

Knies, Gundi; Sala, Emanuela; Burton, Jonathan

Working Paper

Consenting to health record linkage: Evidence from the British household panel study

ISER Working Paper Series, No. 2011-27

Provided in Cooperation with:

Institute for Social and Economic Research (ISER), University of Essex

Suggested Citation: Knies, Gundi; Sala, Emanuela; Burton, Jonathan (2011) : Consenting to health record linkage: Evidence from the British household panel study, ISER Working Paper Series, No. 2011-27, University of Essex, Institute for Social and Economic Research (ISER), Colchester

This Version is available at:

<https://hdl.handle.net/10419/65985>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Consenting to health record linkage: Evidence from the British Household Panel Study

Gundi Knies

Institute for Social and Economic Research
University of Essex

Emanuela Sala

University of Milano Bicocca

Jonathan Burton

Institute for Social and Economic Research
University of Essex

No. 2011-27

November 2011



INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

Non-technical summary

Linkage of individual-level administrative data to survey data provides a wealth of information that allows academic and policy researchers to investigate in much more detail than otherwise would be the case how well the people in our society fare. In some countries (including the UK), survey and administrative data can only be linked when the survey respondents have given their informed consent. Willingness to give consent is not universal in that some respondents may choose to withhold their consent. This may be a problem for analysis. On the one hand, there may not be enough linked data to analyse, and on the other hand, when consenters are different to non-consenters on some measures of interest, the analysis may provide inaccurate results that cannot be generalised.

In this paper we provide empirical evidence on consent to link administrative health records to the British Household Panel Study (BHPS). The research documents the procedures that were implemented to obtain consent. It furthermore evaluates whether consenters and non-consenters differ on key demographic and socio-economic variables and health indicators, and provides an estimate of the number of the linked survey data-administrative records available for the analysis.

Our research shows that 40% of the respondents gave permission to link their records. Consent is not biased with respect to socio-economic characteristics and health conditions. Consenters are more likely to be British/Irish White or to be young. However, when looking at respondents' socio-economic characteristics and health conditions we do not find differences that are marked enough to be significant in a statistical sense. Consent is biased, however, with respect to utilisation of health services. Recent users of General Practitioners' services, for instance, are underrepresented among consenters. Due to the low consent rate obtained, only a minority of BHPS records can be linked to health records, although this could, in principle, give more than 5300 matched survey-administrative records.

Consenting to Health Record Linkage: Evidence from the British Household Panel Study¹

Gundi Knies (ISER University of Essex)
Emanuela Sala (University of Milano Bicocca)
Jonathan Burton (ISER University of Essex)

Abstract

The British Household Panel Survey (BHPS) is the first study of its kind to have asked for permission to link to a range of administrative health records. Multivariate analysis is applied to investigate whether there is consent bias. We find that consent on the BHPS is not biased with respect to socio-economic characteristics or health; recent users of GP services are underrepresented among consenters. Whilst consent rates are lower than on previous surveys with a more medical focus, the problem of bias is less of an issue.

Keywords: data linkage, consent, BHPS, selection bias, health inequality

JEL Classification: I19, C83, Z00

¹ We are grateful for comments received on an earlier version of the manuscript from Stephanie McFall and anonymous reviewers. This project is funded by the ESRC grant RES-586-47-0002.

1. Introduction

The British Household Panel Survey (BHPS) is a renowned general topic social survey covering areas such as demographics, household composition, employment, education, training, health, values, opinions and finances. The survey started in 1991 and the same individuals and their households have been followed over time.

The BHPS assists with understanding the long-term effects of social and economic change, as well as policy interventions designed to impact upon the general well-being of the UK population. It is widely used in a number of different disciplines, but use in health research has been limited by the lack of objective measures of health. The BHPS collects self-reports of utilisation of health services and of health status which give a holistic summary of the health condition of an individual. However, self-reports and medical records do not always match-up (Newell et al. 1999; Okura et al. 2004; Ritter et al. 2001; Sauver et al. 2005). Whilst medical records contain objective measurements and confirmed diagnoses, they exclude diseases for which the patient has not consulted and they do not have much background information on the individual patient (e.g., education, employment status, financial well-being etc). As a result, wherever possible it is best to use both approaches, i.e., survey data augmented with administrative records.

