Ymchwil gymdeithasol

Social research

Number: 71/2015

www.cymru.gov.uk

Review into the capability of the 'Secure Anonymised Information Linkage' (SAIL) Databank to provide data for the Social Services National Outcomes Framework for people who need care and support and carers who need support

Final Report

AUTHORS

Karen Tingay. Swansea University
Martin Heaven, Swansea University
Sarah Lowe, Welsh Government

For further information please contact:

Name: Sarah Lowe
Department: Knowledge and Analytical Services
Welsh Government
Cathays Park
Cardiff
CF10 3NQ

Tel: 02920 826229
Email: sarah.lowe@wales.gsi.gov.uk

Welsh Government Social Research, 17 December 2015

ISBN 978-1-4734-5489-7

Table of Contents

Chapter 1 Introduction

Background

1.1 Programme for Government sets out the Welsh Government's manifesto commitments and how it plans to deliver on these commitments. "It represents a real commitment to delivery, measured by the impact government is actually having on people's lives.[1]" Sustainable Social Services for Wales (SSSW): A Framework for Action builds on the Programme for Government commitments in the social care area and puts in place a framework for meeting the challenges facing social services in the next decade and beyond, and sets out the priorities for action.

1.2 The Social Services and Well-being (Wales) Act 2014 sets out a definition of well-being, that covers all aspects of a person's life. Well-being is defined in the Act as:

- physical and mental health and emotional well-being;
- protection from abuse and neglect;
- education, training or recreation;
- domestic, family and personal relationships;
- contribution made to society;
- securing rights and entitlements;
- social and economic well-being; and
- suitability of living accommodation.

1.3 Section 8 of the Act requires Welsh Ministers to issue a statement of well-being outcomes to be achieved under each aspect of well-being, defined in the Act. A draft well-being statement[2] for people who need care and support and carers who need support has been published in advance of the commencement of the Act in April 2016.

1.4 The well-being statement forms a part of the national outcomes framework for social services. This framework will set national direction and promote the well-being of people in Wales who need care and support, and carers who need support. It will also provide greater transparency on whether care and support services are improving well-being outcomes for people in Wales, using consistent and comparable indicators.

1.5 The framework describes the well-being outcomes for people who need care and support and carers who need support as defined in the Social Services and Well-being (Wales) Act, and outlines how the Welsh Government will measure whether these outcomes are being achieved. The well-being outcomes apply to all people who need care and support and carers who need support, including children, young people, adults and carers.

1.6 The framework will continue to be developed through consultation with stakeholders and with people and will be formally issued by Welsh Ministers and laid before the National Assembly on commencement of the Act in April 2016.

Project aims and objectives

1.7 The aim of the Project was to identify whether additional outcome indicators can be identified to support the National Outcomes Framework for Social Services using the Welsh Government-funded 'Secure Anonymised Information Linkage' (SAIL) databank at Swansea University. This could either be through existing administrative data or by linking together administrative and/or survey datasets held in the databank.

---

[1] Programme for Government

[2] The well-being statement

1.8 In addition, the project was intended to work with internal and external colleagues to build on the work that has already begun with local authorities, third sector and other service delivery partner organisations to bring additional social care data into SAIL. The Project was also required to liaise with the project to create a Pan-Wales IT Specification for Health and Social Care data and other relevant projects.

1.9 The Project was designed to identify the extent to which existing data stored at the anonymised individual level in SAIL could be used to identify the groups of people that need care and support, and carers who need support, as well as evidence about the well-being of those groups.

1.10 In addition, the project was intended to investigate governance issues around sharing of local authority social care data with the SAIL databank in Swansea University.

1.11 The research and analysis for the Project took place between August 2013 and November 2014; the information contained in the report is correct as of November 2014.

1.12 In order to help navigate this document each of the main chapters begins with a summary of its key points, with further detail following in the body of the chapter and, where applicable, recommendations are summarised at the end of each chapter:

- **Chapter 2** introduces the concepts and challenges of using multiple data sources linked at the level of the individual to deliver additional outcome indicators, to allow indicators to be reported separately for social care service users or by user group as well as summarising some further lines of enquiry that could contribute to the wider social care evidence base.
- **Chapter 3** describes what is known about social services data collected by local authority social services systems, in particular through an examination of a historical dataset from the City and Council of Swansea.
- **Chapters 4 and 5** explore how survey and administrative data can be used to define and report on additional outcome indicators and user groups and to allow indicators to be reported separately for social care service users or by user group.
- **Chapter 6** summarises the further work that would be necessary to report on the additional outcome indicators and user groups and how it could be delivered.

1.13 Technical information for the report is provided as a series of appendices:

- **Appendix 1** provides a summary of the proposed additional outcome indicators and notes the further data acquisition or development work that will be required before reporting can begin.
- **Appendix 2** describes data linkage and the SAIL system in place at Swansea University.
- **Appendix 3** summarises what is known about the existing social care information systems in local authorities in Wales.
- **Appendix 4** summarises related research or data acquisition projects.
- **Appendix 5** provides an overview of the historical social care datasets for the City and Council of Swansea available in SAIL.
- **Appendix 6** provides a summary of survey datasets available in SAIL.
- **Appendix 7** provides a summary of the administrative datasets available in SAIL

Chapter 2 How can linked data be used to deliver additional outcome indicators for the national outcomes framework?

Key points

- Linked data can provide a depth of information not found when using a single dataset. In terms of providing evidence to underpin the well-being statement, two main types of data are available for linking – survey data and administrative data. Information about both outcome indicators and user groups may be found in data of either type. In some cases, a combination either of survey and administrative data or of multiple administrative datasets will be necessary to create an outcome indicator or flag a user group. This Chapter describes the key advantages and disadvantages to using linked data, summarises the current availability of social care data for linking and documents future data acquisition plans.

Opportunities

- In principle, linked data offers some distinct advantages in terms of its ability to report additional outcome indicators and define user groups:
    - where multiple administrative datasets are required to define an outcome indicator or user group, linking will be the best solution;
    - where indicators or user groups are derived from individual level social care records for all Wales, this will only be possible by linking records from each local authority in Wales; and
    - where analysis of outcome indicators is required separately either for known social care users or by user group, linking will be the best solution for some indicators.
- Data linking allows some additional outcome indicators to be proposed (all of which would require additional development work before reporting could begin) but also opens up further lines of enquiry that would help services to be targeted more effectively:
    - pathways between social care and health services can be examined;
    - service users can be followed up over time to examine their long-term well-being outcomes across a range of areas e.g. health, education and housing;
    - the complexity of well-being can be explored more fully - identifying how many people – and what kinds of people (e.g. service users) - have ‘multiple’ low well-being, i.e. low well-being across more than one of a range of indicators;
    - the dynamics of well-being can be explored – identifying how many people – and what kinds of people (e.g. service users) – suffer from persistent low well-being; and
    - where surveys include questions about both subjective well-being and satisfaction with services, if numbers allow, the relationship between service use, subjective well-being and satisfaction with services can be explored.

- SAIL already contains large numbers of routine health records, data for a growing range of topics including housing and education, and data for some surveys. Several research projects are already beginning to drive the acquisition of additional data. A follow-on project, if funded, could continue to liaise with those projects and look for synergies.

Challenges

- Nevertheless, there are significant challenges associated with using linked data to create indicators, some of which can be overcome with further development work.

- Many of these challenges relate to the fact that administrative data is collected for operational and not for research purposes. The limitations of the data are explored in detail for each proposed outcome indicator in Chapter 5, but in summary:
    - It can take a long time to acquire new administrative datasets for linking and to get the data 'research ready'. Legal barriers and/or risk aversion can also cause delays. However, with the creation of the ADRN, attempts are being made to address legal barriers and risk aversion, the process of acquisition should begin to shorten and ongoing, routine data flows can begin to be established.
    - Administrative data can be missing key information needed to perform robust analysis. However, missing information can often at least partially be replaced by linking to survey or other administrative sources.
    - When developing indicators using large, complex administrative datasets, time and resources can be required to manage the complexity e.g. deciding precisely which definitions to use and how to manage issues around e.g. severity and duration.
    - It can be challenging to validate estimates based on administrative data where a) existing 'best estimates' may be based on surveys or other less robust sources; or b) the definition that can be created using administrative data is qualitatively different to that used for the existing 'best estimate'.
    - Surveys tend to have relatively small sample sizes, limiting the analysis that can be done for survey-derived indicators. However, population estimates based on survey data can be improved by linking to large administrative sources and potentially using modelling methods.

Introduction to Data Linking

2.1 Data Linking is a technique for creating links between data sources so that anonymised information that is thought to relate to the same person, family, place or event can be connected for research purposes.

2.2 In 2006, the Welsh Government National Institute for Social Care and Health Research (NISCHR) funded the creation and development of the Health Information Research Unit (HIRU) at Swansea University. The aim of this unit was to develop a means by which routinely collected health data from many different sources could be utilised in a linked way whilst conforming to international best practice in terms of information security.[3] The process developed was called Secure Anonymised Information Linkage (SAIL). SAIL demonstrated how administrative data from multiple sources could be made available for research purposes (further information about SAIL is provided in Appendix 2).

---

[3] All research proposals using SAIL can only proceed if approved by a group of independent reviewers called the Information Governance Review Panel (IGRP).

2.3 The Welsh Government Knowledge and Analytical Services Programme to Maximise the Use of Existing Data (Data Max) has been working with the UK Economic and Social Research Council to fund various activities to explore precisely how ambitious Welsh Government plans can be in terms of making better use of existing data for Wales, much of which involves improving the availability of linked data for research purposes. As a result of these activities, it became clear that the technique of data linking had the potential to allow additional outcome indicators to be proposed to support the outcomes framework and could open up further lines of enquiry that could contribute to the wider social care evidence base. This Project was therefore funded by the Welsh Government social care policy team and carried out by two part-time researchers seconded from SAIL.

2.4 At the UK level, research funders, government departments and devolved administrations joined forces to form the Administrative Data Taskforce 2012. As a result of the Taskforce, the Economic and Social Research Council and other funders created a UK Administrative Data Research Network, which includes an Administrative Data Research Centre (ADRC) in each country of the UK. A collaborative bid between Cardiff University and SAIL at Swansea University was successful in bidding to become the ADRC in Wales; future data linking projects would therefore be completed at the ADRC in Wales and within the information governance, information security and ethical context of the ADRN.

What are the advantages and disadvantages of using linked data?

Opportunities

2.5 In terms of defining some outcome indicators or user groups, linked data may be the only effective approach, particularly where individuals may be recorded as having the relevant characteristics in a range of datasets relating to a variety of services. For example, individuals with substance misuse problems may not come into contact with specific substance misuse services at all or may use multiple service providers; different individuals may use different combinations of services across e.g. social care, health services or the criminal justice system. By linking together data for a range of services, a more accurate picture of service use can be developed than will be possible when looking at any single source.

2.6 The effectiveness of many policies is monitored and/or evaluated using a patchwork of aggregate estimates derived from a range of unlinked, stand-alone sources. This delivers a very simplistic picture, because:

- Estimates drawn from unlinked sources can’t take into account whether the same individuals are accessing multiple service providers, as in the substance misuse example given above. Linking data across datasets covering a range of providers or topics makes it possible to examine the extent to which the same individuals – or groups of individuals with similar characteristics – tend to be recorded in multiple contexts e.g. whether they have multiple needs or are receiving support from a variety of services.
- Nor may it be possible to examine whether the same individuals are included in multiple user groups in different contexts e.g. as both a ‘vulnerable adult’ in social care records and an ‘individual with mental health problems’ in health records.
- Similarly, only by linking data together will it be possible to examine the extent to which the population is made up of people who have a low level of well-being according to only one indicator or how large a proportion have low well-being across more than one of a range of indicators. Data linking will allow an estimate to be made of the number of individuals there are - and to examine the characteristics of individuals - with ‘multiple’ low well-being i.e. low well-being according to a number of indicators.

- Unlinked data only allows the identification of the outcomes of an intervention for an individual if those outcomes are recorded in the same database as information about receipt of the intervention. Only by linking individual-level records is it possible to look across topics to, for example, examine the health or housing outcomes of a social care intervention.
- By linking across topics, user groups can also be followed - or service users more generally - over time in order to examine their long-term outcomes across a range of areas e.g. health, education and housing.
- Estimates drawn from cross-sectional surveys and some (but not all) stand-alone administrative sources, only allow change over time to be examined at the aggregate, e.g. population, level. Certainly, outcome indicators reported at the aggregate or headline level do not tend to look at change over time below the population level. Linked data tends to be longitudinal in nature, allowing individual-level dynamics to be examined. For example, two point-in-time measures of the proportion of the Wales population using social care services does not allow an estimate to be made of how many individuals were users at both points, or how many individuals either started or stopped using services between two points. Linking data over time allows the dynamics of e.g. individual well-being or use of social care services to be examined. This would allow an estimate to be made of how many individuals suffer from persistent low well-being and an exploration to be made of their characteristics.

2.7 For some outcome indicators and user groups, linked data can allow:

- comparisons of outcomes between social care users and the general population; and
- indicators to be examined separately for each of the user groups e.g. the mental health status of ‘vulnerable adults’.

2.8 Administrative data tends to be a census, whilst survey data is collected for a sample of the population. Estimates derived from survey data and administrative data each have advantages and disadvantages but by linking the two together, some of those weaknesses can be moderated; for example:

- whilst administrative data does not tend to include information about attitudes and behaviours, surveys do. Survey data can therefore be used to enrich administrative data or vice versa through linking. The National Survey for Wales has the advantage of containing a set of self-reported well-being questions. Self-reported well-being as collected in surveys, can be linked to administrative records to, for example, examine the self-reported well-being of a sample of known social care users or explore the well-being of known social care service users (versus non-users) with the service. Whether National Survey for Wales respondents are also social care users can be identified in two ways, either by linking National Survey for Wales data to routine social care records or by asking National Survey for Wales respondents if they are social care users;
- whilst some administrative sources do not contain information about socio-demographic characteristics or the circumstances of individuals, linking to other administrative data can add this missing information e.g. data from the Department for Work and Pensions or HMRC on Benefit receipt or employment status can be linked to routine social care, health, housing or education records;
- the analysis of combined survey and administrative data allows an assessment to be made of the accuracy of estimates derived from survey data; and

- combining multiple administrative and/or survey datasets that contain some of the same information about the same individuals or types of individuals, e.g. carers, can be used to triangulate between sources, making it possible to assess the quality of individual sources and potentially providing a more complete picture of reality.

2.9 By linking together datasets across topics, it would be possible to examine pathways between e.g. social care and health services, in order to examine issues such as 'delayed transfers of care'.

2.10 Analysing the existing survey and administrative data available in SAIL can be cheaper than undertaking new primary data collection. In addition, it can often deliver results more quickly, since the data has already been collected. Even when significant data cleaning and development work is required, it will tend to be cheaper and quicker than commissioning a new piece of primary research e.g. a survey.

2.11 When potential research subjects are likely to have issues around capacity to consent to research or are otherwise vulnerable, are hard to reach or rare individuals, existing primary research methods may be unsatisfactory either ethically or methodologically, leaving the use of existing data as, if not a perfect solution, then the least unsatisfactory alternative. Examples of such groups would be individuals with dementia, individuals with learning disabilities or individuals with substance misuse problems.

2.12 Administrative data tends to be a census of the population of interest – this may be close to the whole population of Wales e.g. the Welsh Demographic Service contains a record for everyone who has been registered with a GP in Wales, or a population of service users e.g. local authorities keep records of every social care user. Subject to an assessment of the robustness of the data, estimates derived from administrative data – or from several linked administrative databases - are potentially associated with relatively little margin of error.

Challenges

2.13 It can take a long time to acquire new administrative datasets for linking and to get data that was not initially collected for research purposes 'research ready'. Depending on which organisation controls the data, the legal framework under which the data was collected and any organisational or legal barriers in place around data sharing, it can take several months of discussion to clarify the data controller's legal ability to share and to get a data-sharing agreement in place.

2.14 The UK Administrative Data Research Network was created specifically to provide an infrastructure to address these challenges and to facilitate the flow of data for research. Although, at point of writing, the Network and its principles and procedures are still under development, timescales for data acquisition are likely to reduce over time. Whilst it can still take a matter of months to acquire data, compared with commissioning a new large-scale survey, data linking will still tend to be less time- consuming and less expensive.

2.15 As noted above, unlike surveys, administrative sources tend not to include behavioural or attitudinal information and can be missing information about the socio-demographic or other circumstances of individuals. Where no direct information is available, attempting to derive indicators using administrative data alone can therefore require the use of less exact proxy indicators when it comes to subjective issues like well-being. However, missing information can often be at least partially replaced by linking to survey or other administrative sources.

2.16 Administrative sources are not collected for research purposes. When developing indicators using large, complex administrative datasets, time and resources can be required to manage the complexity, for example:

- examining large numbers of potential indicators to identify the best proxy for each outcome indicator or user group;
- to refine the definition to be applied, e.g. where conditions included under a definition of mental health as developed for another purpose may not be appropriate for an outcome indicator for the well-being framework; and
- to identify and/or develop methods to manage issues around severity and duration e.g. of mental health problems.

2.17 Estimates based on administrative data can be a challenge to validate where a) existing ‘best estimates’ may be based on surveys or other less robust sources; or b) the definition that can be created using administrative data is qualitatively different to that used for the existing ‘best estimate’.

2.18 As noted above, estimates derived from survey data and administrative data each have advantages and disadvantages. Whilst linked data can help to overcome some of the limitations of individual sources, it cannot overcome them all:

- Surveys tend to have relatively small sample sizes, limiting the analysis that can be done. Where information about the issue of interest is only collected in a survey e.g. self-reported well-being, linking to administrative data will not, in itself, increase the sample size and therefore the accuracy of the estimate. However, if sufficient appropriate information is available in the administrative record, grossing methods can potentially be improved or modelling methods developed, leading to improved population estimates or simulations based on the population contained in the larger, administrative dataset.
- When survey data is acquired for linking it is subject to the respondent giving their consent to the linking. In practice, around 65% of respondents tend to consent. If those who consent are different to those who refuse in terms of the indicator of interest, this can introduce bias into the estimate. However, since a lot of information is available both for those respondents who consent and for those who refuse, both the nature and the magnitude of the bias can be examined and, in some cases, weighting strategies developed in order to correct for it.
- Survey data is based on a sample of the population so, whilst grossing techniques can be used to provide estimates for the population as a whole, such estimates are associated with a margin of error that will vary in size depending on the level of analysis being undertaken.
- Most of the survey data available for Wales is cross-sectional – in other words, it interviews a fresh sample of individuals each time rather than repeatedly re-interviewing the same individuals. In contrast, since administrative data tends to flow regularly into SAIL, it tends to be longitudinal, with information held for the same individual at numerous points in time. This means that, for example, National Survey for Wales can be used to monitor how the subjective well-being of the *population* changes over time but not to examine how the subjective well-being of *individuals* changes over time, because individuals are only asked about their well-being on a single occasion. Whilst administrative data will sometimes provide a less direct outcome indicator, because it is longitudinal, it can be used to monitor change over time for individuals e.g. allowing an examination to be made of whether outcome indicators change after specific events such as the provision of a specific social care service.

2.19 Using linked data to report additional outcome indicators can be a challenge precisely because the administrative data tends to be longitudinal in nature i.e. the information may change over time. So, issues arise with the data itself, where an individual may have been recorded as a 'carer' at one point in time but there is no subsequent information on the basis of which to confirm whether they are *still* a carer at any future point in time – or whether or not they were a carer *beforehand*. The issue of change is also a challenge in terms of defining which individuals to include when defining user groups or reporting outcomes, for example whether the definition of 'individuals with mental health problems' should include those recorded as having a relatively minor mental health problem or for a relatively brief period of time.

