Institute of Education

How consistent is respondent behaviour to allow linkage to health administrative data over time?

Tarek Mostafa and Richard D Wiggins

CLS Working Paper 2015/03

May 2015

Centre for Longitudinal Studies
Following lives from birth and through the adult years
www.cls.ioe.ac.uk

CLS is an ESRC Resource Centre based at the UCL Institute of Education

How consistent is respondent behaviour to allow linkage to health administrative data over time?

Tarek Mostafa and Richard D Wiggins

Centre for Longitudinal Studies, Department of Quantitative Social Science, Institute of Education, University College London

Abstract

This study constitutes the first longitudinal exploration of consent to link survey and administrative data. It relies on a theoretical framework distinguishing between passive, active, consistent and inconsistent consent behaviour. The findings show that, in general, consent behaviours are both passive and consistent. First, consent rates indicate that most respondents behave consistently over time. Secondly, the regression analyses show that for the majority of respondents, consent is not driven by personal convictions but rather depends on the circumstances of the respondent at the time of the interview and on the impact of the interviewers. The findings also show that in longitudinal surveys cross-sectional analyses of consent can be misleading. The changes in the magnitude and in the significance of the results when the temporal dimension of consent is taken into account is a clear indication that consent should be treated as a dynamic phenomenon.

Acknowledgements

This work was funded by the Economic and Social Research Council and the National Centre for Research Methods Research Grant No. R00960-01-001-1. The authors would like to thank Professors John Micklewright and Lucinda Platt and various participants at the Royal Statistical Society's Annual Conference (2014) for their valuable comments.

Keywords: Consent; data linkage; Millennium Cohort Study; longitudinal data.

I- Introduction

Longitudinal surveys face significant challenges due to the rise in survey costs, attrition over time, and non-coverage of the target population. All these challenges have the potential of damaging the quality of the collected data. One method of reducing the costs of data collection and improving quality is to link selected individual administrative information to the survey record. Administrative data linkage leads to shorter interviews, less respondent burden and an overall reduction in costs (Sakshaug et al. 2012) in addition to the gain of valuable information on respondents. However, access to administrative records will suffer from non-consent whenever respondents refuse permission to link their records. Non-consent will obviously result in smaller sample sizes and possibly bias the sample composition if the likelihood of consent is related to the characteristics of the respondents.

The existing literature on consent is dominated by studies coming from the medical profession and epidemiology. Much of the early work (Baker et al. 2000; Dunn et al. 2004; Nelson et al. 2002; Kho et al. 2009; Silva et al. 2002; Huang 2007) focused on linking patients' administrative records to medical surveys. More recently, the work of Jenkins et al. (2006) opened the door to a new wave of studies focusing on multi-topic social surveys. All these studies examined the impact of various socio-demographic characteristics on consent and some focused on particular attributes of the respondents such as their personality traits (Sala et al. 2012; Jenkins et al. 2006; Olson 1999; Woolf et al. 2000; Armstrong et al. 2008; and Sakshaug et al. 2012). On the other hand, there are some studies which explored the impact of interviewers (Sala et al. 2012 ; Sakshaug et al. 2012; Korbmacher and Schroeder 2013; Sakshaug et al. 2013) and others have provided experimental evidence on the effect of question wording and placement (Sakshaug et al. 2013; Sala et al. 2014).

Despite the recent developments in the analysis of consent, the evidence is still scarce. The existing literature focused on the patterns of consent arising in cross-sectional surveys and very little is known about the patterns of consent over time. The importance of the temporal dimension of consent becomes apparent in longitudinal and birth cohort studies. First, these surveys follow the lives of cohort members (CMs) over a long period. Consequently, consent needs to be sought at the different occasions of data collection because it is impossible to ask for all consents at birth for ethical reasons. Secondly, the responsibility for responding to the survey as well as giving consent for data linkage is transferred from parents to CMs once the latter are old enough (typically around age 16). Hence, consents will have to be collected again from CMs to cover the remainder of the survey. Given the necessity to collect the same consents on multiple occasions, it is reasonable to expect that the changes to the circumstances of respondents will affect consent behaviour over time. This warrants the study of the temporal dimension of consent.

From a theoretical perspective, consistency in one's attitudes and behaviours is a central motivator for human conduct (Festinger 1957, Heider, 1958, Newcomb

1953). People in general are inclined to be consistent with what they said or did in the past. Thus, after committing themselves to a particular behaviour, they are likely to act in ways that are congruent with this behaviour especially if it is freely chosen (Cialdini et al. 1999, p. 1244).

In this study, the consistency principle implies that respondents who have consented to link their survey and administrative data in the past are likely to consent in subsequent waves of data collection. However, since all existing studies have focused on cross-sectional surveys, it was impossible to examine the validity of this argument. Furthermore, the nature of this consistency might vary depending on whether consent behaviour is passive or active. The distinction between passive and active behaviour in the context of data linkage has not been studied in the past even though it has been explored in the case of survey response (Roselberg et al. 2000, 2003; Sosdian and Sharp; 1980; Youssefnia 2000). Rogelberg et al. (2003) showed that active non-respondents made conscious and purposeful decisions not to respond to a survey while passive non-respondents were less conscientious (e.g. they are likely to have forgotten about the survey, did not have the time or inclination to co-operate, etc). In the context of consent to data linkage, it is also possible to distinguish between active behaviour reflecting strongly held convictions and passive behaviour resulting from extrinsic influences.

Given the absence of studies focusing on the longitudinal dimension of consent, the consistency principal as well as its limitations have not been explored. This study sets out to examine consent behaviour for the same respondents and the same consent domain: health data linkage over time. It exploits the longitudinal nature of the UK's Millennium Cohort Study (MCS) and the fact that the same consent questions were asked in different waves. It addresses the following research questions:

RQ1: Is consent behaviour consistent over time?
RQ2: Is consent behaviour active or passive?

Consent behaviour is said to be consistent if respondents behave in the same manner over time. Furthermore, behaviour is active if it reflects the existence of strongly held convictions about whether to consent or not. In contrast, behaviour is passive if it results from external influences such as the circumstances of the respondent at the time of the interview or from the impact of the interviewers.

To the best of our knowledge, this is the first study to explore consent to data linkage in a longitudinal survey. The novelty of this paper is that, it develops a theoretical framework which distinguishes between consistent/inconsistent and active/passive consent behaviour. In addition to this, it examines both the cross-sectional variations in consent between different sub-groups (e.g. singles vs. couples) and the variations over time caused by changes in the respondents circumstances (e.g. divorce).

Moreover, the growing popularity of longitudinal cohort studies and the expanding practice of administrative and survey data linkage highlight the value of this study for both data users concerned about non-consent bias and survey professionals interested in improving fieldwork practices.

The paper is organised as follows. Section II discusses consent mechanisms over time. Section III presents the data, consent procedures, and methods. Section IV presents the findings, and the final section concludes.

II- Consent Mechanisms Over Time.

In longitudinal and birth cohort studies, consent for survey and administrative data linkage has to be sought repetitively over time in order to give respondents the chance to make informed decisions about whether to release their administrative records or not. Therefore, it is possible that respondents' consent behaviour will change. Those who have consented in the past might refuse to consent in the future and vice versa. In other words, some respondents will behave consistently over time while others will have inconsistent behaviours. According to Cialdini's consistency principle, people are expected to act in ways which are congruent with their past behaviours. However, this principle assumes that past actions were active, informed decisions that people can remember. As suggested in the introduction and based on existing evidence on survey response (Rogelberg 2003), this is not always the case since consistency could be the result of passive behaviour.

In this study, we are extending Cialdini's framework by testing four scenarios which sub-divide consistency/inconsistency along the lines of activeness/passiveness.

Active Consistency is the case where respondents are aware of their previous choices and are committed to make the same choices on future occasions because of stable beliefs or personality traits (e.g. belief in the importance of scientific research, being a private person, etc).