It is an ethical requirement in the UK that survey participants give informed consent to data linkage before certain information can be combined. Previous studies have shown that 70-90 percent of the survey population allow access to their health records (Baker et al. 2000; Dunn et al. 2004; Gerber et al. 2007). Socio-demographic characteristics and health profiles have been found to be different for consenters and non-consenters, implying that results on health of linked datasets might be misleading. The evidence on consenting and bias is from epidemiological surveys that focus on particular health outcomes, and on experience from data linkage requests on birth cohort studies, which are run by medical research units and have a strong focus on health-related issues. Very little is known on consenting to health record linkage on surveys of the general population which have a much wider scope and survey the whole age range.

This paper is an attempt to expand the current knowledge and exploits the rich resource of the BHPS data about those individuals who gave or withheld their consent. In 2008, in the 18th annual survey of the BHPS, participants were asked for their informed consent to allow their survey data to be augmented with health records held by the National Health Services (NHS) and the Departments of Health. Patterns of consent are analysed and how many linked records may be available for innovative analyses of health outcomes are extrapolated.

2. Methods

The British Household Panel Survey (BHPS) is an annual longitudinal household panel survey, managed at the Institute for Social and Economic Research (ISER) at the University of Essex. The survey started in 1991 with a nationally-representative stratified, clustered sample of 5500 households. Within each household, all those aged 16 and above were eligible for a full individual interview. At Wave 2, in 1992, all individuals were issued for re-interview. From Wave 2 onwards, new entrants to the household are eligible for interview. If a sample member changed addresses, they are followed (within the UK) and interviewed in their new address. Any adults in their new address are also eligible for interview. As a young person reaches the age of 16 they become eligible for a full adult interview. A baby born to a sample household becomes part of the sample. Thus, the sample is designed to be self-perpetuating; whilst some members sample die, new sample members are joining all the time. This longitudinal design offers great potential for those who wish to research the life-course. In 1999 additional booster samples were added to the BHPS in Scotland and Wales (1500 households each) to allow for analysis within and across England, Scotland and Wales. In 2001, a Northern Ireland booster sample was added (2000 households). Our analysis focuses on the 13454 adults who gave a full interview and were living in the 7596 households which were interviewed at Wave 18 of the BHPS (2008).

Interviews are conducted face-to-face using Computer-Assisted Personal Interviewing (CAPI) with interviewers calling on participants in their homes. The individual questionnaire takes around 45 minutes to answer, and questions cover a broad subject range. There are question modules on demographics, education, training,

employment, values and opinions, politics, the environment, finances, receipt of benefits, external transfers and household expenditure. The BHPS also collects a considerable amount of self-reports on health, including, for example, the General Health Questionnaire (GHQ) to measure mental health, whether participants have been admitted to hospital, or diagnosed with a cancer. In waves 9 and 14, the survey includes the 36-item Health Survey (SF-36).

Approximately one week before the start of Wave 18 of the BHPS, sample members were sent an advance letter to inform them that an interviewer would soon be calling on them to ask for an interview. Consent to health data linkage was asked for the first time at Wave 18 and an information leaflet detailing the plans to add administrative health data to the survey was also enclosed with the advance letter. The leaflet set out what information would be added, who will use the information, how long consent lasts and data security. Health data that were specifically mentioned in the information leaflet were admissions or attendance to hospital (including dates, diagnoses, treatments, surgical procedures, waiting times), records of specific conditions such as cancer or diabetes, prescriptions, and health registration information (such as name of Health Authority, NHS number, cause and date of death). There was also information for sample members about how they could revoke their consent at any time and gave a free-post address and a free-phone telephone number so they could contact ISER to ask any questions. Participants had the opportunity to request further information or to opt out of the survey on receipt of the advance letter.

Consents for health data linkages were collected as part of the individual questionnaires from adults (16+ years old) at the end of the interview. The procedure for asking for consent, along with the actual question wording and content of the information leaflet and the consent forms, had to be approved by the Medical Research Ethics Committee (MREC). The MREC aimed to ensure that the consent request was transparent, and that the respondents were fully informed about what they were consenting to.

During the interview the interviewer explained to the participant about the data linkage, gave the participant a permission form to sign and also answered any additional questions the participant may have had about the linkage. Signing the permission form indicated that the participant had read the information leaflet and had

the opportunity to ask questions about the process. On the form, there was a box for the participant to tick to indicate that they gave consent for data linkage for a particular stream of data, differentiated by which authorities hold the information. Participants were thus able to give consent to one stream of linkage and withhold their consent to another. The participants then had to sign and date each form.

