Dadansoddi ar gyfer Polisi

YMCHWIL GYMDEITHASOL Y LLYWODRAETH

GSR

GOVERNMENT SOCIAL RESEARCH

Analysis for Policy

Llywodraeth Cymru
Welsh Government

www.cymru.gov.uk

Ymchwil gymdeithasol

Social research

Number: 77/2013

Data Linking Demonstration Project: Examining Fuel Poverty using Home Energy Efficiency Data (HEED) and routinely collected Health Data

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Martin Heaven, Health Information Research Unit, Swansea University / Data Linkage Fellow, Welsh Government

Sarah Lowe, Programme to Maximise the Use of Existing Data, Welsh Government

Views expressed in this report are those of the researcher and not necessarily those of the Welsh Government

For further information please contact:

Sarah Lowe

Knowledge and Analytical Services

Welsh Government

Cathays Park
Cardiff

CF10 3NQ

Tel: 029 2082 6229

Email: sarah.lowe@wales.gsi.gov.uk

Welsh Government Social Research, 2013

ISBN 978-1-4734-0676-6

Table of contents

Glossary of acronyms

A&E	Accident and Emergency (Data Set)
ALF	Anonymised Linking Field
CESP	Community Energy Saving Program (CESP)
CHS	Child Health System (Data Set)
EASHR	European Age Standardised Hospitalisation Rate
EASMR	European Age Standardised Mortality Rate
EDDS	Emergency Department Data Set
EST	Energy Saving Trust
EWM	Excess Winter Mortality
EWMI	Excess Winter Mortality Index
HEED	Home Energy Efficiency Database
HEES	Home Energy Efficiency Scheme
HIRU	Health Information Research Unit
IGRP	Information Governance Review Panel
LSOA	Lower Super Output Area
NISCHR	The Welsh Government National Institute of Social Care and Health Research
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
WDS	Welsh Demographics Service (GP registration history database)
WG	Welsh Government
WIMD	Welsh Index of Multiple Deprivation

1 Introduction

The aim and objectives of the demonstration project

1.1 This project is being delivered as part of the Welsh Government Programme to Maximise the Use of Existing Data. It aims to demonstrate the unique contribution data linking can make to the evidence base. The suite of three data linking demonstration projects has examined the anonymised data linkage process from acquiring additional data to carrying out analysis on new data sets created by linking existing administrative data. The projects are intended to stimulate engagement of appropriate WG officials with regard to information governance and practical issues around acquiring, processing and analysing new linked data sets. The projects were delivered by a WG-ESRC Knowledge Transfer Research Fellow working on a one-year fellowship so were designed to be small in scale and exploratory in nature. These constraints are reflected in their relatively limited scope and in both the practical and analytical decisions made throughout. The demonstration projects have taken advantage of the fact that the research fellow was seconded from the Health Information Research Unit at Swansea University, and was therefore already authorised, having sought the correct project-specific permissions, to access large anonymised health and other databases.

1.2 This Project specifically investigates the contribution to the fuel poverty evidence base represented by linking various health data sets to Home Energy Efficiency Database (HEED) data. HEED contains intervention details of the Home Energy Efficiency Scheme (HEES), funded by the Welsh Government from 2000 and put in place to help eradicate fuel poverty through the provision of home upgrades such as heating and insulation measures for those people most in need and schemes funded through obligations placed on larger energy suppliers. From 2000, these schemes included the Energy Efficiency Standards of Performance 3 (EESoP3), the Carbon Emissions Reduction Target (CERT) and the Community Energy Saving Programme (CESP), and were focused on reducing carbon emissions through the provision of home heating and insulation improvements

1.3 Given the limited scope of a demonstration project, it was accepted that it would not be possible to address every potential shortcoming of the data and/or analysis. For

example, the rationale for investment in home energy interventions is that a householder is more likely to be in fuel poverty if they live in a home that is very energy inefficient. Fuel poor households are likely to under-heat their home, which puts them more at risk of cold-temperature related illnesses. However, it was not possible within the scope of the Project to examine the extent to which the home energy improvements stopped households under-heating their homes, only to examine whether there were beneficial effects on the health of residents.

1.4 However, in assessing whether data linking can deliver improved evidence about health outcomes, it must be accepted that there would also have been disadvantages to attempting to evaluate health outcomes in other, more conventional ways. The cost of collecting baseline and post-hoc self-reported health information would have been considerable, for example. The kinds of self-reported health status information that could be collected by survey would, in addition, almost certainly be relatively limited. Data collected specifically as part of an evaluation may also be amenable to bias due to residents' positive regard for the intervention and may reflect their satisfaction with the process more than identifying genuine improvements in health outcomes. The practicalities associated with maintaining a surveillance of residents' health over the longer term would also have presented considerable challenges. In many ways, then, investigating changes in health status using administrative data - should it prove to be both practical and achievable - is preferable to the use of existing methods.

1.5 The Project is experimental. Due to the limited project scope, the lessons learned about data quality and the challenges associated with the use of newly-linked data sets, as well as the developmental nature of the methods, the reader must note that **the findings must be viewed with caution**.

1.6 In recognition of the limited scope of the demonstration projects, a secondary objective was to identify the further contribution that might be made to the Fuel Poverty evidence base in the future should work be undertaken to acquire additional data sets and undertake additional analysis (a list of potential future analysis projects is provided in Appendix 7).

1.7 To achieve these objectives, the process of acquiring and anonymising the HEED data was completed in such a way that it could be linked with other existing anonymised demographic and health databases. The information about the homes where improvements were carried out was utilised, through data linkage, to establish the characteristics of the population living in 'intervention addresses' before, at the time of, and after the intervention.

1.8 A major challenge with trying to identify whether an intervention has resulted in real improvements over time is the fact that changes in the health of the population may have occurred over the same time period but for other reasons e.g. the introduction of better medications. In an attempt to take account of this effect, the Project used a 'control' group, comparing the health of individuals in whose homes interventions had already been completed with those where they had yet to be carried out. In practice, this meant comparing the health of individuals in whose homes interventions were completed between 2000 and 2007 and those whose homes had interventions in 2008-12.

1.9 The process, issues, problems and limitations encountered in the Project are documented in this Report. Chapter 3 describes the processes involved in obtaining and anonymising data from EST into the SAIL databank and in preparing and making the data available to the researcher linked to a number of different health data sets. Chapter 4 describes the methodology used to select and compare the two population groups - those who had interventions recorded in HEED between 2000 and 2007 and those who had interventions recorded in HEED between 2008 and 2012. Chapter 5 provides a comparative analysis of the groups with regard to the specific health conditions identified as most likely to be affected by home energy efficiency improvements. Chapter 6 provides a general discussion on the process, the results obtained and further avenues of work suggested by these results. The challenges that emerged during the demonstration process will be explored in more detail in a Lessons Learned report, publication of which is to follow[1].

[1] Publication will be in late 2013 or early 2014; date TBC.

2 Acquisition and preparation of the HEED data

Summary of the anonymised linkage process

2.1 The HEED extract provided by the Energy Saving Trust (EST) contained information on insulation, heating and other home energy efficiency improvement interventions to households in Wales over the years 2000 to 2012. The extract contained details of addresses where interventions had taken place and details of the interventions but no detail about the individuals living in the addresses.

2.2 The data was provided to the Health Information Research Unit (HIRU) at Swansea University, through a process called Secure Anonymised Information Linkage (SAIL). The name of 'SAIL' has become synonymous with the large data bank of health and other datasets that have been processed into HIRU using this process. 'SAIL' will be used to refer to the databank at Swansea University throughout the rest of this report.

2.3 The anonymisation process involved the use of a Trusted Third party, the NHS Wales Information Service, (NWIS), who were provided with only the identifiable components of HEED, in this case the full address of each property. When data is linked, the identifiable data can either be provided at individual person level or at address level. The HEED identifiable data was provided to NWIS only at address level since, as noted above, identifiable information about residents was not held by EST. When the identifiable information is at the address level, NWIS use it to generate a unique number for each property, before destroying the identifiable data so that the unique numbers cannot be linked back to the addresses. Consistent processing by NWIS ensures that an address always generates the same unique property number. In this way, records already held in SAIL relating to the residents of those properties could be linked to the HEED data without either individuals or households being identifiable to researchers.

2.4 NWIS use the Welsh Demographic Service (WDS) data as the 'population spine' or 'template' for its anonymisation process. The WDS is a database of everyone registered with a GP in Wales from 1994 to the present day. It includes an anonymised residential address history – an index of numbers, one for each household 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 WDS data as another index of unique numbers, known as the Anonymised Linking Field (ALF). In this way, it is possible to associate ALFs with RALFs, that is: people to homes.

2.5 The HEED data was anonymised by NWIS using household addresses. Anonymisation of HEED generated a set of linkable RALFs enabling the selection of the ALFs for the people living in the homes where interventions had been completed.

2.6 Population groups were defined for individuals receiving interventions during two different time periods, effectively providing 'case' and 'control' sub populations. This allowed the effects of the interventions to be distinguished from any general underlying population changes that also might have been going on. The population groups were compared in terms of health service activity available in hospital data (inpatients and outpatients), Accident and Emergency admissions and General Practice (GP) data, all of which had previously been processed into SAIL, including an ALF for data linkage purposes. The WDS, mortality and hospital data are all-Wales datasets but the GP data held in SAIL does not cover all of Wales. At point of writing, participating GPs were providing information for about 47% of the Wales population but due to the efforts of the WG National Institute of Social Care and Health Research (NISCHR), SAIL's seed-funders, this figure is increasing all the time.

Information Governance Issues

2.7 The Home Energy Efficiency Scheme (HEES) was funded by the Welsh Assembly Government (now Welsh Government) from 2000 and was put in place to help tackle fuel poverty (where more than 10% of net income is spent on all energy costs). The scheme improved the homes of people meeting certain eligibility criteria. Data from HEES and the various energy supplier obligations were fed into the Energy Saving Trust's (EST) HEED database but the information provided to EST did not contain eligibility details. The Energy Saving Trust (EST) collates data across the UK for home energy efficiency improvements put in place through various schemes and government initiatives, and makes the intervention data available for research purposes through the Home Energy Efficiency Database (HEED). To do so, the EST operates within a policy framework that ensures confidentiality in data use and

compliance with the Data Protection Act. Under the terms of the EST policy, SAIL were required to sign a data sharing agreement with EST (see Appendix 1).

2.8 SAIL follows the data protection guidance provided by the Data Commissioners Office, and operates within the Swansea University Data Protection policy which is in line with all the relevant UK laws. The anonymous nature of data held in SAIL is such that it is not governed by the Data Protection Act, and it has been agreed by the National Research Ethics Service that research carried out within SAIL does not require ethical review. However 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.

2.9 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). IGRP is a panel of independent specialists in informatics governance and lay members that oversee all research taking place within SAIL. Current membership (June 2013) is listed in Appendix 2. An IGRP application contains an outline of the research rationale for creating the links, 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 data sets are all totally anonymised in SAIL, the selection of a really 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 level of detail necessary in order 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).

An example of potentially identifiable information from the HEED data:

The data set contains a very small number of addresses where 'ground source heat pumps' were installed, only one of which was successfully addressed matched. Reporting on the health outcomes related to the installation of ground source heat pumps could therefore easily become disclosive. IGRP would expect to see assurances that such data would not be reported at this level.

2.10 The IGRP application for this Project, (Appendix 3) was successful without any concerns being raised. This may be in part due to the research fellow being very familiar with the application process and having discussed the data maximisation demonstration projects at a prior meeting of the IGRP panel. The process still took eight weeks to complete, mainly due to waiting for individual reviewers to respond. The target time from submission to approval for a project is 4 weeks, but the IGRP panel members perform this role in their own time, which for some of them is a scarce commodity. Researchers should be aware that gaining IGRP approval can be a time-consuming, iterative process requiring adjustments to their research proposal.

2.11 As noted above, publication of a 'Lessons Learned' report is to follow.

Transferring the data

2.12 The normal standard practice for transferring data is to utilise a secure electronic data transfer facility. For NHS organisations transfers into NWIS use such a system based on the Digital All Wales Network (DAWN). For non-NHS data providers, a secure internet based facility is in place, and for the transfer of data into SAIL a separate but similar Internet based facility is available. User accounts and passwords are created for named individuals from the data provider organisation to allow them to access these systems. The timing of downloads for this Project was such, however, that some of these facilities were not in place due to system changes at both NWIS and SAIL. Data was therefore transferred using the older system of posting CDs.

2.13 With data sharing agreements in place, the data was transferred using what is known as the "split file process" as depicted in Figure 1. An index field was added to the HEED data at EST, numbering all the records.

2.14 EST then created two password-protected, encrypted CDs. The first CD contained the index and the address details. This CD was double wrapped (the CD is wrapped once with “Confidential” written on it, then plain wrapped a second time) and sent by recorded delivery to NWIS. This file is always referred to as “File1” by those processing it.

2.15 A second file, “File 2” was copied to CD, double wrapped and sent from the EST directly to SAIL. This file contained the index field and the intervention data, without the address data. In this way, the identifiable data and the intervention data never appear in the same file during the transfer.

2.16 The “File 1” CD received by NWIS was processed to produce “File 3”. File 3 consisted of a table with two columns - the index and the associated RALF generated from each address. This file was transferred to SAIL using a web-based secure file upload and switching service[2]. File 2 and File 3 were linked using the index number. The index number was then discarded, leaving a table of anonymised intervention data linkable to other data using the RALF.

Preparing data for research

2.17 A database “view” is a structured ‘image’ of information stored in the database, including only a subset 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 administrator can very closely control the data each researcher can work with. There is, in addition, no way that a researcher can alter the underlying data table providing the “view”.

2.18 The “view”, tailored specifically to meet the requirements of the researcher’s project, is loaded into the SAIL databank by the SAIL technical team. The SAIL technical team provide the hardware and database management support for research and are not data analysts. Separating the data management and research analyst functions prevents the need for technical team members to understand the data and for researchers to access underlying data tables or any intermediate stage data. Access to views is controlled and restricted to authorised approved researchers. For the Fuel

[2] See secure data transportation in “The SAIL Databank: building a national architecture for e-health research and evaluation” http://www.biomedcentral.com/1472-6963/9/157.

Poverty Project, access was originally restricted to the Author, who is a senior research analyst working in the SAIL databank. Later, access was sought for named WG staff for Quality Assurance purposes.

2.19 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.

2.20 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.

2.21 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.

2.22 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.

Success Rate for Linkage and Quality Assurance of the Data

Address matching

2.23 Initial scrutiny of the HEED data revealed that 692,129 interventions had been recorded at 490,277 addresses. However, 18% of these addresses were not recognised during the anonymisation process, so data for 584,549 interventions at 400,321 (or 82% of) addresses were successfully linked.

2.24 Most datasets processed into SAIL have a matching success of over 90%. The most likely reason for the large number of addresses that were not matched during the anonymisation process is that the recording of addresses during data collection was poor quality. Some of the investigations that could be done for the unmatched addresses are difficult because researchers working in SAIL do not have access to the identifiable data in File 1, i.e. the addresses. An independent piece of investigative work involving EST and NWIS would therefore be required to examine this issue further but this could not be carried out within the limited scope of this demonstration project because both additional capacity and funding would need to be made available in order to involve these organisations. To complete additional investigation would also require the EST to resubmit the address data to NWIS because, as noted above, the original download was routinely destroyed after the matching process.

Developments in address matching

The address matching process used by NWIS is, at point of writing, subject to a review, partly due to changes in the general usage of address identifiers that are supplied by Ordinance Survey. The intention for the future is that SAIL will provide a dedicated computer at NWIS that will be pre-programmed at SAIL to perform address matching. SAIL will program this machine to automatically report back in detail on matching failures. Suppliers can then be advised about format changes for re-submission. It is planned to put this 'appliance' in place at NWIS by March 2014.

2.25 The investigation of unmatched addresses that was possible within SAIL was based on aggregating the data according to various criteria, such as geographical area, year of intervention, intervention type, etc. to see whether data relating to particular areas, time periods or interventions had been systematically excluded. The investigation demonstrated that there was no distinct pattern to the missing RALFs (addresses) by year, intervention type or geographical location. These checks are reported on in Chapters 6 and 7. The process did identify that those people receiving interventions from 2008 onwards aggregated differently by socio-economic status than those receiving the interventions earlier but this was as expected due to changes in targeting of the scheme. The comparability of subgroups of the Project population is discussed further in relevant parts of the report.

2.26 When analysing administrative data like HEED, it is important to remember that the data were not collected for research purposes but were collected by contractors as part of the administration associated with completing the interventions. The recording of addresses may therefore have been at a level of completeness suitable for recording the work in progress, not necessarily to a level suitable for linking. A possible bias that might arise, for example, is where a whole street is receiving a particular type of intervention - perhaps the full address would be captured for the first house and only the house number and some representation of “ditto” might be captured for the rest of the street. This is purely speculation, but serves as an example of the sort of situation that could arise when data is collected in an operational context and with no expectation of being reused for another purpose. As routinely collected datasets become more important for secondary analysis, it is recommended that data collection standards are specified and enforced as part of the tendering process to contractors carrying out interventions such as those funded under the Home Energy Efficiency Scheme (HEES).

2.27 As part of the project management for the Project, regular Project Team Meetings were held with WG Fuel Poverty Team officials and relevant WG analysts. This allowed the Author to engage with policy colleagues in exploring issues such as the unexpectedly large proportion of unmatched cases, and where possible to agree

solutions as the Project progressed. The draft report was reviewed by three WG analysts outside the Project Team as part of project Quality Assurance.

3 Methodology

Rationale for selection of comparison population

3.1 When embarking on this demonstration project it was hoped that information would be available on the eligibility criteria against which each intervention had been granted. Access to the eligibility criteria would have allowed analyses of health outcomes for groups with much more clearly distinguished levels of need. This could start to establish which groups benefit the most from what type of intervention (for a list of potential future projects, see Appendix 7). Unfortunately, this data is not retained by the EST.

3.2 The project aim was to determine whether improving the energy efficiency of housing has a measurable effect upon the health of the residents. Although each improvement happened at a single point in time, the benefits were intended to be long lasting. The length of time over which individuals may have experienced the benefits of effective heating and insulation in their home is therefore an important consideration in interpreting the health outcomes. This means that people who move house during the Project period pose a problem. Individuals may move house shortly after an intervention has been completed and into a non-improved property, or a property for which home energy efficiency information is unavailable. In order to minimise the effects of varied exposure to the benefits as much as possible, the decision was made to exclude people who moved house during the Project period (2000-2012). The numbers of non-movers was sufficiently high (77%) for the analyses to remain statistically robust, and the groups for comparison were therefore made up of people who did not move – i.e. with known exposure to the home improvement benefits. In order to eliminate other sources of partial exposure to the interventions, new births into the groups during the time period 01.01.2000 to 31.12.2012 have also been excluded from this analysis.

3.3 More sophisticated analysis methods could be used in order to take into account partial exposures and migration, and to examine the effects of housing upgrades on the health of children from birth through the early years of life. However, such additional analysis was beyond the limited scope of a demonstration project and

would require additional resources to be identified (see suggestions for further analyses in Appendix 7).

3.4 The changes we might see as a result of energy efficiency interventions may be either positive (e.g. fewer chest infections) or negative (e.g. more injuries due to people feeling able to move around more in a warmer home), so where a combination of the two may be happening within the same group, identifying a sufficiently large, positive effect of home energy efficiency interventions may be difficult.

3.5 Since many health outcomes are age related, the health of our ageing Project Group would naturally be expected to show some deterioration over time, which would, in turn, be expected to reduce any improvement in health that might be created by a successful intervention.

3.6 To identify change that can confidently be attributed to the intervention, then, it is necessary to adjust for any underlying changes in the health of the population over time. In order to achieve this, two groups of population were defined, each based on individuals resident in HEED addresses, all of whom had lived at a single address in the period 2000 to 2012 (or until death). Individuals whose homes received interventions between 2000 and 2007 (First Group) were compared with individuals whose homes didn't have interventions until 2008 to 2012 (Second Group).

3.7 During the period 2000 to 2007, any changes in the health of the Second Group cannot be due to interventions, so those changes must relate to the underlying change in the population; by subtracting any change observed in the Second Group from that observed in the First Group, any remaining change can be attributed to the interventions.

3.8 For the period 2008 to 2012, when interventions are happening in the Second Group, the First Group is the best available comparator group to allow us to control for underlying changes. This is because we have no information either about the energy efficiency of or about any interventions that may have been made in the homes in the rest of Wales. While it is possible that residents in the First and Second Groups may have paid privately for improvements which would therefore not be recorded in HEED, given that some of the recipients within HEED were eligible to receive

improvements due to their benefit receipt, this group may have been less likely than average to be paying for other home energy efficiency improvements themselves. It is important to note that there are reasons why the First Group is nevertheless not an ideal comparator:

- some of the First Group will have had their interventions relatively close to the cut-off point (2007) so may only just be experiencing any resulting health improvements; and
- others may only just be beginning to experience any longer-term effects of the improvements.

3.9 There were three possible ways to divide the population into two comparison groups: using equal time periods, equal numbers of home interventions, or equal numbers of people. As reported in Chapter 6, (see Table 6.3) the interventions were not evenly distributed over time. Dividing the population into two equal time periods would produce very unbalanced groups in terms of both population size and the time periods over which those populations would be exposed to the benefits of the interventions. Experimentation with different choices of cut-off date determined that using the 8 year period 2000-2007 and the 5 year period 2008-2012 created two groups of similar size both in terms of population and the number of interventions,. The First Group had the first intervention on their property between 01.01.2000 and 31.12.2007. Throughout this report they will be referred to as the First Group. Those people who had the first intervention on their property after 31.12.2007 form the Second Group.

3.10 The analyses that follow examine health service use over time for people in the First and Second Groups. We may expect health outcomes to improve following the interventions but within the limited scope of the demonstration Project, it was difficult to demonstrate that the changes over time we observed were due to home energy efficiency interventions rather than being explained by other factors or having occurred purely by chance. Further research is recommended to examine the extent to which other factors or interventions e.g. Communities First, may have influenced the outcomes. In terms of demonstrating whether the changes could have occurred by chance, we would usually use statistical tests to demonstrate that findings were sufficiently substantial not to have occurred by chance. For this Project, it was not

possible to use the kinds of statistical testing that would usually be applied; this was because a) the changes we expect to identify are relatively small; b) the changes would be expected make the Groups more rather than less similar; and c) in order to control for different population structures of the two Groups (for further discussion, see Chapter 4, below), we needed to use an age standardisation process to allow rates to be compared. However, findings where a consistent effect over time was observed are nevertheless worthy of note and suggest some association between the HEED interventions and the health outcomes so are reported using the phrase 'the data suggests'. It should be noted, then, that the Project is experimental and **individual findings should be viewed with caution and as indicative rather than conclusive**. However, because a number of the findings suggest the same general pattern, taken together they represent a somewhat more conclusive picture.

3.11 In order to allow the comparison of interventions across the full Project period, intervention types like Smart meters, introduced in later years, were excluded from the analysis. Therefore, given the limited scope of a demonstration project, only those addresses receiving heating or insulation interventions were included in the analyses.

The Rest of Wales (who did not move house between 2000 and 2012)

Subtracting the First and Second Groups from the 'non-migratory population' of Wales i.e. those who lived at the same address over the full time period (or until death) left a third, much larger population for whom no home improvements are recorded in the HEED dataset held in SAIL.

