Queen's University
Belfast

Large panel-survey data demonstrated country-level and ethnic minority variation in consent for health record linkage

Published in:
Journal of Clinical Epidemiology

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Large panel-survey data demonstrated country-level and ethnic minority variation in consent for health-record linkage

Sharon Mary Cruise[1, 2]

Lynsey Patterson[1, 2]

Chris R. Cardwell[1, 2]

Dermot O'Reilly[1, 2]*

[1] Centre for Public Health, Queen's University Belfast

[2] UKCRC Centre of Excellence for Public Health (NI)

*Corresponding author: Dr Dermot O'Reilly, Centre for Public Health, Institute of Clinical Sciences Block B, Queen's University Belfast, Royal Victoria Hospital, Grosvenor Road, Belfast, BT12 6BJ, Northern Ireland, United Kingdom. Email: d.oreilly@qub.ac.uk

ABSTRACT

Objective: To investigate individual, household and country variation in consent to health record linkage.

Study Design and Setting: Data from 50,994 individuals aged 16-74 years recruited to wave 1 of a large UK general purpose household survey (January 2009 – December 2010) were analysed using multi-level logistic regression models.

Results: Overall, 70.7% of respondents consented to record linkage. Younger age, marriage, tenure, car ownership and education were all significantly associated with consent, though there was little deviation from 70% in subgroups defined by these variables. There were small increases in consent rates in individuals with poor health when defined by self-reported long term limiting illness (adjusted OR 1.11; 95%CIs 1.06, 1.16), less so when defined by General Health Questionnaire score (adjusted OR=1.05; 95%CIs 1.00, 1.10), but the range in absolute consent rates between categories was generally less than 10%. Larger differences were observed for those of non-white ethnicity who were 38% less likely to consent (adjusted OR 0.62; 95%CIs 0.59, 0.66). Consent was higher in Scotland than England (adjusted OR 1.17; 95%CIs 1.06, 1.29) but lower in Northern Ireland (adjusted OR 0.56; 95%CIs 0.50, 0.63).

Conclusion: The modest overall level of systematic bias in consent to record linkage provides reassurance for record linkage potential in general purpose household surveys. However, the low consent rates amongst non-white ethnic minority survey respondents will further compound their low survey participation rates. The reason for the country-level variation requires further study.

Key words: Informed consent, health record linkage, selection bias, survey

Running title: Variation in consent for linkage to health records

Word count:

Abstract: 237

Text: 4439

What is new

Key findings / what this adds to what is known

- This large general purpose household survey showed little evidence of systematic variation in consent to link health records across most demographic and socioeconomic factors

- Consent to linkage was lower in non-white ethnic groups which will further compound the generally lower participation rates in these groups

- Marked variation between countries was evident despite the standardised survey methodology

What is the implication, what should change now?

- The need to maximise survey response rates is emphasised and there may be a need to further enhance the already boosted ethnic minority sample

- Further research is needed to understand and learn from the variations in levels of consent between nations

INTRODUCTION

Cross-sectional studies provide a snapshot of the characteristics of society at a point in time and are an efficient way to determine the prevalence, changing prevalence and social patterning of disease and associated risk factors in the population [1–3]. Repeated cross-sectional demographic and health surveys, using standardised data collection procedures across populations and consistent content over time, have been used to support evidence-based policy development, and in the planning and monitoring of health and development programmes in low- and middle-income countries [4]. A combination of interviews and health examinations can be also be used to investigate the prevalence of diagnosed and undiagnosed disease in the population [5,6].

Survey data can be further enhanced by linking to routine administrative data to create a longitudinal cohort, and as these routine administrative data are collected for other purposes this is an extremely efficient process with no additional burden to cohort members. Within the United Kingdom (UK), both the Health Survey for England [7] and the Scottish Health Survey [8] provide good examples of what can be achieved. For example, data from the Health Survey for England has been used to determine the role of elevated inflammatory markers in the higher rates of cardiovascular mortality amongst passive smokers [9] and the influence of low levels of physical activity on coronary heart disease risk amongst UK-born South Asians [10]. More recently, analysis of pooled data from eleven independent Health Surveys for England (from 1994-2004) has demonstrated that even modest elevations in psychological distress were associated with subsequent all-cause mortality, and mortality from cardiovascular disease and external causes [11]. In

Scotland, health survey data linked with hospital data has been used to determine the social and lifestyle factors related to risk of psychiatric admission [12] and the transgenerational relationship between birthweight and maternal cardiovascular risk [13].

However, in the UK, informed consent from the survey respondent is a prerequisite for subsequent record linkage; therefore, any ensuing variation in consent rates may introduce a selection bias that could affect the external validity of the study findings, and accentuate or confound any participation bias in the original survey. Participation bias has been reasonably well characterised, and studies have demonstrated that respondents participating in surveys are more likely to be female [14]; be of higher socioeconomic status and employed [15–17]; and to be educated [18] and married [15,18]. Some authors have found higher participation in older age groups whilst others did not (see [19] for review). The findings for ethnicity are also unclear, with some reporting higher response rates in whites and others reporting similar response rates across all ethnic groups [18,20]. Furthermore, studies have shown that responders in general are also likely to have better health [21–24] and to have more healthy lifestyles [25,26]. However the salience of the survey is also important as people are more likely to respond if the focus of the survey has particular relevance to them, except where the condition is perceived as being stigmatising [27].

Although some studies have examined predictors of consent to record linkage, they are far fewer than those focused on survey response, and with a few exceptions [14,28] they are either among smaller, non-representative samples, or they are

among samples that are larger but survey-specific (e.g., a female only sample). Factors associated with consent appear to be similar to those associated with survey response, but there are inconsistencies. For example, some studies show a gradient of increased likelihood of consent with increasing age [29,30], while others found that older respondents are less likely to consent [14,31]. Higher socioeconomic position (i.e., socioeconomic status, educational level, car ownership, having private health insurance) is generally associated with higher rates of consent to record linkage [32,33]. However, in a large, nationally representative general survey Knies et al. [28] found no effects for income. There is also evidence that those with health problems and higher levels of primary health care utilisation have higher rates of consent [14,28,30,33]. Given the paucity of studies utilising large population-based samples, the limited numbers of factors considered within studies, and evidence of inconsistencies in the findings, there is a rationale for further research in this area.