A copy of the signed form was left with the participant for their own records. As well as receiving the signed consent form, the interviewer also coded in the CAPI questionnaire whether or not the participant gave consent. Once the fieldwork for Wave 18 of the BHPS was over, the data and the signed consent forms were returned to ISER. The consent forms were then checked against the CAPI data, to ensure that there had been no errors and that a signed consent form exists for all participants expected to have one according to the data. Where the CAPI data indicated consent, but there was no form, the CAPI data was edited to indicate that the form was missing and this was treated as not having consent.

All 13 454 adults who were interviewed were asked for their consent to health data linkage, and 5362 (41 percent) consented. The dependent variable used in this research is whether or not the participant signed a consent form to allow ‘Flagging or tracing on the NHS Central Registers’ or ‘Adding of Administrative health records’. The independent variables are grouped into four blocks and include not only population characteristics that may be readily available in all medical research, say from health registrations (variable set 1, see Table 1), but also characteristics that are not typically collected in medical surveys (variable sets 2-4, see Table 1). We report bivariate associations and estimate multivariate logistic regression models, allowing for spurious correlations in the characteristics. Results are weighted using population weights for the UK in 2008. The weights used are provided in the BHPS and account for unequal selection probabilities of addresses, non-response at the household level and non-response of individuals within responding households (Taylor (ed.) 2010). The consent rates are calculated in the statistical data analysis programme Stata using the command *mean* (StataCorp 2009). Adjusted Wald tests are performed to test for statistically significant differences in group means.

Finally, we estimate how many linked health records may be available, based on achieved consent rates. The calculation assumes that all those who have reported any hospitalisation over the course of the past 18 years do actually have a record on the

respective country's hospital episode database, and, based on experience with linking hospital episodes from England to survey data, that 90% of current year hospitalisations and 75% of earlier hospitalisations can be correctly identified as belonging to the consenter (Dr Foster Unit 2009). Indeed, respondents are likely to have more than one hospital episode per hospitalisation, and they may have been hospitalised repeatedly.

3. Results

Table 1 reports the number of observations and estimated consent rates broken down by population characteristics. Consent is significantly higher among people who live in England (Mean: 42.4; $F(1, 13\ 453)=56.91, p<.001$), participants aged 16-24 (M: 45.9; $F(1, 13\ 453)=10.43, p<.001$), and among those who consider their ethnicity to be British/Irish White (M: 42.4; $F(1, 13\ 453)=33.31, p<.001$). There is some indication that consent is associated with level of qualification. There is a significant difference in consent for people who have a higher degree (M: 49.2; $F(1, 13\ 203)=6.27, p<.05$), A-levels (M: 45.6; $F(1, 13\ 203)=8.51, p<.01$), a commercial qualification (M: 21.7; $F(1, 13\ 203)=8.43, p<.01$) or no qualifications (M: 36.8; $F(1, 13\ 203)=12.36, p<.001$). Consent is also higher among participants not living in standard household types (M: 50.5; $F(1, 13\ 453)=8.54, p<.01$), and among participants whose income falls in the third quartile of the income distribution (M: 43.6; $F(1, 13\ 453)=5.53, p<.05$). Self-reported health is generally not associated with consent. There are two exemptions; diabetes (M: 48.2; $F(1, 13\ 453)=7.67, p<.01$) and obesity (M: 45.5; $F(1, 13\ 453)=8.31, p<.05$) are associated with higher consent. Consent is higher for participants who reported that they have used health services in the previous 12 months (M: 42.8; $F(1, 13\ 453)=6.8, p<.01$). No other indicator of use of health services is associated with consent in the bivariate models.