What data is already available and how can additional data be acquired?

2.20 SAIL already contains the majority of routine health records e.g. GP Event, Hospital Accident and Emergency, in- and out-patient data, plus a growing number of administrative datasets for other topics e.g. housing and education. Most of the routine health datasets are available for the whole population of Wales; the exception is the GP Event data which, at time of writing was available for around 50% of the GP practices in Wales but is increasing all the time. A National Institute for Social Care and Health Research-funded project has the target to increase this to over 70% by March 2015. Data from a range of surveys is also available in SAIL. For further information about the datasets currently available in SAIL, see Appendix 6.

2.21 However, SAIL was created by the Welsh Government specifically to support research in the field of health informatics, and the majority of work before 2011 understandably went into the acquisition of routine health records. Datasets relating to non-health topics have tended to be acquired as a result of specific projects, so the non-health content is variable. SAIL currently contains little data for social care and it does not currently have the capacity to allow the population of social care users to be flagged for research purposes. In order to report on additional outcome indicators therefore, a range of additional datasets will need to be acquired by SAIL/the ADRC-W.

2.22 Chapters 4 and 5 identify which datasets - either currently in SAIL or that can be acquired for SAIL (i.e. further development work would be required before reporting could begin) - are best placed to allow the definition and reporting of additional outcome indicators and user groups. If a follow-on project is funded, a further step would be to identify areas of the outcomes framework where, despite the additional outcome indicators proposed by this Project, relatively few outcome indicators are available and to identify any additional datasets that could be used to propose further outcome indicators.

2.23 The legal frameworks around the acquisition of survey data and administrative data are different. As mentioned above, survey respondents must be asked to consent to their anonymised survey answers being linked to other records held about them and, as part of the consent process, are informed about the process for linking and storing their data. Whether administrative data can be used for research and statistical purposes, including data linking, is dependent on the fair processing information or privacy notice issued e.g. to service users, when their data is collected and, for some datasets, on the regulations under which the data is collected and held. Where necessary, work can be undertaken to add consent to link to additional surveys and to amend privacy notices and/or regulations for additional administrative sources to allow further indicators to be reported in future.

2.24 If follow-on funding is available, it is therefore recommended that a proposal is made to the ESRC-funded UK Administrative Data Research Network (ADRN) for a project designed to deliver an analysis of the linked datasets necessary to report the additional outcome indicators proposed by this Project. As part of their remit in supporting such a project, the UK Administrative Data Service (ADS), a part of the ADRN, will be responsible for acquiring the relevant UK-level datasets for the Administrative Data Research Centre for

Wales (ADRC-W) for the duration of the Project. Although the ADS would be responsible for data acquisition for UK-level data e.g. data from the Department for Work and Pensions, both the acquisition of data from local authorities and the analysis work itself would need to be part of a funded project (for further discussion of follow-on options, see Chapter 6). The follow-on project analyst/team could work with the ADS to ensure that any UK-level data is fully documented and fit for purpose. So, whilst the work to acquire the data itself would be free to Welsh Government:

- there would be a cost associated with funding a project; and
- because it receives its core funding from the ESRC, projects supported by the ADRN cannot consist simply of the reporting of a set of outcome indicators. Although the methodological work to define some of the proposed outcome indicators or user groups would fall within the ADRN remit, the proposal would need to make a wider contribution to the field of research; ideally, some of the more complex lines of enquiry suggested above, where data linking is adding significant value, should be included in order to meet ADRN requirements e.g. examining the dynamics or complexity of well-being among service users.

It should be noted that this Project was carried out within SAIL on an exploratory basis and as such did not require approval from the SAIL Information Governance Review Panel. Any further analysis would require both the approval of the ADRN Approvals Panel and, because it would involve the use of routine health records, of the SAIL Information Governance Review Panel. This process would take a minimum of one to two months and could potentially take longer depending on the availability of resources within the ADS, the number of other projects being proposed and the complexity of the proposal (i.e. the number of clarifications likely to be sought by either the ADRN Approvals Panel or SAIL Information Governance Review Panel).

2.25 The tasks required of this Project in terms of informing the future acquisition of data – in particular all-Wales social care data - were:

- to review the existing social care data in SAIL to learn lessons for future data acquisition and to assess what additional outcome indicators or user groups all-Wales social care data might provide (see Chapter 3);
- to examine the social care data systems in place in the local authorities of Wales (see Chapter 3 and Appendix 3 for further detail) – further work is needed on this task and could be completed if a follow-on project is funded; and
- to liaise with existing projects in order to pass on lessons learned and ensure that any data acquired will meet the requirements of the National Outcomes Framework (a list of existing projects is provided in Appendix 4); although this liaison took place during the August 2013 to March 2014 project period, further funding would be necessary to continue this liaison activity.

Acquisition of administrative data

2.26 In order to provide the opportunity of developing additional outcome indicators for the national outcomes framework and to inform research on social care more broadly, the administrative data generated through user interactions with local authority social care will be a key data source. Only limited social care data is currently available in SAIL/the ADRC-W but it is the most suitable environment within which individual-level records for all local authorities in Wales can be analysed. A Welsh Government National Institute for Social Care and Health Research -funded project is underway (April 2014-March 2015) to document but not to acquire all-Wales local authority social care data. It is therefore essential that arrangements be put in place to acquire all-Wales local authority social care data and to ensure that it is updated on an ongoing basis. As noted above, if Welsh Government were to fund an ADRN project that included the analysis of data not currently held by SAIL/the ADRC in Wales, the ADS would be responsible for acquiring any UK-level data required but not data from local authorities.

2.27 In order to report some of the proposed additional outcome indicators and to define some of the user groups, additional datasets will need to be acquired, including data held by UK Government Departments such as the Department for Work and Pensions/HMRC.

2.28 In terms of administrative datasets, it is often research projects or studies that are the drivers to acquire additional data. A number of projects have commenced, each of which is designed to provide additional social services metadata (descriptive information about the data) or to acquire additional social services data for linking in SAIL/the ADRC-W. These projects are outlined below in brief and described in more detail in Appendix 4.

- mapping health and social care datasets to link to SAIL: a project to produce a high level mapping of datasets and a detailed report for each dataset, including information about the owner, location and purpose as well as technical specifications;
- facilitating the collection of accurate metadata to be linked with SAIL: a project to develop standards for metadata;
- the Social Care and Health of Older People (SCHOOP) project: this project will acquire social services data from the City and Council of Swansea and link it to health records to examine the relationship between use of local authority social care and use of health care by people aged 65+ years; and
- data linkage in social care: a pilot project in Gwynedd to build the complete picture of service provision across health, social care and the third sector. This project will include the anonymisation of Gwynedd Social Services data into SAIL.

Both the Social Care and Health of Older People (SCHOOP) project and the Gwynedd study are expected to deliver additional anonymised data into SAIL during 2014-15. However, the data acquisition for each will be relatively limited in scope, either in terms of the age group or geography to which it will relate. Use of the data beyond the study remit would also need to be negotiated both with the study teams and their data providers.

2.29 Negotiations are underway for the anonymisation into SAIL of several National Administrative databases during 2014-15, including the national dispensing database (all prescriptions dispensed through over the counter pharmacies in Wales), the National Exercise Referral Scheme (GP-prescribed exercise programme), and the ‘Stop Smoking Wales’ database. Welsh Government is keen to provide the Lifelong Learning Wales Record data to SAIL and to use the National Institute for Social Care and Health Research social care project to ask local authorities in Wales about access to datasets beyond social care e.g. leisure swipe card data. Each of these administrative datasets may provide means for improved flagging of user groups or additional outcome indicators, e.g. flagging people on anti-depressant drugs could be used as part of the definition of individuals with mental health problems.

2.30 In summary, although some data acquisition will take place even in the absence of any follow-on work to this Project, it would be limited and without significant input would risk not being fit for purpose in terms of reporting additional outcomes indicators or flagging social care users more broadly or more specific social care user groups.

Acquisition of survey data

2.31 The Welsh Government Programme to Maximise the Use of Existing Data has been working since 2009 to add ‘consent to link’ to key social surveys for Wales as well as working with key UK Surveys operating in Wales to assess whether existing consents are sufficient to allow linking of survey data. So far, ‘Consent to Link’ has been added to the National Survey for Wales, the Welsh Health Survey, the Sports Wales Active Adults surveys 2012 and 2014 and various smaller surveys undertaken by organisations outside Welsh Government e.g. the Cognitive Function and Ageing Study. Work has been completed to confirm that existing consents are sufficient for the Millennium Cohort Study and are underway for Understanding Society, the UK’s largest longitudinal panel survey. Both SAIL and Welsh Government have worked with the UK’s new birth cohort *Life Study*, which is about to begin collecting information about children born in 2014-15, to ensure its consents can be used to link the data into SAIL.

Consented survey data is acquired for SAIL / the ADRC-W when complete years of data become available. Welsh Health Survey and National Survey for Wales data will be added annually in the autumn of each year. Active Adults 2012 and 2014 asked consent to link and an ADRN project examining the health impacts of participation in sports and culture is expected to acquire this data for SAIL. CFAS will deliver data once sufficient cases (i.e. over 1,000) with consents have accumulated, which is expected to be in 2015-16.

Chapter 3 What can the City and Council of Swansea social services data 2004-07 tell us about using linked data to report additional outcome indicators?

Key points

- The City and Council of Swansea data held in SAIL is a rich source of information about social care clients.
- However, there are a number of limitations either to the SAIL Database or to the social care data held by SAIL that will need to be overcome to enable additional outcome indicators to be reported in the future:
    - Social care data is not currently available in SAIL for all local authorities in Wales. This means that reporting for proposed additional outcome indicators cannot be done for the whole population of local authority social care users in Wales without additional data acquisition.
    - Data acquisition will be complicated by the fact that the way social care data is collected not only varies by local authority but, where groups of local authorities use the same software, they may nevertheless use it differently. However, requirements by the Audit Commission suggest that some consistent local authority data collection occurs at the level required for data linkage to occur.
    - Data linking was possible for only 92% of the City and Council of Swansea social care client records held in SAIL, which may introduce bias into the reporting of proposed indicators.
    - For some proposed additional outcome indicators and some user groups, changes to data collection are recommended to deliver more robust data in the long-term e.g. for carers.
- Significant work would be required to prepare the data files for linking and to provide or, more likely where data are administrative and therefore not collected for research purposes, creating from scratch the required metadata[4] It is important to look for ways to minimise any burden on staff when considering future work.
- It is recommended that a follow-on project is funded to allow more work to be done to explore these challenges and to develop and test solutions e.g. ways to reconcile the data once acquired.
- Once social care data has been acquired from all local authorities, the projected client numbers are likely to be sufficient to allow the reporting of a range of additional outcome indicators by the majority of user groups (see Chapters 4 and 5 for more detail).
- In the longer term, the Implementation of the Social Services and Wellbeing Act and the corresponding development of the Welsh Community Care Information System (WCCIS) will provide the opportunity to develop a standardised social care data set for Wales that could potentially be acquired for SAIL for analysis purposes.

Introduction

3.1 The review of the City and Council of Swansea social care data was completed in order to:

---

[4] Metadata is ‘data about data’. There are two metadata types: i) ‘structural metadata’, about the design and specification of data structures; and ii) descriptive metadata about the content of the dataset or of individual fields within the dataset.

- identify lessons about data quality to inform future local authority social care data acquisition; and
- to identify the extent to which additional outcome indicators could be delivered or user groups defined using linked social care data for all local authorities in Wales.

3.2 In 2007, Social Services data from the City and County of Swansea for 2004-07 was anonymised into SAIL. The data was acquired purely as part of the Proof of Concept for anonymised data linkage and no routine data flow was established at that time. This section describes the data that this download contained, to give an initial indication of the potential uses to which, once acquired, all-Wales Social Care data could be put in support of the outcomes framework. The opportunities and challenges with using the City and Council of Swansea data are identified and recommendations are made for future data acquisition and future data collection.

Swansea as a 'representative' local authority

3.3 The City and Council of Swansea data relates to a subset of the Wales population (the City and Council of Swansea includes around 8% of the population of Wales) and, as noted above, to a limited time frame (2004-07). Records were captured for just under 38,000 individuals, just over 19,000 of whom were coded as 'clients' and 18,000 as 'associated people'. 'Associated people' included carers, support workers and others involved with the support of Social Services clients. The clients were listed either as historical (i.e. a closed case, the support having officially ended) or current as of 2007 when the data extract was delivered to SAIL. The City and Council of Swansea data is summarised in further detail in Appendix 5.

3.4 The data matching conducted by NWIS found either a perfect or highly probable match for 92% of the client records submitted to them. It is not clear why some individuals could not be matched but if a follow-on project is funded, further work can be done to investigate.

3.5 For each client, data was captured during an initial assessment and each subsequent support visit by Social Services generated further records. A total of 308 data fields exist in the dataset but not all fields were completed for every client – as for a survey, core information is required for every client but subsections of data are only completed where relevant e.g. the question "Do you have a lifeline pendant?" leads to (and links to) a block of information depending on whether the answer was 'yes' or 'no'.

3.6 Using ONS mid-year population estimates from 2007 to extrapolate from the City and Council of Swansea data, it can be estimated that an all-Wales local authority Social Services client dataset containing three years of records would include around 245,000 clients and 200,000 'associated people'. The clients would be estimated to include around 3,300 people aged 0-16 years, 5,500 people aged 17-24 years, 77,000 people aged 25-64 year olds and 161,000 people aged 65 years and over. With such large numbers, it is likely that analysis for clients could be completed for different age groups.

3.7 The City and Council of Swansea used the PARIS Information System to store social services records but they also used a number of other systems to interrogate the data, including Excel spreadsheets, Access databases, and 'business objects' software, each customised to deliver particular analyses and reporting of the data. Various service-specific datasets (e.g. domiciliary care provided by a private contracting organisation) were also used that were external to the PARIS system.

3.8 A summary of the systems known to be in place in local authorities in Wales is presented in Appendix 3. Early investigation suggests that a mixture of data collection or data management systems are being used by local authorities across Wales and that even those groups of local authorities that are using the same system may be using it differently. Four local authorities are known to use PARIS and eight to use RAISE (three have both). However, requirements by the Audit Commission[5] suggest that some consistent data collection occurs at the level required for data linkage to occur.

Lessons learned on the benefits of and issues with using local authority data

3.9 This section reports on the benefits of and issues the Project has identified with the City and Council of Swansea data. In order to influence future data collection, these lessons will be passed on to future projects attempting to acquire local authority social care data. Lessons learned will also be shared with the joint (NHS Wales and Local Authority) procurement of a Welsh Community Care Information System that will support the needs of both Community Health and Social Care Services.

3.10 The success of further work is dependent on developing a better understanding of the data and a clear understanding of how to interrogate the data. If a follow-on project is funded, a better understanding of the data could be gained from working with both existing and future projects working in the area of social care and potentially from completing additional analysis of both the existing and newly acquired data associated with these projects (see Appendix 4 for further detail but such projects include the Social Care and Health Of Older People (SCHOOP) project and the National Institute for Social Care and Health Research-funded project to document social care data for Wales).

Opportunities

3.11 Routine administrative social care records represent a rich source of data and data linking allows individual service users to be followed up over time and their outcomes to be examined across a range of areas e.g. health and housing. If a follow-on project is funded, more work can be done to identify further outcomes to support the national outcomes framework for people who need care and support and carers who need support.

3.12 Examination of the 2007 Swansea social care data suggests that it is a rich source of information about those people who receive social services from the City and Council of Swansea, which will allow a broad range of analysis to be delivered (for a complete list of information, see Appendix 5). The information includes:

- week of birth, gender, ethnicity, religion and nationality;
- information that can change over time: accommodation (tenure, housing type etc.), marital status and employment status. Each record is accompanied by a beginning and end date indicating the period to which the information applies;
- information relevant to interacting with the client, including legal information, medication details, lifeline pendant issued, known allergies, alerts to staff (ranging from ‘beware of the dog’ through to ‘do not visit alone’); and
- information about carers, staff, neighbours and other individuals who have been involved in some way in the provision of services to the client.

3.13 It should also be noted that due to the continual flows of routine administrative data into the SAIL Databank, individuals with, for example, ‘learning difficulties’ according to the 2007 data can be flagged up anonymously for follow-up using more up-to-date data from other routinely available health records e.g. GP records.

---

[5] The Audit Commission require local authorities to submit identifiable individual-level data to a central database so that potential fraud can be investigated

3.14 Data about the 'associated people' was not sent for anonymisation in 2007 but future projects could capture this information. This would provide the opportunity to flag a group of carers across linked datasets, so that the health and well-being of carers could be examined. However, further advice would need to be sought from local authority information officers about the likely quality of this information and to examine any legal barriers to the acquisition of this data for linking.

3.15 Clients are classified into various 'service user groups' and the group into which an individual fits can change over time. A 'main' and 'current' client group can be recorded at each stage, giving the potential for multiple user groups to be recorded over time. Some of the City and Council of Swansea user groups align well with the potential user groups proposed in Chapter 4 and for other user groups a similar proxy can be identified. For example, the City and Council of Swansea user groups include 'Services for Elderly and/or Disabled Persons (SEDP)' and 'mental health' (for detailed discussion by user group see Chapter 4).

3.16 Some data appear in free text fields i.e. information can be entered in a completely unstructured way.[6] This risks the inclusion of identifying information so is usually excluded when data are being shared. However, free text often includes rich information that would be of interest to researchers. Future data linkage projects will have the benefit of Natural Language Processing, an automated process where an intelligent computer program 'reads' text fields and turns important pieces of information into codes, discarding the rest of the text. The ADRC-Wales funding has included the purchase of such software. The Natural Language Processing software will need further testing but represents a potential way to avoid losing this rich source of data.

3.17 Even when working with anonymised datasets, there is a theoretical risk of re-identification or disclosure. However, SAIL/the ADRC-W does not provide raw data to researchers and the outputs are scrutinised to ensure that the results are presented at a sufficient level of aggregation to prevent both re-identification and disclosure.

Challenges

3.18 The 2007 social care records were held in a number of databases even within the City and Council of Swansea. Different local authorities in Wales use different systems or, even if some groups of local authorities use the same system, they may use them differently. Not all of the databases used by the City and Council of Swansea contained a standard person identifier that would allow them to be linked easily back to other data. For example, gathering data for the Social Care and Health Of Older People project has involved manually matching 200 cases using the PARIS identification numbers because the data did not include a standard person identifier. The extent to which such variation may still exist within the City and Council of Swansea and exists across other local authorities may present a challenge in terms of data extraction.

3.19 For some potential user groups there was no corresponding user group identified in the City and Council of Swansea e.g. there was no record of adopted children (for detailed discussion by user group, see Chapter 4). It is likely that, in order to provide sufficient evidence to support the outcomes framework, local authorities will need to label both carers and adopted children more consistently and will need to collect additional information about the caring responsibilities of carers.

---

[6] Free text data is not normally included in SAIL, as there is a risk that identifiable data might have been typed in. Some free text fields were included in 2007 as part of the SAIL proof of concept, and some attempts were made at the time to remove potentially identifiable information from the data. However these attempts were not fully successful so SAIL will be taking action to either delete these fields or to remove the Swansea data altogether and await new data.

