A review of Data linkage procedures at NatCen

Author: Michelle Gray
Date: September 2009 and updated in November 2010.
Prepared for the National Centre for Social Research.
# Content

1 Introduction ........................................................................................................... 2  
1.1 Background ........................................................................................................... 2  
1.2 Introduction .......................................................................................................... 3  
1.3 Structure of this review ........................................................................................ 4  

2 Current and past practice at NatCen .................................................................... 5  
2.1 The English Longitudinal Study of Aging .............................................................. 5  
2.2 Health Survey for England .................................................................................... 7  
2.3 The National Child Development Study ............................................................... 8  
2.4 The Adult Psychiatric Morbidity Survey 2007 ...................................................... 9  
2.5 Family Resources Survey ..................................................................................... 10  
2.6 Drug Treatment Outcomes Research Study ....................................................... 11  
2.7 Millennium Cohort Study ..................................................................................... 12  
2.8 ‘Understanding Society’ ....................................................................................... 13  
2.9 Growing Up in Scotland Study (sweep 4) ............................................................ 14  

3 Previous research ................................................................................................... 15  
3.1 FRS Data Linkage Request: Findings from a Qualitative study (2007) ............... 15  
3.2 Families and Children study (FACS) cognitive question testing (2005) ............. 16  
3.3 ‘Understanding Society’ questionnaire development and testing (2008) ......... 17  

4 Considerations for the care questions study and for other research .................. 19  
4.1 Written or verbal consent ..................................................................................... 19  
4.2 Matching variables for the linkage ...................................................................... 19  
4.3 Direction of the linkage ....................................................................................... 20  
4.4 Age effects on agreement to data linkage ........................................................... 20  
4.5 Ethical and legal Issues ....................................................................................... 21  

5 References ............................................................................................................ 23
1 Introduction

1.1 Background
This report details a review of data linkage procedures which took place as part of a project with an aim of developing a set of new survey measures about receipt and provision of social care services, both formal and informal, for people aged 65 and over. The main output was intended to be a module of questions on receipt of care services and related benefits, payment for care services and provision and receipt of informal care. The questions will be suitable for longitudinal or cross-sectional population surveys covering community-based care. The data will be available for social research in this area and for use in economic evaluations.

Along side the development of the survey questions, a small review took place exploring the scope for any linkages between survey data and administrative data on provision of care to older people which could complement survey questions. This short report details the outcome of this review, which involved outlining the current data linkage procedures and practices in some of the surveys either currently or previously carried out by the National Centre for Social Research (NatCen). More specifically, the objectives of the data linkage review were to:

a) Review data linkage questions, procedures and which data are matched from some of NatCen's existing and previous surveys
b) Review how data linkage has been used and what ethical permission has been sought and granted, and from whom, for existing surveys
c) Review previous research on this issue such as that conducted by NatCen for the data linkage questions in the FRS and harness the knowledge and experience gained by research teams running studies, which have, or are in the process of obtaining, permission for linkage to administrative data.

Although the intended use of this review was to feed into the care questions study, it is likely to have wider interest and has therefore been made available as resource to researchers.

Parallel research as part of the care questions study identified that there is no consistent administrative data available about use of social care since it is held at a local level. Therefore, in the end no data linkage questions were developed for the new social care module. However in the future as administrative data sources develop, this review will be able to help inform the development of linkage questions.
1.2 Introduction

NatCen currently asks respondents permission to link their survey data to administrative records for a number of different surveys, including the Health Survey for England (HSE), the Family Resources Survey (FRS) and the English Longitudinal Study of Ageing (ELSA) and has done so on previous studies, for example on the Adult Psychiatric Mobility Survey 2007. The request to link data almost always comes at the end of the survey interview and permission is usually obtained through one of three ways:

1. Interviewer is prompted to ask for permission but there is no standardised script (to read out). Respondents are usually given some information to read, explaining the request and what they are consenting to. Verbal consent is then recorded or coded by the interviewer, usually in the CAPI interview. The Millennium Cohort Study (MCS) used this method.

2. Interviewer is prompted to ask for permission but there is no standardised script. Respondents are usually given some information to read, explaining the request and what they are consenting to. Written consent is then obtained. When written consent is obtained, the signature of the respondent and sometimes of the interviewer too, usually follows some information which explains the request. This is known as a consent form. Adult Psychiatric Mobility Survey 2007 used this method.

3. Interviewer reads out a standard script, on the CAPI screen, and the respondent is asked whether they will consent to linkage. Consent is then coded in the CAPI. Written information, in the form of a leaflet or a consent form is often provided too and sometimes respondents are asked to provide written consent as well as, or instead of, verbal consent. The Health Survey for England (HSE) uses this method, with written consent recorded on a consent form.

Figure one below summarises the ways in which data linkage requests can broadly be categorised.

Figure 1: The three main ways data linkage is obtained.
1.3 Structure of this review

The following chapter (chapter 2), looks at the current and past practice at NatCen, using a selection of some of NatCen’s large population surveys. In this chapter we review procedures for nine surveys either currently, or previously, carried out at NatCen:

- The English Longitudinal Study of Aging (ELSA),
- Health Survey for England (HSE),
- The National Child Development Study (NCDS),
- The Adult Psychiatric Morbidity Survey (APMS) 2007,
- Family Resources Survey (FRS),
- Drug Treatment Outcomes Research Study (DTORS),
- Millennium Cohort Study (MCS),
- The UK Household Longitudinal Study (Understanding Society) and
- Growing Up in Scotland Study (GUSS) sweep 4.