Therefore, the aim of the current study is to examine socio-demographic and health factors associated with not consenting to record linkage amongst respondents to a representative general population survey, and to compare these to the factors known to be related to survey response. Of particular interest is additional variation in consent rates between countries in the UK. Published and unpublished data from health surveys have demonstrated that while the consent rates for record linkage have remained high at approximately 90% in Scotland [8], the rates in England have declined from 96% in 1994 to 78% in 2009 (for linkage to cancer and death data) [7], while only 64% of respondents to the equivalent survey in Northern Ireland in 2005/06 consented to record linkage. However, such differences may have arisen due to variation in survey design, the wording or the location of the consent question

within the questionnaire, or the general context of the survey. The present study

utilises one of the world's largest longitudinal household surveys, the Understanding

Society survey, for which data is collected during an annual interview. Respondents

are asked for their consent to linkage of their survey responses to administrative

data related to health, education, economic circumstances and transport. The

advantage of Understanding Society, in addition to its large size, is the use of a

standardised methodology across the four countries of the UK.

METHOD

Data source

The data for this analysis was collected during wave 1 of the main-stage

Understanding Society survey, which was carried out over 24 months between

January 2009 and December 2010. This comprised a general population and an

ethnic minority boost sample which was collected using a proportionately stratified

(equal probability), clustered sample of addresses in England, Scotland and Wales,

and (because of its smaller population) a non-clustered systematic random sample

of addresses in Northern Ireland. Overall, 57.6% of households in the general

population sample took part in the survey and 81.8% of eligible residents aged 16

years or over in these households completed the individual interview. The

equivalent response rates in the ethnic minority boost sample were lower (52.0%

and 71.9% respectively) [34]. Analysis conditional on household response showed

that the response rate was higher in females than in males, increased with age, was

higher in white than non-white residents and was inversely related to car ownership

[34]. In total, 30,169 households participated, with 59,436 respondents aged 16

years and over being eligible for the full interview, and a final total of 50,994

respondents taking part in either the full or proxy interview.

Respondents were asked to consent to linkage of two types of health data; the first

sought agreement that the respective elements from the National Health Service, the

Departments of Health, the General Registration Offices and the Office for National

Statistics could disclose information about the individual's health treatment and use

of health services for future research studies. The second authorised the access of

information about the individual's National Health Service registration from the

National Health Service Central Register so as to follow up on registration and health

status (i.e., about date and cause of death). The request for consent to linkage was

worded as follows, "... *we would like to add some information from administrative*

health records to the answers you have given. This leaflet gives you information

about what we would like to do. Please read it, ask me any questions and sign the

form if you are happy for us to do this". The leaflet detailed the types of information

that could be linked, and explained that personal details would be removed from the

data supplied to potential researchers and that access to sensitive information was

only provided with legally binding licences. Respondents were also informed that

they were free to withdraw at any time and that their current or future dealings with

the health service would not be affected.

The overwhelming majority of respondents gave the same response to the two

consent questions, with 67% agreeing to both and 31% declining to both. The

results that are presented hereafter relate to the authorisation of access to

information on an individual's health treatment and use of health services; the

models for access to registration data were almost identical to those presented below for the consent to linkage to health and treatment data and are available on request from the corresponding author.

Other covariates

A range of individual and household characteristics that are known from other studies to influence initial participation were also included in the analysis. These included the following socio-demographic features: age (grouped into nine-year age bands from 16-24 through to 65-74), gender, ethnicity (white/non-white) and marital status (never married, married, separated/divorced/widowed); and three measures of socioeconomic status: housing tenure (own/rental), the number of cars available to the household (none, one, two or more) and respondent's educational attainment (other, secondary, higher education/degree). Two measures of health status were available: a question on limiting and long term illness (LLTI), which is known to be primarily related to physical aspects of health [35] (with a yes/no response), and the General Health Questionnaire (GHQ-12), which is a well validated, self-report measure of mental and psychological well-being for use in population studies [36]. High scores represent poorer mental health, and a threshold is often chosen to indicate the presence of significant psychological morbidity (i.e., corresponding to the average case that might be of clinical significance). A threshold of three or more was used here, though a sensitivity analysis using a threshold of four or more was also tested. Preliminary descriptive analysis indicated larger proportions of missing data for the GHQ than for other variables, and suggested that in the majority of cases this was as a result of non-participation in the self-completion component of the survey (which includes the GHQ items). Therefore, in order to adjust for any bias

as a result of non-random missingness in the GHQ, all regression models included a variable which indicated whether the respondent had participated in the self-completion component of the survey (no/yes). Body mass index (BMI) was available as a derived variable in the dataset, and was based on self-reported height and weight, and categorised according to the World Health Organisation recommendations (i.e., underweight, normal, overweight and obese). Four separate countries of residence were identified: England, Wales, Scotland and Northern Ireland.