We considered using this group as an additional control group. However:

- this group includes the 18% of HEED households that received interventions but were not 'address matched' into SAIL; and
- too little is known about the remainder of the group – we know nothing about the energy efficiency of these homes or the extent to which they may have paid privately for improvements, so information would not have been recorded within the schemes and initiatives contained in the HEED database.

This group was therefore not included in the analysis for the Project.

Defining Population Groups

3.12 Defining population groups who were living in the intervention homes between 2000 and 2012 was done by linkage of the RALF to the WDS anonymised data contained in SAIL. The WDS address history includes, for each person, all RALFs at which they are recorded as living plus both the 'move-in date' and the 'move-out date'. Where people were still resident at the time of analysis the "move-out date" is set for the year 9999. The WDS data also provides the "date of week of birth" and "date of death" (if applicable) for each individual.

3.13 ALFs (the anonymised representation of people) were to the 'First' Group if they were recorded in WDS as living in one RALF (the anonymised representation of an address) for the full period from 01.01.2000 to 31.12.2012 AND had a first intervention between 01.01.2000 and 31.12.2007. ALFs for residents who were living in the same RALF from 01.01.2000 to 31.12.2012 AND had a first intervention between 01.01.2008 and 31.12.2012 were assigned to the 'Second' Group. The fact that the data was not collected for research purposes introduced some challenges into this process e.g. the presence of duplicate 'move-in' or 'move-out' dates; these issues are described, including any solutions developed, in Chapter 5, below.

Comparison measures

3.14 The following measures were calculated on an annual basis in order to compare changes in health outcomes between the First and Second Groups over time.

EASMR, Excess Winter Mortality Index and EASHR

3.15 The all cause European Age-Standardised Mortality Rate (EASMR) was calculated on an annual basis from 2000 to 2012. The EASMR adjusts for any differences in death rates that are explained by the differences in age profile between the Groups, so that when the rates in the two Groups are compared, any remaining difference is not explained by differing population structures.

3.16 A European Age-Standardised Hospitalisation Rate (EASHR) for several specific causes was calculated for the population Groups. In some cases emergency and elective admissions were analysed separately in order to present a more detailed picture. As with EASMR, the EASHR adjusts for any differences in the age profile of the Groups to eliminate any difference that is explained by differing population structures. EASHRs were calculated for causes theorised to be particularly associated with changes in home energy efficiency.

3.17 The Excess Winter Mortality Index (EWMI) is a comparative measure of deaths in the winter months December through to March, the previous four months and the following four months. It indicates the increase in mortality brought about by the harsh winter conditions.

3.18 The PHOW figures shown in Figures 6.2 and 6.3 are All-Wales figures calculated with the complete population for Wales used as the denominator. The Project Groups are the populations who had interventions and who didn't move house between 2000 and 2012, so that the denominator is made up of the subset of the Wales population that did not move house. For this reason, the rates shown in the following analyses are not directly comparable with All-Wales EASMRs.

Primary care data

3.19 The primary care data within the SAIL databank does not have complete coverage of Wales, but is a rich source of additional information about the health of the population. Linking the Groups to the primary care data set using the ALF field

established the proportions for which primary care data was available. GP data was available for 43% of the First Group and 42% of the Second Group populations. A description of the GP practice data is given in Appendix 1.

3.20 GP data is coded in a hierarchical system called Read Coding. Table 3.1, below, lists the top level of the Read code classification system, a hierarchical coding system of several hundred thousand codes which can be used to record almost anything as a code. Events tend to be recorded when relevant to a specific primary care interaction, rather than routinely. It is not possible in SAIL to distinguish what kind of interaction e.g. a face to face visit to a GP, test carried out by a practice nurse, printing of a repeat prescription etc., generated a particular record. The date of each 'event' is recorded. Each single piece of information recorded in primary care creates a record, so two prescriptions and a blood pressure reading, for example, would create three records for a patient on the same day.

Table 3.1 Read Code categories

History, examination and observations
Investigations
Operations and Procedures
Disorders
Administration
Drug and Appliance Products
ICD10 Disease Codes
OPCS 4 Operative procedure codes

3.21 Due to the 'reactive' (as opposed to proactive) nature of the data collection, it is not easy to create a general measure of primary care service utilisation using the GP data. Within the limited time available for this Project, we have chosen to measure primary care use by using a fairly broad measure, that of counting the dates on which patients had a prescribing event. By counting the date rather than the prescribing event, we do not count people multiple times if more than one recorded prescription was generated on the same day. This is not ideal because it only counts the level of prescribing activity and excludes many of the other activities that would be relevant, such as the results of tests, measurements etc. Further, more detailed work would be required to establish what the ideal indicators might be for the Project but this was not possible within the limited scope and timescale of the demonstration project.

3.22 In the development of the Project a number of issues were identified relating to the quality of the data itself. These are described in Appendix 8.

4 The Characteristics of the HEED data set

4.1 This chapter describes the data contained in the HEED database extract provided by EST for linking into SAIL. The number of records by type of intervention is shown in Table 4.1, below. The majority of interventions - and those most evenly distributed across the time period 2000 to 2012 - were the heating and insulation 'Measure Groups' (see Table 4.3, below, for a full list of interventions completed per year). To ensure sufficient numbers were available for analysis, we chose to include only those homes receiving heating and insulation interventions. The majority of the relevant interventions were loft insulation, cavity wall insulation and combined heating measures. Although the analysis for this Project was done at the level of the 'measure group', it would be possible to undertake further analyses for the specific 'Measure categories' individually. Suggested further work is listed in Appendix 7

Table 4.1: HEED Intervention type and quantity

Measure Group Name	Measure category name	Number of homes*	Address matched in SAIL	Address not matched in SAIL	% not matched
Heating	Heating Measures	24,678	22,589	2,089	8%
	Condensing Boilers	13,436	13,436	<5	0%
	Fuel Switching	9,830	7,740	2,090	21%
	Control Measures	263	140	123	47%
	Solid Fire Conversion Cassette	16	15	<5	6%
Insulation	Loft Insulation Measures	305,364	261,210	44,154	14%
	Cavity Wall Insulation	219,658	195,811	23,847	11%
	Draught Proofing Measures	14,689	13,317	1,372	9%
	Hot Water Tank Insulation	13,078	12,940	138	1%
	Solid Wall Insulation	3,,782	2,534	1,248	33%
Microgeneration	Solar PV Panels	1,656	1,047	609	37%
	Solar Water Heating System	1,043	610	433	42%
	Air Source Heat Pump	110	78	32	29%
	Solar Heating Measures	10	7	<5	30%
	Ground Source Heat Pumps	4	<5	<5	75%
Other	Real Time Display Measures	83,864	52,565	31,299	37%
Total	All measures*	691,481	584,039	107,434	16%

** Homes receiving more than one intervention are counted more than once in this table*

Source: HEED data in SAIL

Number of interventions per home

4.2 Just over 81% of the homes had a single intervention recorded in HEED and around 17% had two (see Table 4.2, below). As the vast majority of homes had relatively few interventions, the earliest intervention date has been used to assign homes into the two intervention Groups.

Table 4.2 Number of HEED interventions per home

Number of interventions recorded	Number of homes	Percentage of homes	Cumulative percentage
1	395,175	81.4%	81.4%
2	81,700	16.8%	98.3%
3	6,836	1.4%	99.7%
4	1,512	0.3%	100.0%
5 or more	<5	0.0%	100.0%

Source: HEED data in SAIL

Analysis of the type of HEED intervention and the pattern of interventions over time indicates that insulation measures form the bulk of interventions, and these increased dramatically by year, peaking in 2009. Heating upgrades began to be rolled out in 2001, with the numbers of homes upgraded per year starting with less than 500 in 2001 and rising to over 5,000 in 2007. Figure 4.1 below indicates how the price of domestic fuels (electricity and gas) changed over the period 1996 to 2012. Domestic coal prices rose in a similar manner to gas over this time period[3]. There were large increases in heating fuel costs from 2005 onwards. These underlying changes in costs will therefore have driven up fuel poverty during the same period of time for which we are investigating the effects of home improvements on health. However, rises in fuel costs may have been harder on the First (i.e. more deprived) Project Group. Assuming that the First Group, being more deprived, had more pressures on their finances than the Second Group, increasing fuel prices may have reduced the positive impact of the interventions on the First Group more than on the Second group. If so, any relative change we observe would be an underestimate of the effects of HEED-recorded interventions

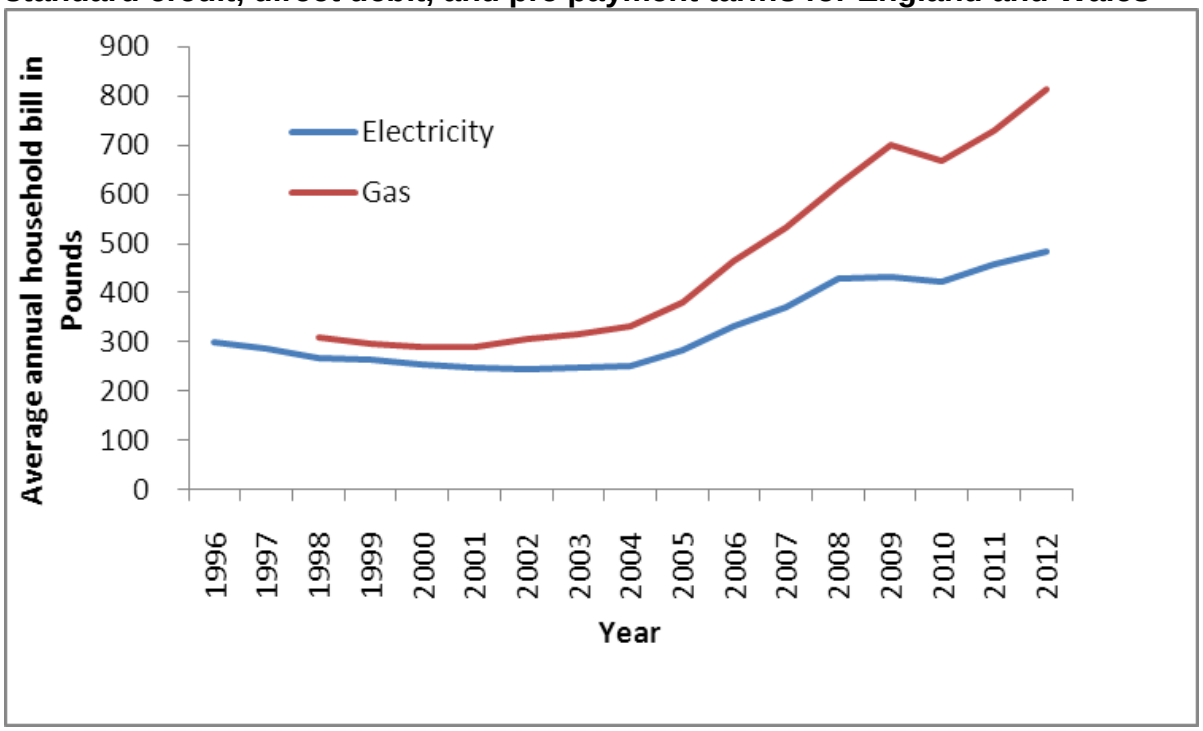
[3] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253878/qep213.xls the tab titled 'monthly prices'.

Table 4.3 Number of interventions per year

Year	Type of intervention				
	Heating	Insulation	Microgeneration	Other	Total
2000		592			592
2001	534	1,979			2,513
2002	3,881	10,918			14,799
2003	4,270	14,276			18,546
2004	4,331	18,770			23,101
2005	1,073	33,653			34,726
2006	766	54,019			54,785
2007	5,001	46,314			51,315
2008	2,838	66,843		2,429	72,110
2009	487	67,049	7	12,757	80,300
2010	515	43,344	1	30,295	74,155
2011	544	31,515	1,695	7,092	40,846
2012		6,892			6,892
Total	24,240	396,164	1,703	52,573	474,680

Source: HEED data in SAIL

Figure 4.1 Average domestic fuel bills 1996-2012[4] based on averaging standard credit, direct debit, and pre payment tariffs for England and Wales



Source: DECC Tables 'QEP 222' & 'QEP 232'

[4] https://www.gov.uk/government/statistical-data-sets/annual-domestic-energy-price-statistics

5 The demographic characteristics of the Project Groups

5.1 The two Groups of individuals selected for the purposes of this Project (i.e. who were resident in the same address for the complete period from 01.01.2000 to 31.12.2012 (or death)) included 244,162 individuals who had interventions in the 'first' period and 236,737 who had interventions in the 'second' period. Some residents died during the Project period; by 2011, this left 185,189 individuals in the 'First' Group, 187,702 in the 'Second' Group. Although the people who died during the Project had variable exposure to the benefits of their improved homes, these people have not been excluded from the Project because death itself may be a relevant outcome.

5.2 Annual population counts were calculated for both Project Groups. The age profile of the First and Second Groups was very similar at the beginning of the Project period (i.e. in 2000)[5] (see Table 5.1, below), with quite a high proportion of older people in both Groups compared with the general population of Wales. This is as we would expect, because older age was an eligibility criterion for some of the schemes included in HEED.

5.3 As the years progress, because we are not adding people who are migrating in or being born into these Groups, so the populations will age and the size of each Group can only diminish through mortality.

Table 5.1: Project Group populations, 2000 and 2011

Age Group in 2000	First Group		Second Group	
Years	N	%	N	%
0-19	31,552	13	30,010	13
20-44	47,511	18	47,546	18
45-64	84,301	35	83,546	35
65+	80,798	33	75,635	32
All ages	244,162	100	236,737	100
Age Group in 2011	**First Group**		**Second Group**	
Years	N	%	N	%
10-19	16,082	9	14,801	8
20-44	33,676	13	31,925	13
45-64	56,094	30	61,020	33
65+	79,967	43	79,956	43
All ages	185,819	100	187,702	100

Source: WDS Registered Populations in HEED intervention defined Groups: SAIL databank

[5] Populations have been compared at the beginning of the study even though the interventions didn't start in the Second Group until 2008, because the rates in the two Groups are being compared throughout the study period.

5.4 By 2011, all survivors in each Group have aged such that no one under the age of ten years remains; therefore where an age group of 0-19 years exists in 2000, by 2011 the Group contains no one aged under 10 years, so the corresponding age group is labelled 10-19 years.

Geographical distribution

5.5 It is possible within SAIL to determine which geographical area each RALF falls into at the level of Lower Super Output Area (LSOA). This is a small area based on Census Output Areas and contains on average 1,500 people. A broad geographical distribution of the Project Groups was created by adding up the populations in each Lower Super Output area and summing this to Local Authority level. Table 5.2, below, shows the numbers and proportions of interventions completed in each Local Authority area. The proportion of the Wales population residing in each Local Authority in 2007[6] is included for reference. The 2007 estimates were chosen for this comparison as being approximately half way through the Project period. The Project Group populations of course lived here from at least 2000 to 2012. The overall impression from the table is that the interventions are reasonably evenly spread across Wales geographically.

5.6 It should be noted that the 18% of HEED addresses that did not match into SAIL are missing from Table 5.2. If for some reason these happened to be concentrated in one area of Wales, the Table might show a slightly different geographical distribution. However, as noted above, further investigation of this issue falls outside the scope of this Project.

[6] http://wales.gov.uk/docs/statistics/2013/130430-mid-year-population-estimates-local-authority-revised-2002-2010-en.pdf

Table 5.2: Geographical Distribution of (Matched) Heed Interventions compared with the Wales population

Local Authority (LA)	Population			Proportion of Group			Proportion of LA Population	
	Wales 2007 ONS MYE	First Group	Second Group	% of First Group in LA	% of Second Group in LA	% of Wales in LA	% of LA in First Group	% of LA in Second Group
Blaenau Gwent	69,685	7,127	4,241	3%	2%	2%	10%	6%
Bridgend	135,949	13,122	11,730	6%	5%	5%	10%	9%
Caerphilly	174,987	11,510	13,631	5%	6%	6%	7%	8%
Cardiff	328,196	18,236	25,753	9%	11%	11%	6%	8%
Carmarthenshire	181,314	14,898	18,965	7%	8%	6%	8%	10%
Ceredigion	75,326	2,862	3,816	1%	2%	3%	4%	5%
Conwy	113,778	6,693	7,988	3%	3%	4%	6%	7%
Denbighshire	94,530	5,623	7,298	3%	3%	3%	6%	8%
Flintshire	150,816	11,866	13,529	6%	6%	5%	8%	9%
Gwynedd	119,398	9,559	8,410	5%	4%	4%	8%	7%
Isle of Anglesey	69,700	6,100	8,963	3%	4%	2%	9%	13%
Merthyr Tydfil	57,173	4,369	3,747	2%	2%	2%	8%	7%
Monmouthshire	89,592	4,633	6,914	2%	3%	3%	5%	8%
Neath Port Talbot	138,957	6,403	7,422	3%	3%	5%	5%	5%
Newport	141,376	12,741	12,396	6%	5%	5%	9%	9%
Pembrokeshire	119,640	6,295	9,306	3%	4%	4%	5%	8%
Powys	131,982	7,240	6,466	3%	3%	4%	5%	5%
Rhondda, Cynon, Taff	234,471	17,874	21,940	8%	9%	8%	8%	9%
Swansea	232,460	23,737	17,527	11%	7%	8%	10%	8%
The Vale of Glamorgan	124,732	7,576	10,497	4%	4%	4%	6%	8%
Torfaen	90,974	8,042	7,502	4%	3%	3%	9%	8%
Wrexham	131,263	4,682	7,256	2%	3%	4%	4%	6%
Wales	3,006,299	211,188	235,297	100%	100%	100%	7%	8%

Source :Ons Population, WDS and Heed data in SAIL

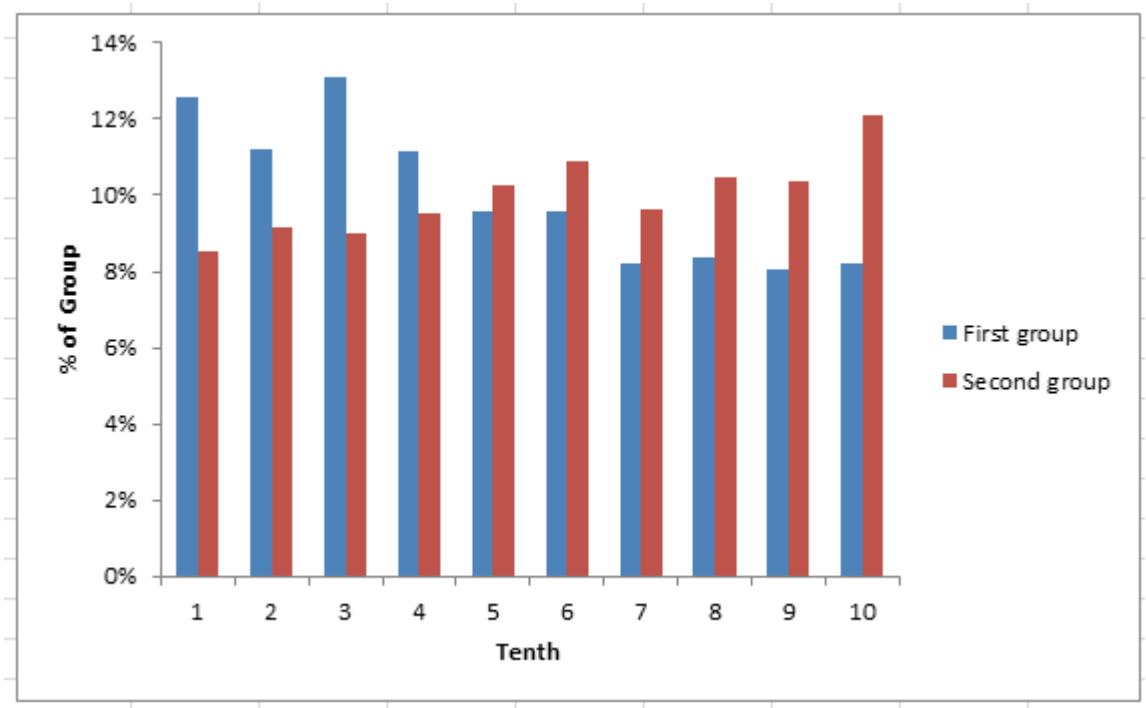
Deprivation

5.7 Using the LSOA associated with each RALF also allows the assignment of an area-based deprivation score to each household. The Welsh Index of Multiple Deprivation (WIMD) ranks Welsh LSOAs according to area based deprivation indicators from a number of domains e.g. income and housing. We chose to use the 2008 WIMD rather than 2011 because it was calculated on data based around the mid-point of the Project period. The problem with using area based scores assigned at an individual level is that an average area score is assigned to the individuals when in reality they may be either relatively more or less deprived than the average for their neighbourhood. Figure 5.1, below, shows the percentage of each group by ten

equally sized groups of the Welsh population ranked and divided by WIMD deprivation score.

5.8 The 'First' Group included a considerably higher proportion of people resident in overall deprived LSOAs and a lower proportion of people resident in less deprived LSOAs compared with the 'Second' Group. Figure 5.1 indicates accurate targeting of interventions, at least to the areas defined as being most in need, even if due to the use of an area-based measure we are not able to assess whether the individual residents were those most in need. The time period covered by the Project means that there is data from a number of different schemes that have different eligibility criteria, with area based schemes being introduced in later years, such as the Community Energy Saving Program (CESP) introduced in 2009. These changes in eligibility criteria over time could produce the effect as seen in Figure 5.1, below.

Figure 5.1: Project Group Population Proportion by WIMD Tenth (1 = most deprived)

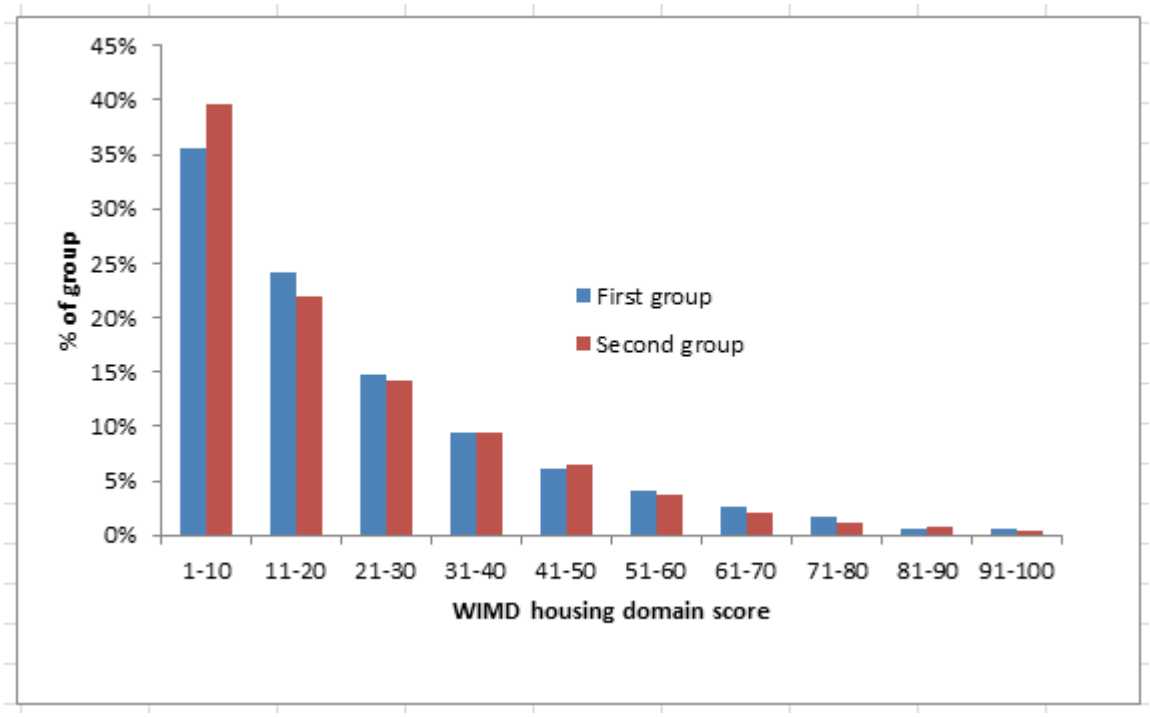


Source: WDS registered populations, 2008 WIMD scores and HEED based Group in the SAIL Databank.