The amount of information detailed for each survey, such as consent rates broken down by subgroups, differs depending on the availability of the information from the NatCen research teams. We have chosen these nine surveys as they reflect the range of different studies carried out by NatCen (e.g. longitudinal, cross sectional and birth cohorts), however there are other surveys that NatCen carries out which also have data linkage procedures. For more details of all of NatCen’s surveys, visit our website [www.natcen.ac.uk](http://www.natcen.ac.uk).

In chapter 3 we review previous research on data linkage and finally chapter 4 addresses some of the key considerations for developing a data linkage consent question alongside, and to be used with, the new module of social care questions. The considerations set out in chapter 4 can be more widely applied to other survey research practice.
2 Current and past practice at NatCen

This chapter outlines the procedures and processes, in relation to data linkage, for some of NatCen's large population surveys.

2.1 The English Longitudinal Study of Aging

The English Longitudinal Study of Ageing (ELSA) looks at the lives of people in England who are aged 50 and over (and about their partners). This study covers a broad range of topics such as people's health, economic situation and quality of life. ELSA has three data link requests (each discussed below). Because this study is longitudinal, the same respondents are interviewed multiple times over a period of time, interviewers try and gain consent to data linkages in the first interview they do with respondents. If they can’t gain consent in the first interview, they will try again when they go back to conduct the second of a series of interviews. If they still have not got consent they will not ask again.

Firstly, ELSA requests permission (in the form of written consent) for respondents’ survey data to be linked to the National Health Service Cancer Register (NHSCR), if the respondent has not already agreed to this linkage when they took part in the Health Survey for England (HSE)\(^1\). For the NHSCR link, respondents are asked for permission to send their name, address and date of birth to the NHSCR. A consent form is given to respondents containing information about the data linkage between ELSA and the NHSCR and information on how they can withdraw consent. The form explains that the data linkage would create a flag on the respondent’s health record to show that they took part in the ELSA study. A copy of this consent form can be found in the appendix.

ELSA also requests permission to link to the Hospital Episodes Statistics (HES) database and Her Majesty’s Revenue and Customs (HMRC) / Department for Work (DWP) records on National Insurance contributions, benefits, tax records and pension. This consent request asks for permission for the administrative body to disclose the respondent's data to NatCen. In practice however, some information about our respondents still needs to be passed to DWP/HMRC in order for them to make the link between their data and NatCen’s: hence the subtle difference between the NHSCR and HES and HMRC/DWP links.

A second type of consent form is used on ELSA to create these different data links with the Department of Health (DH), the Department for Work and Pensions (DWP) and Her Majesty’s Revenue and Customs (HMRC). The consent forms are written in accordance with relevant data protection laws.

---

\(^1\) The survey sample is drawn from respondents to the Health Survey for England (HSE) - a study conducted jointly by the Department of Epidemiology and Public Health, UCL, and the National Centre for Social Research, on behalf of the Department of Health. Around 12,000 respondents from three separate years of the HSE survey were recruited to provide a representative sample of the English population aged 50 and over.
Majesty’s Revenue and Customs (HMRC). The form is split into two sections and respondents can choose to give consent for all of the departments to be data linked or for just DH or just DWP and HMRC. This form asks permission to view additional information such as NHS health records and information on National Insurance contributions, benefits, tax records, savings and pensions. A copy of this consent form can be found in the appendix.

Tables 1 and 2 below show the numbers of respondents who have consented to linkage of their details to the NHSCR or the Hospital Episodes Statistics databases. The most up-to-date versions of these frequencies are taken from the data that was fed forward into the Wave 4 interview (2008/2009), which in turn only contains those respondents who were issued for this wave (so it doesn't include everyone who has ever been asked for their consent).

**Table 1: Number of respondents aged 16+ who consented to data linkage to the NHSCR in 2008/2009**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consent given</td>
<td>17036</td>
<td>95.4%</td>
</tr>
<tr>
<td>Consent refused</td>
<td>818</td>
<td>4.6%</td>
</tr>
<tr>
<td>Total</td>
<td>17854</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

**Table 2: Number of respondents aged 16+ who consented to data linkage to the HES in 2008/2009**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Valid Percent</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consent given</td>
<td>11017</td>
<td>82.8</td>
</tr>
<tr>
<td>Consent refused</td>
<td>2228</td>
<td>16.7</td>
</tr>
<tr>
<td>Consent withdrawn</td>
<td>57</td>
<td>.4</td>
</tr>
<tr>
<td>Total</td>
<td>13302</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The ELSA research team also looked at levels of consent between Waves, between men and women and between those aged under the age of 60 and those aged 60+ (tables 3 and 4 below). This age was roughly the 50% split. A very small percentage of non-white respondents respond to ELSA so a comparison between different ethnicities would not be meaningful. At the time this review was started, these figures were 'current' consents as at ELSA Wave 2 (2004 – 2005) as this was the most up-to-date information that was easily available at the time. The percentages represent the consents from everyone interviewed at Wave 2 and those interviewed at previous waves (but who were not interviewed at Wave 2). These figures are only for those interviewed up to and including Wave 2 of ELSA.
We can see that the HMRC/DWP linkage has the lowest level of consent overall, women have a lower level of consent than men and those aged under 60 have a higher level of consent to those aged 60 and over.