Analytic strategy

Individuals who refused (n=1), responded 'don't know' (n=91) or for whom the response to the question on consent to health record linkage was missing (n=162) were removed from the dataset[1], as were those where a proxy respondent had been used (n=3,085) and for whom no consent was possible. For all covariates (with the exception of the missing responses in the GHQ-12 and BMI), responses that were coded as missing, not applicable, refused or 'don't know' were also removed (n=319). This approach was adopted because they represented a small proportion of the data and deleting them was unlikely to affect any inferences made. The missing category was retained for the GHQ-12 and BMI because these categories represented a larger proportion of the respondents (15.9% and 6.6% respectively). For the GHQ-12 and BMI we compared respondents with data missing to those with complete data. Covariates were compared using the t test for continuous variables and χ^2 tests for categorical variables (unweighted data). For the GHQ-12,

---

[1] We re-ran the analysis including refusals and 'don't know' cases as non-consenters. However, this made little difference to results; therefore, the analysis presented here excludes these cases. Results including the refusals and 'don't know' cases are available from the corresponding author on request.

respondents with missing data were more likely to be non-white, be tenants, be separated, have no cars, no education, have LLTI, have missing BMI, and reside in England or Northern Ireland. For BMI, respondents with missing data were younger, more likely to be male, never married, non-white, have no cars, have no education or secondary education and to be missing GHQ-12 data. The final dataset included 43,709 respondents (92.3% of the original).

A descriptive analysis of characteristics of individuals who agreed to consent to health record linkage was presented, using weighted data to give population representative proportions. To examine the association between the predictors and consent to health record linkage a multi-level logistic regression model was used which incorporated the hierarchical structure of the data (individuals nested within households). Initially, univariate models were performed before including all predictors in the full multivariate model. The 'empty' model, which includes only the random parameter, was used to calculate the variance partition coefficient using the linear threshold model [37]. The variance partition coefficient describes the proportion of the total variance in the outcome that is explained by the household level. The proportion of household members with concordant and discordant responses to the health linkage question was also determined. Generalised estimating equation logistic regression population average models were used to make inferences concerning area characteristics. Sensitivity analyses were performed on the following: 1) follow-up on health status as the outcome; 2) one individual randomly selected from each household to remove the household clustering effect; and 3) white ethnicity only. All statistical analyses were conducted using Stata.

RESULTS

Overall, 29,917 (70.7%) of the respondents in wave 1 of the Understanding Society survey gave consent to record linkage. The characteristics of consenters and non-consenters for the total sample and for the sample stratified by country are presented in Tables 1 and 2 respectively. Approximately 70% of most age groups consented to linkage, though this was highest (75%) in the youngest age group (16-24 years). Similar proportions of males and females consented to linkage (71% and 70% respectively) and across all categories of marital status, but the proportions were higher amongst white (72%) than non-white respondents (59%). No gradients in consent rates were evident across either housing tenure or according to the number of cars in the household, but a non-linear relationship was evident for educational attainment, with a greater proportion of those with secondary level education consenting than those with either higher or lower levels of attainment. Seventy-four percent of respondents who participated in the self-completion component of the survey also gave consent to record linkage. Individuals with poor health, reflected in LLTI and a GHQ-12 score of 3 or more, were slightly more likely to consent to linkage than those with better health, and consent tended to increase according to increasing BMI category. Consent rates were particularly low amongst respondents with missing values for either BMI or GHQ-12 scores. There were notable differences in consent across the UK, with over 70% of respondents in England, Scotland and Wales consenting compared to 58% (704 respondents) in Northern Ireland. When stratified by country, the socio-demographic associations for England were broadly aligned with those for the UK as a whole, which is not surprising given the proportion from England. By comparison, the models for Wales

and Scotland showed higher rates of consent across all covariates, whilst the model

for Northern Ireland had the lowest rates of consent (see Table 2).

< INSERT TABLE 1 ABOUT HERE >

< INSERT TABLE 2 ABOUT HERE >

The multivariate analysis accentuates the differences between the youngest age

group (16-24 years) and the rest, with overlapping confidence intervals for all older

age groups. The difference between white and non-white respondents was

attenuated, with non- white respondents over 38% less likely to consent to linkage

(OR 0.62; 95%CIs 0.59, 0.66). After adjustment for age the slightly higher likelihood

of consent amongst currently or previously married respondents became apparent,

and in the fully adjusted models the more affluent, as assessed by housing tenure

and car availability, were less likely to agree to record linkage, though the differences

in odds ratios were all 10% or less. Those participating in the self-completion

component of the survey had an 88% increased likelihood of consenting to record

linkage after adjustment for all other variables (OR 1.88, 95%CIs 1.62, 2.18). The

relationship between poor health and consent was attenuated by adjustment for

other factors, but those with a LLTI were still 11% more likely to consent than those

without (OR 1.11, 95%CIs 1.06, 1.16). Further analysis omitting the respondents

with missing BMI data confirmed a linear trend with increasing likelihood of consent

with increasing BMI (P<0.001).

The univariate analysis showed that respondents in Wales and Scotland were

respectively 18% and 35% more likely than their peers in England to consent to

record linkage, but these differences were approximately halved to 7% and 17% after adjustment for other covariates. The fully adjusted model confirmed the much lower likelihood of respondents in Northern Ireland agreeing to record linkage (OR 0.56; 95%CIs 0.50, 0.63). Northern Ireland was also a little different in that consent was positively associated with higher educational level (OR 1.29, 95%CIs 1.02, 1.65) and negatively associated with poorer mental health (OR 0.77, 95%CIs 0.62, 0.96) (see Table 2). Unlike England, Wales, and Scotland, there were no significant effects for ethnicity in Northern Ireland which is likely to be due to the small proportion of non-white ethnicities in this region with resultant small cell sizes.

Further sensitivity analyses were undertaken and models excluding non-white respondents produced essentially identical between-country differences. In addition, similar effect sizes were observed for all covariates in an analysis restricted to one individual randomly selected from each household. Finally, using consent to follow-up on health (vital) status as the dependent variable, the models were almost identical, as expected. These results have not been presented but can be obtained from the corresponding author on request.

Household level effects

The level 2 variance from the empty model was estimated at 9.90 (95%CIs 9.27, 10.57). A variance partition coefficient of 75% was calculated indicating that the differences between households were important in terms of consent to health record linkage. An analysis of two-member households illustrates this high level of concordance, with 62.5% of the households where both members said yes, 21.9%

where both said no, and only 15.7% where one member said yes and the other said no.