5.9 All householders would have been incentivised by the rapidly increasing prices of electricity, gas and coal from about 2005 onwards illustrated in Figure 5.1, above, but less deprived home owners would have been in a better position to take advantage of the schemes introduced later in the period covered by the Project.

5.10 One of the WIMD domains is 'housing'. The housing domain uses 'lack of central heating' and overcrowding (excluding student households) to create an indicator[7]. The WIMD documentation reports that housing domain scores are not uniformly distributed: the 10% most deprived population have scores from 50 to 100, and the scores of the remaining 90% run from about 0 to 50. This means that in Figure 5.2, below, the five right-most columns relate to the most deprived 10% of Wales. Summing these five groups shows that 11% of the First Group are from the most deprived 10% of Wales in terms of housing, compared with 8% of the Second Group. As noted above, area-based scores may not reflect the circumstances of individual households in the geographical area.

Figure 5.2: Project Group Population by WIMD Housing Domain Tenth.



Source: SAIL databank

Urban Rural distribution

5.11 It is important to consider access to services in any Project that uses any aspect of health service utilisation as an indicator of health. People living in urban areas may have better access to services and hence a greater tendency to use services; this

[7] Welsh Index of Multiple Deprivation 2008 Summary report: http://wales.gov.uk/docs/statistics/2011/111220wimdsummaryreviseden.pdf

tends to be particularly true of 'unscheduled' services such as Accident and Emergency units[8]. Decisions by health service providers, e.g. discharging a hospital patient, may also be influenced by how rapidly an individual can return if the patient's health deteriorates.

5.12 The split between urban and rural areas was created using the ONS Classification published in 2004[9]. As shown in Table 5.3, below, the urban/rural split is similar for both groups, with the 'First' Group containing a slightly larger 'urban' component and a slightly smaller 'Village, hamlet and isolated dwellings' component, when compared with the Second Group. This might effect how services are delivered to a small extent (e.g. a doctor might be reluctant to send someone home if they cannot quickly return to the hospital should an emergency arise, so might decide to keep this person in hospital for a greater recovery period). For this project we are assuming that the differences will not have a great effect on the analyses.

Table 5.3: Project Group by Neighbourhood Type

Neighbourhood type	First Group	Second Group
	%	%
Urban > 10k	70	65
Town and Fringe	19	19
Village, Hamlet & Isolated Dwellings	12	16
Total	100	100

Source: ONS and HEED data in SAIL

[8] Gender differences in adolescent injury characteristics: A population-based study of hospital A&E data, accessed at http://www.sciencedirect.com/science/article/pii/S003335060600103X

[9] The 2004 Classification was used (rather than the 2011 revision) partly because the analysis relates to data for 2000 to 2012 and partly because the 2011 revision was not available until September 2013, when most of the analysis for the Project had been completed.

6 The health status of the Project Groups

6.1 This Project sought to ascertain whether home energy interventions such as improved heating efficiency and insulation have any identifiable effect on the health outcomes for residents. Numerous potential health-related indicators are available using SAIL. For the purposes of this Project, it was a challenge to reduce the indicators to a small set of policy-relevant measures. In discussion with the Welsh Government Energy Efficiency and Fuel Poverty Team, cardiovascular, respiratory and mental health conditions as well as injuries were chosen as the conditions most likely to be affected by living in a cold home. Overall mortality rates and excess winter deaths were also analysed. With the large data set available to the Project it would be possible to focus on more specific conditions known to be affected by fuel poverty e.g. specific cardiovascular diseases, respiratory conditions, arthritis and rheumatism[10]; however, to remain within the limited scope of a demonstration project, the analysis compares data only at the level of these very broad categories.

6.2 The considerable time period over which home energy efficiency interventions were taking place created the opportunity to compare the First and Second Groups over time to allow any effects on health outcomes to be identified. However, as noted above, changes over time are complex to study because other unrelated changes can be going on at the same time, changes would need to reach a certain size before we can be confident they are real and, for some longer-term health conditions, it might be necessary to analyse data over a longer period for such benefits to be observed.

[10] The Health Impacts of Cold Homes and Fuel Poverty, Marmot Report May 2011.

Table 6.1: Summary of Indicators

Source Dataset	Indicator name and description
Population and Deaths data as recorded in WDS database	**European Age Standardised Mortality Rate** All age, all cause age standardised mortality rates from 2004 to 2012
	Excess Winter Death Index: Three year rolling averages 2001-03 to 2009-11
Hospital Data PEDW	**All Circulatory disease:** European Age Standardised Hospitalisation Rates (EASHR) per 100,000 population for: • All Circulatory disease admissions • All elective (planned) Circulatory disease admissions • All emergency Circulatory disease admissions • Ischaemic Heart Disease emergency admissions • EASHR for emergency stroke admissions • All respiratory admissions • Asthma admissions • All mental health admissions • Mood disorder admissions • Injuries and poisonings
GP Practice Data	**GP activity level as indicated by:** The number of prescription days per head of population for: • All prescribing • Prescribing for Respiratory conditions • Prescribing to Infections. • Prescribing rates for the First Group before and after intervention, adjusted for underlying changes using the Second Group (demonstration of an alternative methodology).

What are we looking for?

6.3 If a detectable change in mortality, hospitalisation rates, or the use of GP services is caused by home energy efficiency interventions, we would expect to see relative changes between the First Group and the Second Group as illustrated in the example graphs provided in Figure 6.1, below. These are dummy graphs of the kinds of changes that might be seen if the improvements were having a positive effect on health.

6.4 Case 1 shows a simple scenario where there is no underlying change in population health over time caused by other, external factors – this is very unlikely to be happening in practice. Cases 2 shows a more complex scenario where an underlying change is occurring, in this case an overall increase in admissions. In practice, some underlying change in the health indicators is more likely to be seen.

6.5 Cases 1 and 2 depict a small effect of the interventions. Given the discussion above in Chapter 3, a relatively small effect is what is likely to be observed in practice.

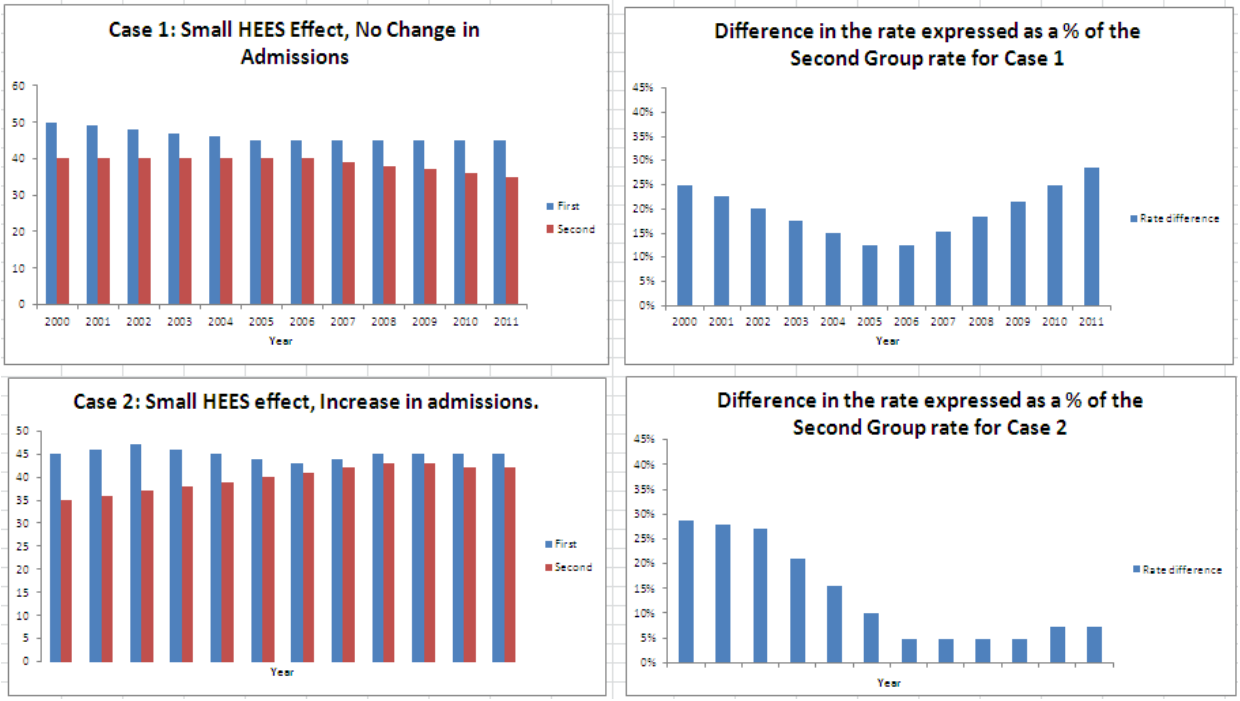
6.6 As noted in Chapter 3 above, the underlying health of the Second Group is better than that of the First (we suggest this is because the First Group is more deprived). In fact, for many of the health indicators, the health of the Second Group was improving faster than that of the First Group, leading to a divergence in their rates over time at the beginning of the study period. This chimes with what is known about increasing inequalities during this period[11]. If there was no change in health caused by the intervention (i.e. if only underlying improvements in population health were affecting both Groups in the same way), we would expect to see an improvement in both Groups, with the gap between the rates for the two Groups remaining about the same over time – or, if health inequalities continued to grow, we would expect to see the gap widening. If an improvement is caused by the interventions, and since the interventions are spread across the whole time period rather than being carried out at a single point in time, what we would expect to see from 2000 to 2007 is a gradual 'excess improvement' in the health of the First Group over time compared with the health of the Second. So, while the health of the Second Group improves only due to underlying improvements in population health, the health of the First Group shows an

[11] http://www.ifs.org.uk/publications/6759

'excess improvement', with its rates gradually moving closer to – or converging with - the rates seen in the Second Group.

6.7 In later years, the example rates would be expected to gradually diverge again – this is because the improving health in the First Group reaches a plateau because no new interventions are taking place but the health of the Second Group begins to improve due to interventions gradually being completed on their homes.

Figure 6.1: Example Effects for HEED-recorded Interventions – What are we looking for?



6.8 To make any divergence or convergence easier to visualise, the percentage difference in rates between the First and Second Groups is depicted in the right-hand graphs for both Case 1 and Case 2. Here, we see the convergence mentioned above as a decrease in the difference over time, followed by the divergence mentioned above, which is seen as an increase. For clarity, both types of graph are shown for the findings reported in this Chapter.

6.9 The example graphs assume that the effect of the interventions on health begins to be seen immediately. This will not be the case. Firstly, this is because, as noted above, the interventions are being completed gradually over time, so a certain proportion will need to have taken place for any health effect to become apparent at the level of the Group. Secondly, some health conditions may take a long time to develop and therefore to improve e.g. cardio-vascular disease, so we might expect

the effect to take longer to become observable in the data. Irrespective of the extent to which there is a delay between the start of the interventions and any observable effect, the pattern we are looking for is the same – a convergence between the rates for the two Project Groups, followed by a divergence. If a delay between the start of interventions and the outcomes emerging is present, the patterns in Figure 6.1 would simply move 'forward' in time and be seen closer to the right-hand side of the graph. If effects are taking a long time to occur, the current series of years (up to 2012) may not be sufficiently long to show the full effect of the HEED-recorded interventions.

6.10 For this Project, the changes over time are also occurring against the background of a worsening economic climate that may have had a relatively greater impact on the health of the more deprived. As noted in Appendix 8, fuel prices increased significantly from 2005 and the financial crisis hit in 2008. Bearing this in mind, over the period 2000 to 2012, we might, in fact, expect to see rates diverging over time, unless the interventions have a protective effect. Any consistent time periods where the First and Second Group bars converge, in no matter how small a way, will therefore be worthy of note.

Findings

6.11 A summary of the findings for each of the indicators considered is shown in Table 6.2, below. Indicators have been considered in terms of the way rates changed over time and the relative changes between the First and Second Groups. A 'positive effect' indicates an 'excess improvement' in the First Group in addition to any improvement seen in the Second Group.

6.12 As discussed in greater detail in Chapter Three, it is difficult to demonstrate that the changes over time we observed were due to home energy efficiency interventions rather than being explained by other factors or having occurred purely by chance. In terms of demonstrating whether the changes could have occurred by chance, it was not possible to use the kinds of statistical testing that would usually be applied; however, findings where a consistent effect over time was observed are nevertheless worthy of note and suggest some association between the HEED interventions and the health outcomes so are reported using the phrase 'the data suggests'. It should be noted, then, that the Project is experimental and **individual findings should be**

viewed with caution and as indicative rather than conclusive. However, because a number of the findings suggest the same general pattern, taken together they represent a somewhat more conclusive picture.

Table 6.2: Summary of findings

Indicator name and description	Finding
European Age Standardised Mortality Rate	No noteworthy impact of HEED recorded interventions
Excess Winter Mortality	Using rolling averages, the data suggest a positive effect of HEED by 2002-2004
European Age Standardised Hospitalisation Rates (EASHR)per 100,000 population for admissions for:	The data suggest …
• All Circulatory diseases	• a positive effect after 2 to 3 years
• All elective Circulatory diseases	• a positive effect after 2 to 3 years
• All emergency Circulatory diseases	• a positive effect after 3 to 4 years
• Ischaemic Heart Disease	• a positive effect after 5 to 6 years
• EASHR for emergency stroke	• a positive effect after 4 to 5 years
• All respiratory diseases	• some limited indication of a positive effect after 4 to 5 years
• Asthma	• a positive effect after 5 to 6 years
• All mental health	• too much fluctuation to be confident of any effect
• Mood disorder	• a possible positive effect after 3 to 4 years
• Injury and poisonings	• a positive effect after 3 to 4 years
The number of prescription days per head of population:	The data suggest …
• all prescribing	• some limited indication of a positive effect after 4 to 5 years
• prescribing for Respiratory Conditions	• some limited indication of a positive effect after 7 to 8 years
• prescribing for Infections	• too much fluctuation to be confident of any effect
Prescribing rates for the First Cohort ‘before and after’ intervention adjusted for underlying changes using the Second cohort.	• a positive effect on prescribing after the exact intervention date.

Mortality

6.13 The all-age, all-cause EASMR is a standard statistic produced by public health bodies. It is usual to use the ONS mid-year population estimates in the denominator; however, within SAIL the population registered in primary care is used because this is the population spine all the data linkage is based on.

6.14 EASMR calculations would also usually be done for a 'whole population' i.e. unlike the Groups created for this Project, they include both new births and a migratory component. The EASMRs for this Project are based on the population surviving each year, including no new births or migrants. For this reason, the calculations in this report are not directly comparable with routinely published European Age Standardised Rates.

6.15 European Age Standardised Mortality Rates (EASMR) were calculated for the Project Groups using a custom spreadsheet template provided by the Public Health Wales Observatory (PHWO)[12]. The European Age Standardised Mortality Rate in Wales has been reducing for a number of years in line with the well-documented reduction seen in the UK as a whole[13]. Published trend data, taken from "Measuring inequalities: Trends in mortality and life expectancy in Wales"[14], is shown in Figures 6.2 and 6.3, below. The graphs show the variation in EASMR between the least and most deprived fifths (based on WIMD 2008) of the population of Wales against the All-Wales EASMR. They clearly show a consistent downward trend in EASMR for both the least and most deprived groups in Wales but a consistently higher rate in the most deprived compared with the least deprived. The trends are the same for women and men, but with higher rates for men.

[12]PHWO spreadsheet entitled 20130112_Methods_09_DirectStandDobsonCIs.xlsx

[13]Public Health Wales Observatory http://www.wales.nhs.uk/sitesplus/922/page/58384

[14]http://www2.nphs.wales.nhs.uk:8080/PubHObservatoryProjDocs.nsf/($All)/BA402B3D53C6A33D8025795E00556236/$File/InequalitiesProfiles_AllWales_Final_English_v1.pdf?OpenElement

Figure 6.2 European age standardised mortality rate per 100,000 population: Wales three year rolling average - females

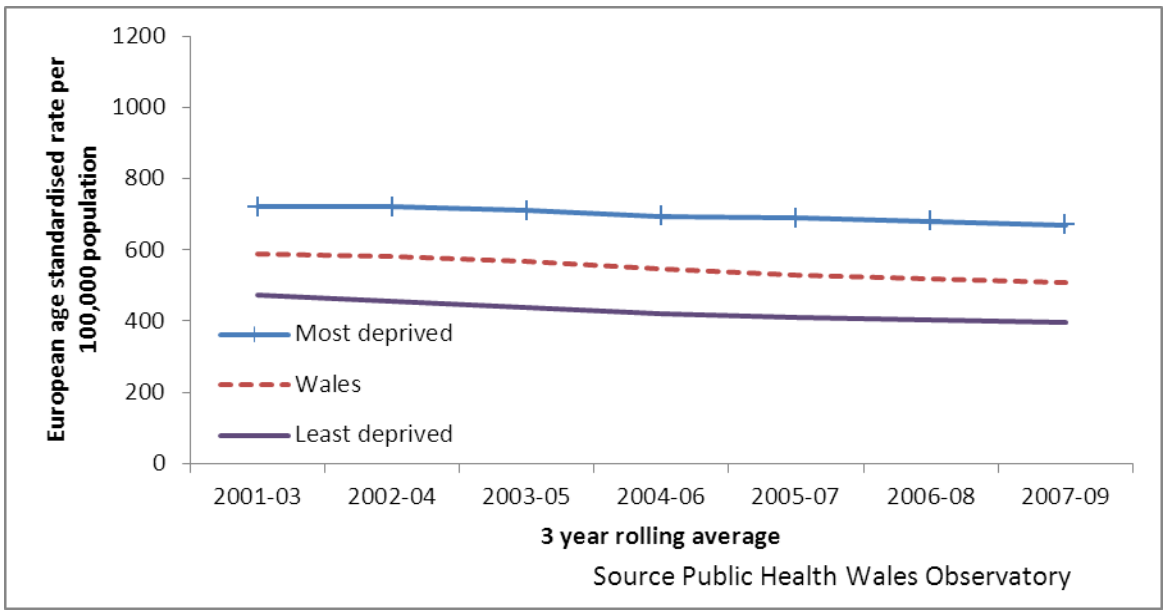
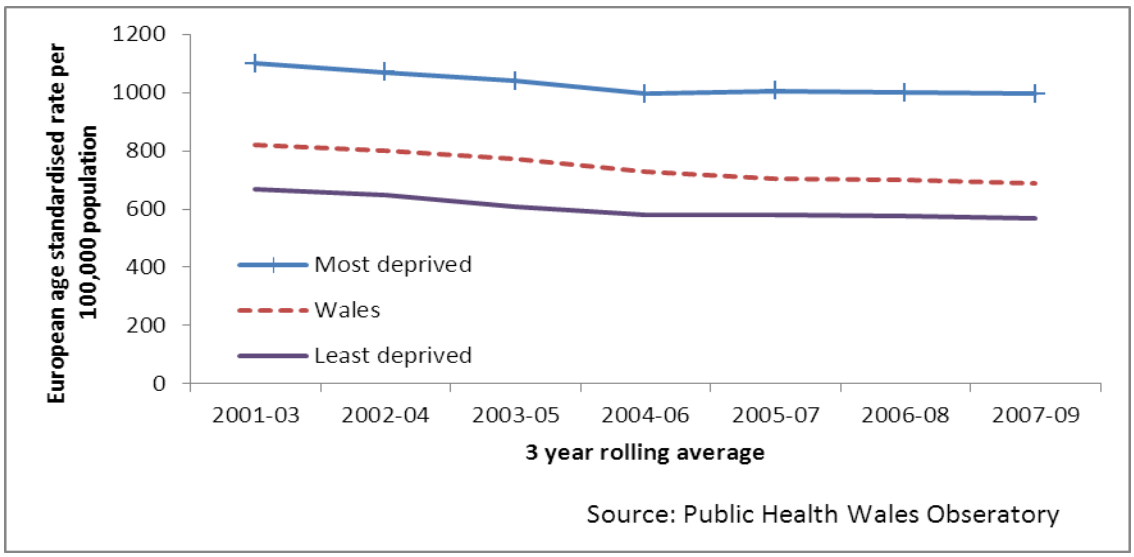


Figure 6.3 European age standardised mortality rate per 100,000 population: Wales three year rolling average - males



6.16 The European age-standardised mortality rates for 2000 to 2011 for the First and Second Group are shown in Figure 6.4a. The First Group has higher rates than the Second throughout the Project period, with the rates increasing up to 2008, followed by a steady decrease. The confidence intervals in Figure 6.4a are calculated using Dobson's method for EASRs, as executed in the PHWO spreadsheet. The First and Second Groups have significantly different rates except for 2000 and 2003; we can therefore conclude that the First Group has a significantly higher mortality rate than

the Second. The Second Group shows a relatively steady rate over time. Figure 6.4b shows how the difference between the rate in the First Group and the Second Group changed year on year. This indicates that the rates diverged until 2008, then levelled off, and may have started to converge in the last two years. However, this effect is unconvincing due to the relatively large amount of year-on-year fluctuation. A more detailed study would be required to establish whether any positive effect of home energy efficiency interventions could be verified. Note that for Figure 6.4a, at point of writing, mortality data were only available for part of the year 2012.