<table>
<thead>
<tr>
<th>Table 3: ELSA Data linkage consent by Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>% who consented to data linkage request for NHSCR</td>
</tr>
<tr>
<td>Men</td>
</tr>
<tr>
<td>Women</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4: ELSA Data linkage consent by Age Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>% who consented to data linkage request for NHSCR</td>
</tr>
<tr>
<td>Those aged 50-59</td>
</tr>
<tr>
<td>Those aged 60+</td>
</tr>
</tbody>
</table>

2.2 Health Survey for England

Each year the Health Survey for England (HSE) covers core topics on general health and key lifestyle factors that influence health. In most years it also focuses on an additional health topic such as cardiovascular disease, physical activity or respiratory diseases and in some years it focuses on a particular demographic group. The HSE covers the adult population aged 16 and over living in private households in England and has included children since 1995. Each year, the HSE team are required to get approval, for the Survey, from the National Research Ethics Service (NRES) which has replaced MREC.

On the HSE, respondents are asked to consent to NatCen using their name, address and date of birth to link to data on mortality and cancer from the NHS Central Register, and to link to Hospital Episodes Statistics (HES). The respondent details are used to establish the respondent’s NHS number, and this is used to put a flag on the person’s record on the Central Register to say they took part in the HSE study; if they contract cancer or die, the type of cancer or cause of death can be linked to their survey answers. The process is slightly different for HES, where the NHS number is used to identify HSE respondents in the HES database; a system of flagging is not used and a matching process is carried out each time data are needed. The only new information that the administrative records
holders receive from NatCen, in order to identify the NHS number and add the marker to people’s records, is that the person took part in the Health Survey for England.

The consent forms, like ELSA’s, provide respondents with information about which information will be linked and why. They also explain how respondents can later cancel their permission.

The consent rates to the data linkage request have varied over the years. In most years, the HSE adds a boost sample for a specific subgroup and this may well have an effect on the consent rates. Tables 5 and 6 below compare the ‘general population’ sample with two boost samples: a minority ethnic boost in 2004 and an older person’s boost (age 65+) in 2005. It is notable that the rates for ethnic minority respondents were lower than the general population in 2004. The figure for older people in 2005 was slightly lower than the general population. Although not statistically significant, the data in Table 9 (later in the report) shows that this is part of a consistent pattern. From 2006 and onwards children were boosted in the HSE sample and HSE only asks for data linkage to those over the age of 16. The consent rates from 2006-2008 have been around 90%.

### Table 5: HSE 2004: Ethnic boost year

<table>
<thead>
<tr>
<th>% who consented to data linkage request for NHS register only</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population (i.e. all adults aged 16+)</td>
</tr>
<tr>
<td>Minority ethnic boost</td>
</tr>
</tbody>
</table>

### Table 6: HSE 2005: Older persons boost year

<table>
<thead>
<tr>
<th>% who consented to data linkage request for NHS register only</th>
</tr>
</thead>
<tbody>
<tr>
<td>General population (i.e. all adults aged 16+)</td>
</tr>
<tr>
<td>Older people (aged 65 and over)</td>
</tr>
</tbody>
</table>

### 2.3 The National Child Development Study

The National Child Development Study (NCDS) is a longitudinal study which took a sample of all babies born in Great Britain during one week in March 1958. This study has followed the sample’s physical, educational, social and economic development.

On NCDS consent is asked to link to routine economic records (held by DWP and HMRC) and routine health records. Like HSE, on NCDS interviewers read out a short CAPI script to respondents (only adults aged 16 and over in the HSE), before handing them a form to read and sign (if they are willing to consent to linkage). Additionally NCDS respondents
were given a leaflet that provided information and a Centre for Longitudinal Studies (CLS) freephone number should they have further questions. Interviewers were briefed to encourage respondents to spend time reading through the forms and to explain the nature of the consents to the respondent: why they were being asked for consent and what records would be linked to. If the respondent had questions that the interviewer felt they were not able to answer, the interviewer directed them to the CLS freephone number.

If the respondent had a partner living in the household, the interviewer also explained the consents to the partner and asked them to consent to link to routine economic and health records. The interviewer was provided with a separate consent form for the partner.

If the respondent (and partner) were happy to give consent, they signed their individual form and handed them back to the interviewer (both the consent forms were produced in duplicate so a copy was left behind for the respondent and partner). If the Cohort Member or their partner did not want to sign their form at that stage, the form(s) were left behind at the household with a pre-paid, pre-addressed return envelope so they could return the form(s) at a later date.

Interviewers (and respondents in cases where the forms were left behind) were asked to return the form(s) whether or not the respondent and partner had given consent. Interviewers were also required to code whether consent has been given (or not) in the CAPI program.