DISCUSSION

The study showed that overall levels of consent for linkage to health registration and medical records in this general purpose population survey were moderately high (71%) with little variation amongst survey respondents across most of the demographic, socioeconomic and health characteristics, the only exception being ethnicity and country of residence. This should provide some reassurance and suggests that once an individual responds to the survey the decision to consent to further linkage is unlikely to compound any initial bias resulting from variation in response rates.

The overall consent rate observed in the current study is consistent with findings from a recent systematic review which found that 8 out of 11 studies reviewed had proportions of consent greater than 72% [38], the other three having proportions of 53% or lower. We found the highest consent rate in the 16-24 age range, with little difference across the rest of the age spectrum, and although the association between age and consent in other studies has been inconsistent [38], this may reflect a greater willingness to share information secondary to greater exposure to and use of social media among this age group. The study also confirms the positive association between poorer health and consent to record linkage [14,30], which may arise because individuals with poor health, who are utilising the health service, can readily appreciate the benefits of linkage or want to give something back to the service.

One of the larger differences observed in this study was between white and non-white respondents. The absolute difference was approximately 13% but in the fully adjusted model non-white respondents were 38% less likely to consent to record linkage than their white peers. This is likely to further compound the difficulties the Understanding Society study has in achieving sufficient numbers from non-white populations, as both the household and individual survey response rates in the ethnic boost sample (52% and 72% respectively) were lower than for the general population sample (58% and 82%) [34]. The lower consent in ethnic minority groups has been noted in some other studies [39,40], and while this could be attributed to cultural differences, including a suspicion of health research, a more detailed examination of the underlying causes was beyond the scope of the present study. However, further research is planned that will explore reasons for the lower consent rates among non-white ethnic minorities. This analysis will utilise information available within the Understanding Society dataset that relates specifically to ethnic minorities (e.g., immigrant generation, whether the interview was carried out in respondent's first language, date when respondent arrived in the UK), as well as providing a more detailed examination by using a less aggregated ethnicity group variable than was used in the present study.

Marked differences in rates of consent were also evident between the four UK countries, with levels highest in Scotland and lowest in Northern Ireland. The reasons for these differences are not immediately apparent, though they mirror the rates of consent to linkage within the country-specific health surveys. However, while the variations in consent rates between national health surveys might have

been explained by variations in survey content or design, the standardised approach and questionnaire used in the Understanding Society survey rules out local methodological variation as a cause. However, a number of studies have established that specific interviewer variables (e.g., age, education, quality indicators) can contribute to variance in consent to data linkage rates [41–44]. Analysis to determine the role of interviewer bias as a non-random source of regional variation was not considered in the present study but is the focus of ongoing research. The higher proportions consenting in Scotland and to a lesser extent Wales may reflect a greater awareness, appreciation and investment in data with the SHIP (the Scottish Health Informatics Project) [45] and SAIL (Secure Anonymised Information Linkage) [46] initiatives. Other factors, such as variations in the individual's privacy concerns are known to be important [42,43] and this may be particularly relevant to the Northern Ireland population which might have been sensitised about use of personal information as a result of 30 years of civil unrest, colloquially known as the Troubles. Further qualitative studies would be needed to unravel these differences.

Strengths and limitations

Understanding Society has some significant strengths and limitations that need to be mentioned. It is one of the largest studies of its kind in the UK and has been designed with the aim of capturing much additional information through record linkage. The size ensured robust estimates of known attributes associated with consent as well as some new ones such as country of residence. The evidence of high consent rates in Scotland sets a target for the rest of the UK and suggests that there are elements of good practice which could be shared. The uniformity of survey

methodology across the UK addressed some of the questions posed by individual health surveys. Unlike most other studies in the field, the analysis took account of clustering within the household, and the high levels of agreement between household members is a reminder that such survey data should also be analysed with a recognition of the clustered nature of the data.

On the negative side, the analysis was necessarily limited to the respondent characteristics captured in the original survey. However, one important facet that could help to explain inter-country variation is the potential influence of the survey interviewers who are a critical link between survey design and implementation. From a methodological perspective it is important to estimate how much of the observed variation was occurring at this level; therefore, this is the focus of ongoing research using interview/interviewer characteristics data available in Understanding Society. This study was focused on consent by respondents for linkage of survey responses to their own health data. However, respondents are also asked for consent to linkage to their children's health and education records, and research examining factors associated with parental consent using a general purpose household survey such as Understanding Society in comparison with bespoke birth cohort studies such as the Millennium Cohort Study is warranted. No attempt was made to account for the design of the survey in the analytical approach because of the mixed methods used to derive the study sample within each country. However, the multi-level models used will adjust the standard errors making the inferences slightly more conservative than a standard logistic regression model. Finally, these findings relate to a general survey where health was only one of a range of dimensions explored and may not necessarily apply to health-specific studies.

Implications

One of the most frequently expressed concerns about differential loss to follow-up is that it may threaten the validity of results from observational studies [14], as biased estimates of association could arise if follow-up was related to both exposure and outcome. The inference is that representativeness is therefore as important for the follow-up cohort as it is for the initial survey, where the aim was to estimate the occurrence of a disease or risk factor in the population. However, follow-up studies are more usually concerned with questions of aetiology and the importance of representativeness in this context is hotly debated [47–52]. In practice, many [53,54], though not all [55], studies find that estimates of effect size amongst respondents consenting to follow-up/linkage differ very little from those found in the full population. However, low rates of consent to linkage, combined with the effects of initial low response rates, can substantially diminish sample size and potentially erode the robustness of any subsequent research and the ability to make clear statements about important sub-populations, such as ethnic minorities. It is therefore important that as much as possible is done to maximise both response rates to the initial study [19] and consent rates amongst those who are recruited. The potential for selection bias invalidating the study findings has also prompted some researchers to question the universal need for mandatory consent for linkage to medical and other records [38,56,57].