Passive consistency is the case where respondents make consistent choices over time even though the decision making process is passive. This means that consent decisions do not reflect an active adherence to well-defined beliefs but rather external influences such as the respondents' circumstances at the time of the interview and the impact of the interviewers.

Active inconsistency is the case where respondents are aware of their previous choices and intentionally behave in inconsistent ways. This change in behaviour could be the result of a change in convictions. For instance, a past consenter might actively decide to withhold consent after a breach to data confidentiality.

Passive inconsistency is the case where respondents switch from consenters to non-consenters or vice versa. This switch is not the result of changes in convictions but rather the result of changes to the circumstances of the respondent (e.g. divorce, acute health problems), changes to the interviewers over time (e.g. persistence in pursuing consent), and the fact that respondents could have forgotten what they did in the past. In all cases, the respondent has a passive role and the changes in consent behaviour are caused by extrinsic factors.

In the next section, we present a brief outline of our data source, the Millennium Cohort Study, the consent procedures, and the chosen methods designed to test the plausibility of the aforementioned scenarios.

III- Data, Consent Procedures and Methods.

The Millennium Cohort Study

The Millennium Cohort Study (MCS) is the most recent of the British Cohort studies. It follows the lives of a nationally representative sample of more than 19,000 children born in the UK in 2000-01. The primary sampling unit is the electoral ward. These were disproportionally stratified to ensure adequate representation of all countries of the UK (i.e. England, Scotland, Wales and Northern Ireland), of disadvantaged areas and of areas with high concentration of ethnic minorities. Survey data has been collected on five occasions when the CMs were 9 months, three, five, seven, and eleven years old. The main respondents (MRs) were mostly the mothers although some have swapped with the fathers or other members of the household over time.

MCS has a complex survey design (Plewis 2007). The sample is stratified by country, clustered at the electoral ward level, and has oversampled minorities and disadvantaged groups. The sample also experienced attrition over time. All these design features are accounted for in the analysis which follows. The original study had 19,244 families who were interviewed at least once in waves 1 and 2, some of which had twins and triplets. Our analytical sample consists of 12,165 MRs who were present in waves 1,2 and 4. Presence in wave 3 was not a qualifier for inclusion as the consent question was an exact repeat item from wave 2. Therefore wave 3 was discarded. Any MR with twins or triplets was also excluded. The participating MRs were interviewed by 328 interviewers in wave 1, by 335 interviewers in wave 2, and by 443 interviewers in wave 4.[1]

Consent Procedures

Written consent was sought from MRs for linking their children's health records in three waves (at age 9 months, 3 and 7 years). Consent was never sought directly from the CMs because they were too young. Prior to the interview, leaflets explaining

[1] For more information on sampling, response, and on how to use MCS refer to: the MCS technical report on sampling, the MCS technical report on Response, and the MCS user guide for analysing MCS data in Stata.

what consent to administrative data linkage consists of were posted out to the MRs. All interviews were face to face, and all consent questions were administrated at the end of the main interview. Respondents who were willing to give consent were asked to tick a box containing two options: ‘yes’ or ‘no’, then sign print their names and date the form. The wording and the content of the consent question changed between waves 1 and 2. In wave 1, consent was sought to link information on pregnancy and birth and to follow the CM’s National Health Service (NHS) registration. In wave 2, consent was sought to link health records from birth to age 7. All consent forms made it clear that respondents can refuse to participate or withdraw from any part of the survey by simply expressing the wish to do so. All consent questions included a confirmation statement.[2] The procedures, the leaflets and consent forms are presented in detail in the technical report on Ethical Review and Consent (2012). The outcomes of interest are presented below in table 1:

Table 1: Health Consent outcomes

Consent	Wave	Notes
CM's health records	MCS1 age 9 months	Consent for linking information on pregnancy and birth and for following the baby’s National Health Service (NHS) registration.
CM's health records	MCS2 age 3 years	Consent for linking health records (hospital admissions and records held by the NHS) from birth to age 7.
CM's health records	MCS4 age 7 years	Consent for linking health records (hospital admissions and records held by the NHS) from birth to age 14.

Wave 5 (age 11) was not included in the sequence of consent above because the health consent question was not asked in this wave as the consent obtained in wave 4 was valid until age 14. Also as mentioned earlier, wave 3 consent question was a repeat of wave 2 and therefore was not included.

In terms of fieldwork organisation it is worth noting that the survey agency carrying the fieldwork changed between waves 1 and 2 (NatCen to GfK NOP). This disruption might have affected the levels of consent since the interviewers and the survey management procedures would have changed from one agency to another.

[2] Example of wave 4 confirmation statement: I have read or heard the information leaflet about information from other sources and have had the opportunity to ask questions. I understand the information released will be treated in strict confidence in accordance with the Data Protection Act and used for research purposes only. I understand that this consent will remain valid unless revoked by me in writing and that I may withdraw my consent at any time by contacting the Child of the New Century in writing to the address below, without giving any reasons. (MCS4 consent forms).

Finally, it is worth emphasising that the consent patterns condition on the same consent outcome (i.e. health) for the same respondents over time.

Methods

The methodologies used in this study are designed to explore the hypothesised mechanisms of consent and to distinguish between consistent/inconsistent and passive/active behaviours. On the one hand, consistency is ascertained based on the proportion of MRs who have the same behaviour in all three waves (i.e. they consented or did not consent in all waves). On the other hand, the behaviour is said to be active if it reflects the existence of strongly held convictions. In contrast, the behaviour is passive if it is influenced by the MRs circumstances at the time of the interview or by the impact of the interviewers.

A simple descriptive statistic measuring the proportion of MRs who have behaved in the same way over time is enough to determine whether behaviour is consistent or inconsistent. However, in order to distinguish between passiveness and activeness, three analytical approaches are needed. First, since the data does not provide information on the direct driving forces behind consent (i.e. personal convictions, etc) we have adopted a multivariate probit model to measure the existence of an individual latent propensity to consent reflecting such convictions. Secondly, a number of logit models are used to measure the association between the MRs observable characteristics and consent as a binary outcome. Thirdly, linear probability models with interviewer fixed effects are used to measure any change in the explanatory power of the model (i.e. rise in R-squared) when interviewer effects are accounted for. It was decided that logit models are more familiar to our readers than alternative models and easier to interpret.

The **first analysis** consists of a joint estimation of the three consent outcomes using a multivariate probit specification (i.e. three consent equations estimated jointly). This analysis allows for the computation of the cross-equation correlations: the strength of the association between the unobserved factors (error terms) explaining each consent outcome. The M-equation multivariate probit model is the following:

$$y_{im}^{*} = \beta_{im}' x_i + \varepsilon_{im}, m = 1, \dots, M$$

$y_{im} = 1$ if $y_{im}^{*} > 0$ and 0 otherwise

where y is the binary consent outcome for respondent i and consent outcome m with $m = 1, \dots, 3$. x is a vector of independent variables for respondent i. ε_{im}, are error terms distributed as multivariate normal, each with a mean of zero and a variance-covariance matrix V, where V has values of 1 on the diagonal and values different to 1 off-diagonal (Cappellari and Jenkins 2003). The model is estimated using a similar approach to the one used in Cappellari and Jenkins (2003 and 2006) and Mostafa

(2014). The procedure was adapted to take into account the complexity of the MCS survey design through the use of the *svy* command in Stata 13.

Since the unobserved circumstances of the interview are not the same over time, it is possible to attribute these associations to the existence of a latent propensity to consent. In other words, the existence of significant associations between the latent parts of the different outcomes indicates the presence of unobserved factors (e.g. strong belief in the importance of scientific research, etc) affecting consent over time. In this case, behaviour is said to be active.