2.4 The Adult Psychiatric Morbidity Survey 2007

The Adult Psychiatric Morbidity Survey (APMS) series provides data on the prevalence of both treated and untreated psychiatric disorder in the English adult population (aged 16 and over). This survey was the third in a series and was conducted by the National Centre for Social Research in collaboration with the University of Leicester for the NHS Information Centre for health and social care. The previous surveys were conducted in 1993 (16-64 year olds) and 2000 (16-74 year olds) by the Office for National Statistics, and covered England, Scotland and Wales.

Respondents to the APMS 2007 were asked for their survey data to be linked to the NHS Central register. The data linkage request was not in the form of a scripted question in the CAPI interview, instead interviewers were just promoted to ask respondents. Interviewers then obtained written permission, using a consent form.

The APMS requests agreement to pass the respondents’ details to an external body (NHS central register is held by the NHS Information Centre, previously by ONS). For the purpose of this review, the research team kindly cross tabulated agreement rates by ethnic group and sex (tables 7 and 8 below) and a quick scan of the results (without significance testing) suggests the following. Compared with the general population there was a higher refusal rate among minority ethnic groups, especially black and little variation in consent rates by sex.
### Table 7: APMS 2007: Data linkage consent by ethnic group

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>% who consented to data linkage request for NHS Central register</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>86%</td>
</tr>
<tr>
<td>Black</td>
<td>67%</td>
</tr>
<tr>
<td>South Asian</td>
<td>73%</td>
</tr>
<tr>
<td>Mixed or other</td>
<td>73%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>85%</td>
</tr>
</tbody>
</table>

### Table 8: APMS 2007: Data linkage consent by Sex

<table>
<thead>
<tr>
<th>Gender</th>
<th>% who consented to data linkage request for NHS Central register</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>85%</td>
</tr>
<tr>
<td>Women</td>
<td>84%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>85%</td>
</tr>
</tbody>
</table>

#### 2.5 Family Resources Survey

The Family Resources Survey (FRS) is a continuous survey of private households and is commissioned by the Department for Work and Pensions (DWP). Adults eligible for inclusion in the survey are asked a wide range of questions about their circumstances including questions about: income and state support, tenure and housing costs, assets and savings, occupation and employment, health and ability to work, pensions and insurance, and childcare and carers.

The FRS has a data linkage consent question which is electronically recorded in the CAPI interview (so agreement or refusal is only recorded at one question in the CAPI). There are no consent forms to sign – the request is accompanied by materials, such as a flow chart, which are aimed at helping respondents understand what is being requested, enabling them to make an informed choice. A data linkage leaflet provides information about the request and gives information on how to withdraw consent. The data linkage consent does not have a time limit. If consent is given, the respondent’s name, gender and date of birth, collected as part of the interview, and the postcode is supplied from the sample details and supplied to DWP. The current FRS data linkage request was designed following a cognitive testing exercise undertaken by the Questionnaire Design and Testing (QDT) Hub at NatCen. Details of which can be found in section 4.
Data services at DWP conduct the data matching and therefore the FRS survey data is passed on to DWP by ONS and DWP conduct the data matching (i.e., the link with their administrative data). The direction of the link (i.e. the respondent’s name and address will be passed onto the Department for Work and Pensions) is made clear to respondents in the FRS data linkage question and supporting materials. Data linkage for the FRS involves survey answers being added to DWP’s Work and Pensions Longitudinal Study (WPLS), which includes DWP’s administrative data and some details supplied by HMRC on Tax and savings. Ethical approval was given by NatCen’s in-house Research Ethics Committee which reviews all projects conducted their own requests for ethical and legal approval of the wording of the request. The Office of National Statistics (ONS), which conducts half of the fieldwork for FRS, also used their in-house procedures for gaining ethical and legal consent for the request. The combined ONS and NatCen consent rate was around 61% in 2008/9.

DWP have carried out data analysis to examine possible factors which are related to the consent rate and have found that it is only ethnic group (i.e. white vs. other) that shows any difference in consent rate. Consent rates for the non-white group are lower than for the white group. Overall, the research team have found that, in practice, people either refuse outright (because they think they’ve given enough information already or because they have data security concerns), or are happy to link because they ‘have nothing to hide’ or consider the linking a good idea.

On FRS, sex, date of birth, postcode and first and last name are used for matching. Self NINO (National Insurance Number) is not collected as respondents were extremely concerned about being asked to give this information the first year the linkage was asked (resulting in a 35-40% consent rate). DWP undertakes the matching exercise to link the two data sources together and there is a traffic light system to indicate how good the match is.

### 2.6 Drug Treatment Outcomes Research Study

This longitudinal study was about people entering drug treatment programs in England. It aimed to find out what sort of things affect a person’s chances of success in drug treatment programs. The DTORS survey included 3 interviews over 12 months. The data linkage request was in the first interview, and these took place between March 2006 and January 2007.