3.20 A total of 92% of the 19,275 individual records were able to be anonymised into SAIL (using the split file process, as described in Appendix 2), which means that it was not possible to link around 8% of cases to other records about those individuals. Matching based on personal details tends not to be as successful as matching on a standard identification number such as an NHS number. Typing errors, the use of nicknames and alternative address formats can all cause matching failures. Further investigation would be required, in co-operation with the City and Council of Swansea and NWIS, to identify the issues preventing successful matching and to develop methods to improve the matching rate in future. With the development of the UK Administrative Data Research Network – and of an Administrative Data Research Centre (ADRC) in each country of the UK - it should be possible to consider the adoption of standard identification numbers such as National Insurance Numbers or NHS numbers more widely in client-based systems in order to improve the data matching process. Software is under development at SAIL that can be installed where the data is being collected - this will allow standardised anonymisation at the data source. This system is due to be tested in April 2014 and, when fully developed and tested, will be rolled out to local authorities in Wales thereafter.

3.21 Very little metadata (information about the structure of the data or the definitions associated with the database) was provided with the dataset.[7] Significant work is required to prepare the data files for linking and to provide or, more likely where data are administrative and therefore not collected for research purposes, create from scratch the required metadata. For the future capture of all-Wales local authority data into SAIL, it will therefore be important to look for ways to reduce the burden on local authority staff (e.g. provision to a central database might negate the need for some individual local authority reporting returns). It is also recommended that all projects seeking to acquire local authority social care data should work together to ensure that all requirements are combined into a single negotiation with each local authority. The National Institute for Social Care and Health Research-funded project will seek to deliver improved metadata for local authority social care datasets.

3.22 The City and Council of Swansea social care data is a complex matrix of information which is both ‘client type’ and date dependent. Date information is essential in interpreting the data, as many items of data such as client status can change over time. The longitudinal nature of the data makes it richer because issues such as the movement of clients through the system, and the length of time clients remain in the system, can be examined. However, complex longitudinal analysis methods are needed to examine this kind of data and this limits the pool of available analysts who can work on the data. If a piece of follow-on work is funded, this kind of analysis could be completed but there could be a risk that a suitably skilled analyst may not apply for the post.

3.23 A significant number of errors was identified in the data e.g. spelling or coding errors. For some variables it was not clear whether the data had been collected using a pre-defined list or free text; some information may have been recorded using an editable pre-defined drop-down list, leading to the addition of spelling errors. It is often possible to re-code data or to create an algorithm to address spelling or coding errors automatically. However, this is unlikely to address the full range of problems identified in this dataset. Ideally, changes should be made to up-stream data collection. More consistent data entry at source may take time to achieve; however, so developing algorithms to correct errors could form part of the work of any follow-on project, if funded. Improvements to data collection recommended are, broadly:

---

[7] For example, the **data set** labelled ‘alert’ appears to contain a mixture of information designed to protect the Social Care staff (e.g. "Previous aggressive behaviour" or "Do not visit alone" codes) and the client (e.g. "Serious Health Issues" or "Vulnerable Adult" codes).

- To introduce fixed selection drop-down menus with, where necessary, a free text ‘other’ code(s).
- To introduce the use of variables with a standardised format where appropriate e.g. for date and gender.
- Where possible, to introduce checks to ensure valid entries e.g. for dates.

3.24 As noted above, the risk of re-identification is a central concern to SAIL. The City and Council of Swansea data includes a record of the client’s ethnicity. Since classifying people by ethnicity can produce some small, relatively identifiable clusters of people, ethnicity data is considered a re-identification risk. In fact, existing data agreements between SAIL and other, non-social care data providers (e.g. GP practices and hospitals) specify that data on ethnicity, religion and nationality will not be made routinely available for research through SAIL. There is, however, an increasing interest in data relating to characteristics protected under the Equalities Act 2010, including ethnicity. The challenge of making ethnicity data available within the information governance and privacy controls that SAIL applies is an issue that needs further attention from SAIL/the ADRC-W.

Recommendations

3.25 The City and Council of Swansea social care holds rich information about social care clients. The acquisition of an all-Wales local authority dataset, whilst challenging, would contribute significantly to the broader evidence base available for research on the topic of social care.

3.26 In the longer term, the Implementation of the Social Services and Wellbeing Act and the corresponding development of the Welsh Community Care Information System (WCCIS) will provide the opportunity to develop a standardised social care data set for Wales that could potentially be acquired for SAIL for analysis purposes. However, the migration process will take a number of years to complete and it is not yet clear how much historical data will be migrated as part of the process. Any new dataset relying on WCCIS to provide data may therefore be limited, at first, as to how many cases would provide a sufficiently long timeline of information to allow change over time to be reported for the relevant indicators.

3.27 In the meantime, it is recommended that a follow-on project is funded so that the following key activities can take place:

- joint working with ADS and with other relevant projects to ensure that any data acquired is both correctly documented and that the specific data needed to report additional outcome indicators is acquired. This would include assisting with reconciliation between the different data collection systems used by the different Local Authorities in Wales;
- the identification of further/improved definitions of outcome indicators;
- the identification of further/improved definitions of user groups;
- further development and validation work to assess the robustness of estimates for both outcome indicators and user groups generated using data extracted from administrative records;
- further work to increase the linking rate;
- further research to examine the range of data management systems being used by local authorities in Wales and potential solutions proposed for managing this complexity (in conjunction with other projects where appropriate); and

- ensure that the lessons identified during the acquisition and analysis of the Social Services data from the City and Council of Swansea are shared in order to inform future data acquisitions.

3.28 Future data sharing agreements should include the capture of data for ‘associated people’, including carers.

3.29 Every effort should be made to ensure that all projects seeking to acquire social care data work together to ensure that all requirements are combined into a single negotiation with each local authority. Improved metadata should be acquired alongside the data.

Chapter 4 How can linked data be used to define service user groups?

Key points

- A more complete picture of the well-being of the population of Wales would be gained by examining the well-being of specific user groups separately.
- With some further development work, the following user groups can be 'flagged' for analysis using all-Wales data that is already available in SAIL:
  - individuals with mental ill-health (needs further work to agree, refine and validate the definition);
  - individuals with substance misuse problems (needs further work to agree, refine and validate the definition);
  - older people aged 65 years and over;
  - disabled people* (needs further work to agree, refine and validate the definition; depending on the definition, may need further data acquisition); and
  - people with dementia* (needs further work to refine and validate the definition).

  The following user groups can be 'flagged' for analysis using all-Wales data that can be acquired for SAIL:
  - 'people who need care' – either by assuming everyone in the population of Wales is a potential user or met partially, by flagging all local authority social services clients (needs further data acquisition), to be used either with or without an exercise to identify 'unmet need';
  - carers associated with local authority social care clients* (needs further data acquisition);
  - looked after children (needs further data acquisition and liaison with local authorities); and
  - people with learning disabilities* (needs further work to refine and validate the definition and further data acquisition for social care).

  * These categories can be reported separately for young adults, adults and older people although for some groups, 'young adults' may suffer from small numbers.
- Data could not be identified to allow the flagging of a user group of adopted children. However, further investigation could be undertaken to identify additional sources/indicate whether changes to data collection for available sources could provide a 'flag' for this user group.
- It is recommended that a follow-on project should complete the further development work required:
  - to seek out additional datasets needed to improve the definitions of user groups;
  - to ensure that any further UK-level data acquired by ADS is fit for purpose; and
  - to refine and validate some of the proposed user group definitions e.g. to agree which conditions comprise 'mental (ill) health'.

Introduction

4.1 The national outcomes framework sets out well-being outcomes for people who need care and support and carers who need support. The broad population of 'all people who need care' consists of several specific service user groups. A more complete picture of the well-being of the population of interest would be gained by examining the well-being of each user group separately.

4.2 In order to use the outcome indicators to estimate the well-being of specific user groups, work would be required to 'flag' certain user groups in SAIL using datasets that are either currently available in SAIL or that could be acquired by SAIL. For members of a user group to be 'flagged', the definition would need to be developed using information that is already available about those individuals in SAIL (or using information that can easily be acquired). For example, it is a simple matter to flag individuals based purely on their age. However, some of the user groups proposed in this Chapter are defined in different ways in different contexts, e.g. individuals with mental ill-health, and the use of narrower or wider definitions would result in different numbers of individuals being flagged.

4.3 In the context of flagging the user groups, the key strengths in using data linking are that:

- if necessary, definitions can use information from multiple data sources; and
- analysis can be completed even where the data required to flag a user group is not available in the same dataset as the indicator itself.

4.4 Potential user groups are:

- people who need care and support;
- carers who need support;
- looked after children (local authority);
- adopted children;
- individuals with mental ill-health;
- individuals with substance misuse problems;
- older people aged 65 years and over;
- individuals with physical disabilities;
- individuals with learning difficulties; and
- individuals with dementia.

4.5 For some user groups it is relatively simple to create a 'flag' that can be used in the analysis of the outcome indicators. This would be true of any user group that is based on a specific age group. For example, flagging all citizens of Wales aged 65 years and over can be done using the Welsh Demographic Service, which is the population 'spine' used by SAIL when linking datasets together. If any of the user groups described in the remainder of this Chapter are required to be split by age group e.g. for adults and children separately, this can therefore also be done relatively easily.

4.6 For other user groups, the definitions provided above are not sufficiently precise to allow a 'flag' to be created.

4.7 This Chapter summarises the challenges associated with defining and 'flagging' potential user groups. For information about which indicators it will be possible to analyse by the example user groups described here, please see Chapter 5.

4.8 As noted in Chapter 2, both survey and administrative data was reviewed when considering which sources contained useful information for defining the user groups. Administrative data can be preferable to survey data because the data is likely to be more complete, i.e. to provide information about all members of the user group as opposed to just a sample of them. Where administrative data is lacking, surveys are proposed as an alternative but in the long run – and where possible - improved administrative data collection is recommended.

4.9 There are advantages and disadvantages to using each data source that might be used to define the potential user groups. For some user groups, different sources will need to be used for different indicators. Where this is the case, issues tend to arise regarding the quality, coverage etc. of each source so that, for example, the size and composition of the user group may be different when derived from different data sources. For example, there is evidence that self-reported estimates of physical activity are different when collected using different methods e.g. detailed diary-keeping versus questions about what the respondent did in the last week. Where possible, existing statistics have been included in the discussion below in order to assess the extent to which the administrative data is successfully capturing the user group and the limitations of each proposed user group definition are discussed.

People who Need Care and Support

4.10 This group can be defined in two ways:

- People already receiving care i.e. the past and present clients of social care services across Wales. For people in receipt of services through a local authority, this group can be flagged in SAIL by acquiring the client datasets for all local authorities in Wales for SAIL.
- The whole population of Wales, since although a proportion of the population are already receiving social services – i.e. individuals recorded as either recent or current clients at the point of data capture - there will be some aspect of unmet need in the population – i.e. future clients. As noted above, the Welsh Demographic Service is the population spine used by SAIL and includes all individuals who have been registered with a GP in Wales, which, for the purposes of this study, can be assumed to include the whole population of Wales.

4.11 Data linking will allow:

- all indicators to be examined for recent and current local authority social care clients or for the whole population of Wales; and
- all indicators to be compared between recent local authority social care clients, current social care clients and the whole population of Wales.

4.12 It should be noted that the Project focused on identifying indicators based on individuals who were already in contact with social care services. An exercise to assess unmet need could be completed if a follow-on project is funded. This exercise would examine the characteristics of Social Services clients and seek to identify others in the population with similar characteristics (for example, in terms of health service use – use of accident and emergency, evidence of frailty or falls) but who are not currently clients.

Carers

4.13 The 2011 census found that around 370,000 people in Wales reported that they 'provided unpaid care to family members, friends, neighbours or others because of long-term physical or mental ill health or disability, or problems related to old age'.[8] A self-reported measure is more likely to highlight the true figure of carers and gives some indication of

---

[8] http://www.nomisweb.co.uk/census/2011/KS301EW/view/2092957700?cols=measures

the scale of caring in Wales and therefore the number of carers that might be flagged using administrative data.

4.14 The challenge when trying to flag carers is that the individuals they are caring for may be receiving a variety of services from different providers or may be receiving no services at all. Further, even if there is a record of the individual who is receiving the service, there may not be a record that they have a carer or, if they do, of who that carer is. There will therefore inevitably be some proportion of carers that cannot be identified from routine administrative records.

4.15 Given that carers are most likely to be caring for an individual in receipt of either health or social care services, three possible approaches can be taken to flagging the population of all carers in Wales. These would involve identifying whether a record exists of:

- the carers associated with social care service users (local authority and non-local authority);
- the carers associated with health care service users; and
- all carers, independent of the individuals they are caring for.

4.16 The review of the City and Council of Swansea Social Services data has demonstrated that a record did exist of the carers associated with local authority social care service users. As discussed in Chapter 3, the City and Council of Swansea data contains information for people who were specifically labelled as being a carer (this was around 700 individuals). However, the same data field includes information about family members and other individuals who may also be/have been carers but because only one label is permitted, ‘carer’ and ‘family member’ could not appear together. This suggests that family members who were carers were not distinguishable from family members who were merely key-holders or contacts e.g. in case of emergency. Further discussion with information officers and/or social care staff would be required to establish whether, in practice, recording an individual as a ‘carer’ was given precedence over all other codes – if this was the case, then the ‘carer’ label may be relatively reliable. Longer-term, it is recommended that a relatively simple change to data collection should be made to allow multiple labels to be attached to the ‘associated people’ so that a more robust flag for carers could be created.

4.17 Based on the numbers of carers recorded for the City and Council of Swansea but assuming that no changes are made to the coding of ‘associated people’, an all Wales local authority social services database with three years of data would yield an estimated 8,750 carers for flagging. Compared with the Census figure, it is clear that either the majority of carers associated with social services clients are not formally recorded as carers in local authority records, or that the majority of carers are caring for individuals who do not receive a service from their local authority social services department.

4.18 Although routine hospital records may identify patients that are carers, there is no facility in the administrative health data to record information about the carer(s) associated with a patient. It is therefore not possible to use routine hospital records to flag carers.

4.19 With regards to defining a population of carers independent of the people they are caring for, there are two possible approaches – firstly, using administrative data and secondly using survey data:

- As noted above, a review of administrative health records confirmed that the GP Event data contains a field for GPs to code whether patients are carers. However, on average around 1,500 carers per annum were found over the period 2010 to 2012 within the approximately 50% of practices in Wales included in SAIL at time of writing. If SAIL had data for all GPs in Wales, this would be likely to increase to around 4,000 carers per annum. When compared to the Census figure, this suggests that the carer code is not being used consistently by GP practices. This is not surprising, as other GP codes e.g. ethnicity, tend only to be used when they are relevant to a health condition and not as a matter of routine.

- The Welsh Health Survey asks questions specifically about whether respondents are carers and the amount of time involved. In 2011, just under 3,000 Welsh Health Survey respondents reported that they were carers. Around 60% of Welsh Health Survey respondents tend to give their consent to link for their data to be included in SAIL. The Welsh Health Survey would allow a flag to be created for a subset of around 2,000 self-reported carers for 2013-14. As more years of Welsh Health Survey data are added to SAIL, more detailed analysis would be possible. Because the Welsh Health Survey is a statistically representative survey, with some changes to the weighting scheme to account for differential ‘consent to link’ rates, these findings could be grossed to provide estimates for all Wales.

4.20 As noted above, groups of carers created using different sources are likely to differ in size and composition because they will be made up of different sorts of individuals. Those who self-report as a carer in the Welsh Health Survey are likely to be the same sorts of people who self-report as a carer on the Census, but those recorded as carer in the routine datasets will be a much smaller subset of these.

4.21 There are two options with regard to flagging a user group of carers, each of which may be preferred for different types of analysis:

- flag only the carers associated with local authority social care clients, knowing that the definition of carer will be narrower (and the numbers therefore smaller) but that the definition would nevertheless be relatively consistent across Wales; or
- flag carers from all available sources i.e. Welsh Health Survey, GP Event data and Social Care data, knowing that the definition of carer is wider (and the numbers therefore greater) but that there would be no consistent criteria for inclusion. Sensitivity analysis could be used to assess the extent to which the indicators vary by data source so that appropriate caveats can be reported alongside the estimates.

4.22 It should be noted, however, that even using local authority social care records to identify a group of carers would require some further investigation and potentially changes to data collection by local authorities.

Looked after children

4.23 Discussions with the Welsh Government suggest that capturing information from the Wales Children in Need Census would be the most accurate solution to flagging looked after children.

4.24 Local authorities annually provide aggregate data to Welsh Government about children who have a care plan in place at the end of the year. Since local authorities extract this data from background operational information, it would potentially be possible for Welsh Government to work with SAIL to negotiate regular data flows into SAIL from each local authority. Arrangements could then be put in place with SAIL to report the required aggregates to Welsh Government, potentially reducing the collection burden on local authorities in future, and supporting greater validation of the information held by SAIL. The presence of education data in SAIL might also support a more seamless means of joining the social care data to education data via the unique pupil number, which is often a burdensome manual aspect of the existing Welsh Government collection.

4.25 Further work would be required to assess whether this solution is feasible and agreeable to both Welsh Government and the local authorities, especially from the perspectives of data security, data protection and the need for a co-ordinated and practical approach to the validation of data on supply. However, if successful this approach could potentially be extended to cover other current and future requirements such as similar data for adult service users or the more detailed collection on care episodes and outcomes data for the subset of those children who are looked after by local authorities.

4.26 Using this method, the same numbers of looked after children would be ‘flagged’ as are reported in the Wales Children in Need Census, avoiding any inconsistencies in reporting. Since between 5,000 and 6,000 children in Wales are recorded in the Census each year, this would allow quite complex analysis to be completed, depending on the indicator of interest (for further discussion, see Chapter 5).

Adopted children

4.27 The GP Event data was the only dataset in SAIL in which adoption is specifically coded. However, as with the ‘carer’ code, GPs are unlikely to routinely code this, but only to code it if relevant to a health issue, such as the hereditary nature of some illnesses.

4.28 In the City and Council of Swansea data, adoptive parents are not specifically coded among the ‘associated’ people for clients. However, it was apparent that the PARIS system could be adapted for the future recording of whether clients are adopted, if required.

4.29 In summary, no reliable source was identified to define a user group of ‘adopted children’ but further investigation could be completed if a follow-on project is funded to examine whether changes to PARIS would offer a quick win solution.

Individuals with mental ill-health

4.30 As noted in Chapter 2, SAIL already contains a relatively comprehensive selection of routine health records. It is therefore entirely feasible to create a user group of individuals with mental ill-health using linked routine health records. The challenge with flagging this user group is in creating a suitable definition.

4.31 Mental ill-health, as a definition, is very broad, ranging from temporary low mood or mild anxiety through to pharmaceutically-managed chronic psychosis. This variation in the duration and severity of difficulties certainly needs taking into account when developing definitions and ideally needs including in the definition in some way, if for no other reason than that it would impact on service use. For example, an elderly person dealing with bereavement will require different levels of care to a young adult experiencing a schizophrenic episode. If these different requirements are not measured in some way, it will be difficult to interpret the indicators for this user group.

4.32 The Welsh Government approach to mental health is outlined in “Together for Mental Health”,[9] a delivery plan for 2012-16. This refers to the ‘All Wales Mental Health Promotion Network’ which defines mental health as: ‘an integral part of health, more than the absence of mental illness and intimately connected with physical health and behaviour’.[10]

4.33 However, terms used in the Well-being Statement, such as ‘I am happy’ or ‘I can engage and participate’ are likely to mean different things to people experiencing different levels of mental health. To an individual suffering a chronic psychiatric condition, levels of ‘happiness’ and ‘engagement’ are likely to be contextual and may change over time. It is therefore necessary to reflect the duration of a mental health condition when developing a flag.