CONCLUSION

Cross-sectional and cohort studies increasingly incorporate record linkage as an efficient and cost-effective means of capturing additional information about

respondent attributes and outcomes. This study shows that the levels of consent for

such linkages are high and generally homogenous across most socio-demographic

and health factors, providing reassurance that levels of bias are minimal. However,

the low levels of consent amongst non-white ethnic minorities, aligned with their

lower survey participation rates may under-power later analysis. The variation

between countries is unexplained, but this might provide opportunities for improving

overall consent rates by learning from areas where they are highest.

REFERENCES

[1]   Zaninotto P, Head J, Stamatakis E, Wardle H, Mindell J. Trends in obesity
        among adults in England from 1993 to 2004 by age and social class and
        projections of prevalence to 2012. J Epidemiol Community Health
        2009;63:140–6.

[2]   Stamatakis E, Wardle J, Cole TJ. Childhood obesity and overweight
        prevalence trends in England: evidence for growing socioeconomic disparities.
        Int J Obes (Lond) 2010;34:41–7.

[3]   Stamatakis E, Zaninotto P, Falaschetti E, Mindell J, Head J. Time trends in
        childhood and adolescent obesity in England from 1995 to 2007 and
        projections of prevalence to 2015. J Epidemiol Community Health
        2010;64:167–74.

[4]   Corsi DJ, Neuman M, Finlay JE, Subramanian S V. Demographic and health
        surveys: a profile. Int J Epidemiol 2012;41:1602–13.

[5]   Shahab L, Jarvis MJ, Britton J, West R. Prevalence, diagnosis and relation to
        tobacco dependence of chronic obstructive pulmonary disease in a nationally
        representative population sample. Thorax 2006;61:1043–7.

[6]   Roth M, Roderick P, Mindell J. Chapter 3. Kidney disease and renal function.
        In: Craig R, Hirani V, editors. Heal. Surv. Engl. 2009, Leeds: NHS Information
        Centre: 2010.

[7]   Mindell J, Biddulph JP, Hirani V, Stamatakis E, Craig R, Nunn S, et al. Cohort
        profile: the health survey for England. Int J Epidemiol 2012;41:1585–93.

[8]   Gray L, Batty GD, Craig P, Stewart C, Whyte B, Finlayson A, et al. Cohort

Profile: The Scottish Health Surveys Cohort: linkage of study participants to

routinely collected records for mortality, hospital discharge, cancer and

offspring birth characteristics in three nationwide studies. Int J Epidemiol

2010;39:345–50.

[9]   Hamer M, Stamatakis E, Kivimaki M, Lowe GD, Batty GD. Objectively

measured secondhand smoke exposure and risk of cardiovascular disease:

what is the mediating role of inflammatory and hemostatic factors? J Am Coll

Cardiol 2010;56:18–23.

[10]   Williams ED, Stamatakis E, Chandola T, Hamer M. Physical activity behaviour

and coronary heart disease mortality among South Asian people in the UK: an

observational longitudinal study. Heart 2011;97:655–9.

[11]   Russ TC, Stamatakis E, Hame M, Starr JM, Kivimaki M, Batty GD. Association

between psychological distress and mortality : individual participant pooled

analysis of 10 prospective cohort studies. BMJ 2012;4933:1–14.

[12]   Stewart C, Titterington D, Leyland A. The association between the GHQ-12

and psychiatric hospital admission in Scotland [poster]. 2007.

[13]   Leyland AH. Making an Impact: The importance of routine data sources for

population health research. 2007.

[14]   Dunn KM, Jordan K, Lacey RJ, Shapley M, Jinks C. Patterns of consent in

epidemiologic research: evidence from over 25,000 responders. Am J

Epidemiol 2004;159:1087–94.

[15]   Shahar E. The effect of nonresponse on prevalence estimates for a referent population: Insights from a population-based cohort study. Ann Epidemiol 1996;6:498–506.

[16]   Eagan TML, Eide GE, Gulsvik A, Bakke PS. Nonresponse in a community cohort study: predictors and consequences for exposure-disease associations. J Clin Epidemiol 2002;55:775–81.

[17]   Hille ETM, Elbertse L, Gravenhorst JB, Brand R, Verloove-Vanhorick SP. Nonresponse bias in a follow-up study of 19-year-old adolescents born as preterm infants. Pediatrics 2005;116:e662–6.

[18]   Partin MR, Malone M, Winnett M, Slater J, Bar-Cohen A, Caplan L. The impact of survey nonresponse bias on conclusions drawn from a mammography intervention trial. J Clin Epidemiol 2003;56:867–73.

[19]   Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007;17:643–53.

[20]   Sykes LL, Walker RL, Ngwakongnwi E, Quan H. A systematic literature review on response rates across racial and ethnic populations. Can J Public Health 2010;101:213–9.

[21]   Bisgard KM, Folsom AR, Hong CP, Sellers TA. Mortality and cancer rates in nonrespondents to a prospective study of older women: 5-year follow-up. Am J Epidemiol 1994;139:990–1000.

[22] Kauppi M, Sokka T, Hannonen P. Survey nonresponse is associated with increased mortality in patients with rheumatoid arthritis and in a community population. J Rheumatol 2005;32:807–10.

[23] Gray L, McCartney G, White I, Rutherford L, Katikireddi S, Leyland A. A novel use of record linkage: resolving non-representativeness in health surveys and improving alcohol consumption estimates to inform strategy evaluation. Lancet 2012;380:S42.

[24] Vercambre M-N, Gilbert F. Respondents in an epidemiologic survey had fewer psychotropic prescriptions than nonrespondents: an insight into health-related selection bias using routine health insurance data. J Clin Epidemiol 2012;65:1181–9.

[25] Cunradi CB, Moore R, Killoran M, Ames G. Survey nonresponse bias among young adults: the role of alcohol, tobacco, and drugs. Subst Use Misuse 2005;40:171–85.