On Drug Treatment Outcomes Research Study (DTORS), respondents were asked to provide written permission for linkage to the National Drug Treatment Monitoring System (NDTMS) for details of the respondent's drug treatment history; and Criminal Justice System (CJS) records for any conviction and sentencing details. A consent form was used to obtain permission from respondents and can be found in Appendix A. This gave

---

Fieldwork is carried out jointly by the Office for National Statistics (ONS) and the National Centre for Social Research (NatCen) using computer-assisted personal interviewing.
information about the data linkage such as what information the researcher would be looking at. There were separate sections to sign for the NDTMS and CJS links. Ethical permission had to be obtained from COREC (previously called MREC and now called NRES) to carry out the study. In addition, permission was required from the owners of the data. For the NDTMS - the collaborators had a contract to collate the data for the NHS, for the CJS data the sponsors were the Home Office.

DTORS’ data linkage request involved adding information from the administrative sources to the survey dataset. In order to get the details, the NatCen research team had to tell the administrative data holders who had taken part in the survey (which effectively informed the administrative data holders of those individuals who therefore had drug problems and were on drug treatment). Despite this, the agreement to data linkage was relatively high. At the baseline interview, respondents were asked for their signed permission for researchers to access their criminal justice, and drug treatment records, to enhance the research. This permission was voluntarily given - and did not affect participation in the rest of the survey, or receipt of the incentive voucher. 93% of respondents gave permission for both types of information to be accessed, 2% for drug treatment records only, and less than 1% for criminal justice records only (DTORS technical report http://www.natcen.ac.uk/study/drug-treatment-outcomes-research-study).

2.7 Millennium Cohort Study

The Millennium Cohort Study (MCS) is a panel study of parents of children and the children themselves, born in the year in the UK in 2000/1.

NatCen last carried out the survey in 2007 (when the children were age 7) and parents/guardians were asked if their data can be linked to health and economic records. Respondents were given a detailed leaflet (which can be found in the appendix), explaining how data linking works and which information the researcher team would like to link. The respondent was then asked to sign a consent form and provide their national insurance number which was used to create the link. The respondent was also asked to give consent on behalf of their child, allowing the child’s teacher to complete a questionnaire on the child's school progress and for the child's health and education records to be data linked.

The survey research team who worked on MCS believe that as panel respondents, they may have been more willing than respondents on one off surveys to sign documents. Trust had been established with the respondents and they were used to signing other forms in the interview. Around 80% of adults gave consent for their National Insurance Numbers to be used in data linkage and the numbers were higher for their children’s data to be linked to school records (over 90%).
2.8 ‘Understanding Society’

The UK Household Longitudinal study (UKHLS), known to respondents as ‘Understanding Society’, was set up to follow 100,000 individuals in forty thousand households and question them about a wide spectrum of areas relating to their working and personal lives.

This study, at Wave 1, asked for consent to link health and educational administrative records for adults and for children. The respondent was given a consent form to sign and to keep and a carbon copy (of this form) was returned to NatCen’s operations department. The interviewer asked the respondent to tick the relevant boxes, sign, print their name and date the relevant section of the consent form. The interviewer coded in CAPI whether consent was given.

Consent for linking for children was obtained from their responsible adult, and consent for several children could be collected on one form. If a child’s natural mother was in the household, she was defined as the responsible adult. If not, then the child's natural father became the responsible adult if they were in the household. If a child had no natural parents in the household, the interviewer was asked to record who the responsible adult was during the enumeration of the household. The first two consents asked were for linking to health records. The blue form was for adults and yellow for children (see appendix A for copies of the forms). There was also an information leaflet that was given to respondents and it explained which health records we would be linking to, which in summary were:

- Hospital admission and attendance;
- Information about specific conditions such as cancer or diabetes;
- Prescriptions;
- Where respondents are registered;
- Mortality records.

The second set of consents were for linking to educational records, although note that consent for adults was only for those aged 16-24 who went to school in the UK. The pink consent form was for adults and the green one for children (see appendix A for copies of these). There was also an equivalent information leaflet for linking to educational records which was given to the respondent. Links would be made to:

- National tests and formal assessments e.g. SATS/GCSEs/Standard Grades
- Basic information on pupils e.g. demographic characteristics or Special Educational Needs
- Courses taken.

Respondents were also shown a ‘Consent flowchart’, which showed how links are made to their health and educational data and what happens to the data once the researchers have linked to it. The purpose of this tool was to reassure respondents about the confidentiality and anonymity of the data. The data have not been made available yet so we are unable to present any figures in this review.
2.9 Growing Up in Scotland Study (sweep 4)

This study follows the lives of a national sample of Scotland's children from infancy through to their teens. It will provide information that will help develop policies affecting children and their families in Scotland. The study started in 2005, with two groups of children; 1) babies born between June 2004 and May 2005 and 2) Children aged between two and three years old born between June 2002 and May 2003. The study is now in its 6th sweep. Permission has been sought to link survey data with health and education records.