The **second analysis** consists of three models:
<u>Cross-sectional logit:</u> is a logit model separately estimated for each consent outcome (i.e. for each wave apart). This model allows us to measure the within-wave effect of the correlates. The model is written in the following form:

$$log\left(\frac{\mathrm{p}(y_i)}{1-\mathrm{p}(y_i)}\right) = \beta_0 + \beta_1 x_i$$

where $\mathrm{p}(y_i)$ is the probability of giving consent and x_i is a vector of characteristics of respondent i.
<u>Logit with pooled data:</u> uses pooled data over the three waves (i.e. takes into account the time dimension). The model is written in the following form:

$$log\left(\frac{\mathrm{p}(y_{it})}{1-\mathrm{p}(y_{it})}\right) = \beta_0 + \beta_1 x_{it} + \beta_2 z_i$$

where $\mathrm{p}(y_{it})$ is the probability that respondent i gives consent in wave t. $x_{it}$is a vector of time varying characteristics of respondent i in wave t. $z_i$is a vector of time-fixed characteristics of respondent i. This model includes wave dummy variables to ascertain whether the probability of consent varies from wave to wave.
<u>Conditional logit:</u> models the switch of behaviour between two consecutive waves. The dependent variable is a binary variable taking the value of 0 if the respondent had the same behaviour over two consecutive waves (e.g. was a consenter in wave 1 and remained a consenter in wave 2) and 1 if the respondent switched behaviour (e.g. was a consenter in wave 1 and became a non-consenter in wave 2). The right hand side variables are the respondent's characteristics in the initial wave (i.e. waves 1 or 2 depending on the model).

The three models are designed to measure the impact of the respondent's characteristics on the likelihood of consent and on the likelihood of switching behaviour from wave to wave. If there are strong associations between the MRs' characteristics and consent, it is possible to conclude that what drives consent are the circumstances of the respondent rather than his/her convictions. In this case, consent behaviour is said to be passive. In addition to this, the three models allow us to assess separate cross-sectional effects of correlates as well as test for uniformity of effect based on pooled data using wave dummies.

The first and second analyses do not take into account interviewer fixed effects in order to avoid unnecessary complications. In the first analysis, multivariate probit models become very complex and computational time rises dramatically when several hundred fixed effects are included in the model. For the second analysis, it is impossible to know whether the interviewers were the same from wave to wave since their identifiers are not consistent over time. This precludes the use of the interviewers' IDs in a longitudinal approach. Hence, interviewer fixed effects are left for a separate linear probability analysis.

The **third analysis** consists of three linear probability models designed to measure the rise in the models' explanatory power after the inclusion of interviewer fixed effects (i.e. rise in R-squared). However, any rise in the R-squared cannot be completely attributed to the impact of interviewers because their allocation is implemented on a 'nearest-to-home' basis. Therefore, interviewer effects will be confounded by the characteristics of interviewer assignment areas (e.g. assignment areas with large proportions of minorities, high levels of poverty or unemployment, etc). A similar approach to Mostafa (2014) is used to overcome this challenge. It consists of controlling for the characteristics of assignment areas. These are computed as the average of respondents' characteristics at the interviewer's level. Three different models are estimated:

<u>Base model:</u> is a linear probability model with the same correlates as in the first analysis and without interviewer fixed effects.
<u>Model with assignment area characteristics:</u> is identical to the base model and includes additional variables measuring the interviewer's assignment area characteristics (i.e. proportion of minorities, proportion unemployed, log average income, and social class composition). These are computed as averages of MRs' characteristics at the level of the interviewer.
<u>The fixed effects model:</u> is equivalent to the base model and includes interviewer fixed effects.

All models take into account the MCS survey design features: clustering at the electoral ward level, stratification at the country level, oversampling of minorities and disadvantaged groups in the base sample and attrition over time. Oversampling and attrition were accounted for through the use of sampling and unit non-response weights.

When it comes to the choice of the correlates, a wide range of socio-demographic characteristics were included. The choice was motivated by previous literature (Mostafa 2014, Sakshaug et al. 2012, and Jenkins et al. 2006) and by the fact that these characteristics are expected to vary over time. For instance, after seven years in the life of the survey, adult respondents are expected to have higher incomes, higher positions in their jobs, and a growing professional experience. It is also likely that some respondents have experienced divorces/breakups and have started new relationships. Similarly, the number of house owners is expected to grow as young parents grow older. In terms of health, respondents are likely to have more health issues as they age while the reverse is true for children since most of the health problems happen after birth and progressively decline. In addition to time-varying socio-demographic characteristics, selected time-invariant characteristics are also included in the model including gender, ethnicity, personality traits (i.e. being private person), and response history on the survey. Response history is a binary variable taking the value 1 if the respondent was absent in at least one wave, it is used as a proxy for the respondent's willingness to cooperate. For an in-depth description of the motivation behind the choice of the correlates, refer to Mostafa (2014).

IV- Findings.

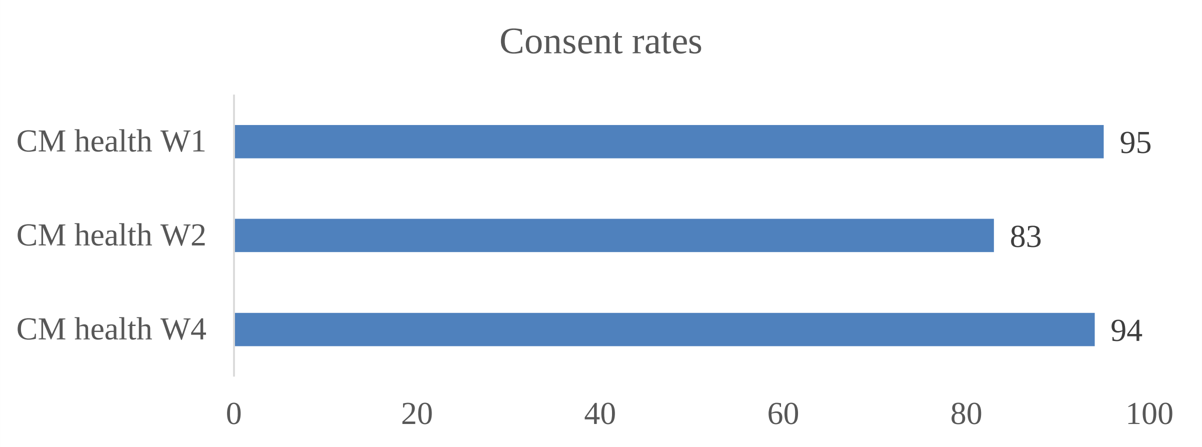
In what follows, we begin with a description of consent patterns over time, followed by the change in the characteristics of the sample, and finally by interviewer's individual success rates in obtaining consent. These descriptive accounts are followed by the regression results presented in the same order as outlined in the previous section.

Consent and sample characteristics

Figure 1 shows the existence of variations in consent over time. Consent rates for linking the CMs' health records are the highest in wave 1 followed by wave 4 and the lowest in wave 2. There are a number of possible explanations. Firstly, there may be a tendency for less fieldwork effort to be put into obtaining consent per se when all of the focus of fieldwork management is to minimise unit non-response. This might well have been exacerbated by a change in fieldwork agencies between waves 1 and 2. Secondly, since most MR's were mothers, they were probably willing to talk about their giving-birth experience because it was a recent and exciting life event and consequently, they were happy to link the pregnancy and birth records of the child but less willing to co-operate as they got older. Thirdly, the drop in consent in wave 2 could also be attributed to the change in the content of the consent question. Wave 2 was the first time MRs were asked to link their children's hospital records over a long period from birth to age 7. In wave one it was only birth records, and NHS registration used for tracing purposes.

The percentage of those MRs who consented in all three waves is 76%. Only 0.5% of respondents were non-consenters in all three waves.

Figure 1: Health record consent rates for the three successive outcomes.



CM stands for cohort member and W denotes the wave of data collection.

