or for services occurring within 48hrs of an acute emergency admission)(ISD, 2017). Details of all variables listed in UCD are available in the background paper published by ISD (2017). In addition to service use, UCD flags presence of any of 14 long-term health conditions in any of the above datasets and, additionally, any acute admission from 1981 onwards.

1.5 Timeline

Describe timeline of approvals process - major hurdles and barriers. **Maybe in methods??**

1.6 Conclusion

Brief summary of chapter here

References

- Alvarez-Madrazo, S. *et al.* (2016) ‘Data resource profile: The scottish national prescribing information system (pis)’, *International journal of epidemiology*, p. dyw060. Available at: <http://ije.oxfordjournals.org/content/early/2016/05/09/ije.dyw060.extract>.
- Atherton, I. M. *et al.* (2015) ‘Barriers and solutions to linking and using health and social care data in scotland’, *British Journal of Social Work*, 45(5), pp. 1614–1622. doi: 10.1093/bjsw/bcv047.
- Bell, J. *et al.* (2016) *Legal issues for adrn users*. Report. Administrative Data Research Network. Available at: https://adrn.ac.uk/media/174198/legal_guide_final.pdf.
- Doidge, J. and Harron, K. (2018) ‘Demystifying probabilistic linkage: Common myths and misconceptions’, *International Journal of Population Data Science*, 3(1), pp. 1–8. Available at: <https://ijpds.org/article/view/410/384>.
- Fleming, M., Kirby, B. and Penny, K. I. (2012) ‘Record linkage in scotland and its applications to health research’, *Journal of clinical nursing*, 21(19pt20), pp. 2711–2721. Available at: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2702.2011.04021.x/full>.
- Harron, K. (2016) *An introduction to data linkage*. Report. University of Essex.
- Harron, K. *et al.* (2017) ‘Challenges in administrative data linkage for research’, *Big Data and Society*, 4(2). Available at: <http://journals.sagepub.com/eprint/wghywxV6WvbGZtRzjE5R/full#articleShareContainer>.
- Hashimoto, R. E. *et al.* (2014) ‘Administrative database studies: Goldmine or goose chase?’, *Evidence-based spine-care journal*, 5(2), p. 74. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4174180/pdf/10-1055-s-0034-1390027.pdf>.
- ISD (2010a) *EDRIS frequently asked questions*. Report. Available at: <http://www.isdscotland.org/Products-and-Services/eDRIS/FAQ-eDRIS/index.asp#d4>.
- ISD (2010b) *Information services division*. Report. Available at: <http://www.isdscotland.org/>.
- ISD (2010c) *Use of the nss national safe haven*. Report. Available at: <http://www>.

isdscotland.org/Products-and-services/Edris/Use-of-the-National-Safe-Haven/.

ISD (2017) *Urgent care data mart (ucd) - background paper*. Report. Available at: http://www.isdscotland.org/Health-Topics/Emergency-Care/Patient-Pathways/UrgentCareDataMartBackgroundPaper_20171002.pdf.

Lowthian, P. and Ritchie, F. (2017) *Ensuring the confidentiality of statistical outputs from the adrn*. Report. Available at: https://www.adrn.ac.uk/media/174254/sdc_guide_final.pdf.

Mazzali, C. and Duca, P. (2015) ‘Use of administrative data in healthcare research’, *Internal and emergency medicine*, 10(4), pp. 517–524. Available at: <https://link.springer.com/article/10.1007/s11739-015-1213-9>.

NRS (2017) *About us*. Report. Available at: <https://www.nrscotland.gov.uk/about-us>.

NRS (2018a) *NHS central register (nhscr)*. Report. Available at: <https://www.nrscotland.gov.uk/statistics-and-data/nhs-central-register>.

NRS (2018b) *Statutory registers of deaths*. Report. Available at: <https://www.nrscotland.gov.uk/research/guides/statutory-registers/deaths>.

NSS (n.d.) *How nss works*. Report. Available at: <https://nhsnss.org/how-nss-works/>.

Pavis, S. and Morris, A. (2015) ‘Unleashing the power of administrative health data: The scottish model’, *Public Health Research and Practice*, 25(4), p. e2541541. Available at: <http://www.phrp.com.au/issues/september-2015-volume-25-issue-4/unleashing-the-power-of-administrative-health-data-the-scottish-model/>.

Scottish-Government (2012) *Guiding principles for data linkage*. Report.

Scottish-Government (2016a) *2016-17 settlement. grant aided expenditure green book*. Report. Available at: <http://www.gov.scot/Resource/0049/00499184.pdf>.

Scottish-Government (2016b) *Social care services, scotland, 2016*. Report. Available at: <https://www.gov.uk/government/statistics/social-care-services-scotland-2016>.

Scottish-Government (2017a) *Inpatient census, 2017*. Report. Available at: <http://www.gov.scot/Resource/0052/00524621.pdf>.

Scottish-Government (2017b) *Social care survey*. Report. Available at: <http://www.gov.scot/Topics/Statistics/Browse/Health/Data/HomeCare>.

Scottish-Government (n.d.) *Public benefit and privacy panel for health and social care*. Report. Available at: <http://www.informationgovernance.scot.nhs.uk/pbphsc/>.

SILC (2017a) *Data linkage safeguards*. Report. Available at: <http://www>.

datalinkagescotland.co.uk/data-linkage-safeguards.

SILC (2017b) *EDRIS*. Report. Available at: <http://www.datalinkagescotland.co.uk/edris>.

SILC (2017c) *Home*. Report. Available at: <http://www.datalinkagescotland.co.uk/>.

UBDC (2017a) *Controlled data service guide for researchers [last updated 01.02.2017]*. Report. Available at: <http://ubdc.ac.uk/media/1445/ubdc-controlled-data-services-guide-for-researchers-v5.pdf>.

UBDC (2017b) *Overview [accessed 14.11.2017]*. Report. Available at: <http://ubdc.ac.uk/about/overview/>.

UBDC (2017c) *Research approvals committee*. Report. Available at: <http://ubdc.ac.uk/about/overview/research-approvals-committee/>.

Walesby, K., Harrison, J. and Russ, T. (2017) ‘What bid data could achieve in scotland’, *Journal of the Royal College of Physicians of Edinburgh*, 47(2), pp. 114–119. doi: 10.4997/JrCPe.2017.201.

Walraven, C. van and Austin, P. (2012) ‘Administrative database research has unique characteristics that can risk biased results’, *Journal of clinical epidemiology*, 65(2), pp. 126–131. Available at: <https://www.sciencedirect.com/science/article/pii/S0895435611002484>.

Witham, M. D. *et al.* (2015) ‘Construction of a linked health and social care database resource—lessons on process, content and culture’, *Inform Health Soc Care*, 40(3), pp. 229–39. doi: 10.3109/17538157.2014.892491.

Wood, R. *et al.* (2013) ‘Novel cross-sectoral linkage of routine health and education data at an all-scotland level: A feasibility study’, *The Lancet*, 382, p. S10. Available at: <http://www.sciencedirect.com/science/article/pii/S0140673613624356>.