Table 1 - Characteristics of consenters in study sample

Participant characteristic	Total N	% Informed Consent=yes	95% Confidence Interval	
			Lower bound	Upper bound
<i>Set 1: Standard demographics</i>				
Country of residence				
<i>England</i>	6633	42.4	41.2	43.7
<i>Wales</i>	2357	33.4	31.3	35.6
<i>Scotland</i>	2282	35.7	33.5	37.9
<i>Northern Ireland</i>	2182	36.2	34	38.3
Age group				
<i>Age 16-24</i>	1965	45.9	42.9	48.9
<i>Age 25-39</i>	3200	41.9	39.6	44.1
<i>Age 40-49</i>	2546	40.1	37.5	42.6
<i>Age 50-59</i>	1984	39.7	36.9	42.5
<i>Aged 60 or over</i>	3758	40.5	38.5	42.6
Gender				
<i>Female</i>	7385	40.7	39.2	42.2
<i>Male</i>	6069	42	40.3	43.6
Ethnic Identity				
<i>British/Irish White</i>	12 077	42.4	41.2	43.6
<i>Other</i>	1377	32.1	28.8	35.4
<i>Set 2: Socio-economic characteristics</i>				
Education				
<i>Higher degree</i>	425	49.2	42.9	55.6
<i>First degree</i>	1593	42.9	39.8	46.1
<i>Diploma in higher education</i>	4220	40.8	38.9	42.7
<i>A-levels</i>	1711	45.6	42.5	48.8
<i>O-level or equivalent</i>	2130	41.4	38.6	44.2
<i>GCSE</i>	663	41.3	36.3	46.2
<i>Commercial qualification, no O-level/GCSE</i>	67	21.7	8.5	34.9
<i>None/still at school</i>	2395	36.8	34.2	39.5
Household type				
<i>Single</i>	1930	41.9	39	44.7
<i>Couple without children</i>	4015	39.9	37.9	41.9
<i>Couple with children</i>	5799	41.2	39.5	42.9
<i>Lone parent</i>	1259	42.5	38.8	46.3
<i>Other</i>	451	50.5	44.2	56.8

Continues next page

Table 1 - Characteristics of consenters in study sample (continued)

Participant characteristic	Total N	% Informed Consent=yes	95% Confidence Interval	
Quartile of household income				
<i>Bottom quartile</i>	3325	41.3	39	43.5
<i>2nd quartile</i>	3425	39.9	37.7	42.1
<i>3rd quartile</i>	3524	43.6	41.4	45.8
<i>Top quartile</i>	3180	40.5	38.3	42.7
Set 3: Health conditions				
Considers self to be disabled	1397	40.1	36.6	43.5
Has health problem(s) relating to...				
<i>Arms</i>	3615	41.1	38.9	43.2
<i>Sight</i>	727	44.5	39.7	49.3
<i>Hearing</i>	1183	44.2	40.5	47.9
<i>Allergy</i>	1580	43.8	40.7	47
<i>Chest</i>	1793	41.3	38.3	44.3
<i>Heart</i>	2528	41	38.5	43.5
<i>Stomach</i>	1046	43.3	39.2	47.3
<i>Diabetes</i>	642	48.2	43.2	53.3
<i>Anxiety</i>	1119	38.2	34.3	42.1
<i>Alcohol</i>	74	40.1	24.4	55.8
<i>Epilepsy</i>	104	39.9	27.9	51.9
<i>Migraine</i>	924	44.5	40.2	48.7
<i>Cancer</i>	208	46.7	37.7	55.8
<i>Stroke</i>	158	43.3	33.3	53.3
<i>Other health problem</i>	689	40	35.1	44.9
Weight (BMI categories)				
<i>Underweight</i>	246	46.7	38.4	55
<i>Normal weight</i>	4915	40.3	38.5	42.1
<i>Overweight</i>	4121	40.7	38.8	42.7
<i>Obese</i>	1946	45.5	42.5	48.5
Set 4: Use of health services				
Covered by private medical insurance	1859	42.9	40.1	45.6
Used health services (last 12 months)	6235	42.8	41.2	44.5
Had a health check	12 142	41.5	40.3	42.6
Has visited GP in the last 12 months	10 252	40.7	39.4	42
Out-patient in the last 12 months	2333	42.4	39.7	45
Hospital in-patient (last 12 months)	1315	43.8	40.2	47.5
Hospital in-patient (previous 17 years)				
<i>Yes - at least once</i>	6536	41.9	40.4	43.5
<i>Yes – but never for childbirth</i>	5163	41.7	40	43.4
Total N	13 458	41.3	40.2	42.4

Source: British Household Panel Study Wave 18.