---

[9] http://gov.wales/docs/dhss/publications/121203planen.pdf

[10] http://www.publicmentalhealth.org/page.cfm?orgid=749&pid=30860

4.34 To illustrate the kinds of approaches that have been used elsewhere, two different approaches to flagging individuals with mental ill-health are summarised below in Tables 4.2 and 4.3.

4.35 The categorisation used by the United States Agency for Healthcare Research and Quality[11] (see Table 4.2, below) is based on the patient's diagnosis. A flag based on this kind of approach could be achieved using data already available for linking in SAIL.

Table 4.2 United States Agency for Healthcare Research and Quality categorisation of mental health - based on diagnosis (uses data already available in SAIL)

Category	**Diagnoses**
Severe	Psychoses, Bipolar I and II disorders, Drug Dependence, Obsessive-Compulsive Disorder, Dysthymia, Oppositional Defiant Disorder, other 'severe' ICD-10 codes
Moderate	Panic Disorder, Separation Anxiety Disorder, Attention Deficit/Hyperactivity conditions, Agoraphobia without panic, Conduct conditions, Post-traumatic Stress Disorder, Drug Abuse, Alcohol Dependence, any impulse control conditions, Generalised Anxiety Disorder, Major Depressive Disorder (not dysthymia)
Mild	Social Phobia and other specific phobias, Alcohol Abuse, Intermittent Explosive Disorder, any anxiety disorder, any other mental health disorder, related 'mild' ICD-10 codes.

4.36 The categorisation summarised in Table 4.3, below, is based on the 'guiding principles for Community Mental Health Teams in delivering community mental health services in Wales', issued by Welsh Government in 2010.[12] This categorisation is based on the type of support required by the patient. This data is not currently available in SAIL but is a priority for acquisition for the National Institute for Social Care and Health Research-funded Project, so it is expected to be acquired by March 2015. Either of these could therefore be chosen depending on the requirements of policy colleagues.

Table 4.3 Welsh Government Community Mental Health Team guidance categorisation – based on type of support (uses data due to be acquired for SAIL by March 2015)

Category	**Description**
Severe	Tier 4: specialist inpatient hospitalisation, residential care or secure accommodation
Moderate	Community specialist mental health care, Assertive Outreach, supported accommodation.
Mild	Primary care support, counselling, mainstream support services.

4.37 A project by Dr Ann John of Public Health Wales is currently underway to develop a 'Mental Health Data Set' using a variety of health records already available in SAIL. Once

---

[11] http://www.hcup-us.ahrq.gov/reports/SOI.jsp

[12] https://www.rcpsych.ac.uk/pdf/The%20role%20of%20community%20mental%20health%20teams%20in%20delivering%20community%20mental%20health%20services%20guidance.pdf

this definition has been finalised, validated and published,[13] this would represent a third option for flagging a mental health user group for analysis purposes. It is recommended that a definition based on this one would be the simplest to apply, given that it will have been developed within SAIL and using the data available by data linking. However, further negotiation would be needed to gain permission to use the definition.

4.38 Whichever definition is used, further work would be required to manage issues around both the severity and duration of mental health conditions.

Individuals with substance misuse problems

4.39 ‘Substance misuse’ is a relatively vague term potentially covering a number of substances, legal or otherwise, which could cause harm if misused. The official statistics for Wales on this topic are based on referrals to substance misuse treatment centres from the Welsh National Database for Substance Misuse. Around 25,000 referrals were reported in 2012-13, excluding 23 referrals where there was no gender information provided and a further 5,691 referrals who had a DNA prior to assessment.[14]

4.40 All service providers in receipt of Welsh Government funding to deliver substance misuse treatment services are required to comply with the reporting requirements of the Welsh National Database for Substance Misuse. However, most of the data[15] is recorded at the level of referrals, not individual clients – this is because many clients do not wish to disclose their personal details and/or providers do not wish to discourage attendance by insisting on collecting personal details. This means that this Database is not suitable to allow the ‘flagging’ of all individuals in Wales with substance misuse problems.

4.41 A number of GP READ CODES are specifically used for ‘safeguarding’ purposes i.e. practices are encouraged to complete them to ensure that GP out of hours services are informed or to ensure that the information is included when patients move between practices. These codes include the following:

- family history of alcohol misuse;
- family history of substance misuse; and
- substance misuse.[16]

Further work would be required to assess the extent to which estimates derived using the GP data are likely to be accurate, since not all individuals with substance misuse problems will be evident to GPs or will be formally coded. When comparing estimates derived using the GP data with data derived from the Welsh National Database for Substance Misuse, the problem will be that a single individual may receive multiple referrals. However, where Abertawe Bro Morgannwg University reported to the Database at the client level rather than referrals, it may be possible to undertake some limited validation that can be used to assess the likely accuracy of the Wales-level estimates; an assessment could also be made of the extent to which individual level records are held by LHBs, even if they are not passed on to the Database.

---

[13] As of October 2014, Dr John is working with specific projects to validate these codes and they have found that the suitable set varies with different projects. Some will be freely available once the initial work has been published.

[14] http://gov.wales/docs/dhss/publications/131031profilesmtreatmenten.pdf

[15] With the exception of the Abertawe Bro Morgannwg University area, where they report on a client and not referral-based system.

[16] http://www2.nphs.wales.nhs.uk:8080/childprodocs.nsf/85c50756737f79ac80256f2700534ea3/abdd23b7bd37876e80257b83004eec07/$FILE/GP_Read_Codes_Report_amended%20May%209%202013.pdf

4.42 By linking to GP and other data sources, SAIL can identify those who have been recorded as heavy drinkers or as misusing substances whether or not they have been referred to a substance misuse treatment centre.

4.43 Misuse of each of the substances listed below is available as one or more READ CODES in the GP Event data within SAIL, so the list can be treated as a menu, from which policy colleagues can choose which 'substances' to include in their definition. N.B. the Public Health Wales report on Substance Misuse (2006) included all of the following 'substances' in its definition.

- Alcohol
- Amphetamines
- Anti-depressants
- Benzodiazepines
- Cannabis
- Cocaine
- Crack
- Ecstasy
- Heroin
- Methadone
- Solvents
- Steroids
- 'Other drugs' (this category is very vague and is only used where diagnostic codes are for 'Substance misuse – other' or similar).

4.44 Although further work would be required to refine the definition and to validate the resulting estimates, this does suggest that a group of individuals with substance misuse problems could be flagged in SAIL. Groups could be defined both for individuals who had been referred to treatment centres and for those who had not.

Older People aged 65 years and over

4.45 As noted above, the Welsh Demographic Service provides the week of birth of all individuals for whom an administrative record exists so a definition based purely on age can easily be flagged in SAIL and can be refined or redefined relatively easily.

Individuals with physical disabilities

4.46 'Disability' can be defined according to a 'medical model' or a 'social model'. Administrative health records would allow the creation of a user group based on the medical model. Survey data i.e. the National Survey for Wales and Welsh Health Survey, could be used to flag a group of people according to the social model.

4.47 To flag a group of disabled people using the social model, data from both the National Survey for Wales and Welsh Health Survey can be used. Both surveys ask respondents whether they are suffering from a long-term limiting illness.[17] The National Survey for Wales 2012-13 could be used to flag a group of 4,639 people with a self-reported disability for analysis purposes so an estimated 4,500 cases per year would be expected to be added. Using data from the National Survey for Wales, it would be possible to flag: young adults with disabilities, older people with disabilities and adults with disabilities, although

---

[17] Respondents are asked if they have any physical or mental illnesses lasting or expected to last for 12 months or more and the extent to which the illness affects the respondent's day to day activities.

the first group will be too small for detailed analysis (25, 2,111 and 2,814 respondents respectively). Consent to link was only asked from April 2013 on Welsh Health Survey so data is not yet available in SAIL; however, where around 33% of the 15,000 respondents per year[18] tend to report a limiting long-term illness on Welsh Health Survey and around 60% of respondents agreed to their survey answers being linked, it is likely that a further 3,200 disabled individuals per Welsh Health Survey year could be flagged.

4.48 To flag a group of disabled people using the medical model, a range of routine health records can be used. The health data in SAIL would allow the flagging of individuals suffering from various disabling conditions e.g. spina bifida, cystic fibrosis. The GP READ CODE for 'vulnerable adult' may also, in combination with other condition-specific codes, allow an assessment to be made of whether the disability is limiting and/or lifelong.

4.49 Unlike for mental health, no existing project has attempted to use routine health records to create a definition of disability. It is recommended that advice be taken from the Farr Institute[19] on whether any projects plan to create such a definition but if not, significant development work would be required to identify a list of appropriate conditions in order to generate and validate a useful definition. One possible way to create the definition would be to use the conditions reported by individuals who self-report as disabled in surveys to generate a definition that could then be applied using the appropriate codes in the administrative records.

4.50 As for other user groups, this category could be split into different age ranges e.g. young adults, adults and older people.

4.51 A longer-term solution to flagging disabled people would be to acquire data from the Department for Work and Pensions for disability-related benefits. This definition would relate mainly to working age individuals and to those suffering from conditions of a relatively greater severity and/or longer duration. This data acquisition may be completed as a result of other Welsh Government-supported projects.

4.52 Because the social model of disability uses survey data and is self-reported, it would be the simplest solution. However, because it uses survey data, numbers would be relatively small. Using a medical model would involve greater development work (unless, as noted above, any other projects propose to complete this work) but could be created essentially for the whole population.

4.53 Although numbers will be small for some indicators, particularly when using survey data to focus on the social model of disability, for other indicators estimates could be provided for both the medical and social definitions of disability, if required.

Individuals with learning difficulties

4.54 There are three main contexts in which a learning difficulty may be recorded – within health records, social care records and education records. Health and social care records will be more appropriate for flagging adults with learning difficulties, while education records combined with health records may be the best way to flag children with learning difficulties.

4.55 Both the GP Event READ CODES and the PARIS social care system include codes for 'learning disability'. Some recording of learning difficulties may also be found in data relating to Accident and Emergency visits or hospitalisations. How learning difficulties are

---

[18] http://gov.wales/statistics-and-research/welsh-health-survey/?lang=en

[19] A parallel research institution, also supported by SAIL in Wales, designed to deliver high-quality, cutting-edge research linking electronic health data with other forms of research and routinely collected data, as well as build capacity in health informatics research http://www.farrinstitute.org/

recorded under Accident and Emergency or hospital assessment circumstances may vary by hospital unit and according to whether the regular carer was present.

4.56 Searching for people with a learning disability in the linked health datasets indicated a range of possible coding. Individuals may either have a learning disability explicitly recorded or can be inferred to have a learning disability from the recording of other relevant health conditions or information about social care needs. There may be some level of subjectivity in whether conditions are recorded at all, particularly for less severe conditions.

4.57 Learning disabilities may not be recorded consistently, or individuals with a disability may have specific conditions recorded but not be coded as having a 'learning disability'. As with the concept of 'disability, an exploratory piece of work would therefore be undertaken to examine whether 'learning disability' can be defined by identifying where explicit mention is made of the term 'learning disability' and by using those individuals to identify a list of relevant conditions. The list could be used to identify additional cases where the same conditions occur, but without a specific 'learning difficulties' flag. It is recommended that, if follow-on work is funded, a terminology specialist in NWIS should be consulted to determine whether work has already been done to define this concept.

4.58 When using routine health records to define learning disability, policy colleagues will need to provide guidance on the level of severity they wish to include – not every learning disability requires either health or social care input.

4.59 The National Pupil Database can be used to identify pupils with Special Educational Needs, which are categorised as 'School Action', 'School Action Plus', 'Statutory assessment', or 'Statemented'. It can also be ascertained whether children attend a 'Special' class under 'class type'. 'Class Type' indicates whether a pupil is attending a nursery/reception or ordinary class and which also has a category of 'Special class'. A third field lists the type of Special Educational Need e.g. Autistic Spectrum Disorder, learning difficulties, Epilepsy and Hearing or Visual Impairment. As noted above, National Pupil Database data can only be used for research on educational attainment.

4.60 By using a combination of health records and - for children - education records, a flag for 'learning difficulties' can be developed. Further work would be required to validate the flag and manage any issues around defining severity.

Individuals with dementia

4.61 By using routine health records, it will be possible to flag a user group of people with dementia. Both GP READ CODES and ICD-10 codes for in-patients exist for dementia. However, further work would be required to refine the definition to exclude individuals with less severe 'memory problems'.

4.62 Overall, just over 6,000 people with a dementia diagnosis were found in the GP data for 2010-2013,[20] and almost 34,000 dementia patients are recorded in PEDW for 2010-2013.[21] Further work would be required to assess the extent of crossover between the two groups.

4.63 It should also be noted that 'consent to link' has been added to the Cognitive Function and Ageing Study (CFAS), which will interview a sample of 5,000 individuals aged 65 years and over in two areas in Wales (Gwynedd and Swansea local authorities) between 2013

---

[20] The READ Codes used were: Vascular dementia (Eu01. and sub-categories), Senile/pre-senile dementia (E00. and sub-categories), Dementia in other diseases classified elsewhere (Eu02. and sub-categories), Alcoholic dementia NOS (E012. and sub-categories), Alcoholic dementia NOS (Eu107).

[21] The ICD-10 codes used were: F00, F01, F02 and F03 and sub-categories.

and 2016 and will, in future, provide a better survey source than National Survey for Wales about individuals with dementia.

Recommendations

4.64 It would be possible to use the proposed additional outcome indicators (after the further development work identified above) to estimate the well-being of specific user groups. It is possible to develop definitions that could be used to flag a range of user groups in SAIL, some examples of which have been proposed in this Chapter. However, in order to do so, further work would be needed:

- to address the various methodological challenges outlined above; in particular:
    - where some definitions are dynamic, e.g. mental ill-health, decisions need to be made about the level of severity and the duration of condition that are included when creating the 'flag'; and
    - where the data in SAIL is longitudinal, it is necessary to decide which point in time to choose for analysis since individuals may move into or out of categories over time e.g. individuals will move from 'child' to 'young adult' to 'adult' to 'older adult' over time.
- to assess the robustness of the figures thus produced; and
- to identify any potential additional data sources that could help to flag the user groups.

The extent to which the proposed additional outcome indicators can be used to estimate the well-being of specific user groups is discussed in Chapter 5.

Chapter 5 Identifying potential outcome indicators

Key points

- The Project identified a number of potential outcome indicators that could be used in addition to the current set of outcome indicators in the outcomes framework.
- Additional headline indicators available without linking data are as follows (N.B. if reporting is required separately by user group or for social care users, data linking WILL be required):
  - Percentage of young people in receipt of Education Maintenance Allowance Wales (EMA) or Assembly Learning Grant (Further Education) (ALG FE).
  - Percentage of adults with a record in the Lifelong Learning Wales Record (LLWR).
  - Percentage of adults, 'younger adults' and 'older adults' participating in a range of sports and physical recreation activities, volunteering in sport, with sports club membership, participating in competitive sport or using leisure and fitness centres.
- Additional headline indicators that **may** be available without linking data are as follows:
  - Percentage of children in receipt of Flying Start meeting the expectation that they would attain Level 2 in each subject at age 7 years/Key Stage 1.
- Additional headline indicators available with relatively little additional acquisition or development work but only by linking data are as follows:
  - Percentage of patients given one of 20 advice leaflets or any of several kinds of verbal advice by their GP practice.
  - Percentage of Flying Start eligible children assessed to be fully prepared for school on entry to Foundation Phase (data not available until 2016).
  - Percentage of Flying Start eligible children meeting the expectation that they would attain Level 2 in each subject at age 7/Key Stage 1.
- Additional headline indicators available with significant additional data acquisition or development work and only by linking data are as follows:
  - Objective, physician-reported status of physical and mental health.
  - Average self-reported 'happiness' score for social care users (compared with the general population).
  - Percentage of people who have experienced domestic violence receiving social care services or supported housing services e.g. Supporting People.
  - Percentage of people who: i) report feeling that they belong in their neighbourhood: ii) think that their local area is a place where people from different backgrounds get on well together: or iii) think that people in their local area treat each other with respect and consideration and who are in receipt of: a) social care services; and b) supported housing services e.g. Supporting People.
  - Percentage of people who report being satisfied with: i) 'personal relationships': and ii) 'the well-being of (their) children', and who are in receipt of: a) social care services; and b) supported housing services e.g. Supporting People (would require additional data acquisition).
  - Percentage of social care users in receipt of benefits.
  - Percentage of children in receipt of Flying Start assessed to be fully prepared for school on entry to Foundation Phase (data not available until 2016).
  - Percentage of former Children in Need receiving Adult Social Care services.

    - Percentage of former Children in Need receiving services through supported housing services e.g. Supporting People and Extracare.
    - Percentage of private residences meeting the Local Government and Housing Act 1985 overcrowding definition.
- Most of the indicators can be reported separately for known social care users as well as, if required, some or all of the proposed user groups discussed in Chapter 4.
- It is recommended that a follow-on project is funded to allow more work to be done to complete the further development work required:
    - to report on those proposed additional outcome indicators for which only relatively small amounts of additional work would be required;
    - to seek out additional datasets needed to improve the proposed additional outcome indicators or to identify further outcome indicators;
    - to ensure that any further data acquired by ADS is fit for purpose; and
    - to refine and validate some of the proposed outcome indicators.

N.B. As discussed in Chapter 2, data linking can open up further lines of enquiry that would help services to be targeted more effectively. The focus of this Project was on proposing additional outcome indicators but if a follow-on project was funded, these further lines of enquiry could be explored.

Introduction

5.1 The well-being statement develops the definition set out in the Act and defines what well-being means to people under each aspect of well-being defined in the Social Services and Well-being (Wales) Act 2014. Well-being is defined as:

- physical and mental health and emotional well-being;
- protection from abuse and neglect;
- education, training or recreation;
- domestic, family and personal relationships;
- contribution made to society;
- securing rights and entitlements;
- social and economic well-being; and
- suitability of living accommodation.

5.2 The national outcomes framework sets out the outcome indicators that will be used to measure whether well-being, set out in the well-being statement, is achieved. This Project has identified potential additional outcome indicators that could be delivered using the SAIL Databank to underpin the well-being statement.

5.3 It should be noted that most of the potential additional outcome indicators listed in this Chapter will require further development work or further data acquisition before reporting can begin. This Chapter also explores the extent to which each headline indicator can be reported separately by the user groups discussed in Chapter 4, and describes the data acquisition and/or development work required to allow reporting of each potential indicator.

5.4 Although some of the proposed additional indicators identified below may be available without linking, they have been included in the Report on the basis that if reporting is required separately by user group or for social care users, data linking will be required.

5.5 It should be noted that this project did **not** look at the existing indicators and assess whether they could be provided for individual user groups – it is therefore recommended that this should be done if a follow-on project is funded.

5.6 It should also be noted that a variety of other projects are being undertaken to improve the broader evidence base for policy-making using linked data. It is therefore recommended that liaison is maintained with other projects where there is potential to identify further indicators to support the well-being outcomes framework.

5.7 It should also be noted that (as discussed in Chapter 2) data linking can open up further lines of enquiry that would help services to be targeted more effectively e.g. following service users up over time to examine their long-term outcomes, identifying the extent of 'multiple' low well-being or persistent low well-being and exploring the relationship between service use, subjective well-being and satisfaction with services. The focus of this Project was on proposing additional outcome indicators, and possibly on the ability to report them separately for user groups or social care users. If a follow-on project was funded, these further lines of enquiry could be explored further.