Table 2 presents the changes in consent between each consecutive waves. From wave 1 to wave 2, 3.5% of non-consenters became consenters and 15.4% of consenters became non-consenters. From wave 2 to wave 4, 15% of non-consenters became consenters and 4.7% did the opposite. Note that respondents who switched from consenters to non-consenters between waves 1 and 2 are almost the same as those who did the opposite between waves 2 and 4. Based on figure 1 and table 2, it is possible to say that there are sufficient changes in consent behaviour over time to warrant exploring its temporal dimension. Moreover, most respondents had a consistent consent behaviour over time (i.e. most of them consented in all three waves). Therefore, it is possible to rule in favour of the consistency assumption.

Table 2: Changes in health record consent from wave 1 to wave 4

	Change	Percent
Change MCS W1 to MCS W2	Yes to No	15.4
	No to Yes	3.5
Change MCS W2 to MCS W4	Yes to No	4.7
	No to Yes	15.0
N		12,165

Table 3 presents the changes in the distribution of correlates, in terms of unweighted frequencies, over the three waves to clarify the exact nature of changes in sample composition.

Table 3: Changes in the correlates over time.

Wave		W1 Age 9 months	W2 Age 3 years	W4 Age 7 years	Total
Marital Status	Single	1,680	1,901	2,765	6,346
	Couple	10,486	10,265	9,401	30,152
SES	Managerial and professional	5,406	5,542	5,681	16,629
	Intermediate	1,647	1,702	1,634	4,983
	Small employers and self-employed	863	981	1,207	3,051
	Lower supervisory and technical	1,358	1,233	1,095	3,686
	Semi-routine and routine	2,892	2,708	2,549	8,149
Employment status	Both in work	5,619	5,665	6,265	17,549
	One in work	5,726	5,822	5,297	16,845
	Both not in work	821	679	604	2,104
Housing tenure	Own	7,811	8,188	8,423	24,422
	Rent	3,683	3,578	3,510	10,771
	Other	672	400	233	1,305
Religion	Christian	5,764	5,756	4,816	16,336
	Non-Christian	1,272	1,265	1,268	3,805
	None	5,130	5,145	6,082	16,357
Language spoken at home	English	10,559	10,443	11,243	32,245
	English and other languages	1,205	1,370	468	3,043
	Other	402	353	455	1,210
Interview translated	No	11,742	11,879	11,818	35,439
	Yes	424	287	348	1,059
MR health	Excellent	3,800	3,709	2,816	10,325
	Good	6,323	6,335	7,888	20,546
	Poor	2,043	2,122	1,462	5,627
CM health	Some problems	5,086	1,953	1,581	8,620
	No health problems	7,080	10,213	10,585	27,878
Total		12,165	12,165	12,165	36,495

Because our analytical sample is constant over time the changes in the distribution of the correlates over time represent genuine changes in status or context. The results show that the number of single MRs increased over the 7 years. In the first wave (at age 9 months) most MRs were living in a couple, but by age 7 more than 1,000 MRs became single (representing a 65 per cent increase between wave 1 and 2). When it comes to the socio-economic status, the numbers have slightly changed with an increase in the proportion of MRs exercising managerial and self-employed jobs, and a decline in the proportion of those doing routine and technical jobs. This indicates that, over time, MRs have moved to better jobs possibly because of their rising experience.

The number of households where both parents are unemployed or at least one is employed has declined whereas the number of households with two working parents has increased. This shows that as the CMs have grown older, more parents have joined the labour market (especially mothers). Similarly and as expected, after 7 years, the number of house owners has increased while the number of those renting, living with parents, or living free of rent has declined.

The number of MRs reporting that they are Christians has declined while the number of non-religious MRs has increased. The number of non-Christian respondents remained the same. Moreover, the number of MRs reporting that they spoke only English at home has increased. This indicates that those who reported speaking other languages in wave 1 (i.e. mostly immigrants) have adopted the English language. Similarly, the number of translated interviews has declined over time.

When it comes to health, the number of MRs reporting excellent health has declined by just over 25%. This shows that over time more MR's have begun to report health problems. The opposite happens for the CMs, where health concerns are more frequent in infancy (i.e. by age 9 months) and tend to subside in early childhood (i.e. after age 9 months).

In general, table 3 shows that the marginal distribution of our correlates have changed over time in a direction that might have been expected. The existence of variations in consent and in the correlates warrants the study of consent over time.

Table 4 provides weighted estimates of percentages of consenters for some of the key variables included in the analyses. Note that the percentage of non-consenters for each category is equal to 100 minus the percentage of consenters. All comparisons and significance tests are done within wave and all figures in bold are significant at the level of $p<0.01$, $p<0.05$, and $p<0.1$.

The results show that consenters are more likely to exercise managerial jobs (significant in waves 1 and 2), to belong to households with two employed parents (significant in waves 1 and 2), to own their homes (significant in waves 1), to belong to the ethnic majority group (significant in waves 1 and 4), to speak mainly English at home (significant in all waves), and to report excellent health (significant only in wave 1). In contrast, they are less likely to have translated interviews (significant in all waves).

Based on these results a number of observations can be made. First, the likelihood of consent seems to vary according to a number of socio-demographic characteristics. The results are in-line with previous studies (Mostafa 2014, Sakshaug et al. 2012, Jenkins et al. 2006) where respondents belonging to lower social groups and ethno-linguistic minorities being less likely to consent. Secondly, the associations between the correlates and the likelihood to consent vary in significance and magnitude from wave to wave. These variations could be the result of the change in the sample composition over time.

Table 4: Health record consent rates for each socio-demographic group

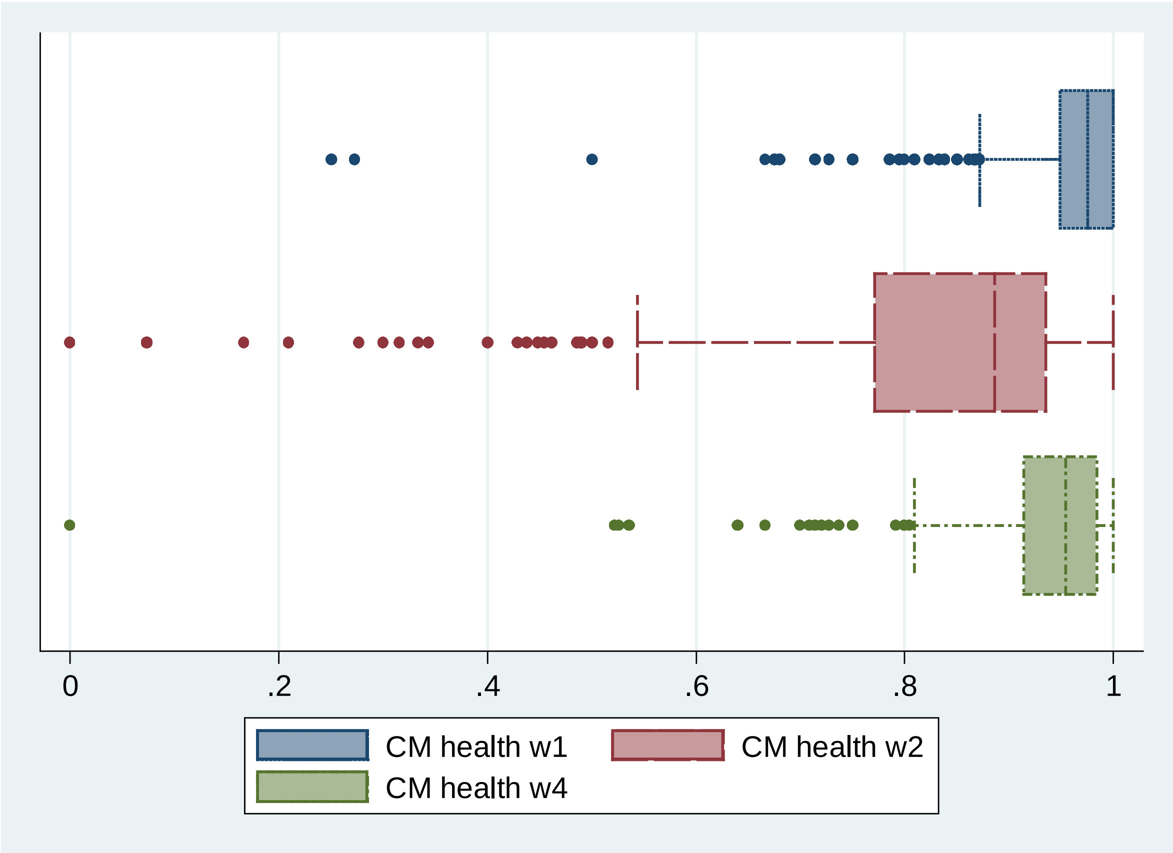
Wave	W1 (Row %)	W2 (Row %)	W4 (Row %)
	Consenters	Consenters	Consenters
Living in a couple in wave 1			
Single	93.9	81.9	93.6
Couple	95.2	83.3	93.2
Socio-economic status in wave 1			
Managerial and professional	**95.9**	**85.0**	93.0
Intermediate	**95.7**	**84.4**	94.2
Small employers and self-employed	**94.2**	**82.5**	94.0
Lower supervisory and technical	**94.5**	**79.6**	92.9
Semi-routine and routine	**93.4**	**80.0**	93.2
Employment status in wave 1			
Both in work	**96.3**	**84.1**	93.9
Only one in work	**94.3**	**82.4**	92.6
Both not in work	**91.2**	**79.6**	93.8
Housing tenure in wave 1			
Own	**95.7**	83.7	93.4
Rent	**94.1**	81.4	93.2
Other	**92.6**	85.5	91.8
Ethnic group			
White	**96.4**	83.3	**94.0**
Non-White	**84.2**	80.7	**88.4**
Religion in wave 1			
Christian	95.9	83.4	93.5
Non-Christian	82.1	82.3	89.9
None	96.5	82.8	93.7
Language spoken at home in wave 1			
English	**96.2**	**83.3**	**93.5**
Half English half other	**86.2**	**81.6**	**89.9**
Other	**80.1**	**79.0**	**89.1**
Whether the interview was translated in wave 1			
No	**95.5**	**83.1**	**93.4**
Yes	**75.7**	**77.4**	**87.4**
Main respondent's health status in wave 1			
Excellent	**95.8**	83.3	93.4
Good	**95.0**	82.7	93.0
Poor	**93.6**	83.5	94.6
Cohort member's health status in wave 1			
Some problems	**95.9**	83.5	**90.9**
No problems	**94.3**	82.9	**93.7**
N	12,165		