Figure 6.4a: European age standardised mortality rate per 100,000 people for all deaths, by Project Group 2000-2012

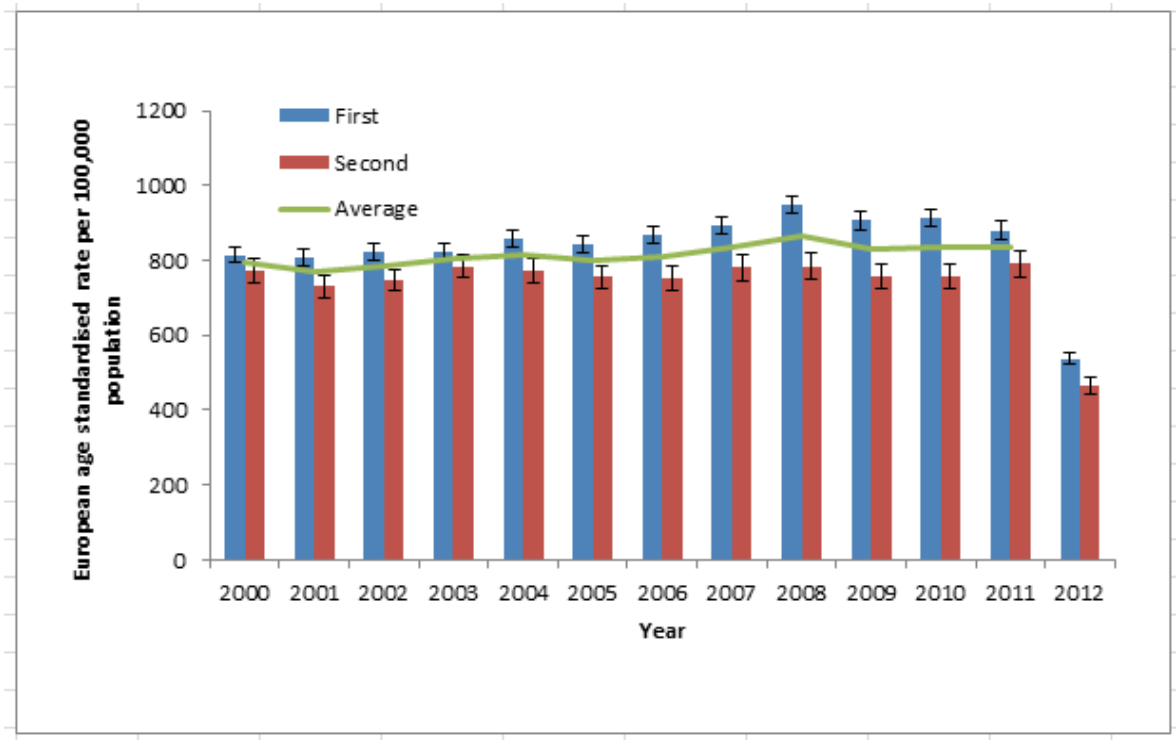
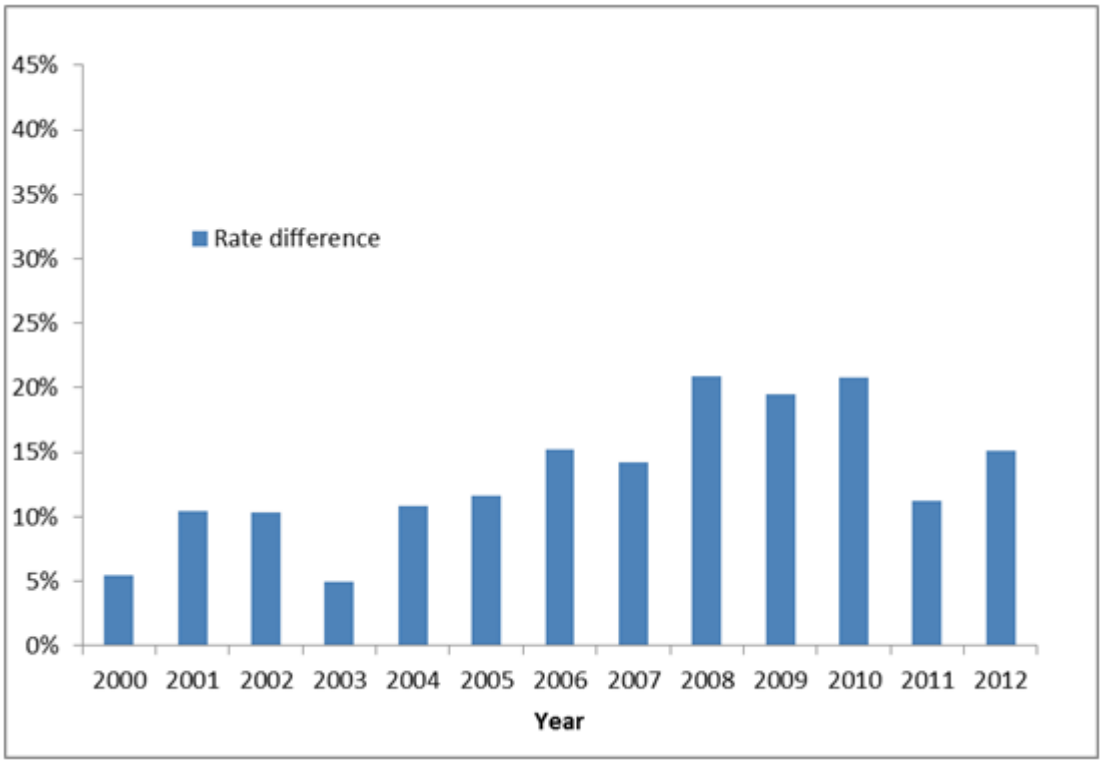


Figure 6.4b: Rate difference as a percentage of Second Group: EASMR for all deaths, by Project Group 2000-2012



Excess winter death Index

6.17 Excess Winter Mortality (EWM) relates to deaths in individuals aged over 65 years; it is calculated as:

$$\text{Deaths in December to March} - \frac{(\text{Deaths in August to November} + \text{deaths in April to July})}{2}$$

Deaths in the winter months December through to March are compared to the previous four months and the following four months.There is some fluctuation in the rates for our Group populations so three year rolling averages have been applied for smoothing [15]. However, there remains considerable variation across the years. The Excess Winter Mortality Index (EWMI) is defined as the EWM divided by the average

[15] ONS Excess winter mortality calculation: http://data.gov.uk/dataset/excess_winter_mortality

non-winter deaths. This was calculated for the two Groups with the results shown in Figure 6.5a, below, alongside the comparable ONS published data.

6.18 The Wales trend shows us the underlying variation in the Index. Both the First and Second Groups follow the underlying trend. However, the differences between the two Groups (Figure 6.5b) suggests a positive effect on the First Group from 2001 onwards followed by an excess improvement in the Second Group for 2008-10 onwards i.e. when improvements are happening in Second Group homes.

Figure 6.5a Excess Winter Mortality Index, three year rolling average

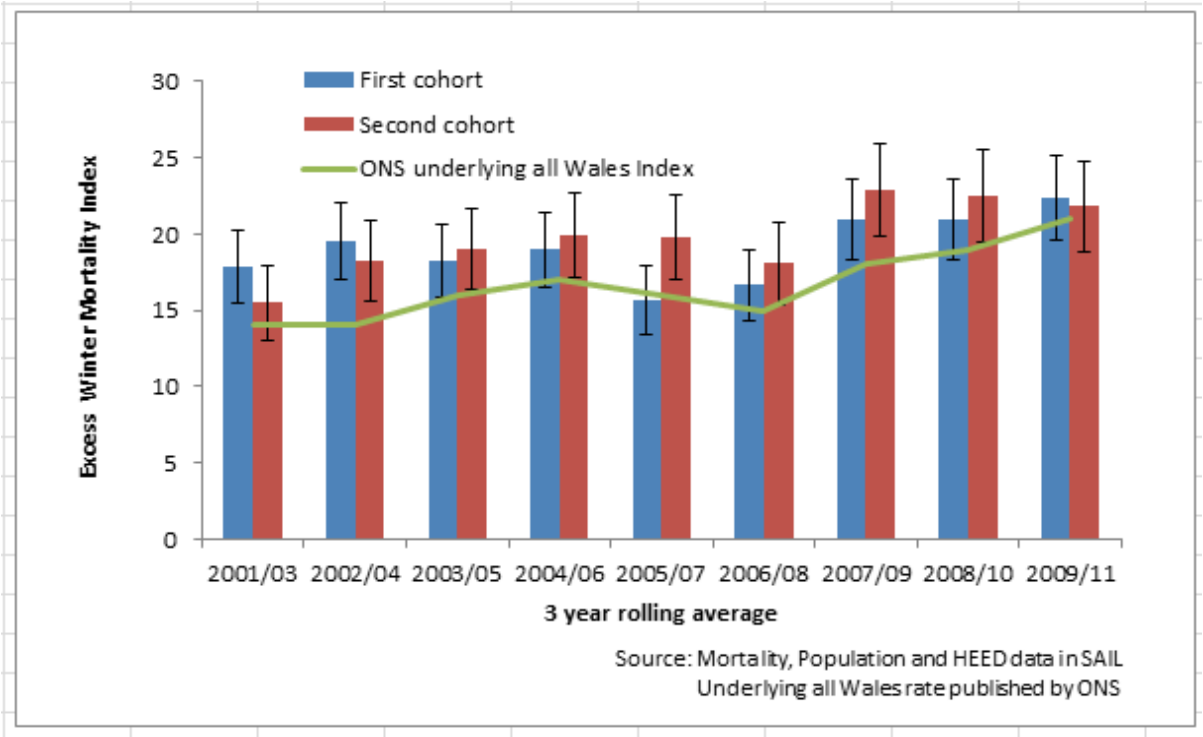
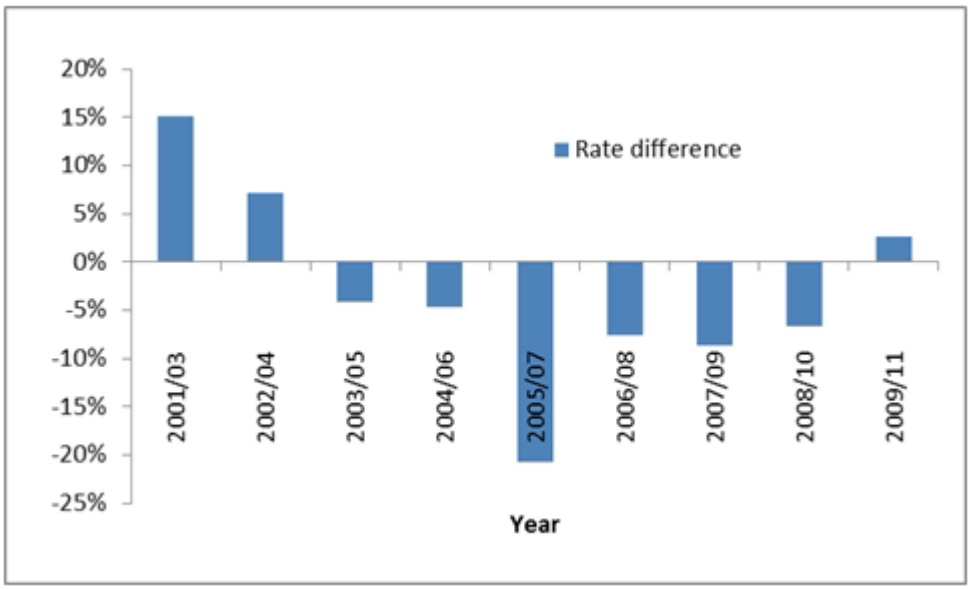


Figure 6.5b: Rate difference as a percentage of Second Group: EWMI, by Project Group 2000-2011



Hospitalisation

6.19 Linking to the Patient Episode Database for Wales (PEDW) was undertaken in order to compare hospitalisation rates between the Project Groups. A description of the anonymised PEDW database within the SAIL databank is given in Appendix 1.

6.20 The level of detail available in the PEDW data allows analysis to be carried out on any health condition that requires hospitalisation, provided that there are sufficient numbers of admissions to allow meaningful comparisons. As noted above, within the limited scope of the Project, it was decided to focus on cardiovascular disease, respiratory disease, mental health and injuries as these were thought to be the most likely to be affected by home energy efficiency interventions. As noted above, records were selected on the basis of any mention of the relevant diagnosis when the patient was admitted.

6.21 A European Age-Standardised Hospitalisation Rate (EASHR) for each cause was calculated for the two Groups, in some cases for emergency and elective admissions separately. The following sections summarise the findings by cause of hospitalisation.

Circulatory diseases

6.22 The EASHR for all circulatory diseases (ICD10 I00=I99) is shown below (Figure 6.6a). The First Group had a consistently higher rate than the Second Group. For both Groups, the rates increased from 2005 onwards, reflecting an underlying trend affecting both. Figure 7.6b shows the relative change in rates, indicating diverging rates until 2003 followed by a consistent closing of the gap that started to level out in 2009. This suggests that the interventions had a positive effect on Circulatory disease admissions in the First Group from 2003 onwards, i.e. around three years after interventions started being rolled out, and may have started to have an effect in the Second Group from 2010 onwards.

Figure 6.6a: European Age Standardised Hospitalisation Rate for All Circulatory Diseases, 2000-11, by Project Group

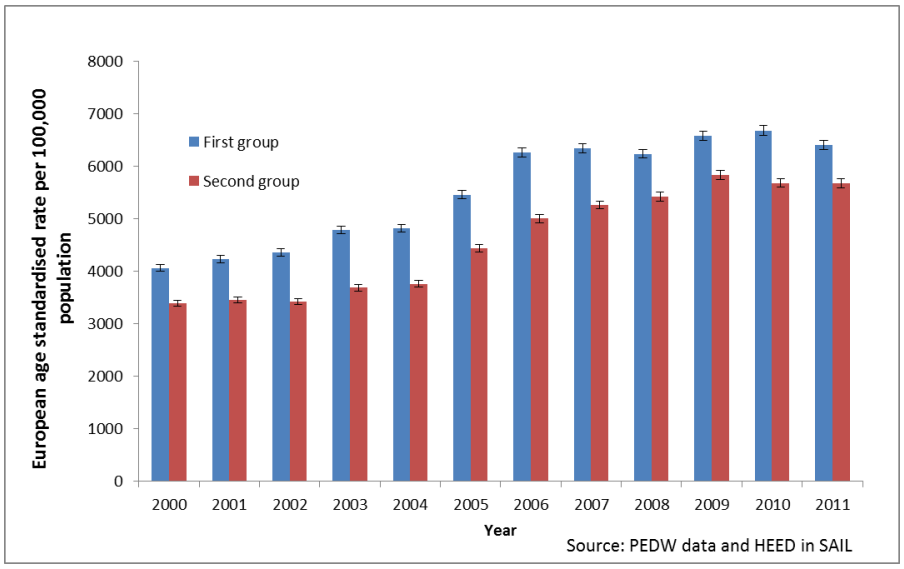
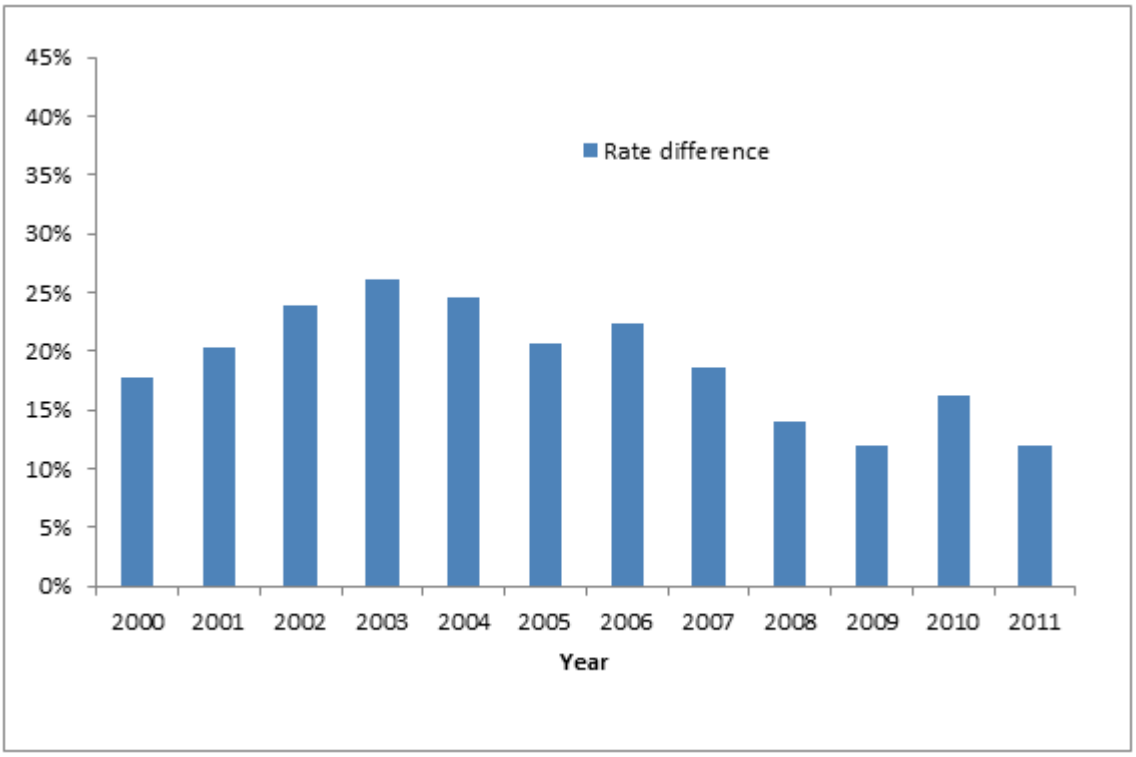


Figure 6.6b: Rate difference expressed as a percentage of the Second Group rate: EASHR for All Circulatory Diseases, 2000-11, by Project Group.



6.23 To investigate this further, the elective and emergency components of the Circulatory disease admissions have been analysed separately, and the component 'cause groups' Ischaemic Heart Disease and Stroke were examined.

6.24 The elective admission rates (see Figure 6.7a, below) for all circulatory diseases show that considerable change occurred in elective admissions for Circulatory diseases after 2004. This may reflect changes in treatment regimes that required shorter but more frequent visits. An analysis of overall length of stay might answer this question, but this has not been possible within the limited Project scope. The relative difference in rates between the First and Second Groups is shown in Figure 7.7b. There was a steady convergence in rates between 2003 and 2009 followed by a divergence in 2010. This would indicate an 'excess' improvement in the First Group compared with the Second, followed by an 'excess' improvement in the Second, consistent with an effect being observed two to three years from the point where HEED interventions began to be completed in each Group.

Figure 6.7a: European Age Standardised Hospitalisation Rate for Elective Circulatory Disease admissions, 2000-11, by Project Group

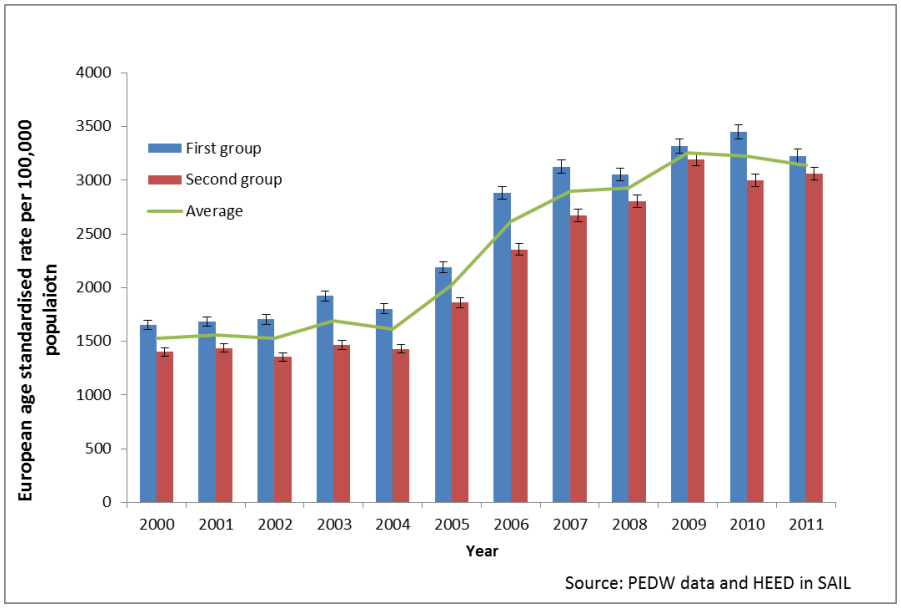
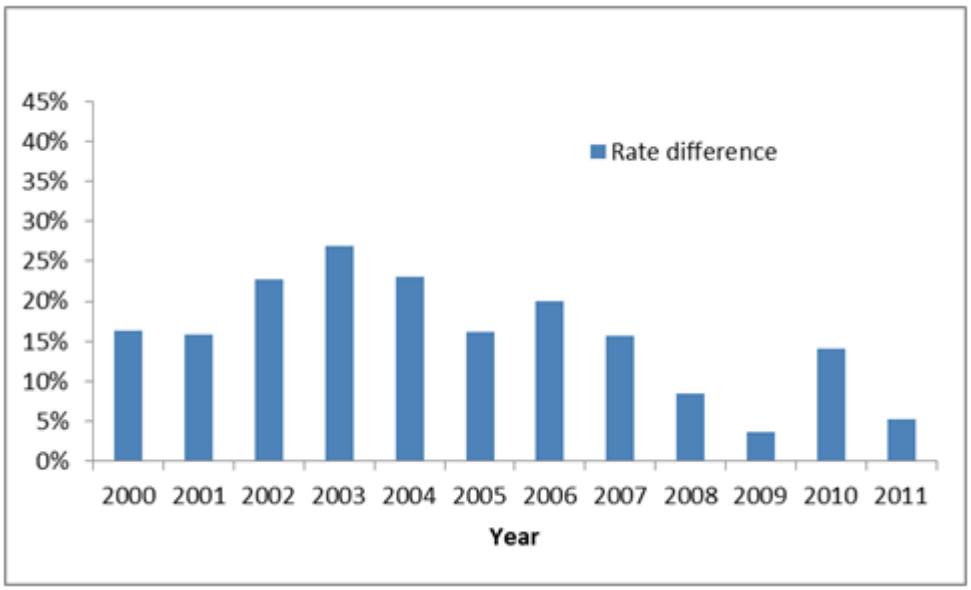


Figure 6.7b: Rate difference expressed as a percentage of the Second Group rate: EASHR for All Elective Circulatory Diseases, 2000-11, by Project Group



6.25 The EASHR for all emergency admissions for Circulatory diseases is shown in Figure 6.8a, below, with corresponding rate differences in Figure 6.8b. The rates diverged until 2004 followed by a continued convergence. This suggests that improvements can be observed in both emergency and elective admissions but for emergency admissions they only reached an observable level around three to four years after the interventions began to be completed.

Figure 6.8a: European Age Standardised Hospitalisation Rate for Emergency Circulatory Disease admissions, 2000-11, by Project Group

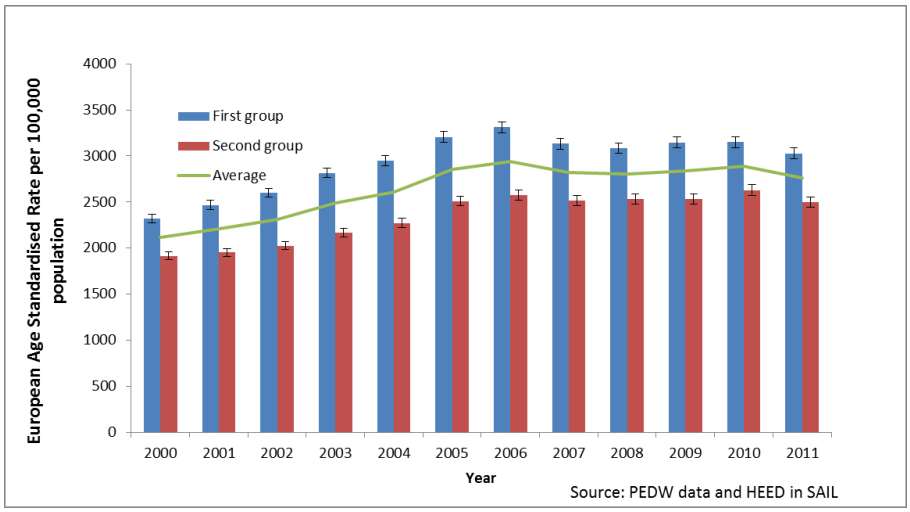
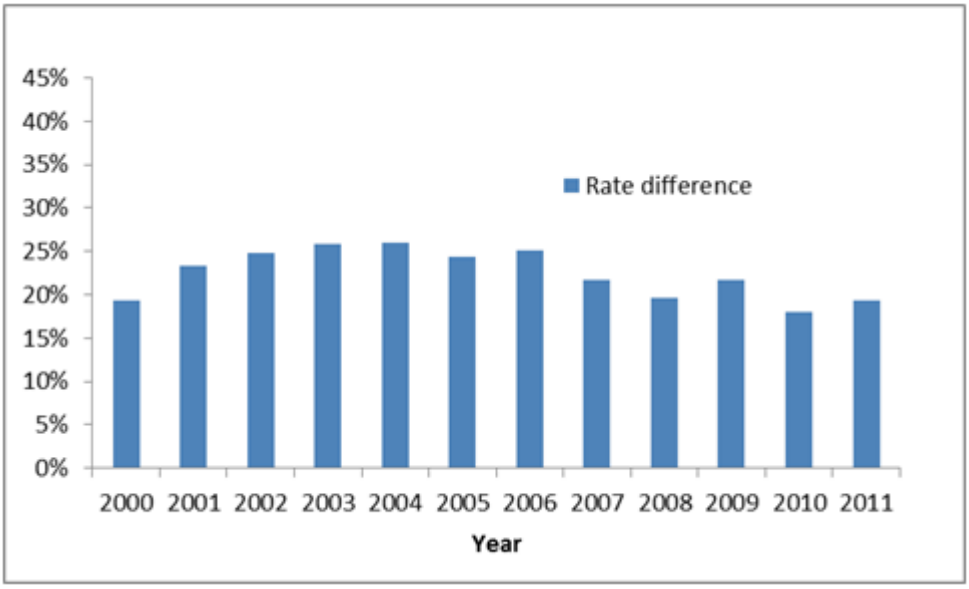


Figure 6.8b: Rate difference expressed as a percentage of the Second Group rate: EASHR for All emergency Circulatory Diseases, 2000-11, by Project Group



6.26 The emergency admissions for a major component of Circulatory diseases, Ischaemic heart diseases, was also calculated (Figure 6.9a with relative rate changes in 6.9b). The rates diverged until 2006 followed by convergence for 2007 onwards. If this effect is due to home energy efficiency interventions in the First Group there was a delay of around 5 to 6 years between the point where interventions begin to be completed and the effect becoming observable.