[26] Stang A, Moebus S, Dragano N, Beck EM, Möhlenkamp S, Schmermund A, et al. Baseline recruitment and analyses of nonresponse of the Heinz Nixdorf Recall Study: identifiability of phone numbers as the major determinant of response. Eur J Epidemiol 2005;20:489–96.

[27] McNutt LA, Lee R. Intimate partner violence prevalence estimation using telephone surveys: understanding the effect of nonresponse bias. Am J Epidemiol 2000;152:438–41.

[28] Knies G, Burton J, Sala E. Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population. BMC Health Serv Res 2012;12:52.

[29] Jacobsen SJ, Xia Z, Campion ME, Darby CH, Plevak MF, Seltman KD, et al. Potential effect of authorization bias on medical record research. Mayo Clin Proc 1999;74:330–8.

[30] Beebe TJ, Ziegenfuss JY, Jenkins SM, Haas LR, Davern ME. Who doesn't authorize the linking of survey and administrative health data? A general population-based investigation. Ann Epidemiol 2011;21:706–9.

[31] Huang N, Shih S-F, Chang H-Y, Chou Y-J. Record linkage research and informed consent: who consents? BMC Health Serv Res 2007;7:18.

[32] Young a F, Dobson a J, Byles JE. Health services research using linked records: who consents and what is the gain? Aust N Z J Public Health 2001;25:417–20.

[33] Harris T, Cook DG, Victor C, Beighton C, Dewilde S, Carey I. Linking questionnaires to primary care records: factors affecting consent in older people. J Epidemiol Community Health 2005;59:336–8.

[34] Lynn P, Burton J, Kaminska O, Knies G, Nandi A. An Initial Look at Non-Response and Attrition in Understanding Society. 2012.

[35] Cohen G, Forbes J, Garraway M. Interpreting self reported limiting long term illness. BMJ 1995;311:722–4.

[36]   Goldberg D, Williams P. Users guide to the General Health Questionnaire. Windsor: NFER-Nelson; 1988.

[37]   Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. J Epidemiol Community Health 2006;60:290–7.

[38]   Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. BMJ 2009;338:b866–b866.

[39]   Tate AR, Calderwood L, Dezateux C, Joshi H. Mother's consent to linkage of survey data with her child's birth records in a multi-ethnic national cohort study. Int J Epidemiol 2006;35:294–8.

[40]   Woolf SH, Rothemich SF, Johnson RE, Marsland DW. Selection bias from requiring patients to give consent to examine data for health services research. Arch Fam Med 2000;9:1111–8.

[41]   O'Muircheartaigh C, Campanelli P. A multilevel exploration of the role of interviewers in survey non-response. J R Stat Soc Ser A 1999;162:437–46.

[42]   Sakshaug JW, Couper MP, Ofstedal MB, Weir DR. Linking survey and administrative records: Mechanisms of consent. Sociol Methods Res 2012;41:535–69.

[43]   Sala E, Burton J, Knies G. Correlates of obtaining informed consent to data linkage: respondent, interview, and interviewer characteristics. Sociol Methods Res 2012;41:414–39.

[44]   Korbmacher JM, Schroeder M. Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer. Surv Res Methods 2013;7:115–31.

[45]   SHIP. ScottisH Informatics Programme (SHIP) 2013.

[46]   SAIL. Secure Anonymised Information Linkage Databank 2013.

[47]   Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013;42:1012–4.

[48]   Richiardi L, Pizzi C, Pearce N. Commentary: Representativeness is usually not necessary and often should be avoided. Int J Epidemiol 2013;42:1018–22.

[49]   Nohr EA, Olsen J. Commentary: Epidemiologists have debated representativeness for more than 40 years--has the time come to move on? Int J Epidemiol 2013;42:1016–7.

[50]   Elwood JM. Commentary: On representativeness. Int J Epidemiol 2013;42:1014–5.

[51]   Ebrahim S, Davey Smith G. Commentary: Should we always deliberately be non-representative? Int J Epidemiol 2013;42:1022–6.

[52]   Rothman KJ, Gallacher JEJ, Hatch EE. Rebuttal: When it comes to scientific inference, sometimes a cigar is just a cigar. Int J Epidemiol 2013;42:1026–8.

[53]   Reijneveld SA, Stronks K. The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data. Int J Epidemiol 1999;28:1134–40.

[54]   Winding TN, Andersen JH, Labriola M, Nohr EA. Initial non-participation and loss to follow-up in a Danish youth cohort: implications for relative risk estimates. J Epidemiol Community Health 2013.

[55]   Greene N, Greenland S, Olsen J, Nohr EA. Estimating bias from loss to follow-up in the Danish National Birth Cohort. Epidemiology 2011;22:815–22.

[56]   Macleod U, Watt GCM. The impact of consent on observational research: a comparison of outcomes from consenters and non consenters to an observational study. BMC Med Res Methodol 2008;8:15.

[57]   Da Silva MEM, Coeli CM, Ventura M, Palacios M, Magnanini MMF, Camargo TMCR, et al. Informed consent for record linkage: a systematic review. J Med Ethics 2012;38:639–42.

Table 1: Socio-demographic and socioeconomic characteristics of Understanding Society Wave 1 respondents and the likelihood of them consenting to linkage of administrative health records to survey responses