Figure 2 presents a boxplot depicting interviewer success rate in obtaining consent to CM health data linkage in the three waves. The success rate is defined as the number of obtained consents for each interviewer divided by the number of conducted interviews. The number of conducted interviews by interviewer ranged from 2 to 108 in wave 1 with an average number of interviews of 50; 2 to 151 in wave 2 with an average number of interviews of 59; and 2 to 76 in wave 4 with an average number of interviews of 34. The success rates varied between 0 and 100 percent. The 12,165 MRs were interviewed by 328 interviewers in wave 1, by 335 interviewers in wave 2, and by 443 interviewers in wave 4. It is worth noting that very few interviewers had a workload lower than 5 interviews.

Figure 2 shows that there are substantial variations in success rates between interviewers. First, the bottom quartile of interviewers has the highest level of dispersion while the top quartile has the lowest. This could reflect the dispersion in interviewers' experience, with the less experienced having more variations in their success to obtain consent. Alternatively, interviewers with a limited number of interviews were more likely to have low success rates. Secondly, for all three consents, the outliers belong to the lowest quartile. Thirdly, the success rates in wave 2 are more dispersed for all quartiles than in the two other waves. The reason is that the consent rate was 10% lower in wave 2 allowing for more variations in success rates.

In summary, it is possible to say that the existence of between interviewer variations in success rates warrants the use of interviewer fixed effects to measure their impact on consent.

Figure 2: Interviewers' success rates in obtaining health record consent for waves 1,2 and 4.



Regression findings

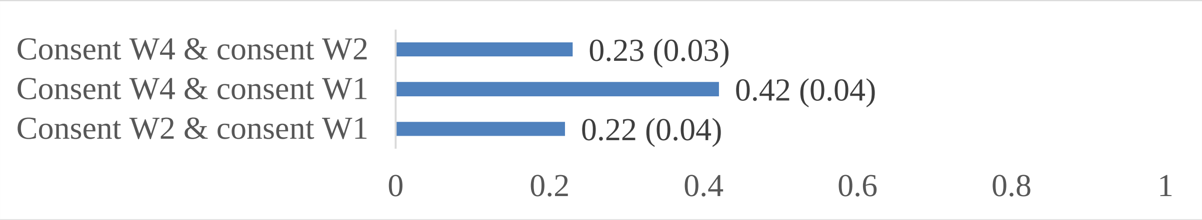
This section presents the results from the regression analyses. In figure 3, the estimated cross-equation correlations from the first analyses are presented. As mentioned earlier, these correlations are obtained through the joint modelling of the three consent outcomes using a multivariate probit procedure. The correlations measure the strength of the association between the unobserved factors explaining each consent. Since the consent outcomes were sought in different waves, the circumstances surrounding the interviews were different. Therefore, these correlations can be attributed to the existence of a latent propensity to consent reflecting stable respondent characteristics such as strongly held convictions. Furthermore, the existence of strong associations between consent outcomes over time is an indication that consent behaviours tend to be active.

The figure shows that the correlations between consent outcomes are not very strong ranging between 0.2 and 0.4 across adjacent and non-adjacent waves, even though they are statistically significant at $p<0.01$. Hence, a weak latent propensity to consent exists and consent behaviours appear to be mostly passive over time. A possible explanation is that respondents have forgotten what they did in the past especially that consent is not a consequential decision in their lives and/or interviewers did not remind them of their previous answers. If consent was motivated by strongly held latent beliefs (e.g. belief in the importance of scientific research), the correlations should have been much higher.

However, in spite of that, the presence of a weak latent propensity indicates that some respondents might still behave actively. This activeness reflects latent characteristics and predispositions such as being cooperative/uncooperative, belief in the importance of scientific research, etc. However, these predispositions seem to be held by a minority of respondents given the weakness of the cross-equation correlations. One can argue that since I am not controlling for personal convictions, it is not possible to ascertain whether they affect consent or not. However, it is reasonable to assume that these convictions are stable over time. If they exist and they were relegated to the unobservable part of the equation, their existence would have led to strong cross-equation correlations. This is not apparent here.

In summary, it is possible to rule in favour of the inconsistency in consent behaviour assumption. However, whether this inconsistency is active or passive depends on the impact of the correlates and the interviewers. If, in the next two analyses, we find that the respondents' circumstances and interviewers fixed effects have a strong impact on consent, it will be possible to conclude in favour of passive inconsistency.

Figure 3: Cross-equation correlations based on multivariate probit models.



All correlations are significant at level of P<0.01. Standard errors are in parentheses.

Moving on to the second analysis, table 5 presents the results of four logit models. The first three are estimated separately for each health consent outcome, and the fourth is estimated with pooled data from all waves. The first three models are cross-sectional and ignore the time dimension. The fourth, on the other hand, exploits the temporal variations in consent and in the correlates and includes wave dummy variables.

Table 5: Odds ratios of logit models for each health consent outcome.