Figure 6.9a: European Age Standardised Hospitalisation Rate for Emergency Ischaemic Heart Disease admissions, 2000-11, by Project Group

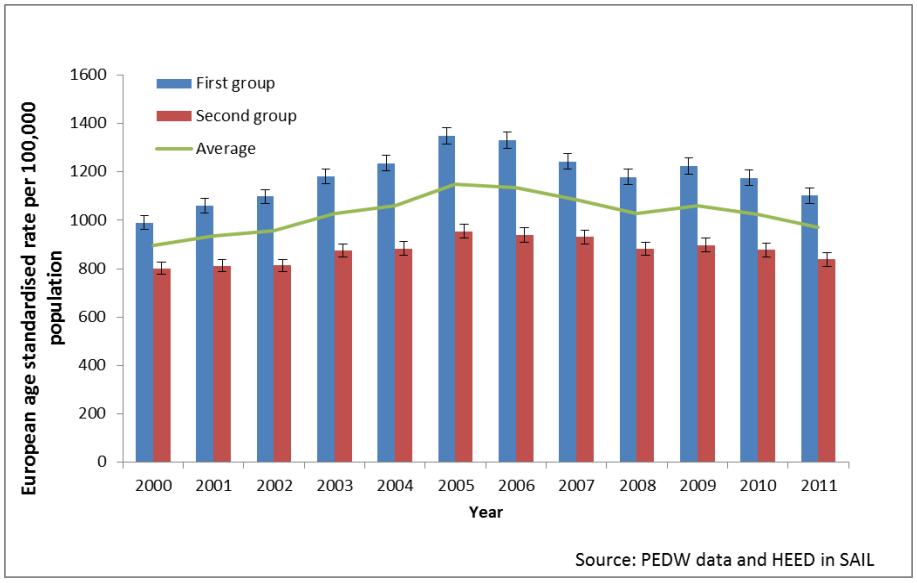
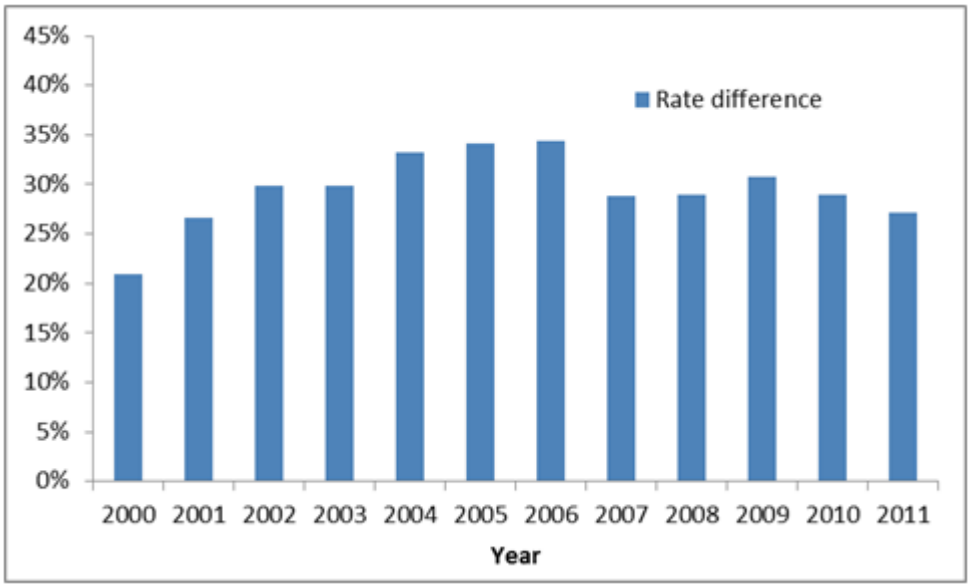


Figure 6.9b: Rate difference expressed as a percentage of the Second Group rate: EASHR for Ischaemic Heart Disease, 2000-11, by Project Group



Stroke

6.27 The rates of emergency stroke admissions (ICD 10 I60 – I 69 Cerebral Infarctions) are shown in Figure 6.10a, below. The rates in the First Group increase until 2006 and then fall. The difference in rates between the Groups is shown in Figure 6.10b, which shows some divergence from 2001 through to 2006 then convergence thereafter. If this convergence is due to HEED-recorded interventions then we are seeing a four to five year delay between interventions starting to be rolled out and an

observable effect being seen in the First Group. Several more years' data would be required to see if the rates for the Second Group showed the same effect.

Figure 6.10a: European Age Standardised Hospitalisation Rate for Emergency Stroke admissions, 2000-11, by Project Group

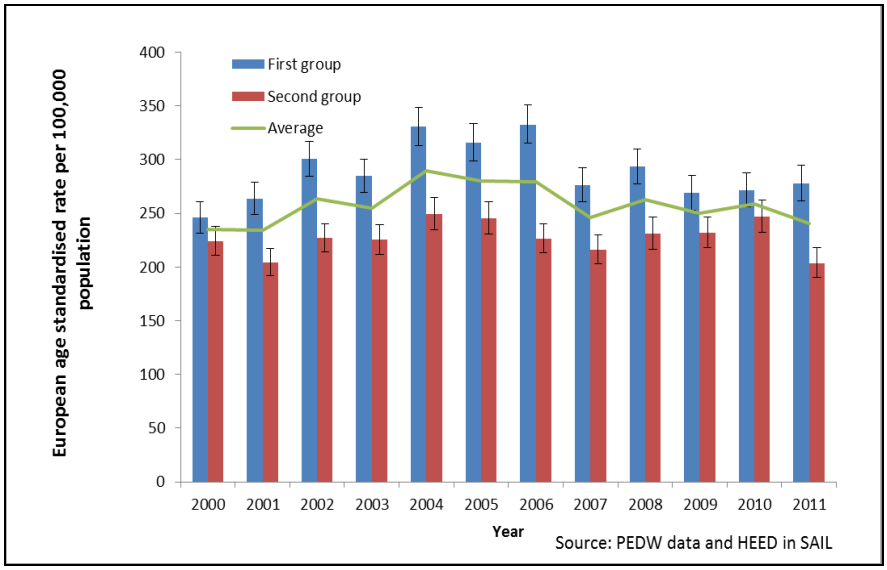
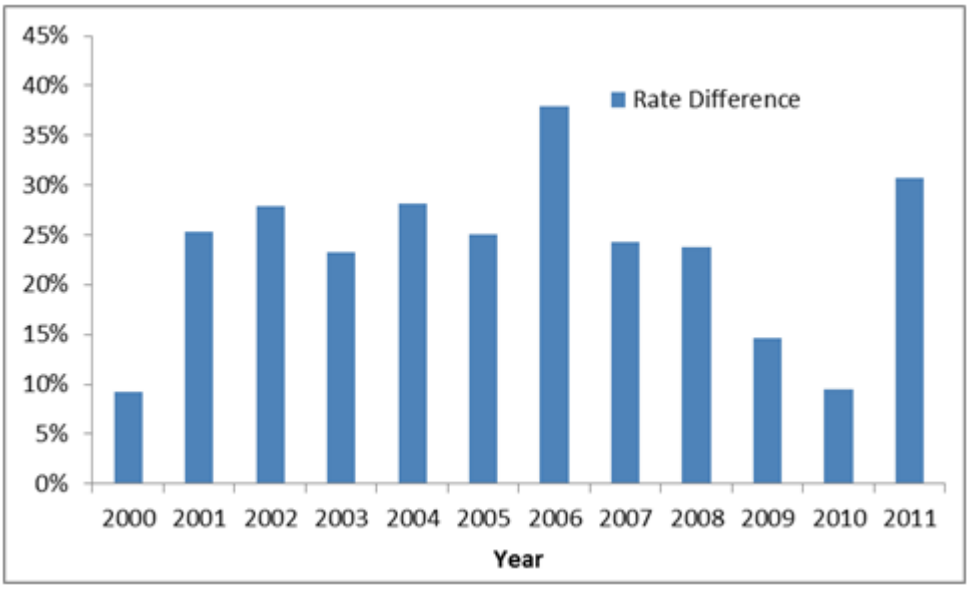


Figure 6.10b: Rate difference expressed as a percentage of the Second Group rate: EASHR for Strokes 2000-11, by Project Group



Respiratory Disease

6.28 European Age-Standardised hospitalisation rates for admissions for 'all respiratory diseases' (ICD 10 J00- J99) are shown, with rate differences, in Figure 6.11a and 6.11b. Significantly higher rates are found throughout the years in the First Group

compared with the Second Group .The rates diverged until 2005 then there is some limited indication of a narrowing of the gap between the two Groups, indicating the possibility of an 'excess' improvement that takes four to five years to become observable, although, as noted above, this cannot be tested statistically.

Figure 6.11a: European Age Standardised Hospitalisation Rate for All Respiratory Disease admissions, 2000-11, by Project Group

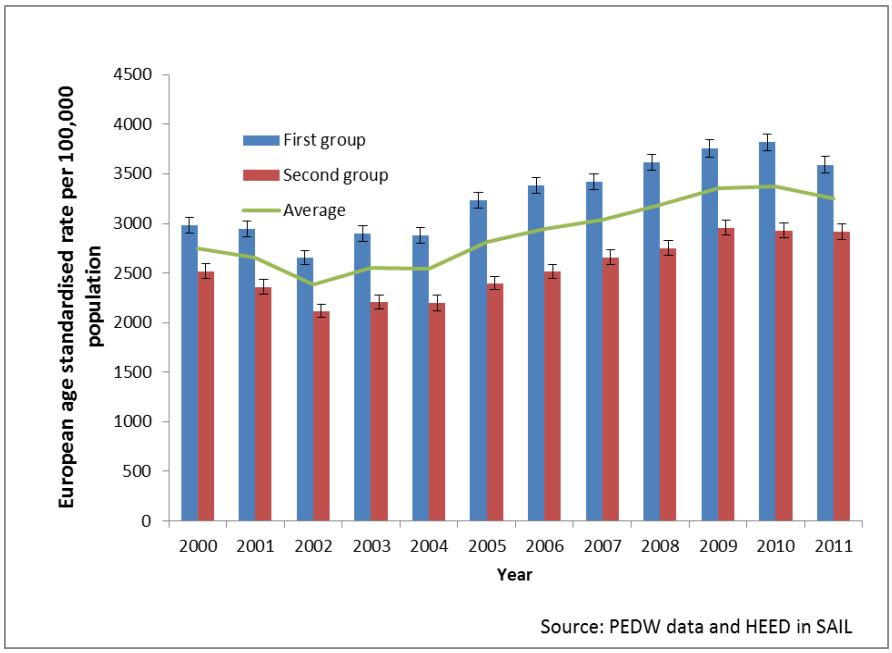
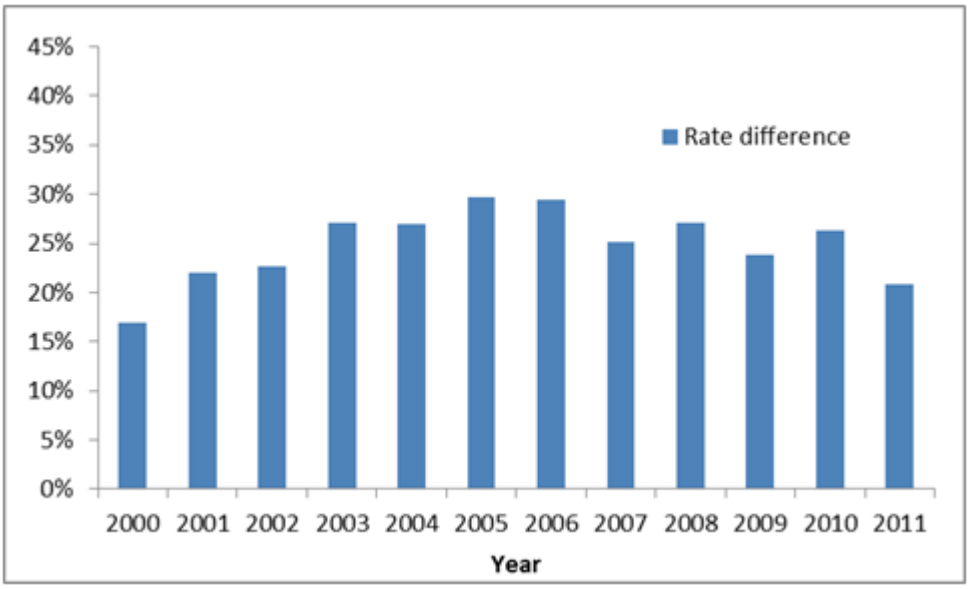


Figure 6.11b: Rate difference expressed as a percentage of the Second Group rate:EASHR for All Respiratory Diseases, 2000-11, by Project Group



6.29 When focusing specifically on asthma admissions, (ICD 10 J45) (see Figures 6.12a and 6.12b, below), a definite convergence in rates occurred from 2006 to 2009. This does suggest an effect of home energy efficiency interventions that becomes observable after five to six years. However, there was a divergence in rates for 2010 - this appears to be due to an increase in Asthma admissions from the First Group rather than a decrease in Rates in the Second.

Figure 6.12a: European Age Standardised Hospitalisation Rate for All Asthma admissions, 2000-11, by Project Group

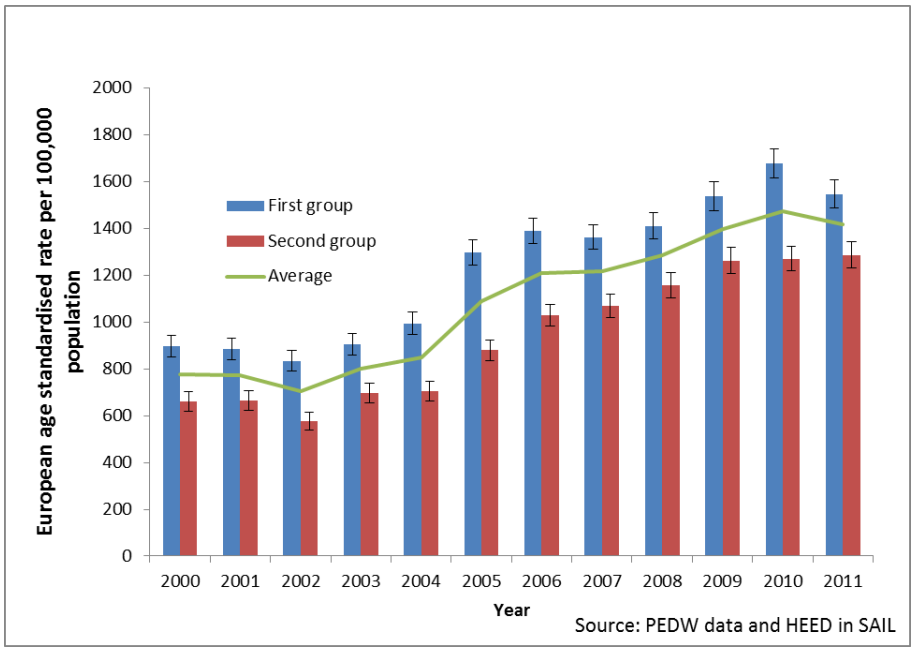
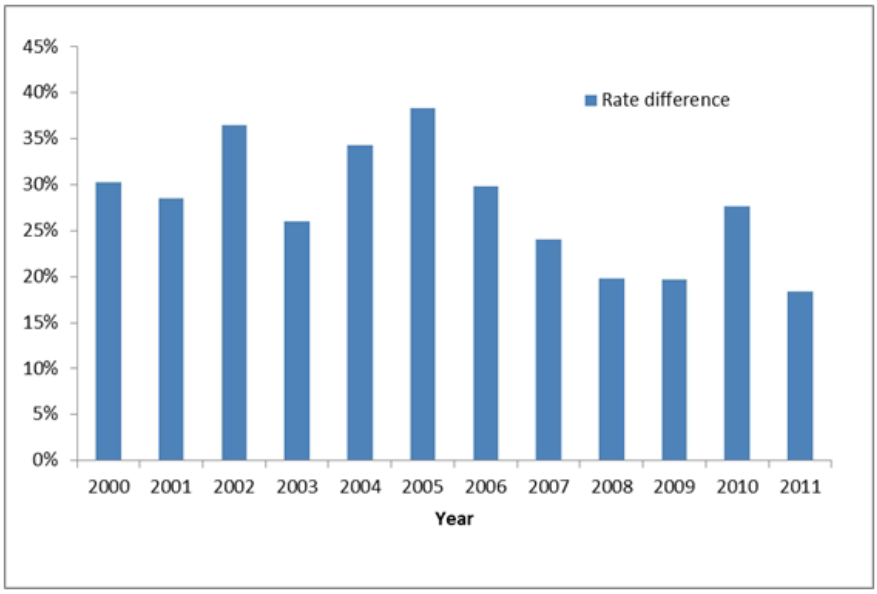


Figure 6.12b: Rate difference expressed as a percentage of the Second Group rate:EASHR for Asthma, 2000-11, by Project Group



Mental health

6.30 In the corresponding analysis for admissions due to mental health problems (ICD 10 F00-F99), shown in Figure 6.13a and 6.13b, below, there is an overall increasing trend in both Groups. The difference in rates suggests an 'excess' reduction after 2003 but because the rates are generally fluctuating, it is difficult to be confident in this pattern and therefore about whether HEED might be having a positive effect.

Figure 6.13a: European Age Standardised Hospitalisation Rate for All Mental Health admissions, 2000-11, by Project Group

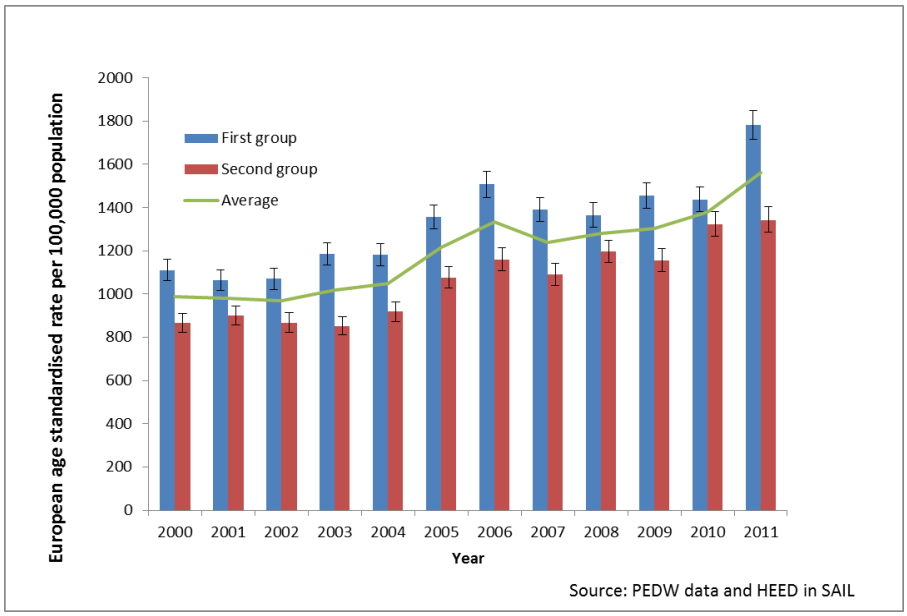
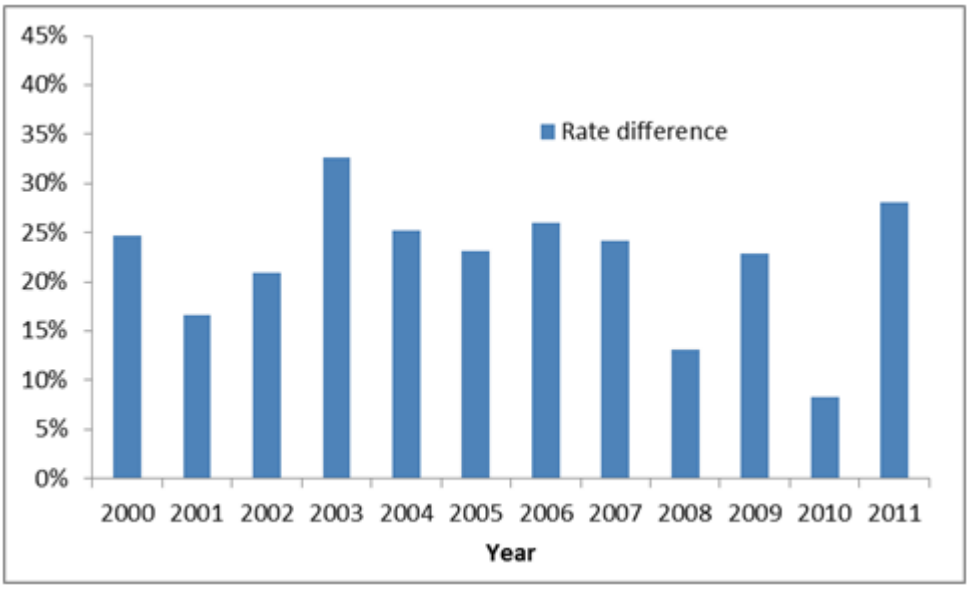


Figure 6.13b: Rate difference expressed as a percentage of the Second Group rate:EASHR for Mental Health, 2000-11, by Project Group



6.31 ‘Mental health disorders’ covers a wide spectrum of different conditions only some of which might be improved by the provision of a warmer home environment. To try and establish a more plausible relationship between the interventions and mental health EASH rates we also calculated for ‘mood disorders’. The results are illustrated in Figures 6.14a and 6.14b, below and suggest, despite some fluctuation, a general trend of convergence after 2004.