Health data linkage consent was collected at sweep 4 which came out of field in May 2009. The mother of the child was asked in the fourth interview to consent to a data link being formed between health records and the information she had given in the first and subsequent interviews about her pregnancy, birth and her child’s health. The information requested from respondents in order to create this link was name, address and date of birth (if the respondent was the child’s birth mother), and those of the child. A written consent form was given to the mother with information on it, a CAPI question was included to give the consent form to the mother and her answer was also coded in CAPI. The direction of the data linkage is not made explicit in the interviewer’s script, however on the consent form itself, and specifically next to where the respondent is required to sign, it refers to ScotCen passing on the respondents’ name, address and date of birth to the Information Division of NHS Scotland, and for the respondent’s information to be passed back to ScotCen. The consent rate for linkage with children’s health records was 91% and for linkage with mothers’ health records was 93%. The research team have not felt the need to break these figures down any further, i.e. age of mother etc, as they had no real concerns due to high consent rates. Copies of these documents can be found in the appendix.

Consent to link to education records is being collected for children in the birth cohort at sweep 6, which is currently in field and will remain so until May 2011. A written consent form is given to the respondent, usually the child’s mother with a CAPI question included as a reminder for the interviewer. The consent form requests permission to pass the child’s name, gender, and date of birth to ScotXed, the Scottish Government’s central education data administration arm and for the child’s information to be passed back to ScotCen. Early data suggests the consent rate to be around 95%.

The Growing Up in Scotland Study had problems when first applying for ethical permission to use the data linkages. The study went through MREC to gain ethical permission. At sweep 1 the committee said that the research team had to detail and get consent for every individual item that they wanted to link to. The study team felt this was unfeasible, particularly at sweep 1 where there was no established relationship with the participants. In addition, MCS had already managed to pass MREC with blanket consent. The team tried again at sweep 4, just putting the consent in the substantial amendment, and this time permission was given for the blanket health linkage consent without question.
3 Previous research

Various research studies, that we are aware of, have explored the issue of data linkage; however few have actually explored how survey respondents view the data linkage request and procedures (Calderwood & Lessof, 2009). The following three small pieces of NatCen research can shed some light on such issues.

3.1 FRS Data Linkage Request: Findings from a Qualitative study (2007)

A qualitative study carried out internally at NatCen, which explored the FRS data linkage, concluded with a number of key pointers concerning communicating the message to respondents:

- The message needs to be clear: Is the request to pass on personal details or to add survey data to administrative records, for example? The information given to respondents, before they decide whether to give consent, needs to be clear and explain what exactly they are consenting to.
- The importance of conveying who carries out the data linking process: this study, for example, found some indications to suggest that respondents may find it more generally acceptable for a survey organisation to carry out the data linking rather than the DWP. Respondents seemed less wary of the DWP having access to anonymous data than they were of the DWP using their personal data to carry out the data link.
- Communicating the message: It is essential that the message is clear and uncomplicated. It may be necessary to provide respondents with a hard copy of the information the interviewer is reading out so that they can really absorb and understand the information. It was suggested, from this study, that answering how and why questions respondents may have, such as ‘why are more than one piece of personal information needed to create the link?’, is important so the respondent fully understands and is not left with any questions.
- Avoid official language, jargon and vague terms: Respondents in this qualitative study, for example, found the words ‘personal records’ and ‘information DWP holds’ ambiguous and confusing. When such terms are used, it is important to provide some definition of such terms but only sufficient to clarify what is meant.
- Avoid acronyms and euphemisms: Respondents do not always understand terms such as the DWP and this acronym can raise suspicion among respondents. It was found from the study that respondents thought that it was some kind of ‘catch’ in the question. It is vital to reduce concerns in any way in order to try to increase the number of participants who do give their consent.

---

3 This study was not published, the findings were shared with DWP and the current FRS data linkage request was slightly adjusted as a result.
- Adopt a ‘friendly’ tone: Respondents showed a preference for a friendly tone, in the qualitative study. There should be a balance between giving enough information for the respondent to understand and make an informed decision about giving consent and giving too much information and making the request seem too formal.

- Consider breaking the request down into smaller parts: Recommendations from the qualitative study point to the importance of feeding respondents the information in a series of digestible parts, which build on from the one another logically. It is important to build up the information, starting with contextual background, moving on to the central message and finally giving a clear explanation of the processes involved, in order to encourage consent.

3.2 Families and Children study (FACS) cognitive question testing (2005)

Another helpful study, which can contribute to our understanding of how respondents react to and make sense of requests to link data, involved some cognitive interviews with respondents in preparation for wave seven of the Families and Children study (FACS)⁴.

In one of the two data linkage requests, asking to link survey answers with administrative data held by the DWP, respondents did not like the opening paragraph of the request and preferred the bulleted format that is used later in the document. The term ‘linking data’ was not fully understood and respondents took it to mean a number of things, for example linking their survey answers with other respondents’ survey answers. Respondents showed concerns that the data which was linked could be used for other purposes. In addition, respondents were unsure of how they could withdraw consent, which also created concerns. Despite these concerns, respondents were (still) willing to give consent to the data linkage. The sample for this cognitive interviewing study was very small so these findings cannot necessarily be generalised to a larger population.

Following findings from the cognitive interviews, researchers made the following suggestions for how the data linkage request could be improved:
- change the introductory paragraph to a bulleted format which was the preferred layout for the respondents;
- consider adding a term such as ‘we would like to look at your answers alongside information…’ in addition to the term ‘link’ which could clarify the linkage process for respondents; and
- re-order some of the bullet points in order to assure respondents that their data will not be used for any purposes other than those stated in the request and consent materials.