	Wave 1 Logit		Wave 2 Logit		Wave 4 Logit		Pooled Logit	
CM's gender, reference: Girl								
Boy	1.00	(0.101)	0.97	(0.052)	0.91	(0.073)	0.96	(0.035)
Main respondent's marital status, reference: Single								
In a couple	1.43*	(0.312)	0.91	(0.096)	0.83	(0.130)	0.93	(0.055)
Highest socio-economic status, reference: Managerial and professional								
Intermediate	1.05	(0.191)	0.96	(0.083)	1.22	(0.187)	1.04	(0.064)
Small employers and self-employed	1.15	(0.257)	0.87	(0.102)	1.29	(0.232)	1.02	(0.074)
Lower supervisory and technical	1.11	(0.219)	0.70***	(0.066)	1.08	(0.215)	0.85**	(0.058)
Semi-routine and routine	0.99	(0.186)	0.72***	(0.069)	1.20	(0.171)	0.88**	(0.053)
Combined labour market status, reference: Both in work								
At least one in work	0.93	(0.122)	0.99	(0.067)	0.76**	(0.086)	0.92*	(0.042)
Both not in work	0.64*	(0.169)	0.95	(0.149)	0.94	(0.214)	0.89	(0.079)
Housing tenure, reference: Own								
Rent	0.93	(0.143)	1.05	(0.089)	1.01	(0.131)	1.02	(0.053)
Other	1.03	(0.230)	1.27	(0.210)	0.92	(0.272)	1.09	(0.113)
Main respondent's ethnic group, reference: White								
Non-White	0.45***	(0.096)	0.82	(0.148)	0.47***	(0.092)	0.60***	(0.051)
Main respondent's religion, reference: Christian								
Non-Christian	0.54	(0.175)	1.47**	(0.241)	1.36	(0.415)	1.17	(0.119)
None	1.28*	(0.164)	1.00	(0.064)	1.06	(0.108)	1.03	(0.043)
Language spoken at home, reference: English								
English and other languages	0.95	(0.353)	0.90	(0.114)	1.24	(0.327)	1.08	(0.101)
Only other	0.81	(0.398)	0.84	(0.201)	1.41	(0.478)	1.03	(0.131)
Was the interviews translated? reference: No								
Yes	0.65**	(0.132)	0.83	(0.198)	0.92	(0.299)	0.72***	(0.086)
Main respondent's health status, reference: Excellent								
Very good, good	0.90	(0.119)	1.01	(0.066)	1.01	(0.104)	0.97	(0.042)
Fair, poor	0.75*	(0.122)	1.14	(0.099)	1.32	(0.232)	1.04	(0.062)
CM's health status, reference: Some problems								
No problems	0.78**	(0.095)	0.95	(0.082)	1.15	(0.149)	0.90**	(0.043)
Main respondent: I am a very private person, reference: Strongly agree								
Agree	1.43	(0.368)	1.25**	(0.139)	0.95	(0.193)	1.21**	(0.092)
Neither	1.21	(0.296)	1.19	(0.143)	1.23	(0.238)	1.21**	(0.095)
Disagree	1.38	(0.357)	1.13	(0.123)	1.15	(0.238)	1.18**	(0.091)
Strongly disagree	1.24	(0.432)	1.27	(0.214)	1.85**	(0.570)	1.36***	(0.154)
Can't say	0.62	(0.276)	1.03	(0.252)	0.64	(0.320)	0.84	(0.136)
Other	1.04	(0.314)	1.16	(0.215)	0.41***	(0.115)	0.82*	(0.091)
Response history, reference: Participated in all waves								
Absent in at least one wave	1.38	(0.307)	0.87	(0.135)	0.61**	(0.118)	0.89	(0.081)
Log OECD adjusted income	0.90	(0.115)	1.13**	(0.062)	0.84	(0.097)	1.02	(0.039)

Main respondent's age	1.00	(0.011)	0.99	(0.006)	1.01	(0.008)	1.00	(0.004)
Wave of data collection, reference: wave 1								
Wave 2							0.26***	(0.013)
Wave 4							0.75***	(0.047)
N	12,165		12,165		12,165		36,495	

Exponentiated coefficients; Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In general, the results in table 5 show that socially disadvantaged groups (those MRs with low SES and members of ethno-linguistic minorities), are less likely to consent. This is in line with previous empirical evidence (Mostafa 2014, Sakshaug et al. 2012, and Jenkins et al. 2006) and with the reasons given by Sheldon et al. (2007) on why disadvantaged respondents tend to be less cooperative. In sum the reasons are purported to be disengagement from government and official institutions, low literacy, and communication barriers in the case of some ethnic minorities.

Respondents living as a couple are 43% more likely to consent (at p<0.1) in wave 1. By contrast, in the pooled-data model, marital status was no longer statistically significant. This finding indicates that, initially in wave 1, couples were more likely to consent than single parents (mostly single mothers). However, over time the rise in the number of single parents by about 40% due to divorces and breakups (table 3) resulted in the effect losing any distinguishing importance. A plausible explanation would be that those who became single in subsequent waves are different from the group of mothers who were single at the birth of a CM. In other words, single mothers at birth represent a particular group of socially disadvantaged (i.e. low socio-economic status) respondents who are less likely to consent for various reasons including: lack of understanding of the purpose of the survey, lower motivation to respond, burden raising a child on their own, etc. The characteristics of this group have changed over time due to socially advantaged respondents becoming single. This is reflected by the observation that the SES of singles is much higher in wave 4 than in wave 1. This finding highlights the value added by considering changes in the sample composition over time. This feature would be lost in a cross-sectional analyses of consent outcomes.

When it comes to the social and economic status of respondents. Table 5 shows, that only in wave 2, those exercising routine, technical and supervisory jobs are less likely to consent than those doing managerial and professional jobs. The presence of enough variation in the dependent variable (wave 2 has the lowest consent rate of 83%), could be the reason why the results were only significant in this wave. When the data are pooled, the effects of SES are similar to those in wave 2 but smaller in magnitude.

The effect of employment status is mostly non-significant except in the cross-sectional model for wave 4 (at p<0.05) and in the pooled data model (at p<0.1) where families with one employed parent are less likely to consent. As seen in table

3, the number of families where both parents are employed has increased. Therefore, after seven years, families who remain with only one employed parent are likely to be the most disadvantaged and the most likely to refuse to consent.

Ethnicity had the expected effect in both the separate logit models and the pooled data model. Since ethnicity is fixed over time and does not have any between-wave variations, the effects of the separate logit and pooled logit models were similar in terms of significance and magnitude. The three reasons advanced by Sheldon (2007) to explain why minorities are less likely to cooperate in surveys are: disengagement from government, low literacy, and communication barriers.

When it comes to religion, none-Christians were found to be more likely to consent in wave 2, while the effect in the pooled logit is non-significant. Those who had a translated interview were less likely to consent in wave 1 and in the pooled regression. The non-significant effects in waves 2 and 4 reflect the fact that the number of translated interviews declined over time. The negative effect of this variable is another indication that communication barriers might hinder consent.

The CM's health status has a significant effect only in wave 1 in the separate logit regressions. MRs who report no health problems for the CM in wave 1 are 22% less likely to consent than those who report some problems. In the pooled data regressions, those who report some problems are 10% less likely to consent and the effect is significant at $p<0.05$. This finding is in line with the results of Mostafa (2014) and Sakshaug et al. (2014). MR's with CMs suffering from health problems have previous experiences with the healthcare system (i.e. the institutions holding the health records). Therefore, they are more likely to cooperate since providing access to their children's medical records might help advance medical research and improve services.

The significance of the effect only in wave 1 and the decline in its magnitude when the data is pooled is the result of variations in health status over time. As shown in table 3, more CMs suffer from health issues during the first 9 months after birth. These concerns tend to subside over time. Hence, fewer CMs had health issues in waves 2 and 4. The decline in variations in health status over time meant that the variable lost significance in waves 2 and 4 and the magnitude was diluted in the pooled regression.

When it comes to privacy concerns, the effects in the separate logit models are only significant in wave 2 and wave 4. When the data are pooled, the effects become significant for most categories with all categories being more likely to consent than those who report that they are very private (i.e. strongly agree). The wave dummy variables were both significant at $p<0.01$. Respondents in wave 2 and 4 were less likely to consent than in the first wave. One possible reason is the change in the survey agency between waves 1 and 2 and the change in the wording of the consent question between these two waves. This has led to a decline in consent which then rose again between wave 2 and 4.