	n (%)†	Consent (%)	Odds Ratios (95%CIs)ⱦ	
			Unadjusted	Fully adjusted$
TOTAL	42307	29917 (70.7)		
Age				
16-24	6897 (16.3)	5152 (74.7)	1	1
25-34	7495 (17.7)	5244 (70.0)	0.86 (0.80, 0.91)	0.83 (0.77, 0.90)
35-44	8300 (19.6)	5759 (69.4)	0.85 (0.80, 0.90)	0.77 (0.71, 0.83)
45-54	7846 (18.6)	5537 (70.6)	0.88 (0.83, 0.93)	0.75 (0.70, 0.82)
55-64	6809 (16.1)	4749 (69.8)	0.87 (0.81, 0.93)	0.71 (0.65, 0.78)
65-74	4959 (11.7)	3476 (70.1)	0.94 (0.87, 1.01)	0.75 (0.68, 0.83)
Sex				
Male	20992 (49.6)	14964 (71.3)	1	1
Female	21315 (50.4)	14952 (70.2)	0.98 (0.95, 1.01)	0.98 (0.95, 1.02)
Ethnicity				
White	38094 (90.0)	27435 (72.0)	1	1
Non-white	4214 (10.0)	2482 (58.9)	0.57 (0.54, 0.60)	0.62 (0.59, 0.66)
Marital Status				
Never married	10772 (25.5)	7628 (70.8)	1	1
Married	26952 (63.7)	19045 (70.7)	0.98 (0.94, 1.02)	1.12 (1.06, 1.20)
Sep/div/widow	4584 (10.8)	3243 (70.8)	1.02 (0.96, 1.09)	1.15 (1.06, 1.25)
Housing tenure				
Rental	13440 (31.8)	9581 (71.3)	1	1
Own	28868 (68.2)	20335 (70.4)	0.98 (0.93, 1.03)	0.90 (0.84, 0.95)
Number of cars				
0	7161 (16.9)	5036 (70.3)	1	1
1	16692 (39.5)	11755 (70.4)	1.01 (0.95, 1.07)	0.93 (0.87, 1.00)
2+	18454 (43.6)	13125 (71.1)	1.07 (1.00, 1.14)	0.95 (0.87, 1.03)
Education				
None	9819 (23.2)	6781 (69.1)	1	1
Secondary level	18744 (44.3)	13630 (72.7)	1.15 (1.09, 1.20)	1.09 (1.04, 1.15)
Higher /degree	13745 (32.5)	9506 (69.2)	1.02 (0.97, 1.07)	1.00 (0.94, 1.06)
LLTI				
No	28213 (66.7)	19724 (69.9)	1	1
Yes	14094 (33.3)	10193 (72.3)	1.13 (1.08, 1.17)	1.11 (1.06, 1.16)
Self-completion				
No	5086 (12.0)	2520 (49.6)		1
Yes	37222 (88.0)	27397 (73.6)	2.56 (2.42, 2.71)	1.88 (1.62, 2.18)
GHQ score				
<3	28132 (66.5)	20611 (73.3)	1	1
≥3	8497 (20.1)	6410 (75.4)	1.06 (1.01, 1.11)	1.05 (1.00, 1.10)
Missing	5678 (13.4)	2895 (51.0)	0.43 (0.41, 0.45)	0.79 (0.69, 0.91)
BMI				
Missing	2510 (5.9)	1523 (60.7)	0.71 (0.66, 0.77)	0.73 (0.68, 0.79)
Underweight	978 (2.3)	692 (70.7)	0.98 (0.88, 1.09)	0.95 (0.85, 1.07)
Normal	17514 (41.4)	12410 (70.9)	1	1
Overweight	13980 (33.0)	9944 (71.1)	1.01 (0.97, 1.06)	1.03 (0.99, 1.08)
Obese	7326 (17.3)	5348 (73.0)	1.09 (1.04, 1.15)	1.11 (1.05, 1.17)
Country				
England	35423 (83.7)	24998 (70.6)	1	1
Wales	2041 (4.8)	1499 (73.5)	1.18 (1.05, 1.33)	1.07 (0.95, 1.20)
Scotland	3636 (8.6)	2715 (74.7)	1.35 (1.23, 1.49)	1.17 (1.06, 1.29)
Northern Ireland	1208 (2.9)	704 (58.3)	0.63 (0.57, 0.71)	0.56 (0.50, 0.63)

Note:
† weighted numbers; may not add up to totals because of rounding
ⱦ univariate/adjusted analysis based on unweighted data
$ fully adjusted for all other variables in the table

Table 2: Socio-demographic and socioeconomic characteristics of Understanding Society Wave 1 respondents and the likelihood of them consenting to linkage of administrative health records to survey responses stratified by country