⁴ Details of the data linkage procedures will be added to section 3 of this review in due course.
3.3 ‘Understanding Society’ questionnaire development and testing (2008)

Cognitive testing was also used to test a data linkage request question for the UK Households Longitudinal Study (‘Understanding Society’). The aim of this request was to provide information and request consent in a way that was least likely to alarm respondents. Before testing the data linkage question, researchers had concerns that non-UK born respondents may be less willing to consent due to concerns about being matched to official records. Evidence from other studies suggests that levels of agreement to linkage may be lower among ethnic minority respondents.

The cognitive interviewers who carried out the testing of the request for data linkage were told to read a preamble, verbatim, before giving a paper consent form to the respondent to read. Respondents were then left to make a decision about the data linkage: consenting to the government departments involved (DWP and HMRC) passing on records to NatCen and ISER, using name, sex and date of birth. The direction of the linkage, however, was not made explicit in the consent forms.

If a respondent queried the information given to them, interviewers were instructed to read further verbatim text to alleviate concerns. This data linkage request was tested on general population and ethnic minority respondents (as the first wave of the survey was planned to include an ethnic boost).

Through cognitive testing it was found that this data linkage request provoked mixed reactions in respondents, and raised various concerns. It was also found that those from minority ethnic groups could have specific concerns. However because cognitive testing has such a small sample, this cannot be generalised. Interestingly people from minority ethnic groups were the only ones to give their consent because they felt they had to. Overall, those who were happy to give their consent gave reasons such as:

- they had nothing to hide;
- they had an appreciation for social research;
- they knew the information would be confidential; and,
- they assumed the information could be accessed anyway, even if they did not give consent.

Those respondents who refused consent to the data linkage did so because there was concern, confusion and distrust about the data linkage process. More specifically, there were a number of common ‘concern’ themes which emerged. The first was related to fear of loss of data. The interviews were carried out around the time when there was publicity about a government department losing child benefit records which could have affected respondents’ response to the data linkage request. Another key concern was that repercussions would occur such as their data being passed on to other companies or taxes being increased.
With regards to the wording of the data linkage request, respondents unsurprisingly found the term ‘revoked by me in writing’ confusing and concerning. It was also suggested that the names of the government departments could be listed in the request, rather than it specifying ‘different government departments’: this would make the request more transparent, official and could increase consent rates.

Based on the findings from the cognitive testing of this data request, researchers made a number of suggestions about how it could be improved, including that:

- Interviewers should make sure that respondents are aware that their survey answers will at no time be disclosed to anyone, only their details which will be used to create the link will be passed on;
- The data linkage request needs to specify, so that respondents understand, that the link will be from DWP/ HMRC to NatCen and not the other way round.
- The researchers suggested that more information about the procedures in place to protect data, such as encryption, needs to be given to respondents before they make their decision to give consent to reduce the risk of concern about data going missing.
- The words ‘This consent will remain valid until revoked by me in writing’ should, if possible, be reworded to avoid confusion.

These suggestions were taken on board and edits were made to the original data linkage request. This included changing the phrase ‘This consent will remain valid until revoked by me in writing’ to instead read ‘Your permission will stay in place unless you write to us to say you want it removed.’
4 Considerations for the care questions study and for other research

There are a number of key issues that will need to be taken into account when designing data linkage procedures. These are outlined, in brief, in this section.

4.1 Written or verbal consent

It will be important to consider the form in which the data linkage consent question will take. For example, will interviewers be instructed to read out an explanation of the request and then simply ask respondents for their verbal consent, which would then be coded in the CAPI, or will respondents instead be given consent forms (and possibly leaflets and other materials) to read and then sign. Written consent tends to be more common in practice and should be considered best practice for ethical and obvious legal reasons. Written information should also be given to the respondent which should help inform them of their decision and ensure that people really understand the process and therefore, if they agree to do so, give genuinely informed consent. Any written information given to the respondent should also ideally be left with the respondent and detail instructions on what they need to do if they change their mind at a later date (i.e. how to retract their consent).

On deciding on the nature of data linkage request, it will be important to consider the procedures of the survey on which the care questions will be asked. If asked on ELSA or HSE, for example, it would make sense for the consent to be requested in a written form (as are the other data linkage requests). Decisions about this should also consider ethical issues: both the ethics of informed consent and also the practical issue of obtaining ethical approval to request permission for data linkage from respondents.

4.2 Matching variables for the linage

An important factor to be aware of when designing a new data linkage request is which variables are used to implement the link between respondents in the survey and records in the administrative source. Will sex, date of birth, and post-code be used, or sex, date of birth, first name and family name or will the records simply be matched on self-reported National Insurance Number (NINO)? The advantages of matching using a NINO are clear: a NINO is a unique personal identifier and virtually all adults in Britain have one. There are, however disadvantages to solely relying on the NINO for the matches, when the NINO is self reported in the survey. There is a risk that respondents will refuse to provide their NINO or they may not know what one is and there is potential for measurement error: respondents may not report their NINO correctly and interviewers may record them incorrectly (Jenkins et al 2004). Additionally using NINO as the matching variable will add
time to the interview length as people will inevitably have to go and hunt for it. Other variables such as D.O.B are less prone to these issues. Further research should look at what variables people are happy with in the linkage and whether it makes any difference from the perspective of the respondent.