Figure 6.14a: European Age Standardised Hospitalisation Rate for All Mood Disorder admissions, 2000-11, by Project Group

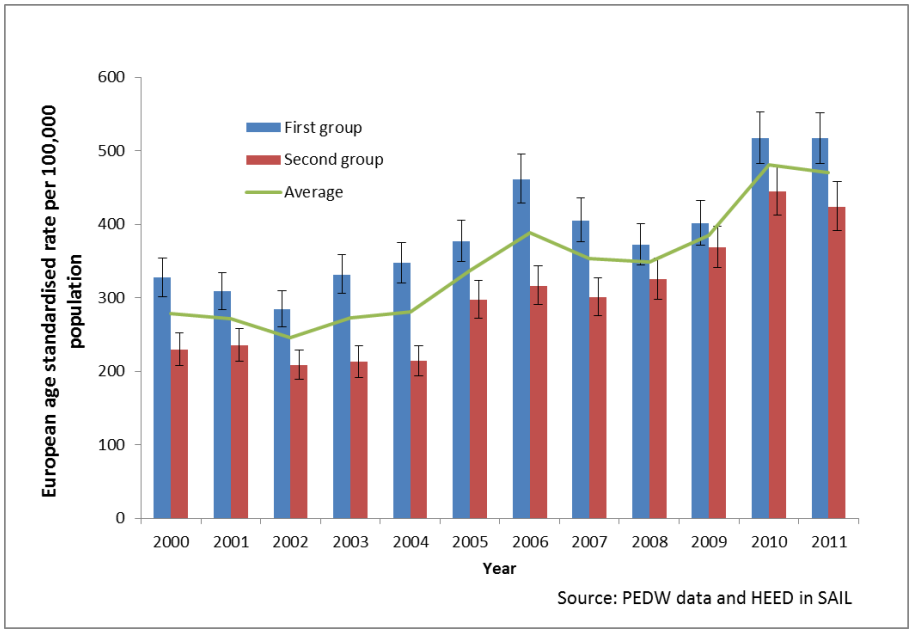
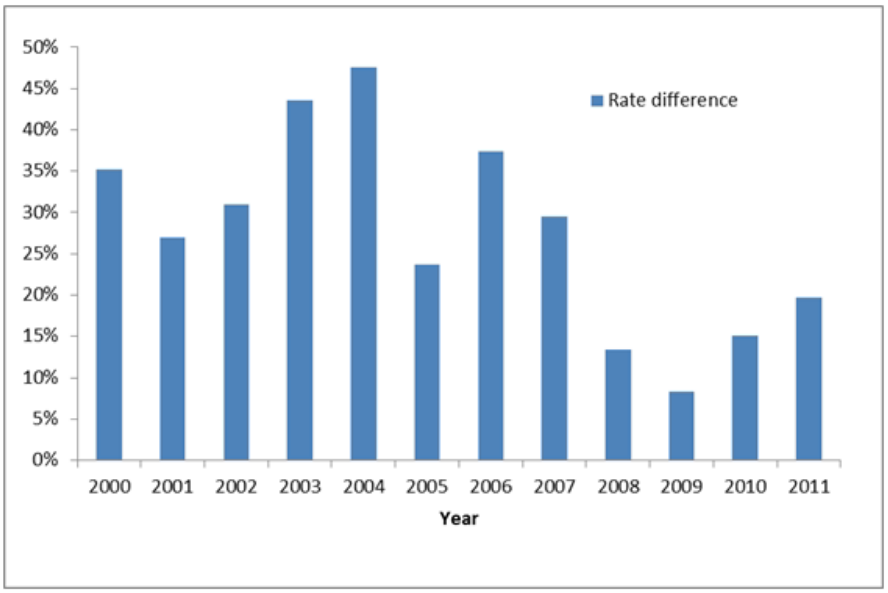


Figure 6.14b: Rate difference expressed as a percentage of the Second Group rate:EASHR for Mood Disorders, 2000-11, by Project Group



Injury and Poisoning

6.32 The diagnosis codes for 'external causes of injuries' (ICD 10 V01-Y98) were used to examine the variation in injury and poisoning based admissions over time for the comparison Groups (Figure 6.15a and 6.15b). The rates suggest a divergence until 2004 followed by a steady convergence in rates until 2010. In 2011 the rates diverged again. This is consistent with an intervention effect occurring in the First Group from 2005 and in the Second Group in 2011. Several more years of data would help to confirm this.

Figure 6.15a European Age Standardised Hospitalisation Rate for Emergency Injury and poisoning admissions, 2000-11, by Project Group

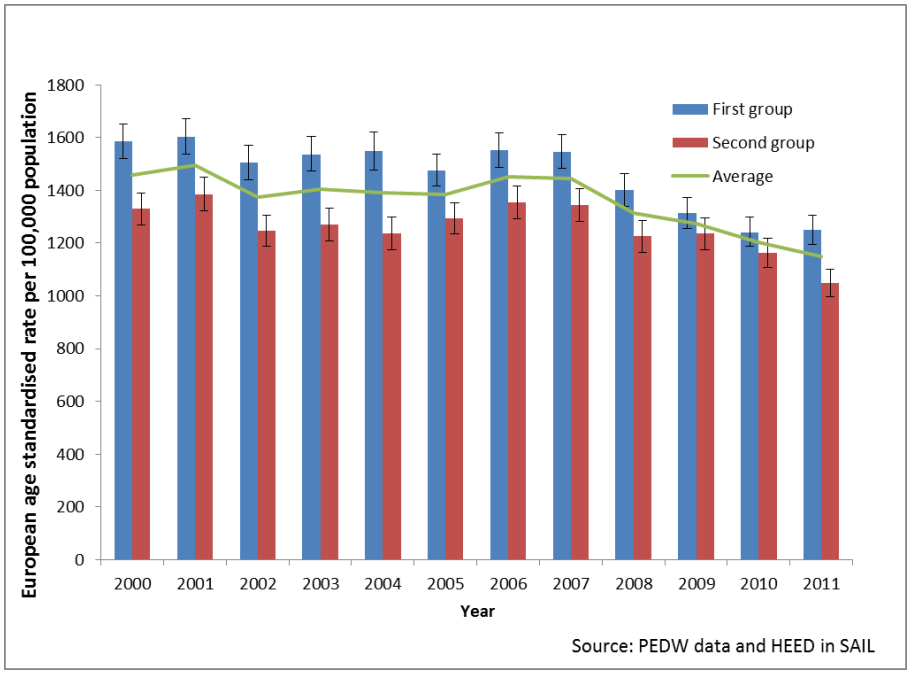
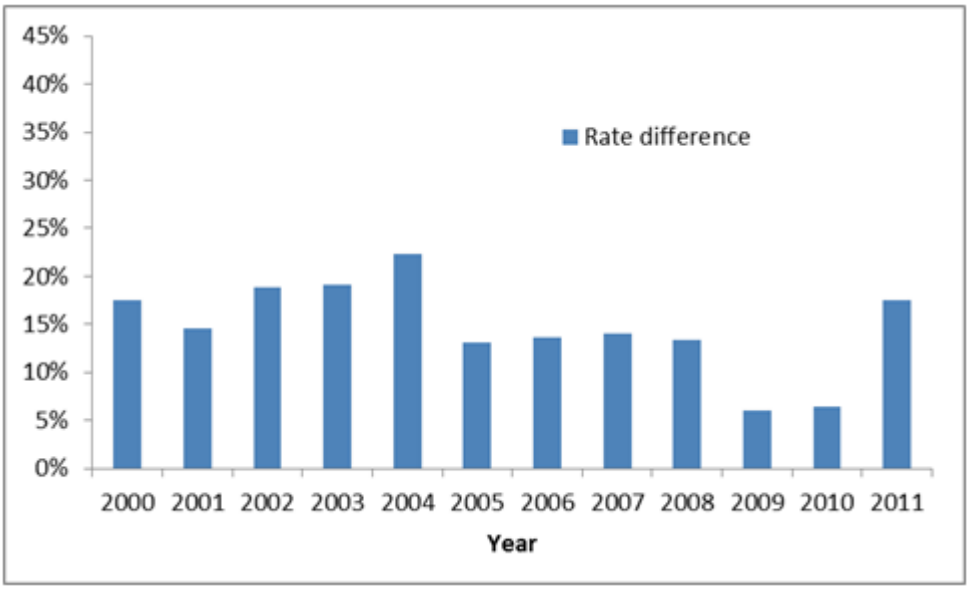


Figure 6.15b Rate difference expressed as a percentage of the Second Group rate:EASHR for Injuries and poisonings, 2000-11, by Project Group



Primary Care Events

6.33 The selection of indicators from the primary care event data is challenging due to the sheer diversity of information available. For the analysis reported in this Section, we have counted the number of dates on which the Group members had prescriptions recorded in order to produce 'prescriptions per head of population'. Figure 6.16a

below shows the comparative rates between the two Groups. The First Group received more prescriptions, and prescribing rates steadily increased over the full time period. The difference in rates shown in Figure 6.16b suggests a diverging rate until 2005 followed by some limited indication of convergence until 2008 and then further divergence, although as noted above, this cannot be tested statistically. This may be consistent with a positive effect being seen in the First Group after four to five years and an effect being seen in the Second Group after two to three years.

Figure 6.16a Prescriptions per head of population 2000-2011

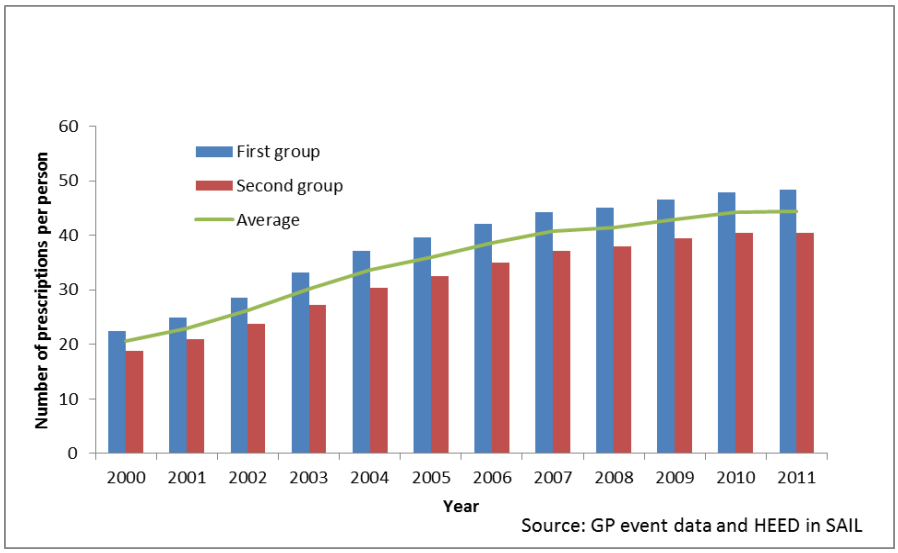
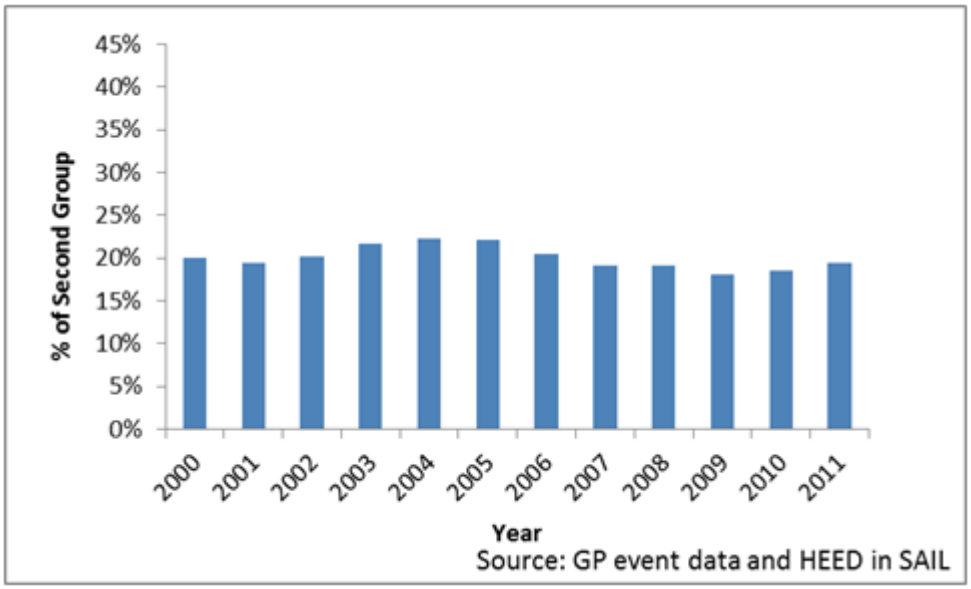


Figure 6.16b Rate difference expressed as a percentage of the Second Group rate:Presciptions per head of population 1999-2011, by Project Group



6.34 Similar analyses were undertaken selecting the populations who were prescribed two specific categories of drugs classified as 'Respiratory drugs' (see Figures 6.17a and 6.17b, below). The rates steadily diverge until 2008 and then there is some limited indication of convergence, although as noted above this cannot be tested statistically; this may suggest a delay of seven to eight years from the beginning of the interventions until a possible effect was observable.

Figure 6.17a Prescriptions for Respiratory conditions per head of population 2000-2011

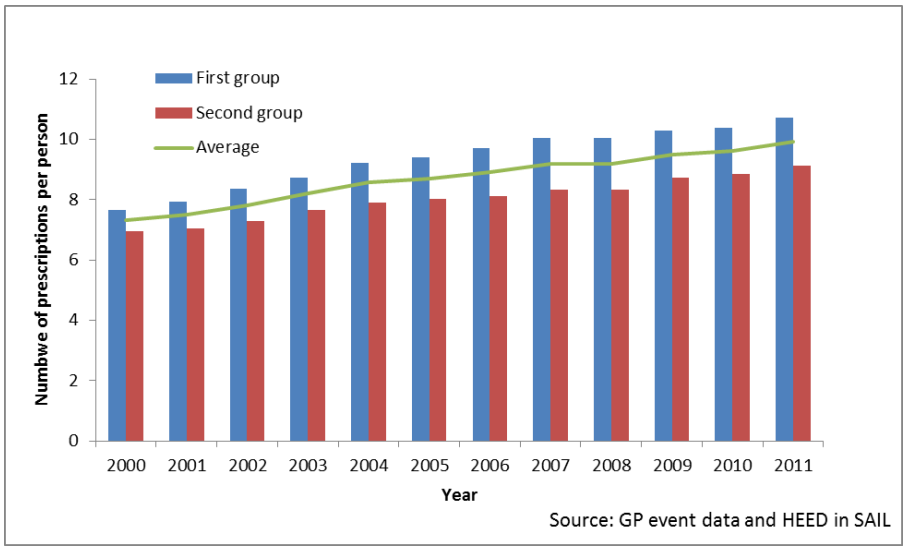
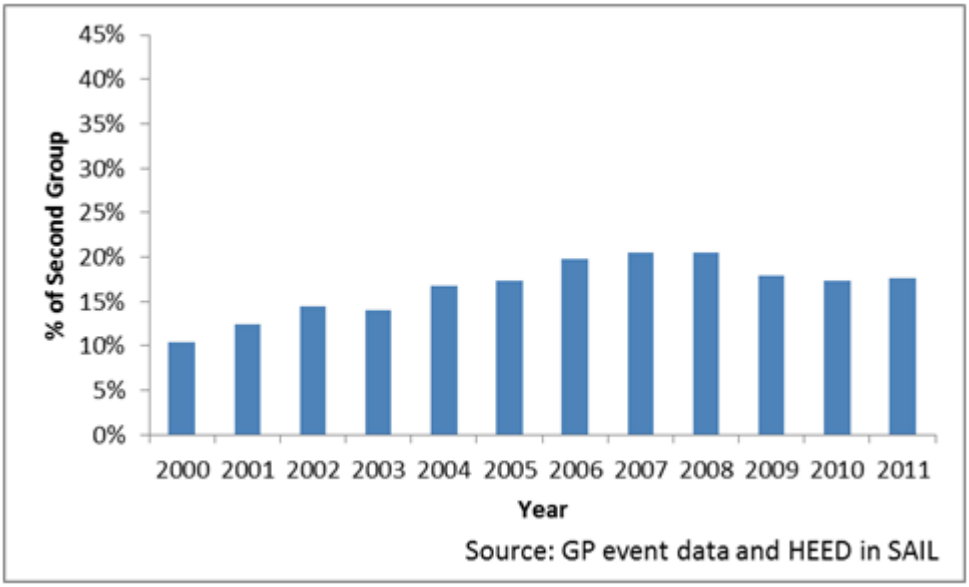


Figure 6.17b Rate difference expressed as a percentage of the Second Group rate: Presciptions for Respiratory conditions per head of population 2000-11, by Project Group



6.35 Prescriptions per head of population were calculated for “Drugs for infections” (Figure 6.18a and 6.18b). This includes antibiotics, steroid inhalers and anti-inflammatory drugs. The rates diverged until 2005, after which there was a relatively small-scale converging trend before they diverged from either 2008 or 2010. The changes observed are small and there is too much fluctuation in the differences between the rates to be confident of a positive effect.

Figure 6.18a Prescriptions per head of population of 'Drugs for infections' 2000-2011.

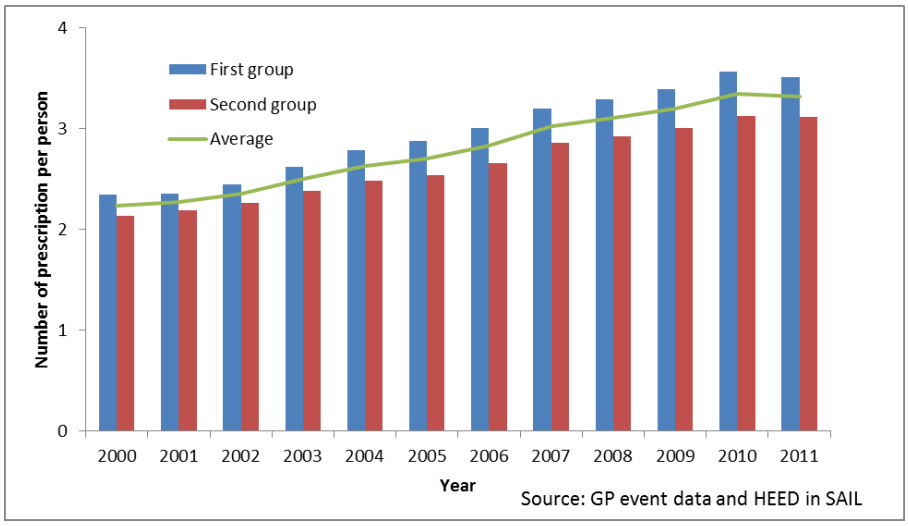
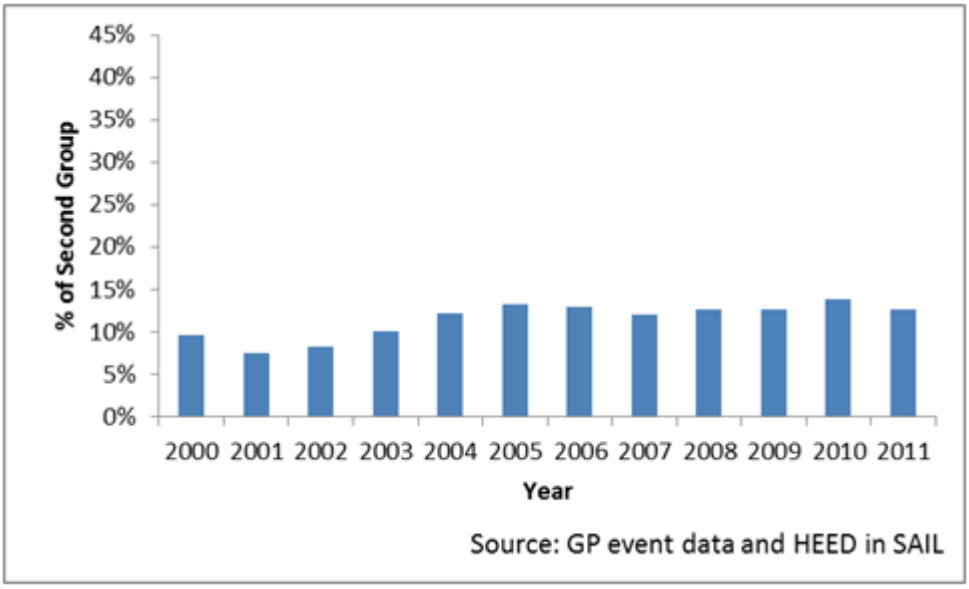


Figure 6.18b Rate difference expressed as a percentage of the Second Group rate:Presciptions for infections per head of population 2000-11, by Project Group



Using GP prescribing to demonstrate an alternative methodology

6.36 In the following analysis, instead of counting events by year, as in the previous chapter the date of intervention is used to determine for each participant in the Project a precise 'before' and 'after' period of time for comparison purposes. Some important considerations are as follows:

- 'After' is always later in time than 'before', so each participant is older in the 'after' period; because death is age related, there will be more losses to mortality in the

'after' period. There will also be increasing levels of age-related illness in the surviving - ageing - population.

- When comparing 'before' and 'after', the reference event e.g. a home energy efficiency intervention, must have happened long enough ago for the 'after' time period both to have happened and to be recorded in the data and for any change in the health condition to take place.
- There may be an underlying trend in the activity or characteristic of interest over time e.g. there has been a steady increase in the prescribing of statins as a preventative measure against stroke and heart disease over the last decade.

6.37 This analysis attempts to answer the question: 'Is there any change in prescribing following the home energy efficiency intervention?' To answer this question, we have compared the number of dates on which intervention Group members had a prescribing event in the five years prior to the first intervention with the number of dates they had a prescribing event in the five years after the first intervention.

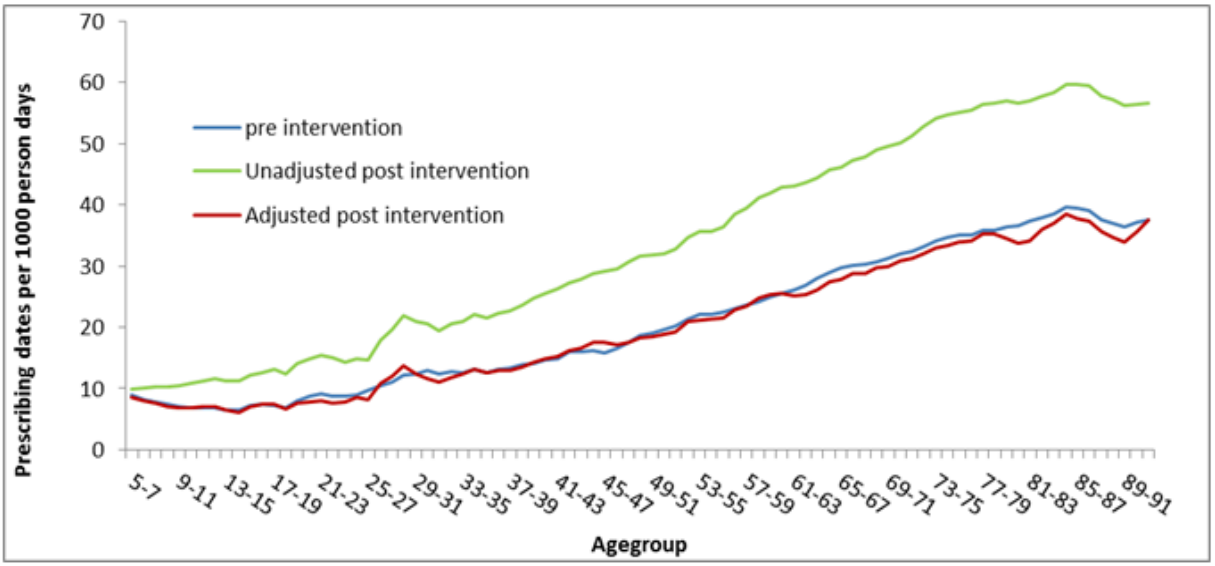
6.38 The total number of person-days was summed for the Group both before and after the interventions. The age of each participant at the time of intervention was used to establish the age-specific number of prescription days per 1000 person-days before and after first intervention. As we noted above, there has been a steady increase in prescription reporting over the Project period, so the rates for all ages were higher following the interventions. The age-specific prescribing rate for the Second Group during the period 2000 – 2005 is the best available indicator of the underlying change in prescribing rate over these years. Assuming that this underlying change would also be applicable to the First Group, it was 'factored out' by simple subtraction.