When we consider all four sets of independent variables in a multivariate logistic regression model, the negative associations between living in Wales, Scotland or Northern Ireland compared to living in England persist (Table 2). Moreover, those who consider their ethnicity British/Irish White have a 75% higher chance of consent. We find that the youngest group of participants (aged 16-24) in the BHPS have a 61% higher chance of giving consent than the oldest group (aged 60 or older). Males are slightly overrepresented among consenters (OR: 1.14, $p < .05$). Whilst having a higher degree is positively associated with consent (OR: 1.52, $p < .001$), household and socio-economic characteristics of the population generally are not associated with consent. Last but not least, out of the 20 subjective indicators of health only one is associated with consent in the multivariate models. We find that participants who report to have been diagnosed with diabetes (OR: 1.30, $p < .05$) are more likely to consent. The association between consenting and utilisation of health services is complex. While people who have recently used GP services are less likely to consent (OR: 0.84, $p < .001$), there is a positive association with use of health services in the last 12 months (OR: 1.15, $p < .05$), and with hospitalisations in a previous wave of the survey, where at least one hospitalisation was related to childbirth (OR: 1.24, $p < .05$). Recent hospitalisations and other indicators of use of health services are not associated with consent.

Table 2 - Multivariate logistic regression of consent to link administrative health records.

Participant characteristics	Odds Ratio	95% confidence interval
<i>Set 1: Demographic characteristics</i>		
Country of residence (comparison group: England)		
<i>Wales</i>	0.7	(0.62 - 0.79)**
<i>Scotland</i>	0.77	(0.68 - 0.87)**
<i>Northern Ireland</i>	0.79	(0.69 - 0.89)**
Age group (comparison group: 60 or older)		
<i>Age 16-24</i>	1.61	(1.26 - 2.06)**
<i>Age 25-39</i>	1.18	(0.99 - 1.42)
<i>Age 40-49</i>	1.04	(0.87 - 1.25)
<i>Age 50-59</i>	1.04	(0.88 - 1.23)
Male	1.14	(1.02 - 1.27)*
British/Irish White	1.75	(1.46 - 2.11)**

Continues next page

Table 2 - continued

Participant characteristics	Odds Ratio	95% confidence interval
<i>Set 2: Socio-economic characteristics</i>		
Has higher degree	1.52	(1.14 - 2.01)**
Household type (comparison group: Single)		
<i>Couple without children</i>	0.91	(0.77 - 1.07)
<i>Couple with children</i>	0.91	(0.76 - 1.10)
<i>Lone parent</i>	0.98	(0.78 - 1.24)
<i>Other</i>	1.16	(0.82 - 1.65)
Quartile of household income (comparison group: 1 st quartile)		
<i>2nd quartile</i>	0.92	(0.79 - 1.07)
<i>3rd quartile</i>	1.07	(0.90 - 1.26)
<i>Top quartile</i>	0.91	(0.76 - 1.09)
<i>Set 3: Health conditions</i>		
Considers self to be disabled	0.94	(0.78 - 1.14)
Has health problem(s)	1.16	(1.00 - 1.35)
Health problem relates to ...		
<i>Arms</i>	0.93	(0.81 - 1.06)
<i>Sight</i>	1.12	(0.90 - 1.40)
<i>Hearing</i>	1.16	(0.97 - 1.39)
<i>Allergy</i>	1.08	(0.92 - 1.28)
<i>Chest</i>	1	(0.86 - 1.17)
<i>Heart</i>	0.94	(0.81 - 1.09)
<i>Stomach</i>	1.07	(0.88 - 1.30)
<i>Diabetes</i>	1.3	(1.03 - 1.64)*
<i>Anxiety</i>	0.83	(0.68 - 1.01)
<i>Alcohol</i>	1.13	(0.50 - 2.57)
<i>Epilepsy</i>	0.86	(0.49 - 1.50)
<i>Migraine</i>	1.08	(0.87 - 1.32)
<i>Cancer</i>	1.31	(0.88 - 1.96)
<i>Stroke</i>	1.06	(0.68 - 1.67)
<i>Other health problem</i>	0.95	(0.76 - 1.20)
Weight (comparison group: <i>underweight</i>)		
<i>Normal weight</i>	0.82	(0.58 - 1.17)
<i>Overweight</i>	0.87	(0.61 - 1.25)
<i>Obese</i>	1.05	(0.73 - 1.53)

Continues next page

Table 2 - continued

Participant characteristics	Odds Ratio	95% confidence interval
Set 4: Use of health services		
Covered by private medical insurance	1.07	(0.93 - 1.23)
Used health services (last 12 months)	1.15	(1.02 - 1.30)*
Had a health check	1.06	(0.88 - 1.27)
Has visited GP (last 12 months)	0.84	(0.73 - 0.95)**
Out-patient (last 12 months)	0.96	(0.82 - 1.12)
Hospital in-patient (last 12 months)	1.05	(0.86 - 1.27)
Hospital in-patient (previous 17 years)		
<i>Yes - at least once</i>	1.24	(1.02 - 1.52)*
<i>Yes – but never for childbirth</i>	0.87	(0.72 - 1.06)

Significant at ***.001, **.01, * .05.