Jenkins et al (2004) did find that high linkage rates can be achieved without using NINO matching and concluded that utilisation of software to clean and pre-process name and address data (such as reviewed by Winkler 2001 for the USA) can help reduce any inconstancies between variables in the survey and administrative data. Finally, the choice of matching variable(s) will also be constrained by which pieces of information are collected on the vehicle (survey) on which the care questions will be asked.

4.3 Direction of the linkage
Will the survey data be passed onto the holders of the administrative records OR will the administrative records be matched to the survey data? We know that respondents may be more willing to consent to the latter type of matching. The choice will in part depend on the type of the administrative data being linked to and who owns it and whether that is the same body as the body which commissioned the survey. However, regardless of what is decided upon, we are aware, from the literature, of the importance of making the direction of the linkage explicit to respondents.

4.4 Age effects on agreement to data linkage
In a study carried out by Jenkins et al (2004), respondents aged 50+ were slightly more likely than respondents aged below 50 to consent to data linkage (79% compared with 76%). In the same study, referring to personal documents to check the NINO was markedly higher among respondents aged 50+ (81.2%) than among respondents aged less than 50 (54.3%). This suggests either that older people were less confident in remembering their NINOs or simply that pension books were more readily available than payslips. Work carried out by Sala et al (2010) also suggests mild associations between respondent socio-demographic characteristics and data linkage consent with ethnic minority respondents and older respondents being less likely to consent (this work uses data from Wave 18 of the British Household Panel Survey (BHPS)). Sala et al found stronger associations however with respondent attitudes; with ‘private’ respondents less likely to consent and more ‘community-minded’ respondents more likely to consent. Another interesting finding from their work was that longer serving panel members are less likely to give consent.

On the HSE, it appears that those aged 65+ consent to linkage at a consistently marginally lower rate than the younger adults in the survey. Table 9 below compares 16-64 year olds consent rates with those over the age of 65 from 2001-2008. The figures for 2009-2010 had not yet been released at the time this review was published.
Table 9: Percentages of 16-64 year olds compared with 65 year olds and over consenting to the HSE data linkage from 2001-2008

<table>
<thead>
<tr>
<th></th>
<th>16-64 year olds</th>
<th>65 years and over</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>88%</td>
<td>83%</td>
</tr>
<tr>
<td>2001</td>
<td>89%</td>
<td>85%</td>
</tr>
<tr>
<td>2002</td>
<td>89%</td>
<td>86%</td>
</tr>
<tr>
<td>2003</td>
<td>89%</td>
<td>82%</td>
</tr>
<tr>
<td>2004</td>
<td>76%</td>
<td>75%</td>
</tr>
<tr>
<td>2005</td>
<td>81%</td>
<td>79%</td>
</tr>
<tr>
<td>2006</td>
<td>83%</td>
<td>81%</td>
</tr>
<tr>
<td>2007</td>
<td>82%</td>
<td>76%</td>
</tr>
<tr>
<td>2008</td>
<td>79%</td>
<td>75%</td>
</tr>
</tbody>
</table>

4.5 Ethical and legal Issues

Most requests for link data require written consent of respondents for the administrative data to be released, usually by asking respondents to sign a consent form (Calderwood & Lessof, 2009). From an ethical point of view, it is essential to ensure that this consent is fully informed. 'Informed consent' means the following pieces of information are explained to respondents, in a way in which they will understand:

1. What the administrative dataset(s) is/are;
2. The nature of the information that will be obtained; and,
3. The reason(s) for requiring the linkage.

One important aspect to any researchers wishing to request for data linkage relates to the first piece of information listed above: which administrative dataset(s) the survey data will be linked to and which government departments or other agencies will be involved. Consent should always be specific, with as much detail as possible, and not open ended and vague.

Respondents in longitudinal studies may be more likely to be compliant than in cross-sectional surveys, as a relationship of trust is built up (over time) between the respondent and the survey organisation, through the interviewer. This will need to be borne in mind if the new care questions are to be asked on both longitudinal studies (such as ELSA) and cross sectional surveys, such as the HSE and FRS.

It is also essential that any data linkage is lawful and unfortunately the legal framework around data linkage in the UK is ill-defined (Calderwood & Lessof, 2009). The main legal
considerations are the Data Protection Act (DPA) 1998 and the Freedom of Information (FOI) Act, both of which contain regulations about the release of personal data. The two acts come at this from opposite perspectives however: the DPA details safeguards against the release of data whereas the FOI Act specifies the time at which data are able to be released.

Additionally there is a UK common-law right to confidentiality and the EU Data Protection Directive requires an individual to sign consent to the release of their data. Finally, when planning to gain consent from respondents to a survey, researchers must approach the holders of the administrative data/records (and possibly their legal advisors) and agree on the wording which will go on any consent forms.
5 References


[www.understandingsociety.org.uk](http://www.understandingsociety.org.uk)


[http://www.iser.essex.ac.uk/pubs/workpaps/](http://www.iser.essex.ac.uk/pubs/workpaps/).