5.8 As with defining the user groups, both survey and administrative sources were considered when looking to identify additional outcome indicators. The proposed additional outcome indicators are summarised in Appendix 1 with a brief indication of the likely timescales for reporting, given the requirements for additional data acquisition and development work. Each proposed indicator is discussed in greater detail in the remainder of this Chapter; as discussed in Chapter 2, each data source has limitations and data linking can overcome some of these limitations. The limitations of the data are explored in detail for each indicator below.

5.9 Some of the proposed indicators use single administrative sources i.e. it is not necessary to link datasets in order to report on the indicator. The analysis would therefore be deliverable as a stand-alone analysis of a single dataset e.g. the Lifelong Learning Wales Record or the National Pupil Database. However, even where this is the case, data linking may be necessary to provide analysis of the indicator(s) separately by user group or separately for social care users. Where the proposed indicator is available without linking data, this has been noted in the commentary below.

5.10 For some proposed indicators, although they can be delivered using single sources, the only way to access the required large scale datasets for all Wales is through SAIL/the ADRC-W, for example the ADRC-W will be the only environment where all social care data for Wales can be brought together for research purposes.

5.11 It will be possible to provide estimates for some indicators by user group and for known social care users. Where this is possible, this will be noted in this Chapter. Where information about different indicators is to be found in different sources, the best source to flag the users groups may vary by indicator. An Excel spreadsheet has been created to document in detail the best source by indicator and user group.

5.12 Appendix 1 provides a brief summary of the proposed additional outcome indicators that could be used to support the outcomes framework. The sections below contain a discussion on the benefits and limitations of each potential outcome indicator.

Well-being: I know and understand what care, support and opportunities are available to me; I get the help I need, when I need it, in the way I want it

Proposed indicator: Percentage of patients given one of 20 leaflets or any of several kinds of verbal advice by their GP practice.

5.13 The GP Event data includes information about advice provided to patients on a variety of subjects. The advice can be coded as various kinds of verbal advice or the GP may have given the patient one of 20 leaflets (e.g. self help, diet, asthma, screening, contraception).

5.14 From 2004 to 2012, 115,000 people were given a leaflet by their GP and nearly two million instances of verbal advice were recorded. The top five verbal advice items were:

1. On suitable footwear;
2. “Advice” (non-specific);
3. On foreign travel;
4. On treatment given; and
5. Patient advice given (non-specific).

Codes that may be of particular interest are: a) written advice about benefits of physical activity; b) lifestyle advice regarding exercise; c) advice about weight; d) lifestyle advice about alcohol; e) lifestyle advice about smoking; and f) advice about domestic violence. Where advice is non-specific, it would be possible to look at other records for the same person on the same day e.g. prescribing, and make an inference.

5.15 As noted in Chapter 2, GP Event data was available in SAIL for around 50% of the population of Wales (as of August 2014) – but is increasing all the time. Some GP Event data can be reported directly by NWIS but this is not possible for all GP Event data so data linking may be the only way to access the required GP Event data information for all Wales. If a follow-on project is funded, or if policy team resources allow, discussions could be had with NWIS to establish whether further indicators could be delivered directly by NWIS.

5.16 Given that there is interest in whether well-being has improved as a result of the care and support received, data linking will allow comparisons to be made of the physical or mental health of individuals before and after receipt of the leaflet and/or advice.

5.17 Although further work would be required to confirm this, given the numbers of advice events reported above, it is likely that this indicator could be reported separately if required for:

- ‘Younger Adults’
- ‘Older Adults’
- Individuals with physical disabilities
- Individuals with mental health difficulties
- Carers
- Social care users (versus the general population)

5.18 The headline indicator can be reported relatively easily using existing data. Further work would be required, as discussed in Chapter 4, above, if the indicator was required to be reported separately for the user groups ‘individuals with physical disabilities’ or ‘individuals with mental health difficulties. Further work would also be required to decide which aspects of physical and mental health to focus on, should a before and after comparison be of interest.

5.19 Reporting for this indicator for local authority social care users would require the acquisition of all-Wales social care data. Acquisition of local authority social care data and making it ‘research ready’ may take as long as 1 to 2 years.

Physical and mental health and emotional well-being: I am healthy; I am happy

Proposed indicator: Objective, physician-reported status of physical and mental health

5.20 The routine health datasets already available in SAIL can be used to develop objective i.e. physician-reported rather than self-reported, indicators of physical and mental health.

5.21 For general physical health, various indicators could be developed but indicators that have been developed for other projects (e.g. the Welsh Government-ESRC Data Linking Demonstration Projects)[22] have focused on the number of contacts with the GP practice or the number of prescriptions received, the number of Accident and Emergency visits, the number of hospital admissions or the number of hospital in-patient days.

5.22 Data for hospital (including Accident and Emergency, out-patient and in-patient care) is available for the whole population of Wales. Where indicators include GP Event data, this was available for around 50% of the population of Wales at the time of writing, although this figure is increasing all the time.

5.23 For mental health, as discussed in Chapter 4, a variety of definitions can be adopted using existing SAIL data to flag individuals with mental health difficulties. A more general indicator of mental health could be developed building on the chosen definition to count, for example, the number of consultations relating to the conditions included in the definition or the number of prescriptions for drugs used to treat those conditions e.g. depression or anxiety.

5.24 Although further work would be required to confirm this, given the numbers of advice events reported above, it is likely that this indicator could be reported separately if required for:

- ‘Younger Adults’
- ‘Older Adults’
- Individuals with physical disabilities
- Individuals with learning disabilities
- Individuals with mental health difficulties
- Individuals with dementia
- Individuals with substance misuse problems
- Children in Need
- Vulnerable Adults
- Carers
- Social care users (versus the general population)

Data linking would be necessary in order to report the indicator separately for some of these user groups.

5.25 The headline indicator for physical health can be reported relatively easily using existing data. The headline indicator for mental health would require further work to agree a definition but, once agreed, could be reported using existing data.

5.26 If the indicator was required separately for the user groups listed above or for local authority social care users, the acquisition of all-Wales social care data would be required. Acquisition of local authority social care data and making it ‘research ready’ may take as long as 1 to 2 years. Some of the user groups would need further work as discussed in Chapter 4.

Proposed Indicator: Average self-reported ‘happiness’ score[23] for social care users (when compared with the general population)

---

[22] http://gov.wales/statistics-and-research/making-better-use-existing-data/?lang=en

[23] Social care questions will be asked in the last National Survey year – it will therefore be possible to report on self-reported happiness score among social care users just using the survey. It is proposed that average score would be used so this is to be consistent with this.

5.27 The National Survey for Wales is the best source of data on self-reported ‘well-being’ for Wales. The National Survey for Wales represents a rich dataset but for a relatively small sample of the population of Wales. Consent to Link was introduced onto the National Survey for Wales in January 2012, with a consent rate for data linkage in 2012-13 of 64% (9,303 cases) and for 2013-14 of 66% (9,795 cases). This means that when required, around 19,000 cases are available for linking and by summer 2015, a grand total of around 28,000 cases will be available.

5.28 As noted in Section 2.5, whether National Survey for Wales respondents are also social care users can be identified in two ways, either by linking National Survey for Wales data to routine social care records or by asking National Survey for Wales respondents directly whether they are social care users. Respondents to National Survey for Wales 2014-15 were asked whether they were social care users and a set of questions relating to their satisfaction with the services they had received. Because the questions about social care use and satisfaction are only being asked in a single year of the National Survey for Wales, using the National Survey for Wales to identify social care users will result in a smaller sub-group for analysis than will be identifiable by linking National Survey for Wales to administrative records. So, in short:

- Self-reported well-being would be available for a group of social care users identified by using routine social care records linked to *all* National Survey for Wales years (i.e. 2012-15). The City and Council of Swansea social care data contained 26 social care users who were also National Survey for Wales respondents, so around 500 known social care users are likely to be found in the National Survey for Wales data already available in SAIL (and as more years of National Survey for Wales are added, this number would increase). This would allow an exploration, for example, of the extent to which the level, type or duration of use is associated with differences in well-being.
- Self-reported satisfaction information (in addition to self-reported well-being information) would be available for those National Survey for Wales 2014-15 respondents who also self-report as being social care users. It is difficult to estimate the number of National Survey for Wales respondents who may self-report as social care users. The Living in Wales Survey 2006 asked approximately 7,500 respondents whether they were social care users and found a total of 116 users. With the greater sample size of the National Survey for Wales, this equates to around 200 respondents who may self-identify as users, giving a smaller group for analysis purposes. The relationship could be explored, for this smaller sub-set of National Survey for Wales respondents, between the level, type or duration of social care use, self-reported well-being and satisfaction with services.

5.29 By linking the survey data to routine social care records, it would also be possible to identify the extent to which survey respondents who reported being social care users also appear in administrative social care records and in what capacity/under what circumstances. In theory, since the administrative records are longitudinal while the National Survey for Wales is cross-sectional and respondents were asked about current use of services, by linking to administrative social care records it would be possible to find a greater number of users than would be found by asking survey respondents to self-report at a single point in time. It would also be possible to explore whether some respondents choose not to self-report as social care users when asked to do so in the context of a survey e.g. because they find the question too sensitive.

5.30 The National Survey for Wales asks four general questions about the respondent’s satisfaction with their life, the extent to which they feel that the things they do in their life are worthwhile, how happy they are feeling and how anxious they are feeling. A set of questions is also asked about satisfaction with different aspects of life e.g. their work situation.

5.31 The National Survey for Wales questions about well-being are as follows:

1. Overall, how satisfied are you with your life nowadays?
2. Overall, to what extent do you feel that the things you do in your life are worthwhile?
3. Overall, how happy did you feel yesterday?
4. On a scale where 0 is 'not at all anxious' and 10 is 'completely anxious', overall, how anxious did you feel yesterday?
5. Overall how satisfied are you with :
    a. your personal relationships;
    b. your physical health;
    c. your mental well-being; your work situation;
    d. your financial situation;
    e. the area where you live;
    f. the amount of time you have to do things you like doing; and
    g. (if respondent has children) the well-being of your children.

The responses to all of the questions were on a scale from 0 (not at all satisfied) to 10 (completely satisfied).

5.32 Reporting for this indicator would require the acquisition of all-Wales social care data. As noted above, acquisition of local authority social care data and making it ‘research ready’ may take as long as 1 to 2 years. Further work will be required, once the routine social care data has been acquired, to identify how many social care users were interviewed for National Survey for Wales but as noted above, this is likely to be at least 500 cases. If numbers are sufficient, it will be possible to report an average ‘happiness’ score for local authority social care users and to compare it to the score for not local authority social care users. Numbers may be small but work could be done to establish how representative the estimate is likely to be and to provide, if possible, a confidence interval.

Protection from abuse and neglect: I am safe and protected from abuse and neglect

Proposed indicator: Percentage of people who have experienced domestic violence receiving social care services or supported housing services e.g. *Supporting People*, *Extracare*

5.33 Reporting for this indicator would require the acquisition of all-Wales social care data. As noted above, acquisition of local authority social care data and making it ‘research ready’ may take as long as 1 to 2 years. A proposed Welsh Government-supported data linking project is currently proposing to acquire data for supported housing schemes; this project will also assess the quality of the data and further advice can then be provided on the feasibility of the supported housing component of this indicator.

5.34 Once the data has been acquired, it would be a relatively simple task to flag all local authority social care users in Wales and, if data quality indicates that it is feasible, to flag recipients of Supporting People and Extracare.

5.35 The simplest analysis would use local authority social care data to identify abuse or neglect as a reason for referral or using other social care data. A more ambitious exercise could be undertaken to identify people who have experienced abuse using, in addition, police data and data from third sector organisations involved in providing services.

5.36 Acquisition of police data by ADS and acquisition of third sector data as part of a follow-on project is likely to take around one year – this estimate is based on the fact that Cardiff Women's Aid have already begun working with SAIL to deliver this data for Cardiff and that this will likely act as a driver for the rest of Wales and may act as a driver for investment in improving linked data by the relevant area of Welsh Government. Previous discussions with South Wales Police and Association of Chief Police Officers Cymru suggest a willingness to share data.

Education, Training and Recreation: I can learn and develop to my full potential; I can do the things that matter to me

Proposed Indicator: Percentage of Flying Start **eligible** children assessed to be fully prepared for school on entry to Foundation Phase

5.37 Children eligible for Flying Start are already flagged in SAIL for the years 2004-2012 and flagging will be updated annually as part of parallel work being delivered to support the evaluation of Flying Start.

5.38 The new baseline 'school preparedness' assessment for the Foundation Phase is due to be introduced from September 2014 and is statutory from September 2015. The data can, if required, become part of a revised National Pupil Database extract provided to SAIL on an annual basis. This includes data for all children in Wales except those in independent schools.

5.39 The baseline 'school preparedness' assessment is statutory from September 2015, with the first collection of data in the summer of 2016. Even though assessments are done at the start of the year, it is more convenient and economical to collect the data with the other assessment data at the end of the year. Following this timetable, data could be made available for linking in autumn 2016.

Proposed indicator: Percentage of children **in receipt of** Flying Start assessed to be fully prepared for school on entry to Foundation Phase.

5.40 The feasibility is currently being assessed of adding a flag for children receiving some component of Flying Start to the National Pupil Database or to the Child Health System. Since both datasets already regularly flow into SAIL, an indicator could be developed that used the more accurate figures of children in receipt of Flying Start rather than the broader measure of those living in an eligible area.

5.41 As for the indicator above, the baseline 'school preparedness' assessment will be available within National Pupil Database from autumn 2016.

Proposed indicator: Percentage of Flying Start **eligible** children meeting the expectation that they would attain Level 2 in each subject at age 7/Key Stage 1

5.42 As part of parallel work being delivered to support the evaluation of Flying Start, children eligible for Flying Start are already flagged for the years 2004-2012 and flagging will be updated annually.

5.43 Educational attainment data is included in the National Pupil Database. This includes data for all children in Wales except those in independent schools.

Proposed indicator: Percentage of seven year old children **in receipt of** Flying Start meeting the expectation that they would attain Level 2 in each subject at Key Stage 1

5.44 The feasibility is currently being assessed of adding a flag for children receiving some component of Flying Start to the National Pupil Database or to the Child Health System. Since both datasets already regularly flow into SAIL, this would allow an indicator to be developed that used the more accurate figures of children in receipt of Flying Start to be used rather than the broader measure of those living in an eligible area. If receipt of Flying Start is flagged in the National Pupil Database, this indicator can be delivered without linking data. However, if receipt of Flying Start is flagged in the Child Health System, linking between the National Pupil Database and the Child Health System would be required to report this indicator.

5.45 Educational attainment data is included in the National Pupil Database.

Proposed indicator: Percentage of young people in receipt of Education Maintenance Allowance Wales (EMA) or Assembly Learning Grant (Further Education) (ALG FE)

5.46 The two main existing sources of data are the Lifelong Learning Wales Record and Student Loan Company records.

5.47 An existing ADRC-W project proposes to acquire the LLWR and SLC data in order to build on the Welsh Government-funded evaluation of EMA and ALG FE, which used an existing linked database. However, the report for the Welsh Government-funded evaluation of EMAs[24] reported that the use of this database raised more questions than it answered. The team therefore recommend creating a linked database using ADRC methods and extending it to later cohorts. There are some concerns surrounding the apparent declining quality of SLC data in terms of consent rates.

Proposed indicator: Percentage of adults with a record in the Lifelong Learning Wales Record (LLWR)

5.48 As noted above, a project is already proposed to acquire LLWR for SAIL so this is likely to take place in the short-term.

5.49 Although further development work would be required to confirm this, it is likely that this indicator could be reported separately if required for:

- 'Younger Adults'
- 'Older Adults'
- Individuals with physical disabilities
- Individuals with mental health difficulties
- Carers
- Social care users

5.50 If the indicator was required separately for the user groups listed above or for local authority social care users, the acquisition of all-Wales social care data would be required. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years. Some of the user groups would need further work as discussed in Chapter 4.

---

[24] http://gov.wales/statistics-and-research/evaluation-education-maintenance-allowance-assembly-learning-grant/?lang=en

Domestic, family and personal relationships: I belong; I have safe and healthy relationships

Proposed indicator: Percentage of people who: i) report feeling that they belong in their neighbourhood; ii) think that their local area is a place where people from different backgrounds get on well together; or iii) think that people in their local area treat each other with respect and consideration and who are in receipt of social care services or supported housing services e.g. *Extracare, Supporting People.*

5.51 National Survey for Wales data is required for this indicator. As noted above, around 19,000 cases are currently available for linking and by summer 2015 a total of around 28,000 cases will be available.

5.52 Local authority social care data would need to be acquired in order to flag all social care users in Wales. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years.

5.53 Local authority social care and/or housing data (depending on where supported housing services data is held) would need to be acquired in order to flag recipients of Supporting People and Extracare. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years. Further work will be required, once the routine social care data has been acquired for all local authorities in Wales, to identify how many users of supported housing were interviewed for National Survey for Wales. If numbers are sufficient, it will be possible to compare e.g. the proportion feeling they belong among recipients versus non-recipients.

5.54 Further work will be required, once the routine social care data has been acquired for all local authorities in Wales, to identify how many social care users were interviewed for National Survey for Wales but as noted in Section 5.23, this is likely to be at least 500 cases. This will make it possible to compare e.g. the proportion feeling they belong among social care recipients versus non-recipients.

Proposed indicator: Percentage of people who report being satisfied with: i) 'personal relationships'; and ii) 'the well-being of (their) children', and who are in receipt of social care services or supported housing services e.g. *Extracare, Supporting People.*

5.55 National Survey for Wales data is required for this indicator. As noted above, around 19,000 cases are currently available for linking and by summer 2015 a total of around 28,000 cases will be available.

5.56 Local authority social care data would need to be acquired in order to flag all local authority social care users in Wales. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years.

5.57 As noted above, local authority social care and/or housing data (depending on where supported housing services data is held) would need to be acquired in order to flag recipients of Supporting People and Extracare. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years. Further work will be required, once the routine social care data has been acquired for all local authorities in Wales, to identify how many users of supported housing were interviewed for National Survey for Wales. If numbers are sufficient, it will be possible to compare e.g. the proportion feeling they belong among recipients versus non-recipients.

5.58 Further work will be required, once the routine local authority social care data has been acquired for all local authorities in Wales, to identify how many social care users were interviewed for National Survey for Wales but as noted in Section 5.23, this is likely to be at least 500 cases. This will make it possible to compare e.g. the proportion feeling they belong among social care recipients versus non-recipients.

Contribution made to society: I can engage and participate; I feel valued in society

Proposed indicator: Percentage of adults, ‘younger adults’ and ‘older adults’ and social care users (versus the general population):

- participating in a range of sports and physical recreation activities;
- volunteering in sport;
- with sports club membership;
- participating in competitive sport; and
- using leisure and fitness centres.

5.59 The items listed above would be delivered using data from Active Adults 2012 and 2014 surveys.

5.60 Reporting for additional user groups may be possible by linking. Data for Active Adults 2012 is already available for linking. Data for Active Adults 2014 will be available in summer 2015. Active Adults surveys around 15,000 adults per wave, of which around 60% give consent to link, giving around 9,000 cases that can be linked to other administrative records held for respondents.

5.61 By linking to administrative health records and where numbers allow, these indicators could be delivered for:

- individuals with physical disabilities; and
- individuals with mental health difficulties

5.62 An indicator of the use of leisure and fitness centres could, in addition, be delivered using local authority leisure card data. Welsh Government Knowledge and Analytical Services is currently investigating the extent to which local authorities in Wales hold individual-level data for local authority-run leisure and sports centres. Where it is unlikely that data could be acquired for non-local authority leisure centres, an exercise is recommended (this could be completed if a follow-on project is funded) to use data from the Active Adults 2012 and 2014 surveys to assess the coverage of local authority leisure card data and therefore the validity of an indicator developed using this method.