All other variables such as CM’s gender, housing tenure, language spoken at home, MR’s self-reported health, income and response history have mostly non-significant effects.

In table 6, the probability of switching consent behaviour between two consecutive waves is modelled using a conditional logit approach. The dependent variable is a binary variable taking the value of 0 if the respondent had the same behaviour over two consecutive waves and 1 if the respondent switched behaviour. Note that the analytical samples were restricted to consenters or non-consenters in the initial wave depending on the model. Table 6 shows that some of the correlates have a strong and significant impact on the likelihood of switching behaviour over time.

Table 6: Cross-sectional logit regressions of consent behaviour changing over time.

	W1 to W2 (Yes to No)=1		W1 to W2 (No to Yes)=1		W2 to W4 (Yes to No)=1		W2 to W4 (No to Yes)=1	
CM's gender, reference: Girl								
Boy	1.04	(0.061)	0.91	(0.211)	1.17*	(0.112)	1.12	(0.17
Main respondent's marital status, reference: Single								
In a couple	1.15	(0.136)	0.98	(0.370)	0.88	(0.155)	1.42	(0.38
Highest socio-economic status, reference: Managerial and professional								
Intermediate	1.07	(0.097)	1.24	(0.512)	0.72**	(0.107)	0.84	(0.21
Small employers and self-employed	1.40***	(0.163)	2.49	(1.389)	0.83	(0.150)	1.28	(0.40
Lower supervisory and technical	1.37***	(0.155)	0.88	(0.444)	0.80	(0.161)	1.41	(0.53
Semi-routine and routine	1.44***	(0.126)	1.05	(0.331)	0.82	(0.136)	0.94	(0.26
Combined labour market status, reference: Both in work								
At least one in work	0.88**	(0.056)	0.71	(0.198)	1.13	(0.155)	0.88	(0.17
Both not in work	0.84	(0.129)	0.80	(0.450)	1.49*	(0.316)	0.52	(0.22
Housing tenure, reference: Own								
Rent	1.09	(0.093)	0.73	(0.234)	0.94	(0.126)	0.47***	(0.11
Other	0.96	(0.146)	0.40*	(0.195)	1.37	(0.371)	0.60	(0.31
Main respondent's ethnic group, reference: White								
Non-White	1.29	(0.229)	2.80*	(1.536)	1.96***	(0.448)	0.64	(0.20
Main respondent's religion, reference: Christian								
Non-Christian	0.59***	(0.118)	1.06	(0.518)	0.56*	(0.192)	0.50	(0.20
None	0.99	(0.067)	1.04	(0.274)	1.00	(0.108)	1.58	(0.29
Language spoken at home, reference: English								
English and other languages	1.20	(0.175)	0.72	(0.283)	1.01	(0.255)	0.96	(0.32
Only other	0.94	(0.227)	0.56	(0.278)	0.72	(0.233)	0.70	(0.37
Was the interviews translated? reference: No								
Yes	1.56**	(0.326)	0.47	(0.220)	0.83	(0.250)	3.66***	(1.66
Main respondent's health status, reference: Excellent								
Very good, good	1.08	(0.079)	0.86	(0.255)	1.12	(0.131)	0.94	(0.16
Fair, poor	0.96	(0.093)	1.65	(0.655)	1.00	(0.167)	1.33	(0.35
CM's health status, reference: Some problems								
No problems	1.11	(0.072)	1.56*	(0.395)	0.84	(0.118)	0.99	(0.24
Main respondent: I am a very private person, reference: Strongly agree								
Agree	0.76**	(0.088)	0.46*	(0.208)	1.19	(0.288)	1.18	(0.41
Neither	0.77**	(0.089)	0.37*	(0.191)	0.75	(0.187)	0.91	(0.31
Disagree	0.86	(0.096)	0.47	(0.225)	0.96	(0.245)	1.31	(0.43
Strongly disagree	0.72*	(0.131)	0.26**	(0.168)	0.57	(0.211)	1.88	(1.03
Can't say	1.00	(0.267)	0.86	(0.670)	1.87	(0.955)	1.78	(1.40
Other	0.89	(0.172)	1.15	(0.702)	4.84***	(1.472)	0.55	(0.26
Response history, reference: Participated in all waves								
Absent in at least one wave	1.15	(0.182)	0.77	(0.429)	1.57**	(0.361)	0.72	(0.28
Log OECD adjusted income	0.97	(0.064)	1.21	(0.322)	1.23**	(0.120)	0.43***	(0.07

Main respondent's age	1.00	(0.006)	0.93^{***}	(0.023)	0.98^{*}	(0.009)	0.99	(0.02
N	11,640		525		10,083		2,082	

Exponentiated coefficients; Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6 shows that respondents from socially disadvantaged backgrounds (lowest three SES categories) are more likely to switch from being consenters to being non consenters between waves 1 and 2. Ethnic minority respondents are also more likely to switch from being consenters to being non-consenters between waves 2 and 4. In contrast, non-Christian respondents are more likely to do the opposite between waves 1 and 2 and waves 2 and 4. The non-religious were more likely to switch from being non-consenters to being consenters between waves 2 and 4. Respondents who had a translated interview were likely to switch from being consenters to being non-consenters between waves 1 and 2 and to do the opposite between waves 2 and 4. All other variables have non-significant or weakly significant effects. In general, it is possible to say that the results are in line with the interpretations given previously in table 5.

Finally moving to the third strand of analyses, three models were estimated using linear probability procedures. All models are cross-sectional and examine each consent outcome apart. The **base model** includes the aforementioned correlates without interviewers fixed effects. The **model with area effects** is identical to the base model and includes the characteristics of the interviewer's assignment area (i.e. proportion of minorities, proportion unemployed, log average income, and social class composition; these were computed as averages and proportions at the interviewer's level). The **fixed effects** model is equivalent to the base model and includes the interviewers' fixed effects while excluding the assignment area characteristics since they are collinear with the fixed effects.

Table 7: Interviewers' effects for health record consent outcomes

Consent outcomes	Base model	Area effects	Fixed effects	N interviewers	N respondents
	R squared				
Wave 1	0.046	0.048	0.120	328	12,165
Wave 2	0.013	0.020	0.180	335	12,165
Wave 4	0.038	0.040	0.172	443	12,165

In table 7, the comparison of the first three columns indicates by how much the explanatory power of the model (as measured by the R-squared) has changed after the inclusion of assignment area characteristics (third column) and interviewer fixed effects (fourth column). The results show that the explanatory power of the base model is limited (R-squared varies between 1.3 and 4.6 percent depending on the wave). When the area characteristics were included, the explanatory power only rose by a small amount. However, when interviewer's fixed effects were added, the explanatory power rose by a large amount (i.e. 3 to 9 times) even though the R-squared is still modest in magnitude. The dramatic rise in wave 2 is possibly the result of the change of the fieldwork agency. In other words, in wave 2, new interviewers from a different agency were contracted to the study. This has resulted in more between-interviewers variations and in a rise in their impact. This rise has persisted in wave 4. Moreover, interviewers were probably incentivised to minimize unit non-response. Therefore, consent was not the main priority and this has led to more between-interviewer variations in obtaining consent.

In summary, it is possible to draw a number of conclusions. First, socially disadvantaged groups (i.e. low SES and ethnic minorities), are less likely to consent (table 5) and more likely to switch behaviour (table 6). Secondly, the combined impact of the interviewers explains more of the variation in consent than the socio-demographic characteristics of respondents. In general, the findings support the passiveness assumption. Respondents' behaviour is influenced by extrinsic factors (i.e. interviewers) and by their own circumstances. In other words, consent is affected by a range of factors which do not reflect actively held convictions.

Finally, it is worth noting that the cross-sectional effects of the correlates deviate from those in the pooled regression (table 5). This is a clear indication that in longitudinal surveys consent should be seen as a dynamic phenomenon and therefore a single cross sectional analysis could be misleading.