6.39 Figure 6.19, below, illustrates the results. The blue line indicates the level of prescribing in the First Group prior to interventions, calculated from summing individual level (which) data before the individual intervention date. The green line shows the prescribing rate in the five years following the intervention date including any general underlying increase in rate over time. The red line is the result of adjusting the post-intervention prescribing rate by subtracting the underlying increase (the corresponding rate for the Second Group). This appears to show a slight drop in prescribing, particularly for individuals in the older age groups, following the

intervention. There are a number of further refinements that could be developed for this analysis, which is in need of further testing. A list of possible further work is included in Appendix 7.

6.40 It would be possible to carry out many of the earlier analyses described in chapter 5 in the manner described for this prescribing, but the process needs to be further developed and tested.

Figure 6.19 First Group Prescription Rates (per 1,000 person days) by Age Group for the First Group: before and after home energy efficiency intervention date



.

7 Conclusions

In summary

7.1 The ideal way to identify any improvement that could be attributed to home energy efficiency interventions would be to create two population Groups that differed only in the timing of the housing interventions received. However, the way the schemes recorded in HEED were applied changed over time so, splitting the non-migratory population receiving interventions into two Groups according to when they received their interventions has created two distinctly different population Groups. When compared over a number of health indicators, the First Group consistently displayed poorer health outcomes than the Second. This confirmed the finding of the comparison by deprivation which found the First Group to be a more deprived population than the Second.

7.2 When interpreting any change in the health status of the two Groups over time, we also needed to bear in mind the different levels of resilience the Groups may have had to external factors influencing the outcome indicators e.g. the economy. This added a layer of complexity and uncertainty to the interpretation of the findings, where we could not be sure, for example, whether improvements might affect the two Groups differently.

7.3 Comparison of the two Project Groups on a number of health indicators suggested some positive impacts of the home energy efficiency interventions on health. The data suggests a reduction in excess winter deaths, admissions for circulatory diseases, ischaemic heart disease, stroke, respiratory diseases, mood disorders and injures. The GP Event data also suggests a reduction in some prescribing by GP Practices. No noteworthy effects were identified on mental health problems as a broad category, prescribing for infections or the European Age-Standardised Mortality Rate.

7.4 Further work would be required to assess whether the time delay between time when interventions started to be made and an observable effect becoming apparent, as suggested for various health indicators, is plausible. As noted above, other changes

may be influencing the relative outcomes of the First and Second Groups, i.e. causing similar convergences and divergences. Further research is recommended to examine the extent to which other factors or interventions e.g. Communities First, may have influenced the outcomes. In terms of demonstrating whether the changes could have occurred by chance, as discussed in Chapter 3, it was not possible to use the kinds of statistical testing that would usually be applied but findings where a consistent effect over time was observed were nevertheless considered worthy of note and suggested some association between the HEED interventions so were reported using the phrase 'the data suggests'. **All findings should therefore be viewed with caution and as indicative rather than conclusive**. However, because a number of the findings suggest the same general pattern, taken together they represent a somewhat more conclusive picture.

7.5 Without further work to build on the methods developed for this Project, in particular the method developed for prescribing rates, the results presented here remain speculative and 'question-raising' rather than substantive. A new and more focused question could be formulated from this work, with a Project question: "Can the convergence in mortality and hospital admission rates between the two Project Groups from 2000 onwards be attributed to home energy efficiency interventions in the First Group taking place from 2000-07?" In answering this, we would need to examine all the elements of change going on in much more detail, from the individual causes of hospitalisation and death to the severity of each winter and detailed effects of economic change. Potential further analyses are listed in Appendix 7.

7.6 Looking beyond Fuel Poverty to the contribution the analysis of linked administrative data can make to the evidence base, the Project has demonstrated that data linking can deliver considerable added value. In particular, demonstrating that linked data can be used to:

- Establish retrospective Project populations;
- Create 'control' Groups for comparison purposes;
- Anonymously flag individuals who received interventions, taking advantage of 'natural experiment' scenarios to identify the long term effects of policy interventions.

Reporting in future years

7.7 There is a lot of scope for additional work. As time goes on, data for additional years will become available, allowing comparison of the Groups over a longer time period, which could verify the findings for the First Group by identifying whether they are also seen - and to similar time scales - in the Second Group.

7.8 WG and other organisations have plans to anonymise datasets for additional topics into SAIL in the coming years, which will provide the ability to monitor a wider range of outcomes. The national recording of heights and weights in children and the data from the stop smoking service are two examples.

7.9 The SQL coding has been fully documented and will be available to SAIL users[16]. This will be modified to utilise a new automated process for determining continuous periods of residence from WDS data which has now been developed for general use in SAIL. One line of SQL code calling this procedure will replace several hundred lines of SQL code currently in place. Appendix 8 includes the location of the secure project area in the SAIL gateway, where all the current work is stored. For further information, please use the contact information at the beginning of the Report. As noted in Chapter One, the challenges that emerged during the demonstration process will be explored in more detail in a Lessons Learned report, publication of which is to follow.

[16] Within the SAIL Gateway working environment, analysts post completed pieces of code to a 'Wiki'-style bulletin board, where a library of concepts is being developed. When reusable code is developed (e.g. to select cases with a specific condition from complex primary care codes), this is made available to other researchers for peer review, and to eventually develop standard methodologies for data manipulation and case selection. These will be shared across International data linkage forums to help develop encourage global standardisation.

Appendix 1: The Data sets used in this project

The PEDW data structure

The Patient Episode Database for Wales (PEDW) is the national repository for in patient and day case data. NHS Wales Trusts are required to download, on a monthly basis, very clearly defined and standardised data from all hospital Patient Administration Systems (PAS). These are collated by the NHS Wales Informatics Service (NWIS) in Cardiff, who also receive details of Welsh patients treated in England through a mechanism known as the NHS switching service, and provide details of English patients treated in Wales to Hospital Episode Statistics (HES), the equivalent system in England.

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. In PEDW, the ICD 10 diagnostic codes are utilised. So, for example, first episodes of care containing a diagnosis in the range I00-I99 relate to episodes of cardiovascular diseases. The example case study provided below illustrates how the data is recorded and coded.

Mrs Smart's Stroke – a fictitious case study but based on realistic PEDW data

Mrs Smart was admitted on 22/10/2009 as an emergency at the request of her GP. She was 79 years old and her birthday was the following day. Her admission was coded under Specialty 'Geriatric Medicine', and she remained in hospital until the 27/10/2009. She was first treated by Consultant A, a 'General Medicine' specialist. Six different diagnostic codes were later coded from Mr A's notes, indicating that she was admitted with a Cerebral Infarction (stroke) but also suffering from Hypertension, Non-insulin dependent diabetes, a history of diseases of the nervous system and a history of diseases of the circulatory system, and that she had a dependent relative needing care at home. This treatment took less than 1 day and the episode length was recorded as 0 days. One operative procedure was recorded at this time (an examination procedure).

Mrs Smart was then moved to a high dependency bed managed by the team of Consultant B, a geriatric specialist for 1 day. She was then moved to a more general ward for the next four days, looked after by Consultant C who is also a geriatric specialist. She left hospital on 27/10/2009, aged 80. She was discharged to her own home.

Each time she was under the care of a different consultant is called a 'finished consultant episode'. Mrs Smart had 3 episodes. The time she was continually in hospital is known as a 'spell' of care. Mrs Smart just received one, 3-episode, spell of care lasting 5 days.

All this information was either recorded as a code or a date, or could be calculated from the other data. E.g. age can be calculated based on the date of birth and the admission or discharge date. So if calculated on admission date she was 79, but if calculated on discharge date she would be 80, and if being grouped into five year age bands this choice would place her in a different age group.

Welsh demographic service data

The Welsh Demographic Service is a dataset 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. The example case study provided below illustrates how the data is recorded and coded.

Mr Smith's GP registration history - a fictitious case study based on real WDS data

Mr Smith was born on or around 28.11.1927 but we don't know where. He was registered with a GP practice in Wales on 29.06.1974 at the age of 46 but the first address registration date recorded for him was 03.03.1987, thirteen years later, when he was 59. The next house move recorded (chronologically) says he moved house on 05.01.1993 but another record indicates he moved back into the same house on the same date. He moved house again on 06.12.2002 and this time it was to a different location. He lived there till 05.01.2012, which is the last known address for him.

Mr Smith died on 13.08.2012. This was also the date recorded as the end of his last registration with a GP. During his life in Wales he was registered with one GP practice, but there are three sets of end dates for that registration, followed by three sets of registration dates with the same practice, none of which coincide with his house moves. These may be due to a change of GP in the practice, but the anonymised data does not hold this detail.

Note that there are two spells of time when Mr Smith was registered with a GP in Wales but for which we do not have his address. The first was for 13 years at the beginning of his registration in Wales and the last was a few months at the end of his life, when perhaps he was living in some kind of care facility.

GP Event Data

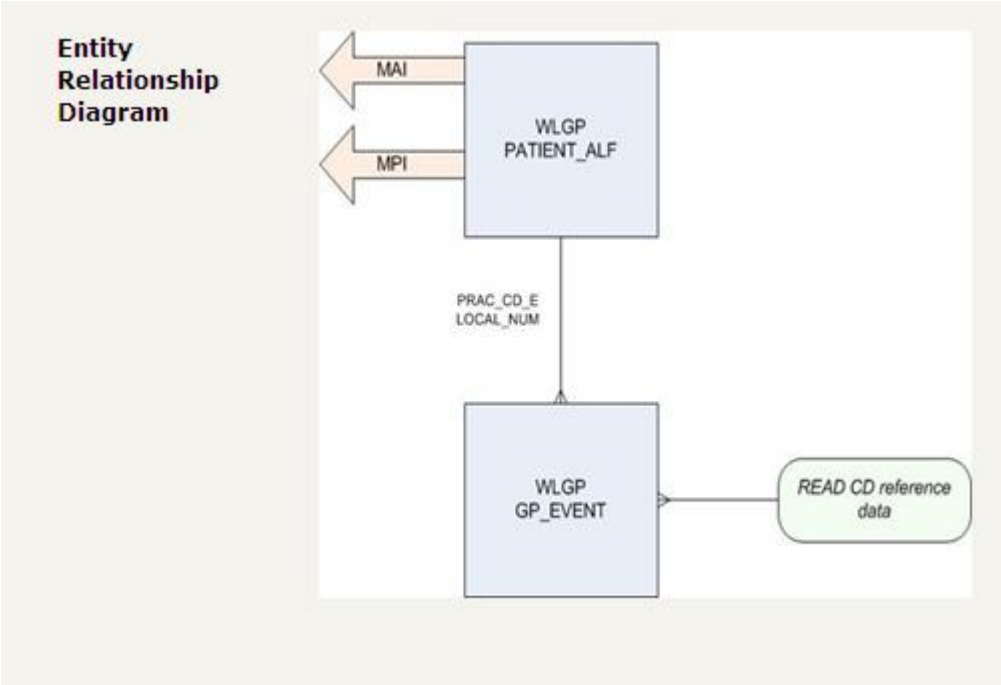
This is data extracted from all Welsh General Practices that have signed up to SAIL. The data is from the clinical information system the practice uses to maintain an electronic health record for each of their patients - capturing the signs, symptoms, test results, diagnoses, prescribed treatment, referrals for specialist treatment and social aspects relating to the patients home

environment. 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 47% of the Welsh population is included in this dataset.

There are no standard rules for recording data within primary care clinical information systems. Therefore, each individual clinician can record information in their own way. The majority use Read Code Terminology, however, sometimes this is applied behind the scenes by the clinical system and sometimes local codes are used. Read codes are not as precise as ICD 10 or OPCS codes. Coding standards have been agreed on for conditions monitored by the QOF (Quality Outcomes Framework) returns. Since the implementation of QOF these conditions have been coded in a more consistent way.

The data format is simple, as presented in the Entity relationship diagram in Figure A.1. Essentially, each item of recorded information is stored in a single row in GP_EVENT, with a Read Code for the item, an event date (when the event occurred) and an optional corresponding event value. So if the event was that the patient's blood pressure was recorded there will be a code for this event, a date when it occurred and possibly a value entered indicating what the blood pressure reading was. The example case study provided below illustrates how the data is recorded and coded.

Figure A.1: Entity relationship diagram for GP Event Data[17]



A small sample case study of GP events for a single patient is shown below.

[17] Entity relationship modelling is a Software Engineering concept, providing an abstract way of describing a database. SAIL comprises of a set of relational databases, i.e. data is logically separated into separate but related tables for efficiency of storage and speed of data manipulation. Some of the data in these tables 'point to' data in other tables – so, for example, a 'person' in the WDS database could point to several entries for each of the 'addresses' they have lived in. For the purposes of the entity relationship diagrams shown below, for example, each 'person' is an entity and each 'address' is an entity and the relationship between the 'person' and the 'addresses' would be 'has an address'. Diagrams created to represent these entities and relationships are called entity–relationship diagrams or ER diagrams. The SAIL entity relationship diagrams provided below all refer to the 'MAI'. The MAI is the Master ALF Index - all the 'person'-based databases in SAIL have a relationship with the 'Master Database' relating to the People of Wales.

Tom Browns knee: fictitious data based on real recording in SAIL

Tom Brown has five records in the GP event data between 4/09/2002 and 9/09/2002. The first two indicate that he had a “complete knee replacement using cement” and “manipulation of the knee”, both on 4th September. The next three records indicate that on the 8th September he was prescribed pain killers, and two ‘breathe easy ‘aerosol inhalers. Then the final record on the 9th indicates he had an influenza vaccination.

This illustrates several features of the primary care database

- Many records can be generated on the same day, one record per ‘event’.
- Records do not necessarily originate in the General Practice, as with the knee replacement, this would have arrived here via a discharge letter from the hospital that carried out the procedure.
- Manipulation of the knee may not be related to the same knee that was operated on. It may have been a check on Tom’s other knee. In general it is not always possible to assume that information recorded on the same date refers to exactly the same condition.
- Records generated around the same time are not always related. Tom was being treated for the flu perhaps. Whether the painkillers for the recovering knee or the flu is unknown, and the Influenza Vaccination on the 9th September may have been scheduled prior to the prescribing events of the 8th September so are coincidentally recorded next to each other chronologically in the database.

When the data are chronologically presented, the total history of Tom’s knee is not neatly captured in one place – a further search revealed he first complained of knee pain in April 1999.

HEED data

Table A.1: HEED data fields

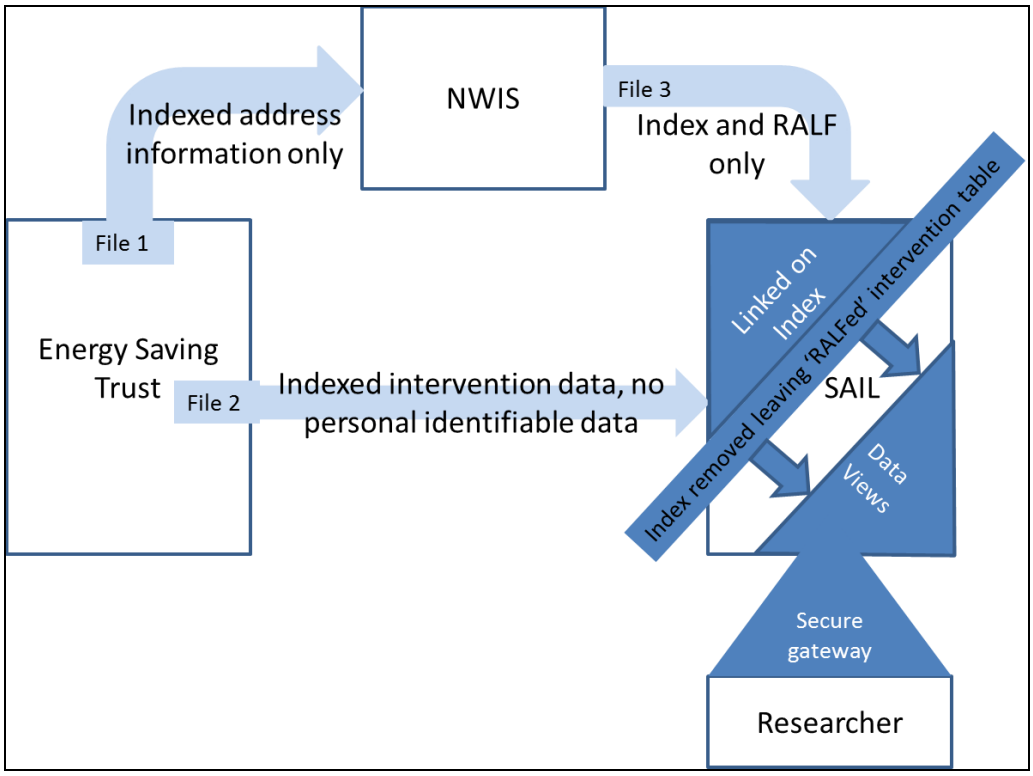
Field name	Description	Explanation
HOME_ID_E	Ten digit link to RALF file	Unique identifier for a home
YEAR_MONTH	YYYYMM	Year and month of intervention
DETAIL_DATE	DD/MM/YYYY	Date intervention was carried out
MEASURE_GROUP_CODE	Measure Group code, a 2 digit number	Code e.g. 63 indicates insulation
MEASURE_GROUP_NAME	Measure Group name, text	Descriptive label e.g. "Insulation"
MEASURE_CATEGORY_CODE	Measure Category Code, a 3 digit number	Code e.g. 259 indicates "Loft insulation measures"
MEASURE_CATEGORY_NAME	Measure Category name, text	Descriptive label e.g. "Loft Insulation Measures"
MEASURE_CODE	Measure code, a 4 digit number	Code e.g.7947 indicates "Loft insulation 100-270mm"
MEASURE_NAME	Measure name, text	Descriptive label e.g. "Loft insulation 100-270mm"
MEASURE_COUNT	Measure count, numeric	Number of measures implemented (usually 1)
AVAIL_FROM_DT	System date the table was created in SAIL	A date generated during the SAIL loading process

Source: HEED data in SAIL

- The “Home ID E” field is a link to the RALF, the “year month” field is the year and month of the intervention and the detail date is where the date of the intervention was recorded. The fields beginning ‘MEASURE’ are a categorisation of the possible interventions into: Heating, Insulation, Microgeneration and ‘Other’ groups. A group and category breakdown is shown in Table A.1, with a detailed table by measure provided in Appendix 5.

Data anonymisation

Figure A.2An Illustration of the Split File Process



Source: SAIL databank

Appendix 2: Data Sharing Agreement EST and HIRU

Energy Saving
Trust
21 Dartmouth Street
London SW1H 9BP

Tel 020 7222 0101

Martin Heaven

Health Information Research Unit (HIRU)

Swansea University

26th June 2012

Dear Martin

You have requested that EST provide data that it holds in HEED relating to the Home Energy Efficiency Scheme (HEES) in Wales for inclusion in HIRU's Secure Anonymised Information Linkage (SAIL) system. This request is supported by the Welsh Assembly Government.

However, prior to permitting you to receive such data, and in consideration of EST providing you with such data, you agree to the following:

1) You shall **only** use the data received from HEED for inclusion in SAIL and research projects utilising the SAIL environment. You must **not** use HEED data to target or market to individual addresses.

2) You must not cross reference any HEED data with any other data in such a way that data provided from HEED contravenes EST's obligations under the Data Protection Act 1998;
   -

3) The use of HEED data must adhere to the protocols and procedures laid out in HIRU's Data Anonymisation Policy and Process (DAPP)

4) You shall not divulge any part of the data received from HEED to any person without EST's prior written consent save that you may divulge such data to persons operating within SAIL and/or HIRU DAPP guidelines in order that they may carry out the permitted uses described in 1.

5) You shall indemnify and keep indemnified EST and any third party that has supplied data to EST that has been included in HEED ("the Indemnified Party") from and against all costs and expenses (including legal costs), claims, damages, demands, liabilities and losses suffered or incurred by the Indemnified Party arising out of your breach of the terms contained in this letter and/or in connection with any claim from any third party that arises out of your act or omission, except to the extent that such costs, expenses or third party claim arises out of or were contributed to by any negligence of the Indemnified Party.

-

6) You acknowledge and agree that any third party that has supplied data to EST that has been included in HEED shall have the right to enforce the terms contained in this letter, under the Contract (Rights of Third Parties) Act 1999.

7) You acknowledge and agree that EST may cease to provide any form of information from HEED forthwith upon any breach of the terms of this letter, and, at the option of EST, forthwith return or destroy any data and/or information in your possession that you have previously received from HEED.

8) While EST will use reasonable endeavours to provide accurate data and reports, to the extent permitted by law, the Energy Saving Trust expressly excludes
   a) any liability for any direct, indirect or consequential loss or damage incurred by any use in connection with HEED Online reporting portals, HEED reports or analysis or HEED data provided in any form, or in connection with the use, inability to use, or the results of the use of HEED data in any form, including without limitation any liability for loss of income or revenue, loss of business, loss of profits or contracts, loss of anticipated savings, loss of data, loss of goodwill, wasted management time and for any other loss of damage of any kind, however arising and whether caused by tort (including negligence), breach of contract or otherwise.
   b) all conditions, warranties and other terms which might otherwise be implied by statute, common law or the law of equity.

Signed

I agree to the terms of this letter and am duly authorised to sign on behalf of Health Information Research Unit (HIRU), Swansea University

Signature Martin Heaven .....

Date: 26/6/2012

Contact Details:
Martin Heaven
Senior Research Analyst
Health Information Research Unit (HIRU)
Centre for Health Information Research and Evaluation (CHIRAL)
College of Medicine
Swansea University
Institute of Life Science 2
Singleton Park

Appendix 3: Membership of IGRP

Membership of the Information Governance Review Panel as at 10/06/2013

The IGRP provides independent advice on Information Governance and reviews all proposals to use SAIL data to ensure that they are appropriate and in the public interest. The current panel is as follows:

Organisation	**Name**
British Medical Association	Dr Tony Calland
National Research Ethics Service	Corrine Scott
Public Health Wales	Dr Judith Greenacre
NHS Wales Informatics Service	Martin Murphy Darren Lloyd
SAIL Consumer Panel	Dr Neil McKenzie Dot Williams

Appendix 4: HIRU Application form (IGRP application)

chiral Centre for Health Information, Research and Evaluation
Canolfan Iechyd Hysbysrwydd, Archwiliad ac Ymchwil

For HIRU use
Ref. No.