Source: British Household Panel Study (BHPS) Wave 18. N=11 021. $F(45, 8376) = 3.28, p < .001$.

Table 3 - Expected sample sizes for linked BHPS and health data.

Country	N	Consent rate [#]	NHSCR	HE## (current)	HE### (retrospective)	HE Total among W18 respondents
England	6613	42.3	2797	247	850	1096
Scotland	2357	35.2	829	73	252	325
Wales	2272	37.9	862	76	262	338
N. Ireland	2217	35.2	780	69	237	306
UK total	13 454	41.3	5362	473	1629	2102

Notes: Estimates are based on rates typically achieved when Hospital Episode Statistics for England are matched with name and address records of the current population. The rates were obtained in private correspondence with members of the Dr Foster Unit, Imperial College, London. Northern Ireland does not maintain a hospital episode database but similar information can be obtained from other sources.

[#] Percent of BHPS Wave 18 population that gives permission to linkage of NHS Central Register and hospital episode (HE) data.

^{##} Assumes that 9.8% of BHPS Wave 18 participants have a current year hospital episode record and 90% of them can be linked.

^{###} Assumes that 40.5% of BHPS Wave 18 participants only have a historic hospital record and 75% of them can be linked.

Table 3 reports the estimated sample sizes for linking to health records. Consent to link health records was obtained from 2860 adults in England, and from roughly 2500 adults in the other countries of the UK. Given these numbers, analysts may expect to find complete linked health and survey records for around 2100 adults. Whilst not the subject of this paper, adults who were ‘responsible’ for a child (aged 0-15) in the household, usually the mother, were also asked for their permission to link administrative health data held on the child(ren) to the survey responses. Around 1500

consents for adding children's health records were obtained. For 90% of them, complete birth-related information would be available.

4. Discussion

This research is the first population-based longitudinal study assessing the consent patterns in a population spanning the whole age range and in the context of a survey that is not specifically focused on health. We find that consent is biased only with respect to socio-demographic characteristics but not with respect to health and socio-economic characteristics. This is in stark contrast to the medical studies which report strong biases on health (Dunn et al. 2004; Gerber et al. 2007; Olson 1999; Woolf et al. 2000).

Consent rates are much lower than in medical studies. Currently only a minority of BHPS records can be linked to health records given relatively low consent rates, although this still gives more than 5300 adults who have consented to data linkage. While consent is not biased with respect to the socio-economic and demographic characteristics of the core BHPS sample, the research potential for health in minority ethnic groups is limited due to relatively low numbers in the sample. In line with previous research, the analysis shows that minority ethnic groups are less likely to consent (Ajzen 2005; Huang et al. 2007; Kho et al. 2009; Tate et al. 2006). Despite the lower consent rates in Scotland, Wales and Northern Ireland, the over-sample in these areas means that the BHPS would be a potentially valuable resource to analyse regional differences. Moreover, the value of the BHPS arises from having followed the same people over a very long period of time and observing their lifestyle and household context. Hospitalisations have been reported by 50% of the respondents to BHPS Wave 18; some utilisations of hospital services may not have been reported (Ritter et al. 2001). The study observes the respondents (in the context of their household) prior to the hospitalisation, and afterwards.

Since most study characteristics are not associated with consent, analysts do not have to worry too much about re-weighting the analysis to adjust for differential consenting in most cases. In a companion paper it was found, however, that many other characteristics of respondents such as markers of their risk aversion and community-mindedness are associated with consent (Sala et al. forthcoming). To the extent that

these characteristics may also be related to health outcomes, analysts may consider more carefully whether re-weighting is necessary.

It is planned to re-approach non-consenters and new entrants to the study with the consent question in a future interview, and knowing who is more reluctant to consent could help targeting resources on groups of the population that currently are underrepresented among the consenters. The study results on diabetes and cancer, two types of disease which were specifically mentioned in the BHPS information leaflet, suggest that consent rates may be boosted by mentioning in the information leaflet specific diseases which have a high prevalence in the study population. The result is in line with findings from the survey research showing that co-operation rates are higher if the study subject is more salient to the participant (Singer 2003).