5.63 Local authority social care data would need to be acquired in order to flag all local authority social care users in Wales; acquisition of local authority social care data and making it ‘research ready’ may take as long as 1 to 2 years. If the current exercise identifies that individual-level local authority leisure card data exists for most or all local authorities in Wales, acquisition can be completed alongside acquisition of social care data so is likely to take 1 to 2 years.

Social and economic well-being: I am supported to work; I have a social life and can be with the people that I choose; I do not live in poverty; I get the help I need to grow up and be independent; I get care through the Welsh language if I need it

Proposed indicator: Percentage of social care users in receipt of benefits

5.64 Acquisition of data from the Department for Work and Pensions would be required to allow the flagging of individuals receiving benefits. Data from the Department for Work and Pensions is likely to be acquired within approximately two years, which would allow the flagging of individuals receiving benefits.

5.65 Local authority social care data would need to be acquired in order to flag all social care users in Wales. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years.

Proposed indicator: Percentage of ex-Children in Need receiving Adult Social Care services

5.66 Local authority social care data would need to be acquired in order to flag all local authority social care users in Wales and to flag historic Children in Need. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years.

Proposed indicator: Percentage of ex-Children in Need receiving services through supported housing services e.g. *Supporting People* and *Extracare*

5.67 Local authority social care data would need to be acquired in order to flag all local authority social care users in Wales and to flag historic Children in Need. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years.

5.68 Local authority social care and/or housing data (depending on where supported housing services data is held) would need to be acquired in order to flag recipients of Supporting People and Extracare. Acquisition of local authority social care data and making it 'research ready' may take as long as 1 to 2 years. Further work will be required, once the routine social care data has been acquired for all local authorities in Wales, to identify how many users of supported housing were interviewed for National Survey for Wales.

Suitability of living accommodation: I have suitable living accommodation that meets my needs

Proposed indicator: Percentage of private residences meeting the Local Government and Housing Act 1985 overcrowding definition

5.69 As noted above, data linking will be part of a range of proposed solutions to meeting Welsh Government's requirements for evidence on house conditions in Wales going forward. This means that further work is likely to be completed to acquire housing data from local authorities, to acquire Valuation Office Agency data and to assess whether Energy Performance Certificate data is worth acquiring. These will include information about the number of bedrooms in the residence. By linking this data to the Welsh Demographic Service (the population spine used by SAIL to link datasets), it will be possible to assess how many individuals are resident against the number of bedrooms. Acquisition of local authority housing data is likely to take 1 to 2 years. Acquisition of Valuation Office Agency data may require primary UK legislation, which may take two years or more.

Recommendations

5.70 As discussed in the preceding sections, some reporting of proposed additional outcome indicators could be completed in the short-term as well as a significant amount of further development work to refine and validate further proposed outcome indicators.

5.71 Liaising with other data linking projects would also be important to ensure that any data acquired by SAIL for those projects is fit for purpose and accompanied by sufficient metadata. As noted in Chapter 4 and in the discussion above, various methodological challenges need to be overcome before some kinds of analysis can be performed.

5.72 As with the user groups, further work could be done to identify additional data sources that could allow additional outcome indicators to be proposed, particularly in areas of the well-being framework for which relatively few indicators are currently available.

5.73 Further work would be needed to allow the reporting of the proposed additional outcome indicators set out in this Chapter. In particular, further work is required:

- to acquire social care data from all local authorities in Wales - plus any additional data required to allow the reporting of the proposed outcome indicators - and to ensure that all data acquired is fit for purpose;
- whilst this project has identified a number of additional outcome indicators, these rely on data that is either already in SAIL or known to the authors and could be acquired for SAIL; further work could be done to seek out further data sources in order to propose further additional outcome indicators; and
- to refine and validate some of the proposed outcome indicators.

Chapter 6 Recommendations for further work

Key recommendations

- Some of the additional outcome indicators proposed in Chapter 5 can be reported without using linked data and some will require datasets to be linked. If reporting is required for the indicators set out in Chapter 5 either for any of the user groups proposed in Chapter 4 or specifically for social care users, data linking will be required. To ensure consistency it would be preferable for all reporting to be done using the same set of linked datasets. Because many of the datasets required are only held in SAIL, it is therefore recommended that the linking is done in SAIL.
- As noted in Chapter 5, most of these indicators will require further development work or further data acquisition before reporting can begin.
- In order to report on the additional outcome indicators proposed in Chapter 5, work will be needed to:
    - acquire data for all social care users from all local authorities in Wales;
    - identify ways to increase the linking rate for future data acquisition;
    - ensure the data acquired and the metadata provided was fit for purpose, completing any data cleaning, validation and initial analysis to ensure it was 'research ready';
    - analyse the all-Wales local authority social care data and any other datasets acquired in order to report on the proposed additional outcome indicators;
    - where appropriate, liaise with relevant research projects to ensure any data acquired by those projects and any definitions developed by those projects are suitable to allow reporting of the relevant proposed indicators;
    - work with appropriate projects and organisations, including WCCIS, to achieve improvements to data collection to allow, in future, more robust versions of the proposed additional outcome indicators to be provided and/or further outcome indicators to be proposed;
    - whilst this project has identified a number of additional outcome indicators, these rely on data that is either already in SAIL or known to the authors and could be acquired for SAIL; further work could be done to seek out further data sources in order to propose further additional outcome indicators; and
    - continue development work to refine definitions and validate estimates for the proposed additional outcome indicators.
- An alternative option is to wait until other projects such as the Social Care and Health Of Older People (SCHOOP) project have acquired additional datasets for SAIL and, at a later stage, revisit the viability of SAIL to provide additional outcome indicators. There are limitations to this option:
    - because they tend to focus on specific user groups, socio-demographic groups and/or geographical areas rather than seeking to acquire all local authority social care data for Wales, relying on other projects to acquire social care data would risk the data not being suitable to evidence any of the proposed additional outcome indicators;
    - even if other projects develop new definitions, these would be unlikely to be exactly what is needed for the proposed indicators;
    - there would be no opportunity to feed into the improved data collection required to create/improve some indicators; and

- there would be no opportunity to identify and acquire additional datasets to allow additional indicators to be proposed in the future.

6.1 A second option would be to wait for all Local Authorities in Wales to migrate to the Welsh Community Care Information System (WCCIS) and to acquire a standardised social care data set for Wales for SAIL for analysis purposes. There are limitations to this option:

- the migration process will take a number of years to complete. The data will not be in a standardised format at the point of migration. Even after the data is migrated, there will be a lengthy and iterative process of standardising critical and common data items. This would delay any further work;
- there would be no opportunity to feed into the improved data collection required to create/improve indicators; and
- it is not yet clear how much historical data will be migrated as part of the process. Any new dataset relying on WCCIS to provide data may therefore be limited, at first, as to how many cases would provide a sufficiently long timeline of information to allow change over time to be reported for the relevant indicators.

Introduction

6.2 To use linked data to report on the proposed additional outcome indicators, the main challenges are:

- the acquisition of any additional datasets required to report on the proposed additional outcome indictors; and
- the need for further development work - beyond the acquisition of data - to refine the definitions of the proposed indicators and, if required, of the user groups in order to validate the resulting estimates and to complete the analysis required to report on the proposed indicators.

The work required to meet these two challenges is summarised in the following two sections.

Data Acquisition

6.3 The biggest gap in the data required to deliver additional outcome indicators is the social care data held by local authorities in Wales. As noted above, work was funded by the National Institute for Social Care and Health Research to document the social care data held by all 22 local authorities in Wales but the project was not designed to acquire the data for SAIL.

6.4 Further work would be required, in conjunction with local authorities, to acquire social care data from all local authorities in Wales and to ensure the data is deposited in SAIL.

6.5 As noted in Chapter 2, because it receives its core funding from the ESRC, projects supported by the ADRN cannot consist simply of the reporting of a set of indicators. Although the methodological work to define some of the proposed outcome indicators or user groups would fall within the ADRN remit, the proposal would need to make a wider contribution to the field of research.

6.6 **Recommendation**: to ensure the proposed follow-on project would meet ADRN criteria, the proposal should ideally include some of the more complex lines of enquiry suggested in Chapter 2, where data linking is adding significant value, e.g. examining the dynamics or complexity of well-being among service users.

6.7 The data acquisition plus any work to get the data ‘research ready’ is likely to take between 1 to 2 years.

6.8 In addition to the social care data held by local authorities, the acquisition of other datasets for the ADRC-W is required for a number of the proposed additional outcome indicators. The timescales for acquisition will vary depending on the data source but if datasets are held by UK government departments and are required for the proposed ADRN project, they will be acquired by the ADS.

6.9 Around 8 per cent of the City and Council of Swansea cases could not be linked and would therefore be excluded from any reporting of outcome indicators. The loss of these missing cases is likely to introduce some bias into the estimates; however, further investigation would be required to look for ways to improve the linking rate, confirm whether significant bias exists and/or identify solutions.

6.10 **Recommendation:** that an ADRN project is proposed and funded. The project would be designed to both acquire and analyse the all-Wales local authority social care data and any other datasets required to report on the proposed additional outcome indicators. The project would also complete the further development work required before the indicators could be reported.

6.11 **Recommendation:** because the project would need to specify precisely which data is required, the project should work closely with the ADS to ensure that the data acquired and the metadata provided was fit for purpose, completing any data cleaning, validation and initial analysis to ensure the data was ‘research ready’.

6.12 **Recommendation:** the proposed ADRN project could identify ways to increase the linking rate for future data acquisition, not only for local authority social care data but for any additional datasets required for the reporting of additional outcome indicators. This would improve the accuracy of the resulting estimates or, where some small proportion of cases remained unlinked, develop appropriate caveats to include in the reporting.

6.13 There are also other research projects ongoing such as the Social Care and Health Of Older People (SCHOOP) project that include some acquisition of social care or other relevant data for SAIL/the ADRC-W.

6.14 **Recommendation:** the proposed ADRN project should include a component of liaison with all relevant research projects to ensure that the data acquired is suitable to allow reporting of the proposed indicators and to gain any added value from those projects e.g. additional data acquisition to develop additional indicators.

6.15 The Project identified some ways in which improved data collection could help either to provide more robust versions of the proposed additional outcome indicators or, in some cases, would allow further outcome indicators to be proposed.

6.16 **Recommendation:** the proposed ADRN project should include a component of working with appropriate projects and organisations (where appropriate with or through the Welsh Government Knowledge and Analytical Services Data Collection function) to attempt to achieve improvements to data collection. This would allow more robust versions of the proposed additional outcome indicators to be provided/further outcome indicators to be proposed.

6.17 The Project has identified a number of indicators based on data currently available for linking or that is known to be available for linking. New datasets are being created or coming to light all the time e.g. as new policies or services are delivered.

6.18 **Recommendation:** if a follow-on project is funded, further work can be done, focused particularly in areas where relatively few outcome indicators have currently been identified, to seek out further data sources in order to propose additional outcome indicators.

Further Development Work in Addition to the Acquisition of Data

6.19 Most of the proposed additional outcome indicators and user groups would require some further development work to refine definitions and validate estimates. Both the Welsh Government-ESRC Data Linking Demonstration Projects and this Project have shown that it can be very time-consuming to develop definitions using administrative datasets that are not designed for research purposes.

6.20 Further challenges arise from the fact that the administrative sources used for linking are by their nature longitudinal, so that issues of change over time need to be taken into account, e.g. the duration of a health condition or state of well-being. This increases the complexity of analysis and reporting for and additional outcome indicators that use linked data.

6.21 In some cases, projects are ongoing that include some attempt to define the concepts required to develop the proposed indicators. With the development of the UK FARR Institute and the UK ADRN as well as the work of the growing International Data Linkage Network (currently chaired by the Director of the ADRC-W), more work than ever before is being done using linked data. Both the FARR Institute and the ADRN have research strands on methodology that may deliver findings that would help in the development of the proposed additional outcome indicators or to identify further outcome indicators.

6.22 **Recommendation:** where further development work cannot be avoided by adopting pre-existing definitions validated by other projects, that work must be completed before the proposed additional outcome indicators can be reported. This additional development work could be completed as part of a proposed ADRN project.

What is the alternative to funding an ADRN project?

6.23 Two alternatives have been identified to funding an ADRN project.

6.24 The first option would be to wait until other projects such as the Social Care and Health Of Older People (SCHOOP) project have acquired additional datasets for SAIL and, at a later stage, to revisit the viability of SAIL to provide additional outcome indicators. There are limitations to this option:

- most research projects are relatively specific, either to geographies or user groups e.g. older people. Relying on other projects to acquire local authority social care data would therefore risk the data not being suitable to evidence any of the proposed additional outcome indicators;
- further development work would be required before most of the proposed additional outcome indicators could be reported. Even if other projects develop new definitions, these would be unlikely to be exactly what is needed for the proposed indicators;
- there would be no opportunity to feed into the improved data collection required to create/improve some indicators; and
- there would be no opportunity to identify and acquire additional datasets to allow additional indicators to be proposed in the future.

6.25 The second option would be to wait for all Local Authorities in Wales to migrate to the Welsh Community Care Information System (WCCIS) and to acquire a standardised social care data set for Wales for SAIL for analysis purposes. There are limitations to this option:

- ➢ the migration process will take a number of years to complete. The data will not be in a standardised format at the point of migration because, although all Local Authorities collect the same data (for example the minimum data set identified in the Integrated

Assessment for Older People), the way those items are coded, named etc. will be different. The data capture will therefore be based on 22 versions of the data reflecting the current practice in 22 different Local Authorities. Standardisation of data is not likely to be achieved until WCCIS is adopted by all 22 Local Authorities. Even after the data is migrated, there will be a lengthy and iterative process of standardising critical and common data items. This would delay any further work;

- there would be no opportunity to feed into the improved data collection required to create/improve indicators; and
- it is not yet clear how much historical data will be migrated as part of the process. Any new dataset relying on WCCIS to provide data may therefore be limited, at first, as to how many cases would provide a sufficiently long timeline of information to allow change over time to be reported for the relevant indicators.

Appendix 1 Summary of potential additional outcome indicators

Well-being: I know and understand what care, support and opportunities are available to me; I get the help I need, when I need it, in the way I want it
Potential Outcome Indicator: Percentage of GP patients given one of 20 leaflets by their GP practice (self help, diet, asthma, screening, contraception etc.) and being given various kinds of verbal advice.
Notes: Physical or mental health can be compared before and after receipt of the leaflet and/or advice.[25] It is likely that the indicator could be reported separately for ‘younger adults’, ‘older adults’, individuals with physical disabilities and individuals with mental health difficulties. This indicator uses GP Event data that is already available in SAIL. The indicator can therefore be reported without further data acquisition but further work would be required to refine the relevant user group definitions; this work may take around six months to complete.

Physical and mental health and emotional well-being: I am healthy; I am happy
Potential Outcome Indicator: Objective, physician-reported status of physical and mental health.
Notes: This indicator could be reported separately for most user groups and for social care users versus the general population. Further development work and further data acquisition would be required to report on these indicators; this work may take between 1 to 2 years to complete.
Potential Outcome Indicator: Average self-reported ‘happiness’ score[26] among social care users (compared with the general population).
Notes: Further data acquisition and development work would be required to report on this indicator; this work may take between 1 to 2 years to complete.
Potential Outcome Indicator: Percentage of children **in receipt of** Flying Start reaching or exceeding developmental milestones at three years.
Notes: Further discussion within Welsh Government Knowledge and Analytical Services will be required to agree a method for flagging receipt of Flying Start services. Further data collection and acquisition would therefore be required to report on this indicator. It is unlikely this indicator could be reported in less than two years.

Protection from abuse and neglect: I am safe and protected from abuse and neglect
Potential Outcome Indicator: Percentage of people who have experienced domestic violence receiving: a) social care services; and b) supported housing services e.g. *Supporting People* and *Extracare.*
Notes: Further data acquisition and development work would be required to report on these indicators. Two approaches are possible, the first of which may take between 1 to 2 years to

---

[25] Data is currently available in SAIL for around 60% of the population of Wales (as of August 2014) – a NISCHR-funded project has the target to increase this to over 70% by March 2015.

[26] Social care questions will be asked in the last National Survey year (2014-15) – in addition to what is proposed here, it will therefore be possible to report a self-reported happiness score for self-reported social care users just using the survey data. In order to be consistent, where average score will be reported for the survey data, average score would also be used for the linked data.

complete and the second (using additional data from police and third sector) that, although more robust, will require additional development work so may take another year to complete.

Education, training and recreation: I can learn and develop to my full potential; I can do the things that matter to me

Potential Outcome Indicator: Percentage of Flying Start (FS) **eligible** children meeting the expectation that they would attain Level 2 in each subject at age 7/Key Stage 1.

Notes: The required data to report on this indicator is already available; this indicator could be reported on with very little further development work and certainly in less than six months.

Potential Outcome Indicator: Percentage of children **in receipt of** Flying Start meeting the expectation that they would attain Level 2 in each subject at age 7/Key Stage 1.

Notes: Further discussion within Welsh Government Knowledge and Analytical Services will be required to agree a method for flagging receipt of Flying Start services. Whether data linking will be required will depend on the method chosen. Further data collection and acquisition may therefore be required to report on this indicator. It is unlikely this indicator could be reported in less than two years.

Potential Outcome Indicator: Percentage of young people in receipt of Education Maintenance Allowance Wales (EMA) or Assembly Learning Grant (Further Education) (ALG FE).

Notes: This indicator can be reported without using data linking.

Potential Outcome Indicator: Percentage of adults with a record in the Lifelong Learning Wales Record (LLWR).

Notes: This headline indicator can be reported without using data linking. However, this indicator could also be reported separately for some adult user groups and for social care users versus the general population. For some user groups e.g. ‘older adults’, the indicator could be reported without further data acquisition. For other user groups e.g. carers, and for social care users, further data acquisition and development work would be required; this work may take between 1 to 2 years to complete.

Domestic, family and personal relationships: I belong; I have safe and healthy relationships

Potential Outcome Indicator: Percentage of people who:

i) report feeling that they belong in their neighbourhood: ii) think that their local area is a place where people from different backgrounds get on well together; or iii) think that people in their local area treat each other with respect and consideration; **and** are also in receipt of: a) social care services: and b) supported housing services e.g. *Supporting People* and *Extracare*.

Notes: Further data acquisition and development work would be required to report on these indicators; this work may take between 1 to 2 years to complete.

Potential Outcome Indicator: Percentage of people who report being satisfied with: i) ‘personal relationships’; or ii) ‘the well-being of (their) children’ and that are also in receipt of: a) social care services and/or b) supported housing services e.g. *Supporting People* and

Extracare.
Notes: Further data acquisition and development work would be required to report on these indicators; this work may take between 1 to 2 years to complete.

Contribution made to society: I can engage and participate; I feel valued in society
Potential Outcome Indicator: Percentage of adults: i) participating in a range of sports and physical recreation activities; ii) volunteering in sport; iii) with sports club membership; iv) participating in competitive sport; and v) using leisure and fitness centres.
Notes: This headline indicator can be reported without using data linking. However, it is likely that this indicator could be reported separately for most adult user groups and for social care users versus the general population. For some user groups e.g. 'older adults', the indicator could be reported without further data acquisition. For other user groups e.g. individuals with mental health difficulties, and for social care users, further data acquisition and development work would be required; this work may take between 1 to 2 years to complete.