V- Conclusion

Despite the growing number of studies dealing with consent to link survey and administrative data, there is very limited knowledge of how consent works over time. This study constitutes the first exploration of consent mechanisms using three waves of data collection from the Millennium Cohort Study spanning 7 years of the lives of the cohort members. The study relies on a theoretical framework that distinguishes between passive/active and consistent/inconsistent consent behaviour and provides evidence in support of passive consistent behaviour.

Firstly, consent rates show that most respondents (i.e. 76.5%) do behave consistently over time. Secondly, the cross-equation correlations from the first analysis show that the unobserved parts of the consent outcomes are weakly associated over time, and therefore, cannot really be held to indicate the existence of strongly held latent convictions. Thirdly, the likelihood of consent and the likelihood of switching behaviour over time are related to the respondents' circumstances, and to the variation in the impact interviewers have on the MRs willingness to consent. These three findings indicate that, for the majority of respondents, consent is not driven by personal convictions but rather depends on the circumstances of the respondent at the time of the interview and on the potential influence of the interviewers. In other words, consent behaviours are passive and subject to circumstantial change even though they are consistent over time.

Our findings also show that cross-sectional analyses of consent for health record linkage could be misleading. Analyses which draw on the dynamics of change reveal that the decision to consent is not a fixed attribute. In this instance however the large majority of the original sample (over 75%) do behave consistently whilst a significant minority do not. Future research needs to focus on the motivational factors which may distinguish the willingness to consent to health record linkage from the willingness to co-operate as respondents in general.

When it comes to using linked survey and administrative data, users need to take into account the sample bias resulting from the correlations between the respondents' characteristics and consent. Moreover, depending on the nature of the longitudinal survey (e.g. age group), it is possible to predict the evolution of the characteristics of the sample and subsequently consent behaviour. In this case, interviewers can be briefed on what to expect and the interviewing approach can be adapted to take into account the possible changes to respondents' circumstances.

References

Armstrong, V. Julie, B. Helen, C. Michelle, M. Moran-Ellis, J. and Shepherd, R. (2008). Public Perspectives on the Governance of Biomedical Research: A Qualitative Study in a Deliberative Context. London, UK: Wellcome Trust.

Baker, R. Shiels, C. Stevenson, K. Fraser, R. and Stone, M. (2000). What Proportion of Patients Refuse Consent to Data Collection from Their Records for Research Purposes?'' British Journal of General Practice 50, 655-56.

Cappellari, L. and Jenkins, S. (2003). Multivariate probit regression using simulated maximum likelihood. *The Stata Journal.* 3 (3), 278–294.

Cappellari, L. and Jenkins, S. (2006). Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *The Stata Journal*. 6(2), 156–189.

Cialdini, R. Wosinska, W. Barrett, D. Butner, J. and Gornik-Durose, M. (1999). Compliance with a Request in Two Cultures: The Differential Influence of Social Proof and Commitment/Consistency on Collectivists and Individualists. *Personality and Social Psychology Bulletin*. 25 (10), 1242-1253.

Dunn, K. Jordan, K. Lacey, R. Shapley, M. and Jinks, C. (2004) Patterns of consent in epidemiologic research: evidence from over 25,000 responders. *American Journal of Epidemiology,* 159, 1087–1094.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.

Huang, N. Shih, S. Chang, H. and Chou, Y. (2007). Record Linkage Research and Informed Consent: Who Consents? *BMC Health Services Research* 7, 18.

Jenkins, S. Cappellari, L. Lynn, P. Jäckle, A. and Sala, E. (2006). Patterns of consent: evidence from a general household survey. *Journal of the Royal Statistical Society Series A*, 169(4), 701-722.

Kho, M. Duffett, M. Willison, D. Cook, D. and Brouwers, M. (2009). Written Informed Consent and Selection Bias in Observational Studies Using Medical Records: Systematic Review. British Medical Journal 338:b866.

Korbmacher, J. and Schroeder, M. (2013). Consent When Linking Survey Data with Administrative Records: The Role of the Interviewer. *Survey Research Methods,* 7, 115–131.

Kreuter, F. and Sakshaug, J. (2014). The effect of benefit wording on consent to link survey and administrative records in a web survey. *Public Opinion Quarterly,* 7(2) 133 – 144.

Knies, G. Burton, J. sala, E. (2012). Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population. *Health Services Research*, 12, 52.

Millennium Cohort Study, Ethical Review and Consent (2012). http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=1601&itemtype=document

MCS Technical Report on Sampling (4th edition, CLS, 2007). http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=409&itemtype=document

MCS Technical Report on Response (3rd edition, CLS, 2010). http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=607&itemtype=document

Mostafa, T. (2014). Variation within Households in Consent to Link Survey Data to Administrative Records: Evidence from the UK Millennium Cohort Study. CLS working paper N 2014/8.

Nelson, K. Elena Garcia, R. Brown, J. Mangione, C. M. Louis, T. A. Keeler, E. and Cretin, S. (2002) Do patient consent procedures affect participation rates in health services research? *Medical Care*, 40, 283–288.

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review*, *60*, 393-404.

Olson, J. A. (1999) Linkages with data from Social Security administrative records in the Health and Retirement Study. *Social Security Bulletin*, 62, 73–85.

Plewis, I. (2007). (Ed). *The Millennium Cohort Study: Technical Report on Sampling* (4th ed.). London, UK: Centre for Longitudinal Studies. Retrieved from http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=409&itemtype=document

Rogelberg, S. Luong, A. Sederburg, M. and Cristol, D. (2000). Employee attitude surveys: Examining the attitudes of noncompliant employees. *Journal of Applied Psychology, 85,* 284–293.

Rogelberg, S. Conway, J. Sederburg, M. Spitzmuller, C. Aziz, S. Knight, W. (2003). Profiling Active and Passive Non-respondents to an Organizational Survey. *Journal of Applied Psychology*, 88, 6, 1104–1114.

Sakshaug, J. Couper, M. Ofstedal, M. and Weir, D. (2012). Linking survey and administrative records mechanisms of consent. *Sociological Methods and Research*, 41(4) 535-569.

Sakshaug, J. and Kreuter, F. (2012). Assessing the magnitude of non-consent biases in linked survey and administrative data. *Survey Research Methods*, 6(2) 113-122.

Sakshaug, J. Tutz, V. and Kreuter, F. (2013). Placement, wording, and interviewers: Identifying correlates of consent to link survey and administrative data. *Survey Research Methods*, 7(2) 133-144.

Sala, E. Burton, J. and Knies, G. (2012). Correlates of obtaining informed consent to data linkage: respondent, interview and interviewer characteristics. *Sociological Methods and Research*, 41(3) 414–439.

Sala, E. Burton, J. and Knies, G. (2014). Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel. *International Journal of Social Research Methodology*, 17(5) 455-473.

Sheldon, H. Graham, C. Pothecary, N. and Rasul, F. (2007). Increasing Response Rates Amongst Black and Minority Ethnic and Seldom Heard Groups. Picker Institute Europe. http://www.nhssurveys.org/Filestore/documents/Increasing_response_rates_literature_review.pdf

Silva, M. S. Smith, W. T. and Bammer, G. (2002) The effect of timing when seeking permission to access personal health services utilization records. *Annals of Epidemiology.* 12, 326–330.

Sosdian, C. and Sharp, L. (1980). Nonresponse in mail surveys: Access failure or respondent resistance. *Public Opinion Quarterly, 44,* 396–402.

User Guide to Analysing MCS Data Using STATA (1st edition, CLS, 2011). http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=1372&itemtype=document

Woolf, S. H., Rothemich, S. F., Johnson, R. E. and Marsland, D.W. (2000) Selection bias from requiring patients to give consent to examine data for health services research. Archives of family medicine, 9, 1111–1118.

Youssefnia, D. (2000, April). *Examining organizational survey response quality with OCB related job attitudes.* Paper presented at the annual convention of the Society for Industrial and Organizational Psychology, New Orleans, LA.