5. Conclusions

There is an interest in the UK in linking survey data to hospital episode records and, in the longer run, to make use of Primary Care Trust and GP data. The results of this study suggest that it may be difficult to obtain consent for such data linkages from recent users of GP services. Therefore, one may want to reconsider plans to link to GP records, i.e., remove any references in the information leaflet to types of health data which people may associate with their GP (here: prescriptions). In the UK it is not currently possible to systematically link surveys to GP records. Institutions that manage surveys may be best advised to design more specific consent forms and survey instruments when such linkages have become feasible.

In 2010, the BHPS sample was incorporated into the new UK household longitudinal study, *Understanding Society*. For more information on this study see www.understandingsociety.org.uk. Future data linkage exercises on the survey will be informed by methodological research carried out on the Innovation Panel of *Understanding Society*. The Innovation Panel offers a unique opportunity to investigate the effect of different designs of survey instruments on survey outcomes, including informed consent to health data linkages.

6. References

- Ajzen, Icek (2005), *Attitudes, personality and behaviour* (Maidenhead: Open University Press).
- Baker, Richard, et al. (2000), 'What proportion of patients refuse consent to data collection from their records for research purposes?', *British Journal of General Practice*, 50 (457), 655-556.
- Dr Foster Unit (2009), 'Personal correspondence on expected matching rates of BHPS consenters to Hospital Episode Statistics (HES) using full name, address and exact date of birth as linking variables'.
- Dunn, Kate M., et al. (2004), 'Patterns of consent in epidemiologic research: Evidence from over 25,000 responders', *American Journal of Epidemiology*, 159 (11), 1087-94.
- Gerber, Yariv, et al. (2007), 'Participation bias assessment in a community-based study of myocardial infarction, 2002-2005', *Mayo Clinic Proceedings*, 82 (8), 933-38.
- Huang, Nicole, et al. (2007), 'Record linkage research and informed consent: who consents?', *BMC Health Services Research*, 7 (18).
- Kho, Michelle E., et al. (2009), 'Written informed consent and selection bias in observational studies using medical records: systematic review', *British Medical Journal*, 822.
- Newell, Sallie A., et al. (1999), 'The accuracy of self-reported health behaviors and risk factors relating to cancer and cardiovascular disease in the general population: A critical review', *American Journal of Preventive Medicine*, 17 (3), 211-29.
- Okura, Yuji, et al. (2004), 'Agreement between self-report questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure', *Journal of Clinical Epidemiology*, 57 (10), 1096-103.
- Olson, Janice A. (1999), 'Linkages with data from social security administrative records in the Health and Retirement Study', *Social Security Bulletin*, 62 (2), 73-85.
- Ritter, Philip L., et al. (2001), 'Self-reports of health care utilization compared to provider records', *Journal of Clinical Epidemiology*, 54 (2), 136-41.
- Sala, Emanuela, Burton, Jon, and Knies, Gundi (forthcoming), 'Correlates of obtaining informed consent to data linkage: Respondent, interview and interviewer characteristics', *Sociological Methods & Research*.
- Sauver, Jennifer L. St., et al. (2005), 'Agreement Between Patient Reports of Cardiovascular Disease and Patient Medical Records', *Mayo Clinic Proceedings*, 80 (2), 203-10.
- Singer, Eleanor (2003), 'Exploring the meaning of consent: Participation in research and beliefs about risks and benefits', *Journal of Official Statistics*, 19 (3), 273-85.
- StataCorp (2009), *Statistical Software: Release 11.0* (College Station, TX: Stata Corporation).

- Tate, A. Rosemary, et al. (2006), 'Mother's consent to linkage of survey data with her child's birth records in a multi-ethnic national cohort study', *International Journal of Epidemiology*, 35 (2), 294-98.
- Taylor (ed.), Marcia F. (2010), 'Weighting, Imputation and Sampling Errors', in John Brice, Nick Buck, and Elaine Prentice-Lane (eds.), *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices* (Colchester: University of Essex), A5-1-13.
- Woolf, Steven H., et al. (2000), 'Selection bias from requiring patients to give consent to examine data for health services research', *Archives of Family Medicine*, 9 (10), 1111-18.