Securing rights and entitlements: My rights are respected; I have voice and control; I am involved in making decisions that affect my life; My individual circumstances are considered; I can speak for myself or have someone who can do it for me.
No new indicators are proposed against this theme at this time but if a follow-on project is funded, further work could be done to identify sources of data for additional indicators.

Social and economic well-being: I am supported to work; I have a social life and can be with the people that I choose; I do not live in poverty; I get the help I need to grow up and be independent; I get care through the Welsh language if I need it.
Potential Outcome Indicator: Percentage of social care users in receipt of benefits.
Notes: Further data acquisition and development work would be required to report on these indicators; this work may take between 1 to 2 years to complete.
Potential Outcome Indicator: Percentage of ex-Children in Need receiving: a) Adult Social Care services; and b) supported housing services e.g. *Supporting People* and *Extracare*.
Notes: Further data acquisition and development work would be required to report on these indicators; this work may take between 1 to 2 years to complete.

Suitability of living accommodation: I have suitable living accommodation that meets my needs
Potential Outcome Indicator: Percentage of private residences meeting the Local Government and Housing Act 1985 overcrowding definition.
Notes: Further data acquisition and development work would be required to report on these indicators. Two phases of data acquisition would be required, the first of which may take between 1 to 2 years to complete and the second (using additional data from the Valuation Office Agency) that, although it may result in more robust estimates, may require primary UK legislation, which may take two years or more.

Appendix 2 Making administrative data available for research in SAIL

Background to SAIL

The NHS in Wales generates a large amount of data collected on a routine basis for the purposes of treating patients and making sure the health service runs properly. Almost none of it is collected for research or statistical purposes. Nevertheless, the data has considerable potential to be used for research. The same is true of other public services, such as education, housing and social care. Data collected by the NHS in Wales is held complete with patient identifiers: NHS number, full name, address, postcode, date of birth and gender. Information Governance law therefore prohibits the use of data in this form for research outside the NHS.

To address this problem, the Welsh Government funded the creation and development of the Health Information Research Unit (HIRU) at Swansea University, from 2006. The aim of this unit was to develop a means by which routinely collected health data from many different sources could be utilised in a linked way, but in such a way that each dataset individually as well as the linked data was rendered completely anonymous. The process developed was called Secure Anonymised Information Linkage (SAIL); it led to the creation of a very large collection of anonymised health and other datasets stored in Swansea University and made available for research. Each anonymised dataset is stored separately and as far as possible in the form in which it is received from the data provider. The name of ‘SAIL’ has become synonymous with both the large data repository and the HIRU. ‘SAIL’ will be used to refer to the databank at Swansea University throughout this Report.

Linkage and the process of creating an Anonymised Linking Field

SAIL use an internationally recognised procedure for safeguarding respondent privacy during data linking. This involves a “split file” process - the separation of all identifying information i.e. name, address, date of birth, NHS number etc. from all other analytical data, whether medical, social, financial, attitudinal etc. in each source to be linked. For each source, this creates two files, the first containing a reference number or ‘index’ plus the identifiable information and the second containing the index plus the analytical data. For each dataset, the identifying information is sent to a ‘trusted third party’ that creates an anonymous linking field. Once the linking field has been created the identifying information is destroyed leaving only the linking field and the index. The index allows the anonymous linking field to be reattached to the analytical data. The analytical data can then be linked to other anonymised datasets without using any identifying variables.

For SAIL, the NHS Wales Information Service (NWIS) act as the ‘trusted third party’ organisation, routinely turning lists of personalised information into anonymised indexes. NWIS utilise the Welsh Demographic Service data as the ’population spine’ or ‘template’ for its anonymisation process. The Welsh Demographic Service is a database of everyone registered with a GP in Wales from 1994 to the present day. It includes an anonymised residential address history – this is an index of unique numbers, one for each residence in Wales, known as the Residential Anonymised Linking Field (RALF). Individual people who have been registered with a GP in Wales, past and present, are represented in the Welsh Demographic Service data as another index of unique numbers, known as the Anonymised Linking Field (ALF). In this way, it is possible to associate RALFs with ALFs, that is: homes with residents.

How data is made available to research analysts

A database “view” is a structured ‘image’ of information stored in the database, including only a sub-set of the complete dataset. A “view” can include data from more than one database, and can be restricted to include specific rows and columns. In this way, the database administrators can very closely control the data with which each researcher is allowed to work. There is, in addition, no way that a researcher can alter the underlying data table providing the “view”.

The database “views” are made available through a secure remote access system, the SAIL Gateway, which can be accessed securely over the internet, using a system where authorised researchers are able to log on to a dedicated computer through a password protected browser. Outputs are ‘locked down’, so that nothing can be copied and pasted out of the gateway, saved to a port or drive on the remote computer, or sent to a printer.

All analysts who are provided with a SAIL gateway account are given access only after both they and their line manager have signed a detailed agreement outlining the researcher’s responsibilities and the agreed usage that can be made of that account. The agreement clearly places the responsibility with the researcher to ensure that no individual could be potentially identifiable from the research outputs. However, in addition, all potential outputs are scrutinised by a SAIL administrator to ensure potentially disclosive information does not leave the secure gateway.

The researcher is required to carry out the analysis within the gateway, in which suitable database, statistical, spreadsheet, word processing, mapping and presentation software are available. The only outputs allowed are summarised or aggregate results. Proposed outputs are processed through the ‘request data out’ link within the gateway. This is the stage at which outputs are scrutinised by a senior research analyst in SAIL, checking for potential disclosure issues such as small numbers. The ‘data out’ process does not check that the analysis has been performed correctly and that results are correct, it merely scrutinises outputs for potentially disclosive situations.

It is not possible to put a process in place that would stop researchers taking photographs of a computer screen, for example, or simply writing down results and not following the ‘data out’ procedure. Given this, the researcher must be trusted to adhere to the terms of the SAIL access agreement. However, when signing the access agreement, researchers and their line managers are agreeing to abide by the statement of procedures in the National Statistics Code of Practice: Protocol on Data Access and Confidentiality, in both letter and spirit, to the maximum extent that they apply. Breaches of these rules would result in penalties and legal action. As part of the creation of the UK Administrative Data Research Network, it is anticipated that researchers accessing databases like SAIL will be required to successfully apply to be part of an official register. Abuse of the privileges of data access would then result in removal from such a register, effectively ending the perpetrator’s research career.

Information Governance Review Panel

All research carried out within SAIL is still managed through a rigorous control structure to ensure that confidentiality is maintained and potentially disclosive outputs are not produced.

One of the controls in place is a requirement for all proposals involving the analysis of linked data within SAIL to obtain approval from the Information Governance Review Panel (IGRP). The IGRP is a panel of independent specialists in informatics governance and lay people that oversees all research taking place within SAIL. Current membership (June 2013) is listed in Appendix 1. An IGRP application contains an outline of the research rationale for creating the link(s), details of any new datasets that would be accessed, and precisely what variables would be required from the linked datasets. Researchers must indicate in the application that they have considered the handling of sensitive data in the research design. Although the datasets are all held completely anonymously in SAIL, the selection of a very specific sub-group based on age and gender at small area (LSOA) level, looking at a specific condition could return small numbers. Small numbers in a published output could be put together with other local knowledge to establish who the statistic refers to. Researchers are given access to the data at the most detailed level required to complete their analysis, but need to ensure that nothing potentially identifiable is revealed in their reporting. IGRP applications must indicate how the analyst proposes to deal with small numbers (e.g. through grouping and aggregation of cases).

Appendix 3 Local Authority Social Services Data Systems in Wales

Due to the commissioning of a National Institute for Social Care and Health Research project to document social care data in Wales, it was agreed that this Project would identify, as far as could be achieved, which data collection systems were in use by each local authority in Wales. It was agreed that this information would feed into the National Institute for Social Care and Health Research project.

In order to establish the size and nature of the task of building a national database for Wales, it is essential to learn if the social care data held by the City and Council of Swansea is typical of all local authorities in Wales.

Several different data collection or management systems are in use in Wales. Table 1 (below) summarises what is known about the situation at point of writing.

The PARIS system is provided by a company called CIVICA.[27] This company provides systems to a large number of Local Authorities, particularly for mental health. However, the system is provided on a tailored basis and may not be being used consistently in these local authorities. CIVICA also provide a system to an NHS organisation in Wales: Cardiff and the Vale NHS Trust.

Another system called RAISE, provided by a company called CareWorks[28] is used in eight local authorities in Wales. The system was a joint purchase, so this system may be used in a more standardised way. The joint purchase was achieved through the setting up of a Wales System Consortium (WSC) in 2002. The eight authorities came together with the Welsh Assembly Government (now Welsh Government) to develop a project designed to bring their social services IT systems into the 21st century. The Welsh Assembly Government was keen to promote a co-ordinated and unified approach to care management and assessment. Joining up information across different teams within the local authority was a key objective offering the prospect of cost savings and more informed decisions through data-sharing. The objective was to create a fully integrated Adults' and Children's Social Care IT system to improve the delivery of social services. The eight authorities that form the Wales System Consortium are Blaenau Gwent, Bridgend, Ceredigion, Gwynedd, Isle of Anglesey, Powys, Torfaen and Wrexham.

Several local authorities appear to have purchased both PARIS and RAISE systems; it is not clear the extent to which they use both.

Social care data for Gwynedd local authority - one of the RAISE users – is about to be anonymised into SAIL for a pilot project (see Appendix 4). This will start to provide comparisons between RAISE and PARIS data collections.

Each local authority is required to submit returns to the Audit Commission as part of the National Fraud Initiative.[29] This includes data matching at the level of the individual, and implies that some robust client recording system is in place in every local authority. The required data for matching include information about housing, housing benefits, council tax, privately-supported care home residents, transport passes and permits, blue badges and concessionary travel. Local Authorities also submit annual individual level data to the Welsh Government for the Children in Need Census. This indicates that some standard flows of information relating to the various potential user groups proposed in Chapter 4 must already be happening.

---

[27] https://www.civica.co.uk/

[28] http://www.careworks.com/

[29] Vale of Glamorgan explanation of its duties to the Audit Commission http://www.valeofglamorgan.gov.uk/en/our_council/achieving_our_vision/national_fraud_initiative.aspx

Table 1 Social Care client recording information systems in use in Local Authorities in Wales

Local Authority	Client Information system
Blaenau Gwent	RAISE and PARIS
Bridgend	RAISE; PARIS is used for Mental Health and Substance Misuse
Caerphilly	SWIFT
Cardiff	OLM (Care Notes) PARIS used for Mental Health
Carmarthenshire	OLM (Care Notes)
Ceredigion	RAISE and PARIS
Conwy	PARIS
Denbighshire	PARIS
Flintshire	PARIS
Gwynedd	RAISE
Isle of Anglesey	RAISE
Merthyr Tydfil	SWIFT
Monmouthshire	SWIFT in the process of developing a bespoke In-house system
Neath Port Talbot	PARIS is used for Mental Health and Substance Misuse
Newport	SWIFT
Pembrokeshire	OLM (Care Notes)
Powys	RAISE and PARIS
Rhondda Cynon Taf	SWIFT
Swansea	PARIS
Torfaen	RAISE
Vale of Glamorgan	SWIFT (PARIS used for Mental Health)
Wrexham	RAISE

Appendix 4 Related Data Linking Research and Pilot Projects

National Institute for Social Care and Health Research-funded research projects

The National Institute for Social Care and Health Research, Welsh Government, launched the Academic Health Science Collaboration in October 2010. The Academic Health Science Collaboration facilitates the collaboration between Health Boards, Trusts and Universities in Wales in order to combine clinical research; both basic and translational research; clinical care and education to create world-leading improvements in healthcare. One of the specific objectives of the Academic Health Science Collaboration is to work with others to develop capacity in the NHS to support high quality research. The Academic Health Science Collaboration Delivery Plan 2012/13 to 2014/15 describes plans for delivering the Academic Health Science Collaboration Strategy. In 2012/13, Academic Health Science Collaboration founded an agreement from the National Institute for Social Care and Health Research, Welsh Government to run a project in collaboration with Swansea University's Health Information Research Unit (HIRU) and NHS information technology departments to enhance the ability of HIRU's Secure Anonymised Information Linkage (SAIL) databank to capture and provide comprehensive and accurate NHS clinical data to be used in future research. This project would also map all existing NHS datasets/registries in Wales, develop stronger systems to manage them and ensure the standardisation of datasets. The project details are given below.

Project 1: Mapping health and social care datasets to link to SAIL

This project was designed to be delivered in two phases. Phase one was required to produce a high level mapping of datasets and define their owner, location and purpose. This report will be utilised as an identification system only. Phase two involved a detailed mapping report to include technological specifications. The detailed report will provide information on the technical aspects of the data such as format of the data, users, protection level (including DPA compliance), any verification/validation currently undertaken and a summary of the robustness of the dataset and its suitability as a candidate for integration.

Project 2: Facilitating the collection of accurate metadata to be linked with SAIL

A key output was designed to be a document detailing standards for metadata and characterisation of datasets. The deliverable would document all standards for the metadata and would provide a full characterisation of the datasets.

There was considerable cross-over between this Project and the projects that National Institute for Social Care and Health Research commissioned around improving resources for linked social care data. It was important that all the projects communicated and worked together.

Pilot Social Care research projects that involve data linkage of local authority social services data in Wales

Two pilot projects involving SAIL at Swansea University were both funded by the National Institute for Social Care and Health Research starting in October 2013. One project was designed to use data from the City and County of Swansea, and the other to use data from Gwynedd County Council

Project 1: Social care and health of older people (SCHOOP) - Swansea

The relationship between use of social care and use of health care by people aged 65+ in Wales before and after the implementation of the Chronic Conditions Management policy - an observational study using routine data

Project summary: There is a link between use of social care and use of health care by people aged 65 and over. Appropriate social care can maintain health and delay or avoid the need for health care such as hospitalisation. Conversely, a spell in hospital can mark the beginning of a need for social care.

The Welsh Government's Chronic Conditions Management policy (2008) sets out a plan for improving care in the community and reducing demand for hospital care. The study will examine the impact of the policy across Wales on emergency hospitalisations of people aged 65 and over and see how this varies with varying social care input.

The City and Council of Swansea Social Services data on individuals was anonymised into SAIL for the detailed analysis.

Project 2: Data linkage in social care: pilot project - Gwynedd

Project summary

Existing data linkage projects in Wales that utilise routinely collected administrative data have focused on health datasets only. To build the complete picture of service provision there is a need to broaden this to include social service provision by Local Authorities, and provision of support by third sector organisations. The tried and tested Secure Anonymised Information Linkage (SAIL) system allows large routine datasets to be presented for research with links at the level of the individual, but with all individuals remaining anonymous.

The work proposed here is a pilot project to test the feasibility of linking datasets from a local authority, the NHS and third sector organisations.

The focus of this work will be on individual-level data from adults who are referred to social services in order to avoid admission to hospital or to facilitate their discharge from hospital. The data linkage will include data from statutory and third sector organisations and services which provide interventions and support in community settings.

Part of the feasibility study will be to assess the range and quality of the data available in each of the relevant organisations and to better understand the utility of data linkage across statutory and third sector organisations, in particular, to inform future service organisation and delivery.

Appendix 5 Details of the City and County of Swansea data held in SAIL

During the development of SAIL, Social Services data from the City and County of Swansea (referred to as the 'Swansea data' throughout the rest of this Appendix) was amongst the data that was put through the anonymisation process as part of the 2007 Proof of Concept. As a result, it has been demonstrated that social care records can be linked to health records at an individual level, to produce a more holistic picture of the overall care received and individuals' interactions with services.

A large number of service users (19,275) were anonymised in 2007, but the data also contains information about other people associated in some way with these service users. This included carers, neighbours and other service agency support workers. No attempt was made to link these 'associated people', resulting in a dataset where the presence of a care network can be associated with some clients, but no linked information can be determined about the carers and other associated people. An opportunity may have been missed to establish a linked cohort of carers. Carers are one of the potential user groups with care needs; therefore, in order to develop a full understanding of carer support, and to support the Well-being Statement analyses, the anonymisation of carers would be essential. The greater the number of individuals whose data can be linked across datasets, the better the coverage for the Well-being Statement analyses. If further work is funded, a priority would be to establish the potential of 'associated people' to provide information about informal and formal care networks, and a defined group of carers for further research. This work would entail the practical detail of ascertaining if sufficient information is recorded for these individuals to be anonymously linked, and to distinguish between professional support workers and family, friends and neighbours,

The core datasets to which both the 2007 Swansea social care data and any social care data acquired for SAIL in the future can be linked at an individual level are described in Appendix 1, with a brief description of the relevance to this project

Table A2.1 below summarises the type of information that is recorded about the individuals in the Social Services data. The 'number of variables' field indicates how many different fields (or variables) exist in each table. The remaining columns indicate the numbers of records and clients there are for each table and the time periods to which they refer. This provides a useful guide when contemplating possible analyses, e.g. for 'Employment', only 22 records exist with 12 variables in each, indicating that this would not provide enough cases for any meaningful analyses to be carried out.

The Characteristics of the 2007 Swansea Social Care client base

In order to establish the geographical distribution of Swansea service provision, the anonymised data about social care clients, associated persons and staff were mapped to Lower Super Output Areas[30] in Wales. While 93% of clients (17,953 out of 19,275) lived in the Swansea area, around 7% lived in the rest of Wales (see Figures A.4.2 and A4.3). As only the clients were anonymised in the data capture process, it is not possible to determine any of the characteristics of those people associated with the clients.

---

[30] Lower Super output areas are Census defined geographical areas containing on average 1600 people, the lowest level of geographical identifier available in the anonymised SAIL data.