		Total	England		Wales		Scotland		Northern Ireland	
		n (%)†	n (%)	OR (95%CIs)‡	n (%)	OR (95%CIs)‡	n (%)	OR (95%CIs)‡	n (%)	OR (95%CIs)‡
Age	16-24	6897 (16.3)	4277 (74.7)	1	294 (79.6)	1	441 (75.8)	1	140 (64.8)	1
	25-34	7495 (17.7)	4395 (69.4)	0.83 (0.76, 0.90)	257 (79.5)	1.02 (0.67, 1.55)	458 (75.4)	0.95 (0.69, 1.31)	134 (59.0)	0.68 (0.48, 0.96)
	35-44	8300 (19.6)	4872 (69.3)	0.78 (0.72, 0.85)	254 (74.6)	0.86 (0.58, 1.30)	496 (72.0)	0.69 (0.50, 0.95)	136 (57.6)	0.56 (0.39, 0.79)
	45-54	7846 (18.6)	4635 (70.8)	0.77 (0.70, 0.84)	248 (69.2)	0.61 (0.40, 0.92)	527 (73.9)	0.78 (0.57, 1.09)	126 (56.7)	0.61 (0.42, 0.88)
	55-64	6809 (16.1)	3946 (69.5)	0.72 (0.65, 0.79)	248 (71.3)	0.62 (0.40, 0.98)	461 (76.4)	0.82 (0.57, 1.19)	95 (53.6)	0.58 (0.38, 0.88)
	65-74	4959 (11.7)	2873 (70.3)	0.77 (0.69, 0.86)	198 (65.7)	0.53 (0.33, 0.86)	331 (75.3)	0.77 (0.51, 1.15)	74 (57.0)	0.66 (0.41, 1.05)
Sex	Male	20992 (49.6)	12498 (71.0)	1	760 (74.9)	1	1347 (75.8)	1	359 (60.4)	1
	Female	21315 (50.4)	12500 (70.2)	0.99 (0.96, 1.03)	739 (72.0)	0.91 (0.77, 1.08)	1368 (73.6)	0.93 (0.81, 1.06)	345 (56.3)	0.96 (0.83, 1.10)
Ethnicity	White	38094 (90.0)	22674 (72.1)	1	1453 (73.8)	1	2619 (75.1)	1	690 (58.5)	1
	Non-white	4214 (10.0)	2324 (58.7)	0.63 (0.59, 0.67)	46 (64.1)	0.49 (0.32, 0.75)	97 (64.2)	0.54 (0.37, 0.78)	15 (50.5)	0.65 (0.34, 1.26)
Marital status	Never	10772 (25.5)	6318 (70.8)	1	397 (76.2)	1	712 (72.8)	1	201 (58.5)	1
	Married	26952 (63.7)	16012 (70.5)	1.10 (1.03, 1.18)	936 (72.5)	1.22 (0.89, 1.69)	1671 (75.3)	1.28 (0.99, 1.64)	426 (58.3)	1.35 (1.01, 1.79)
	Sep/div/wid	4584 (10.8)	2669 (70.5)	1.14 (1.04, 1.24)	166 (72.6)	1.23 (0.83, 1.84)	332 (75.7)	1.22 (0.89, 1.66)	77 (57.7)	1.28 (0.89, 1.83)
Housing tenure	Rental	13440 (31.8)	8020 (70.8)	1	449 (77.7)	1	916 (76.0)	1	196 (60.1)	1
	Own	28868 (68.2)	16978 (70.5)	0.91 (0.86, 0.98)	1050 (71.8)	0.83 (0.62, 1.13)	1800 (74.0)	0.81 (0.64, 1.03)	508 (57.7)	0.81 (0.60, 1.09)
Number of cars	0	7161 (16.9)	4080 (69.5)	1	250 (78.7)	1	581 (75.4)	1	125 (60.5)	1
	1	16692 (39.5)	9772 (70.1)	0.93 (0.86, 1.00)	588 (72.3)	0.74 (0.51, 1.07)	1155 (75.6)	1.09 (0.83, 1.43)	240 (58.7)	0.96 (0.68, 1.35)
	2+	18454 (43.6)	11146 (71.4)	0.96 (0.88, 1.05)	661 (72.6)	0.74 (0.49, 1.12)	979 (73.2)	0.97 (0.71, 1.35)	339 (57.3)	0.83 (0.57, 1.23)
Education	None	9819 (23.2)	5464 (68.9)	1	438 (71.5)	1	660 (74.0)	1	219 (56.8)	1
	Secondary	18744 (44.3)	11426 (72.6)	1.09 (1.03, 1.15)	683 (76.3)	1.11 (0.88, 1.42)	1241 (76.4)	1.08 (0.88, 1.32)	280 (57.6)	1.13 (0.91, 1.40)
	HE/degree	13745 (32.5)	8109 (69.0)	0.98 (0.93, 1.05)	379 (70.9)	0.94 (0.72, 1.24)	814 (72.8)	1.05 (0.83, 1.31)	205 (61.1)	1.29 (1.02, 1.65)
LLTI	No	28213 (66.7)	16550 (69.8)	1	1003 (74.8)	1	1672 (73.0)	1	498 (58.5)	1
	Yes	14094 (33.3)	8448 (72.2)	1.11 (1.06, 1.17)	496 (70.9)	0.90 (0.74, 1.11)	1043 (77.6)	1.28 (1.09, 1.52)	206 (57.9)	1.07 (0.89, 1.28)
Self-completion	No	5086 (12.0)	2132 (49.1)	1	132 (57.9)	1	183 (52.6)	1	73 (42.9)	1
	Yes	37222 (88.0)	22866 (73.6)	1.87 (1.60, 2.20)	1367 (75.4)	2.40 (1.10, 5.23)	2532 (77.0)	1.85 (0.92, 3.72)	631 (60.8)	1.76 (0.91, 3.42)
GHQ	<3	28132 (66.5)	17139 (73.2)	1	1032 (75.4)	1	1946 (76.8)	1	494 (62.1)	1
	≥3	8497 (20.1)	5404 (75.6)	1.06 (1.01, 1.12)	317 (75.9)	1.04 (0.81, 1.33)	560 (78.7)	1.09 (0.89, 1.33)	129 (57.9)	0.77 (0.62, 0.96)
	Missing	5678 (13.4)	2454 (50.7)	0.80 (0.69, 0.93)	150 (59.2)	0.96 (0.46, 2.01)	209 (53.7)	0.64 (0.33, 1.24)	82 (43.0)	0.81 (0.44, 1.50)
BMI	Missing	2510 (5.9)	1246 (59.7)	0.72 (0.66, 0.78)	93 (66.8)	0.82 (0.56, 1.19)	146 (68.5)	0.73 (0.54, 1.00)	38 (54.4)	1.03 (0.72, 1.46)
	Underweight	978 (2.3)	588 (69.7)	0.91 (0.81, 1.03)	32 (81.0)	1.70 (0.78, 3.70)	61 (76.5)	1.21 (0.72, 2.03)	10 (69.3)	1.51 (0.75, 3.05)
	Normal	17514 (41.4)	10463 (70.6)	1	536 (73.6)	1	1125 (76.6)	1	285 (57.9)	1
	Overweight	13980 (33.0)	8280 (71.3)	1.04 (0.99, 1.09)	522 (72.5)	1.00 (0.80, 1.24)	889 (73.2)	0.86 (0.72, 1.03)	254 (59.3)	1.17 (0.97, 1.41)
	Obese	7326 (17.3)	4422 (73.1)	1.11 (1.05, 1.18)	317 (76.4)	1.25 (0.95, 1.64)	493 (74.8)	0.95 (0.77, 1.18)	117 (57.9)	1.11 (0.88, 1.41)

Note:
† weighted numbers; may not add up to totals because of rounding
‡ analysis based on unweighted data and fully adjusted for all other variables in the table