Centre for Health Information Research and Evaluation (CHIRAL)

College of Medicine

Swansea University

Health Information Research Unit (HIRU)

HIRU Enquiry form

Template review chronology

Version no.	Effective date	Reason for change
1.0	29/11/07	N/A
2.0	1/5/08	Establishment of CRS necessitating changes to content and layout
3.0	14/10/09	Recommendations of IGRP
3.1	05/04/11	Annual review

The following form has been designed to collect the information needed from individuals and organisations interested in collaborating with HIRU on work involving the SAIL databank. The information you provide will facilitate consideration of your enquiry. Please complete sections and A & B and provide additional documents as requested.

SECTION A

1a. Contact details of project lead:

Name:

Job title:

Organisation:

Address:

Tel:

Fax:

Email:

1b. The project lead will be the only person accessing the data:

Yes [☐] No [☒]

1c. Please provide contact details of the person(s) who will be accessing the data (apart from the project lead):

Name:

Job title:

Organisation:

Address:

Tel:

Fax:

Email:

2. Does your proposed work with HIRU constitute:

Part of a larger project?

If yes, please complete all questions

The entire project?

If yes, please complete all questions except 3a, 5a and 7a

3a. Full title of the *main* project:

3b. Full title of the (part of the) project involving HIRU (if different):

4a. Who is commissioning the project (if relevant)**?**

4b. Why is the project being done?

5a. Aim of the *main* project, including anticipated outcomes:

5b. Aim of the (part of the) project involving HIRU, including anticipated outcomes (if different):

Please include a copy of the protocol/plan for the proposed work with HIRU, including the contact details of any co-applicants when you return your completed form.

6. Lay summary of the project involving HIRU: (approximately 150 words)

7. Please list the relevant permissions you have obtained or that are being sought:

	Obtained	*Being sought*	*Not required*
Research ethics	[☐]	[☐]	[☒]
Independent peer review	[☐]	[☐]	[☒]
Permission from data-holding organisation to use their datasets	[☐]	[☐]	[☒]

Please state the name of the organisation/committee that is being applied to, or that has given approval, as applicable:

Research ethics:

Peer review:

Data organisation permission:

If you have ticked 'not required' please specify the reasons.

Please note that it is the responsibility of the project lead to ensure that the relevant permissions are obtained.

8a. At what stage is the main project?

Protocol/plan being developed [☒]

Protocol/plan in place but project not commenced [☐]

Project underway [☐]

If underway, what was the start date of the main project (dd/mm/yy)?

8b. Please indicate a prospective start date for the (part of the) project involving HIRU:

(dd/mm/yy)

8c. Over what period do you anticipate you will require the assistance of HIRU?

Start and end dates in dd/mm/yy: [ ] to [ ]

9a. What data do you require for the proposed work with HIRU?

Please list:

The datasets you require information from

The types of variable you need

The datasets that will need to be linked

9b. Will you also be providing other datasets to be incorporated into the SAIL databank?

Yes [☒] No [☐]

If yes, please specify:

9c. Please provide an outline of your analysis plan including the anticipated outcomes

9d. Are the results/methods developed likely to have other potential applications?

Yes [☒] No [☐]

If yes, please specify:

10a. Please indicate your plans for publishing the results of your project, e.g. target journal or intended recipients of report:

10b. What are the potentially sensitive issues that need to be taken into account when publicising the findings of the project?

Please outline the issues and your proposed solutions:

Appendix 5: HEED Data items detailed table.

Measure group	Measure category name	Measure name	Homes benefiting
Heating	Condensing Boilers	Condensing Boiler	10089
	Condensing Boilers	Non-Condensing Boiler	2126
	Condensing Boilers	Condensing Boiler (Intelligent Controls)	643
	Condensing Boilers	Condensing Boiler (Delayed Start Thermostat)	598
	Control Measures	Thermostatic Radiator Valves	129
	Control Measures	Heating controls upgrade: package C/D to F**	11
	Fuel Switching	Fuel Switch - Electric to Gas	5548
	Fuel Switching	Fuel Switch - Solid to Gas	734
	Fuel Switching	Fuel Switch - Electric to oil	490
	Fuel Switching	Fuel Switching - Unknown To Gas	369
	Fuel Switching	Fuel Switch - Coal to Oil	238
	Fuel Switching	Fuel Switch - Partial Coal to Gas	103
	Fuel Switching	Fuel Switch - Oil to Gas	88
	Fuel Switching	Fuel Switching - Solid to Electric	61
	Fuel Switching	Fuel Switch - Partial Electric to Gas	37
	Fuel Switching	Fuel Switching - Gas to Electric	21
	Fuel Switching	Fuel Switching - Partial Coal to LPG	13
	Fuel Switching	Fuel Switching - Coal to LPG	12
	Fuel Switching	Fuel Switch - Partial Coal to Oil	12
	Fuel Switching	Fuel Switching - Electric to LPG	10
	Fuel Switching	Fuel Switching - Gas to Oil	<5
	Heating Measures	Replacement GAS/LPG Boiler	10905
	Heating Measures	Replacement GAS/LPG Condensing Boiler	10136
	Heating Measures	Replacement Oil Boiler	728
	Heating Measures	Replacement Storage Heaters	649
	Heating Measures	Replacement Oil Condensing Boiler	141
	Heating Measures	Replacement Gas Room Heaters	20
	Heating Measures	Replacement Solid Fuel Room Heater	13
	Solid Fire Conversion Cassette	Replacement Solid Fuel Fire Cassette	15
Insulation	Cavity Wall Insulation	Cavity Wall Insulation (pre 1976)	142803
	Cavity Wall Insulation	Cavity Wall Insulation (post 1976)	32872
	Cavity Wall Insulation	Cavity Wall Insulation (Unknown Property Age)	20528
	Draught Proofing Measures	Draught Proofing (General)	13317

	Hot Water Tank Insulation	Hot Water Tank Insulation	12930
	Hot Water Tank Insulation	Properties with HWT Insulation	8
	Hot Water Tank Insulation	Unknown	<5
	Loft Insulation Measures	Loft Insulation 0 - 250mm	66775
	Loft Insulation Measures	Loft Insulation 50 - 250mm	53369
	Loft Insulation Measures	Loft Insulation 0-270mm	35386
	Loft Insulation Measures	Loft Insulation 100 - 250mm	34048
	Loft Insulation Measures	Loft Insulation 100-270mm	14041
	Loft Insulation Measures	Loft Insulation 25 - 250mm	12794
	Loft Insulation Measures	Loft Insulation 25-270mm	12341
	Loft Insulation Measures	Loft Insulation 75 - 250mm	9479
	Loft Insulation Measures	Loft Insulation 75-270mm	8923
	Loft Insulation Measures	Loft Insulation 50-270mm	8853
	Loft Insulation Measures	Loft Insulation 50 - 300mm	1873
	Loft Insulation Measures	Loft Insulation 150 - 250mm	1459
	Loft Insulation Measures	Loft Insulation 100 - 300mm	1257
	Loft Insulation Measures	Loft Insulation 25 - 200mm	497
	Loft Insulation Measures	Loft Insulation 150-270mm	28
	Loft Insulation Measures	Loft Insulation 75 - 200mm	27
	Loft Insulation Measures	Loft Insulation 250 - 300mm	23
	Loft Insulation Measures	Loft Insulation 0 - 200mm	19
	Loft Insulation Measures	Loft Insulation 100 - 200mm	19
	Loft Insulation Measures	Loft Insulation 200 - 300mm	15
	Loft Insulation Measures	Loft Insulation 50 - 200mm	6
	Solid Wall Insulation	External Wall Insulation to U = 0.45	1491
	Solid Wall Insulation	External Wall Insulation to U = 0.37	1208
Microgeneration	Air Source Heat Pump	Air Source Heat Pump	78
	Ground Source Heat Pumps	Ground Source Heat Pump	<5
	Solar Heating Measures	Solar Water Heating (Unknown)	7
	Solar PV Panels	Solar PV Panels	1047
	Solar Water Heating System	Solar Water Heating	610
Other	Real Time Display Measures	RTD Long Lifetime	52565

***Assumes TRVS already installed and room thermostat is replaced with intelligent heating controls*

Appendix 6: Population of groups by age and gender

First Group (Male and Female combined) population by age group by year, (births into this population have not been included).

AGEGROUP	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
00-01	2,662	1,165	-	-	-	-	-	-	-	-	-	-	-
01-04	5,412	5,019	4,404	2,654	1,163	-	-	-	-	-	-	-	-
05-09	10,042	9,961	9,776	9,487	8,975	8,055	6,175	4,399	2,652	1,161	-	-	-
10-14	8,329	8,892	9,286	9,699	9,881	10,033	9,948	9,765	9,475	8,964	8,044	6,166	4,390
15-19	5,107	5,556	6,209	6,827	7,610	8,298	8,855	9,244	9,658	9,839	9,997	9,916	9,744
20-24	4,266	4,292	4,303	4,493	4,747	5,040	5,495	6,143	6,750	7,525	8,217	8,784	9,179
25-29	6,889	6,046	5,387	4,795	4,360	4,181	4,220	4,228	4,428	4,681	4,975	5,427	6,082
30-34	10,431	9,842	9,137	8,361	7,554	6,777	5,941	5,298	4,708	4,293	4,103	4,145	4,142
35-39	12,656	12,308	11,839	11,449	10,932	10,267	9,696	8,997	8,240	7,431	6,666	5,828	5,191
40-34	13,269	13,199	13,007	12,818	12,557	12,466	12,103	11,618	11,219	10,706	10,056	9,492	8,828
45-49	15,358	14,505	14,045	13,669	13,390	12,993	12,939	12,747	12,542	12,278	12,189	11,824	11,340
50-54	20,894	20,145	18,646	16,972	15,784	14,971	14,109	13,639	13,260	12,956	12,584	12,523	12,357
55-59	23,752	22,923	22,855	22,300	21,644	20,196	19,441	17,944	16,286	15,113	14,311	13,478	13,019
60-64	24,297	24,029	23,591	23,397	22,861	22,654	21,783	21,662	21,064	20,390	18,994	18,269	16,842
65-69	23,177	22,922	22,631	22,566	22,590	22,498	22,252	21,830	21,629	21,024	20,789	19,918	19,879
70-74	21,631	21,306	20,995	20,903	20,680	20,640	20,413	20,183	20,038	19,905	19,761	19,574	19,148
75-79	18,463	18,274	18,208	18,062	17,929	17,720	17,547	17,255	17,222	17,087	17,044	16,783	16,535
80-84	9,944	11,078	11,886	12,685	13,121	13,456	13,320	13,349	13,202	13,016	12,813	12,632	12,489
85&+	7,583	7,968	8,348	8,392	8,561	8,879	9,616	10,250	10,665	10,731	10,946	11,060	11,133
All ages	244,162	239,430	234,553	229,529	224,339	219,124	213,853	208,551	203,038	197,100	191,489	185,819	180,298

Second Group (Male and Female combined) population by age group by year, (births into this population have not been included).

AGEGROUP	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
00-01	2,369	998	-	-	-	-	-	-	-	-	-	-	-
01-04	5,135	4,642	3,899	2,362	996	-	-	-	-	-	-	-	-
05-09	9,126	9,210	9,159	8,889	8,406	7,485	5,626	3,893	2,360	995	-	-	-
10-14	7,915	8,295	8,567	8,723	8,849	9,113	9,200	9,147	8,875	8,396	7,478	5,620	3,887
15-19	5,465	5,895	6,434	6,996	7,589	7,885	8,264	8,536	8,693	8,817	9,091	9,181	9,124
20-24	4,079	4,135	4,257	4,603	4,921	5,416	5,838	6,373	6,928	7,520	7,821	8,206	8,485
25-29	6,111	5,472	4,833	4,353	4,101	4,014	4,076	4,207	4,544	4,853	5,334	5,773	6,305
30-34	9,774	8,951	8,311	7,525	6,736	6,015	5,396	4,755	4,269	4,023	3,939	4,003	4,131
35-39	12,841	12,385	11,850	11,246	10,467	9,653	8,830	8,201	7,397	6,612	5,906	5,283	4,647
40-34	14,741	14,367	13,779	13,265	12,990	12,669	12,213	11,671	11,053	10,282	9,481	8,660	8,019
45-49	17,609	16,717	16,190	15,706	15,003	14,461	14,106	13,503	12,999	12,724	12,390	11,934	11,399
50-54	21,195	21,255	20,436	19,108	18,086	17,217	16,327	15,796	15,326	14,640	14,104	13,742	13,143
55-59	19,716	19,331	19,766	20,366	20,696	20,487	20,557	19,762	18,467	17,502	16,666	15,799	15,278
60-64	25,026	23,667	22,159	20,823	19,496	18,734	18,324	18,735	19,347	19,650	19,455	19,545	18,796
65-69	24,195	24,164	24,168	24,171	24,031	23,472	22,196	20,743	19,441	18,167	17,415	17,023	17,437
70-74	20,505	20,959	21,129	21,466	21,572	21,790	21,814	21,910	22,054	21,958	21,423	20,203	18,780
75-79	16,067	15,997	16,251	16,365	16,719	17,056	17,593	17,863	18,228	18,432	18,747	18,829	18,973
80-84	8,556	9,455	10,106	10,728	11,412	11,893	11,846	12,194	12,474	12,881	13,151	13,567	13,794
85&+	6,312	6,745	7,278	7,498	7,676	7,922	8,587	9,051	9,362	9,652	10,042	10,334	10,775
All ages	296,737	232,640	228,572	224,193	219,746	215,282	210,798	206,340	201,817	197,104	192,443	187,702	182,973

Appendix 7: Suggested additional work that could be carried out

Suggested analysis to help confirm the findings

- Calculation of the mortality and hospitalisation rates for the full population of Wales using this data, (which relies on GP registration data for population denominators rather than the ONS mid-year population estimates normally used by Public Health) would provide overall background trends for comparison with the Project group rates.
- Calculation of rates for the component (23%) of the Welsh population who *did* move house during the Project period would be informative. Suitable denominators would need to be devised.
- Calculation of mortality by more specific groups of causes, to determine which causes are contributing to the overall mortality rate.
- Determination of cause of death in the excess winter mortality to investigate whether the causes might be associated with the health benefits of HEED-recorded home improvements.
- Determination of more specific detail about the changes observed in the hospital admission rates, to determine whether differences in outcome exist for specific age- and gender-specific groups.
- Analysis of overall length of stay to compensate for any changes in treatment regimens.
- Further work to establish suitably coded diagnoses, test results, observations or other items regularly recorded by GP practices that could be used as outcome indicators i.e would be expected to change as the result of HEED-recorded interventions.
- Further development of the methodology to take into account the differential length of exposure to the benefits of the interventions, as illustrated in Figure 7.31. Further refinements to this process could be developed to create improved modelling of the exposure times applied to the ‘Second’ group in calculating the underlying change rate. For the purposes of this Project, these were all calculated around the mid-date of the ‘First’ housing interventions - these could be distributed across the whole period, based on the distribution of their actual intervention dates.

- Inclusion of the migratory components of the population. This methodology could provide some person-days based outcomes on the indicators, allowing the existing exclusions to be greatly reduced.
- Establishing a GP record based Project on Asthma. The asthma emergency admissions is an example of a condition that might be sufficiently influenced by the effects of the interventions, but emergency admissions for asthma are small in number compared with the number of records relating to the management of asthma in primary care. Unfortunately, establishing how to 'count' records relating to asthma in GP records is a non-trivial exercise. It is an exercise that is underway though, for a National study, and when that study is complete it may be possible to use the definitions they have developed. There are a number of other categories of GP data that could also be explored, targeting specific prescribing events, records of signs and symptoms, and the results of various diagnostic tests.
- Inclusion of injuries from Primary Care and Accident and Emergency datasets. The examination of Injuries data for this Project was restricted to just those with an admission from the injury. It therefore just includes just the most serious cases. Some exploratory work was carried out comparing the Project groups in terms Accident and Emergency admissions. This proved more complex to analyse than could be managed within the scope of the Demonstration Project, partly because A&E data is not complete for all time periods from all geographical areas. However, the combination of GP based injury recording, Accident and Emergency attendance and hospital admissions would provide a larger dataset to allow injuries to be examined more broadly. With larger numbers of health events involved there may be the possibility to explore home-based injuries such as falls.
- Gender. Some diseases have wide variation by gender, which we have not considered here. Repeating the existing analysis on a gender specific basis may provide insights into more specific groups who benefit from the interventions. This might help in establishing more knowledge about lag times before changes become measurable.
- No analysis of the geographical variation in the impacts of HEED-recorded interventions on health outcomes has yet been carried out. This analysis could focus on variation at LA-level, between Rural and Urban areas and across WIMD Quintiles.

- Family structure. It would be possible divide the population receiving interventions into those living alone and those living in family groups. This may have a bearing on outcomes particularly in the older age groups.

Suggested wider analysis

- Capture of data about eligibility criteria met by the families who received these interventions would allow a much more precise piece of analyses than we have achieved here. The actual population experiencing fuel poverty could be compared to properly defined populations of low deprivation.
- There are a number of other studies underway, e.g. the Carmarthenshire Housing Intervention Project, which is collecting data on improvements to social housing as they are put in place in Carmarthenshire. Bringing together the data from that Project with the HEED data could provide added insight into the relationship between housing and health.
- Energy Performance Certification. Work is underway to anonymise some home Energy Performance Certification data into SAIL to establish if this provides any useful housing indicator information for research work. These certificates remain valid for 10 years so many may be out of date following home improvements. Linkage to HEED data in SAIL would help to establish the extent to which these certificates are up to date.
- Detailed work to determine how to estimate time periods from interventions beginning to outcomes becoming observable, might examine all the elements with potential to change the outcome indicator. This might range from individual causes of hospitalisation and death, the severity of each winter and detailed effects of economic change.
- House fires. Details of house fires are anonymised into SAIL. The linkage to HEED interventions may allow a study that establishes whether the installation of new central heating boilers leads to fewer house fires.

Appendix 8 Data issues

Data Reconciliation

- As noted above, administrative data is not collected for research purposes. Various challenges can therefore arise when multiple data sets collected for purely administrative purposes are linked. These issues are often of wider significance than for the specific analysis being undertaken, since individual administrative data sets may be used to underpin decision making or the allocation of resources. Identifying discrepancies can therefore be helpful in understanding the deficiencies of the individual sources. This section explores the data reconciliation work undertaken and documents any solutions developed. As noted above, publication of a 'Lessons Learned' report will follow.

The "False leaving date" problem

- As noted above, in order to establish groups of individuals who lived in the same address over a period of time the 'move-in date' and 'move-out date' are utilised. However, the WDS data contains thousands of records that indicate that an individual left a property on a specific date, only to move into that property again either on the same date or a few days later. This problem may be due to software and hardware changes to the data collection systems, such as when upgrades occur at a GP practice. These 'false leaving dates' create difficulty when attempting to select people with continuous residence over a specific period of time. The simplest algorithm to identify non-movers would filter the records on the basis of the rule "Select all people who moved in on or before 01.01.2000 and who did not move out until on or after 31.12.2011." However if a person has two records that effectively split a continuous period of residence into two components, as in Table 4.1 below, the algorithm would not select either record, resulting in the individual being wrongly excluded from the analysis.

Table 4.1: Sample moving-in and moving-out dates

ALF	RALF	Move-in Date	Move-out Date	Notes
12412341234	789798748	16.04.1999	04.05.2006	Same person ID and same address ID for both records.
12412341234	789798748	04.05.2006	31.03.2013	

Source: Fictitious data based on SAIL records

- In order to establish whether individuals remained at the same address over a period of time, all the false ‘move-in dates’ and ‘move-out dates’ needed to be identified and removed. This could only be achieved through an iterative process, systematically grouping all the records for each individual-residence pairing together into a true continuous time period. A five-step process was developed for this Project by repeating iterations of the ‘date pairing’ process until only a few unresolved cases remained – these were inspected manually.
- The SQL code for the algorithm was constructed to allow small gaps in residence (up to 30 days) as part of a continuous period of residence. The SQL code for the project is included in Appendix 6.

Conflicting Data

- As noted above, there is considerable duplication and overlap in the recording of periods of residence in the anonymised version of the WDS data as it exists within SAIL. The data extractions from the GP practice include records that differ only in a local (to the practice) code that is included for completeness in the SAIL database but not useable by SAIL. Whilst it is relatively straightforward to deal with records that are entirely duplicated, dealing with situations where conflicting records exist is more complex. The main conflicts identified during this project were for cases where individuals appeared to have lived in two places at once and where individuals were recorded as still being resident at an address after death (up to 10 years in one case!).
- Seven people were excluded from the Project following manual inspection because they were recorded as living at between 2 and 4 places at once and 3 cases were excluded because of the conflicting date of death issue.

Address notification lag

- The data issues so far described can be dealt with by reconciliation or exclusion of conflicting records. This issue cannot be addressed within the anonymised data.
- Counting populations using GP registration databases is known to be problematic due to the slow reporting of address change. Using the linked databases in SAIL, it is possible to eliminate over-counting where an individual is recorded as living in two places at once by selecting the RALF for which the most recent activity has taken place. However, if a person remains registered at a GP practice despite having migrated either temporarily or

permanently away, it is not possible to detect this in SAIL. This issue tends to be particularly pronounced among mobile, young, healthy people (particularly men), who may not need to visit a doctor for long periods and who may migrate for education or employment without registering with a new GP. This problem can create over counting for individual addresses because an individual who has moved out appears to remain a resident in addition to anyone new moving in. The problem is thought to be particularly marked among students[18].

- In England, the London Borough of Southwalk are currently attempting to improve their records by writing to people who have not attended the GP with whom they are registered in the last year[19]. In Wales, the introduction of the Individual Health Record is encouraging people to take a personal interest in how their record is stored by signing up for online access. This on-line approach may prove to be a medium that will appeal to young people so may bring about an improvement in address registration accuracy. Research in Northern Ireland called 'Signs of Life'[20] is having some success in counting populations using a combination of administrative data sets and eliminating from the count those records where no activity has been observed for some time. Benchmarking the findings against the latest 2011 census figures shows that this method provides potential to correct for the people who move on without de-registering.

Potential confounding in prescribing data

- In order to define an indicator of GP Practice activity we have summed for each person and each group the number of dates that a prescribing event occurred. However, changes in prescribing patterns may have occurred after April 1st, 2007 following the introduction of free prescriptions21. When prescription fees were in place, GPs may have prescribed larger quantities to cover a longer time period. This would make repeat prescriptions - and hence payment by the patient - less frequent, making the treatment more affordable. When free prescriptions were introduced in Wales, GPs may have started prescribing monthly rather than quarterly as a means of preventing potential wastage. However, any change would only have affected patients for whom prescription charges were levied i.e. those aged between

[18]Wales Centre for Health *Guide to the use of population data for health intelligence in Wales* accessed at http://www2.nphs.wales.nhs.uk:8080/hiatdocs.nsf

[19] http://www.southwark.gov.uk/info/100010/health_and_social_care/2690/gp_lists/1

[20] http://www.statsusernet.org.uk/Communities/ViewCommunities/GroupDetails/?CommunityKey=ccc5dd32-ef09-4b26-aef5-8e229c2ed623

[21]NHS report on the changes seen in prescribing following the implementation of free prescribing in 2007. http://www.wales.nhs.uk/documents/prescriptions-report-three-years.pdf

26 and 59 years and with an income above a certain level. Free prescriptions were already in place for those people under 25 years, over 60 years, for patients with certain health conditions and eligible under the low income scheme.