Geographical distribution of clients

Figure A4.1. Numbers of Social Care clients in the Swansea data shown by LSOA

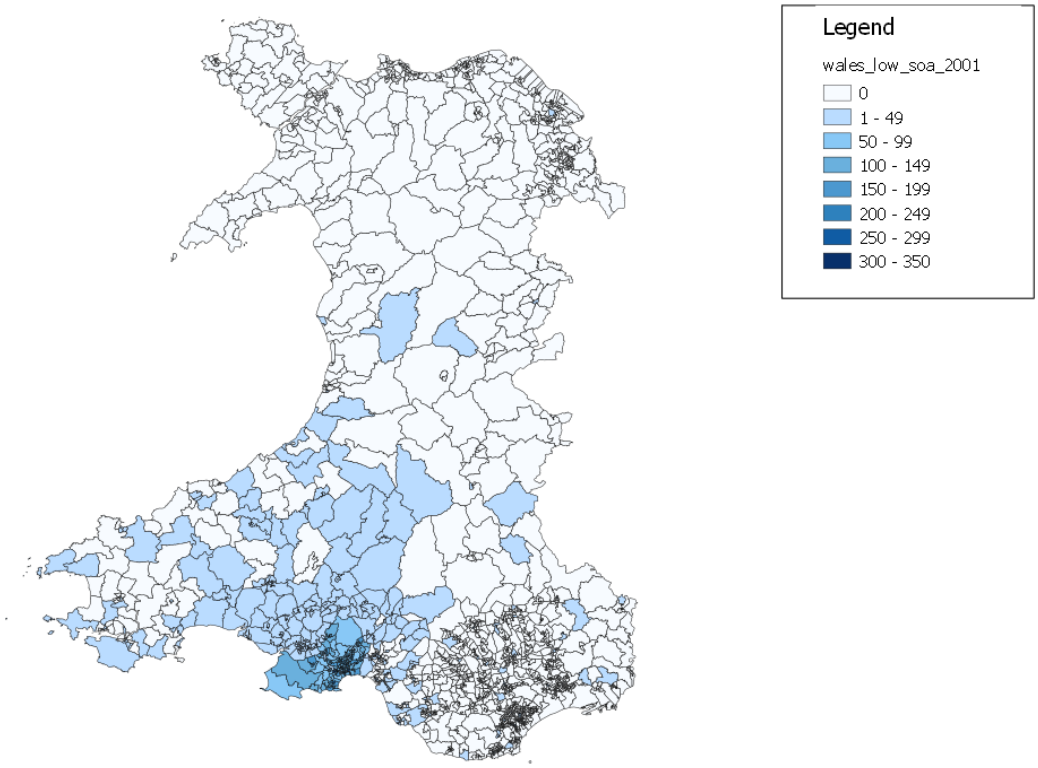

Figure A4.2 Numbers of Social Care clients in the Swansea data in the Swansea area

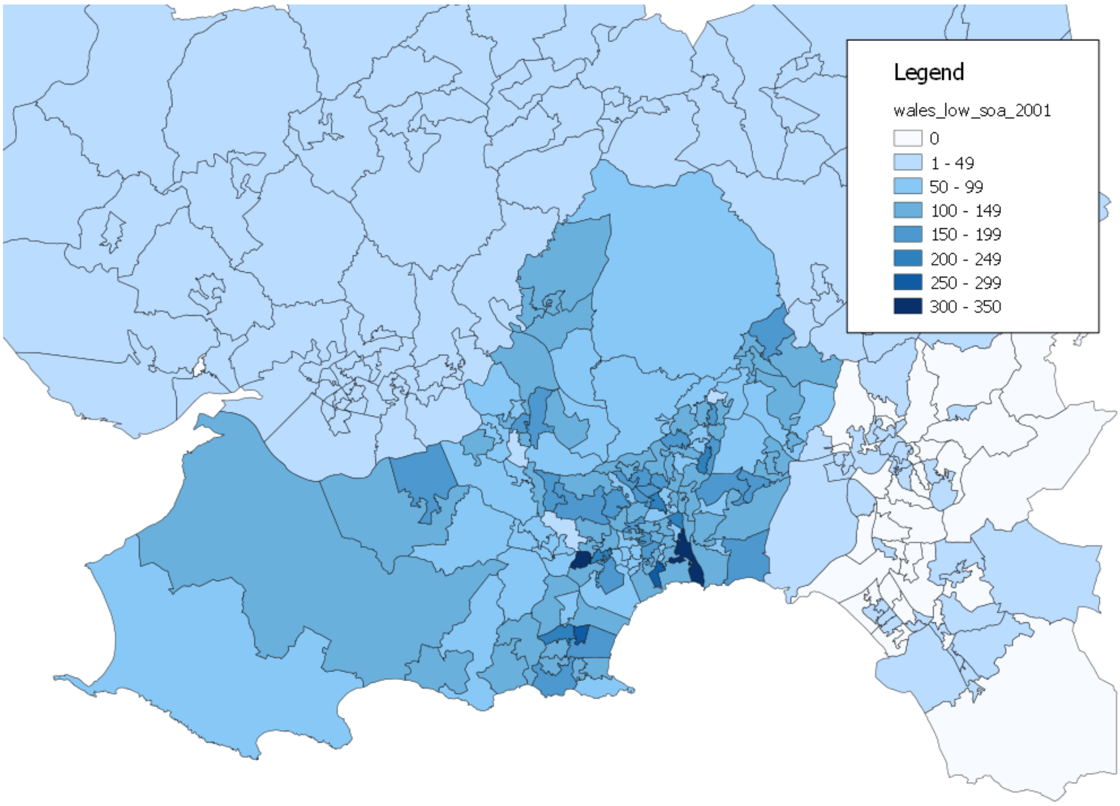

Table A4.1 Summary of Swansea data tables

Table name	Description	Number of variables	Time period covered (total)	Number of records (total)	Number of linkable distinct individuals
Alerts	Alerts for social care staff and clients	4	2003–2007	942	757
Allergies	Any allergies suffered by clients	8	NA	1,963	1,751
Associated people	Other individuals who have a key role in the client's life, e.g. next of kin, carers, etc.	8	NA	15,783	9,410
Care Timetable	A record of the duration of care at each intervention	5	NA	5,447	1,937
Client group	The group the client best fits into	7	1977-2007	26,384	14,731
Client Status	Whether the individual is an associated person, a current client, or a closed case	6	1940-2007	61,299	14,731
Contact Details	Contact details for the client	7	1996–2008	16,028	13,571
Employment	Client's employment details	12	1990–2007	22	20
External agencies	Details of any other agencies involved in the client's care	6	1997–2008	3,944	2,835
Home status	The client's housing situation	5	NA[31]	11,766	8,261
Hospital Action	Whether the client is subject to a Hospital Action	6	1961–2007	794	458
Independence Risk	Risk of (or to?) independence	8	NA	8,261	2,357
Involved Staff	Social Care staff involved in the client's care	10	1997–2015	48,807	12,396
Language	The client's language and communication needs	9	NA	2,107	1,793
Legal Information	Any legal issues relating to the client, e.g. Mental Health Act or refugee status	6	1973–2008	1,835	860
Lifeline	Details of Lifeline alarms	7	1988–2007	6,384	3,603
Lifeline ALF	The linkage table between LIFELINE and other SAIL tables	5	1900–present (Week of birth)	6,384	3,603
Medical details	Details of any medication the client is taking	9	1904–2008	4,641	1,069
PARIS	General demographic details. Contains info also held in other tables.	38	1940–2007 (Week of birth)	14,307	12,892
PARIS ALF	The linkage table between PARIS and other SAIL tables	9	NA	19,725	17,659
Social Services Provision	Type of Social Services support provided to the client	5	1968-present	4,602	2,613
Unified Assessment	Details of any Unified Assessments conducted for the client	128	2003 – 2007	4,104	2,200

---

[31] The data is held in text format only, so it was decided that it would be an unacceptably onerous task to calculate this.

Appendix 6 Summary of Survey Datasets Available in SAIL

The National Survey for Wales

6.26 The National Survey for Wales[32] is a study of adults living in private households across Wales. Each year 14,500 people aged 16 years and over are selected at random to take part in the survey, which is carried out by face-to-face interview. From January 2012, participants have been asked for their consent to link[33] their survey answers to other records held about them and those agreeing to do so have had their survey data anonymised into SAIL.

6.27 Data for National Survey for Wales respondents giving ‘consent to link’ have so far been imported into SAIL for the period of January 2012 to March 2013. The data is held in two separate parts because some questions changed from April 2012. ‘Consent to link’ continues to be asked for the National Survey for Wales, so that data for April 2013 to March 2014 will become available for anonymisation and linking in autumn of 2014. Around 9,000 cases are expected to be added each year.

6.28 The main advantage to using survey data is the richness of the information available about each survey respondent and their household. The National Survey for Wales collects information about environmental factors such as house type (e.g. terraced, detached) and house condition (e.g. in poor condition relative to the neighbourhood) and housing tenure (e.g. owner occupied) as well as the respondent’s socio-economic status and satisfaction with health and educational services. So, although the numbers will tend to be small when compared with those found in administrative datasets, when survey data is linked to administrative data, together they can be used to provide analysis that would otherwise not be possible e.g. whether service use varies by socio-economic status. Conversely, by linking administrative data to the National Survey for Wales data, analysis can potentially be done on topics that were not covered in the survey e.g. the neighbourhood environment, health service use, more specific health conditions. Once the all-Wales social care data is available, analysis could also be done on social care service use and the ability to maintain independent living.

6.29 As noted above, the main disadvantage of using survey data is the relatively small numbers compared with administrative data. Also, as National Survey for Wales is a cross-sectional survey, it cannot be used to monitor how the subjective well-being of individuals changes over time. By comparing the average level of self-reported well-being for the user groups across years, changes can be examined at the population level but it would not be possible to speculate about cause and effect. To examine changes in subjective well-being for individuals over time, only the longitudinal survey Understanding Society could be used (see below).

Understanding Society

6.30 Understanding Society is a study of 40,000 households across the UK, including around 2,800 households containing around 4,000 individuals in Wales. The survey recruited then attempts to follow the same representative sample of individuals by interviewing them every year. Over time, the sample changes since there are rules about including new family members and following individuals into new households as well as individuals who

---

[32] http://gov.wales/statistics-and-research/nationals-survey/?lang=en

[33] Respondents are shown a card which explains data linkage and are asked “Do you agree for us to link your survey answers with other information held about you?”

are lost to follow-up. The Understanding Society survey began collecting information in 2008 but from 2010 it incorporated the remaining sample (much depleted through attrition) for the BHPS or British Household Panel Study which began in 1991. This means that information about the full 4,000 respondents is only available for 2010 onwards but that information about a smaller sample is available back to 1991. The relatively small survey sample size for Wales does limit the potential for data linkage and analysis for this data, but the overall study size can allow UK-level analysis where topics are not as amenable to cross-country variation and/or the policy area is not devolved e.g. receipt of benefits, pensions.

6.31 *Understanding Society* is a multi-purpose and multi-topic survey providing a balance of coverage meeting a wide range of needs from income and employment to childcare. Because the survey is longitudinal, it has a particular interest in change so tends to ask questions about the same topic over time, leaving questions about issues that change slowly e.g. education, longer than questions about issues that change quickly e.g. health and lifestyle. Of particular relevance to ILDSC, questions are included about health and disability, income, welfare and poverty and life satisfaction.

Active Adults 2012 and 2014

6.32 The Sports Wales Active Adults Survey is the most comprehensive source of information about the amount and types of sport and physical activity that adults in Wales do. The survey aims to measure levels of sports participation, club membership, volunteering and other sporting matters among adults in Wales.

6.33 The Active Adults Survey 2012 collected around 22,000 responses from households randomly selected across Wales, with more than half of respondents giving their ‘consent to link’.

6.34 Questions are asked about participation levels in a range of sports and physical recreation activities, about sports club membership and participation, volunteering in sport, participation in competitive sport, usage of leisure and fitness centres and barriers and motivation for participation in sport. For ILDSC, the survey will be a source of evidence in particular for the ‘physical and mental health and emotional well-being’ and ‘contribution to society’ well-being outcomes. Currently, the 2012 data are not available for linking in SAIL but are expected to be acquired within the next reporting period.

The Cognitive Function and Ageing Study, Wales

6.35 The Cognitive Function and Ageing Studies are large UK-based longitudinal multi-centre studies looking at health and cognitive function in older people. ‘Consent to link’ was introduced in Wave 2 of the study which started in September 2013 and will continue until January 2016. The study will interview a representative sample of 5,000 people aged 65 and over in two areas in Wales (Gwynedd and Swansea). Around 88% of respondents already interviewed have given their consent to link.

6.36 Using established and standardised techniques, the study will collect data that will enable the investigation of cognitive impairment, depression, physical disability and healthy active life expectancy for the whole group and within social groups. It will investigate factors that may delay the onset of dementia, specifically focusing on the role of bilingualism and social networks. As the participants reside in a bilingual area, this is a key opportunity.

6.37 CFAS Wales will examine activity and participation, exploring what factors people with and without cognitive impairment believe affect their inclusion and exclusion from community and civic life. It will examine environmental factors, looking at the impact of cognitive impairment on attachment to place, and comparing variations in cognitive functioning across rural and urban areas, including those that are socially deprived. It will examine generational change and estimate changes in the prevalence of cognitive impairment and dementia in Wales, 15 years after the first CFAS cohort was established. It will examine adaptation and resiliency, identifying factors that might have a detrimental effect on resilience and those related to a more favourable outcome. It will examine the relationship between nutrition (vitamin B12) activity and cognitive impairment. In addition to the standard CFAS data collection, a number of additional qualitative interviews will be carried out, based on life history data, 10% of the sample will be asked to provide a blood sample (for vitamin B12 analysis), and a small sub-sample will be interviewed regarding lifestyle and health behaviours.

Welsh Health Survey

6.38 The Welsh Health Survey covers a range of health-related issues, including health status, lifestyle and health behaviours, and health service use. The survey was established in 2003 and runs all year round. Results are published annually. An achieved sample of around 15,000 adults and 3,000 children is aimed for per year, to include a minimum of 600 adults from each local authority area. In 2011-12, work was carried out to test the possibility of introducing data linking to the Welsh Health Survey. This involved developing and testing a new procedure asking respondents for their consent to link their anonymous survey answers to information held about them in other datasets to create a richer information source. As part of the testing process, an additional 1,000 cases asking consent to link were collected to test the effect on overall response; this generated around 500 cases for linking into SAIL.

6.39 Following this work, procedures seeking respondents’ consent to link were introduced from 1st April 2013. Consent to link is an optional part of the survey and respondents choosing not to give their consent to data linking are reassured that their survey answers are still useful in their own right.

6.40 Whilst only 500 test cases are currently available for analysis, a further 7,000 to 8,000 cases are expected to be added each year, beginning in autumn of 2014. The 500 test cases can be used to assess the kinds of analysis that would be possible once the first full year of data is available for linking. Key among the information useful to this project are questions asked about carers.

Appendix 7 Summary of Administrative Datasets Available in SAIL

The Welsh Demographic Service (WDS)

6.41 The Welsh Demographic Service is a database of administrative information about individuals in Wales that use NHS services, such as address and GP practice registration history. It replaced the NHS Wales Administrative Register (NHSAR) in 2009. This dataset contains the full registration history of the population of Wales since 1990, including house moves and changes of registration to different GP practices. This is the core data that is used in linking datasets together in SAIL. Each person's week of birth is recorded and a date of death when known. All residential address information has been anonymised so that it is not possible to locate any address geographically more precisely than at the Lower Super Output Area (LSOA) level – LSOAs are a patchwork of small areas covering the UK, each one of which contains on average 1,600 people.

GP Event Data

6.42 This is data extracted from all the Welsh General Practices that have signed up to SAIL – at time of writing, this was around 50% of the GP practices in Wales and is increasing all the time. A National Institute for Social Care and Health Research-funded project has the target to increase this to over 70% by March 2015. The data is from the clinical information system the practice uses to maintain an electronic health record for each of their patients - capturing the signs, symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the home environment of patients. The majority of the data is entered by the clinician during the patient consultation, though the data also record interaction with other members of the practice team, repeat prescribing, and some test results that are reported back from secondary care systems. The data cover the period from January 2000 to August 2012 (approximately) but this varies by practices. Currently about 40% of the Welsh population is included in this dataset;

Patient Episode Database for Wales (PEDW)

6.43 PEDW is an all-Wales database containing all finished consultant episodes of in-patient or day case care carried out in Wales, and treatments carried out on Welsh residents elsewhere in the UK. A finished consultant episode is defined as a completed 'unit' of care under the care of one consultant. Each episode has provision for a number of diagnosis and operative procedure codes to be recorded. Capturing data on the daily stream of patients entering and leaving hospitals throughout Wales begins with the collation of information from Hospital Patient Administration Systems. There have been very clearly defined data recording standards in place since around 1999, and all hospital activity on a day case or in-patient basis is regularly submitted into NWIS for inclusion in PEDW. Details of dates of admissions, diagnoses and operations carried out are very consistently coded in this data;

Emergency Department DataSet (EDDS)

6.44 Historically, data about Accident and Emergency visits was recorded in SAIL from the All Wales Injuries and Surveillance System. From 2009, this was superseded by the Emergency Department DataSet (EDDS), which contains administrative and clinical information for all NHS Wales Accident and Emergency department attendances (approximately 750,000 per year), which now includes the (AWISS) data. The data covers the time period 2009 to the present day but recording practices vary between some of the minor and major Accident and Emergency units.

National Community Child Health Database (NCCHD)

6.45 NCCHD brings together data from local child health systems for all Wales and holds information collected throughout pregnancy, childbirth and the early years of life for all births in Wales, details of gestation, birth weight, early life testing, and immunisations. Both the anonymised identifier of the child and the mother appear together in this dataset which is the only true familial link that can be established in the anonymised data within SAIL.

Congenital Anomalies Registration Information System (CARIS)

6.46 CARIS collects information about any foetus or baby who has or is suspected of having, a congenital anomaly and whose mother is normally resident in Wales at time of birth. It includes babies in whom anomalies are diagnosed at any time from conception to the end of the first year of life. Multi source data collection methods are used to gather information from antenatal ultrasound, clinical letters, post-mortems and laboratory results, for all babies in whom an anomaly was detected between conception and the end of their first year of life from 1998 onwards.

Education (National Pupil Database)

6.47 This dataset provides information on attendance and attainment for all pupils in schools in Wales from 2004 to 2013. Research utilising this dataset is restricted to research into the factors affecting educational outcomes. This restriction would need to be carefully considered if further work on informing the framework goes ahead.

Wales Electronic Cohort of Children (WECC)

6.48 WECC is not a single dataset, but the results of a project that combines child data from eight SAIL datasets in the creation of a total population anonymised e-cohort from eight datasets for children born or living in Wales from 1990-2008. This child cohort information is being used to study the relationship between pregnancy, birth, childhood and family exposures and health.

The SAIL Lower Super Output Area (LSOA) Dataset

6.49 SAIL has built a reference dataset where each row is an LSOA and columns include a wide selection of measurements about the LSOA, including WIMD domain scores from all three versions of WIMD versions, Rural Urban Split and other scores like Townsend. Geographical characteristics could be added at a later date, from social services delivery catchment to number of bus stops

Home Energy Efficiency Database (HEED)

6.50 This is a database of home improvements carried out between 2000 and 2012, including those completed under various Welsh Government-supported schemes to provide warm, energy efficient homes in Wales. The data has been anonymised at the household level however, through linking, the anonymised population occupying these homes can be determined. An important aspect of the data is that to be eligible for the heating system and insulation upgrades, the homes or their occupants needed to meet certain eligibility criteria, one of which related to benefits receipt, so can provide information on the socio-economic status of some of the home occupants. Details of the eligibility criteria are not yet included in the database. However, if the dataset can be updated with eligibility detail, it could deliver evidence for the 'social and economic well-being' outcome.[34] Work would

---

[34] Further work will be completed as part of Welsh Government-ESRC-funded work on data linking for fuel poverty.

be required to link the HEED benefits data into SAIL and a further project completed to develop a method to select a representative sub-sample of the Wales population - based on the broader socio-demographic profile of residents - to ensure that findings for HEED recipients could be generalised to the population of Wales.

Appendix 8 Glossary of acronyms

A&E	Accident and Emergency
ADRC	Administrative Data Research Centre
ALF	Anonymised Linking Field
CFAS	Cognition Function and Ageing Survey
CHS	Child Health (System Data Set)
Data Max	Welsh Government Knowledge and Analytical Services Programme to Maximise the Use of Existing Data
EDDS	Emergency Department Data Set (A&E records)
ESRC	Economic and Social Research Council
HIRU	Health Information Research Unit
ICT	Information and Communications Technology
ILDSC	Improving Linked Data for Social Care
IGRP	Information Governance Review Panel
LLI	Long-term Limiting Illness
LSOA	Lower Super Output Area
NISCHR	National Institute for Social Care and Health Research
NLP	Natural Language Processing
NWIS	NHS Wales Information Service
ONS	Office for National Statistics
PEDW	Patient Episode Database for Wales
PHW	Public Health Wales
RALF	Residential Anonymised Linking Field
SAIL	Secure Anonymised Information Linkage
SCHOOP	Social Care and Health Of Older People (a pilot project using Social Services data)
SSSW	Sustainable Social Services for Wales
SQL	Structured Query Language
WDS	Welsh Demographics Service (GP registration history database)
WIMD	Welsh Index of Multiple Deprivation