quality and outcomes of person-centred care policy research unit

Patterns and Effects of Nonresponse in the English Adult Social Care Survey

Juliette Malley and Jose-Luis Fernandez

DP Ref 2841

October 2012

University of Kent

The London School of Economics and Political Science

University of Oxford

The Policy Research Unit in Quality and Outcomes of person-centred care (QORU) is a collaboration involving researchers in health and social care from the Universities of Kent, Oxford and the London School of Economics (LSE) funded by the Department of Health.

Our aim is to improve the quality of health and social care of people with long-term conditions through generating high-quality evidence about need, quality and outcomes of person-centred care.

CONTACT

QORU
Personal Social Services Research Unit
University of Kent
Canterbury
CT2 7NF

Email: c.l.heath@kent.ac.uk

www.qoru.ac.uk

Acknowledgements

This is an independent report commissioned and funded by the Policy Research Programme at the Department of Health under the Quality and Outcomes of Person-Centred Care Research Unit. The views expressed are those of the authors and not necessarily those of the Department of Health. We are grateful to Ann Netten for reviewing an earlier version of this paper and to Clara Heath for administrative support.

Journal publication

Considering submitting to Health Expectations or Public Opinion Quarterly

Abstract

Response rates to surveys are falling in the UK and elsewhere. In this paper we explore the factors affecting response propensity for a survey of publicly-funded adult social care users in England and examine, using a weighting approach, the effects of nonresponse on estimates of the quality of adult social care derived from the Adult Social Care Survey. The dataset is particularly interesting because the survey is managed independently by each of the local authorities (LAs) in England, following centrally-set guidance on sampling and data collection. We are therefore able to explore the effect on response rates of variability in the management of the survey across LAs and in the characteristics of service users in the sample. We find that neither LA- nor individual-level factors have a particularly strong effect on response rates, with the regression model explaining very little of the observed variability in response propensity. Nevertheless, a number of individual characteristics and LA-level variables linked to survey management show a small but statistically significant effect on response rates. We also find that weighting for non-response bias has very little effect on the estimates of quality, with the absolute difference in unweighted and weighted estimates being in the majority of cases very small from both a statistical and policy point of view. The results indicate that there might be some potential for LAs to improve response rates by following best practice. However, in the present case and given the range of indicators available in the analysis, little gain is derived from weighting the sample to correct for potentially biasing effects of nonresponse. This does not necessarily mean that subsequent adult social care surveys will not require weighting to correct for non-response bias since samples and LA practices may vary from year to year. Also, a number of theoretically important variables such as disability and receipt of informal care were not available for analysis, and they may explain significant variations in response propensity. Until further research examines the effect of such variables we cannot therefore say conclusively that there is no biasing effect from nonresponse on the estimates of quality observed.

Introduction

In the UK (Martin and Matheson 1999) and elsewhere (de Leeuw and de Heer 2002) response rates to surveys are falling. This is a problem since in the long run such trends will undermine the usefulness of survey data. First, because the greater the loss of data the less precise the survey estimates and second, because loss of data can lead to the unrepresentativeness of samples and thus to biases in the estimates derived from them. Importantly, bias may be present even where response rates are relatively high since the degree of bias is dependent on both the pattern of missingness in the data and the extent of missing data (Rubin 1976; Groves 2006). Specifically, bias occurs where the data are not missing completely at random (MCAR). This can happen when the probability of missingness of data depends on the values of the observed data, a pattern of missing data known as missing at random (MAR), or when the probability of missingness of the data varies in line with factors which are unobserved in the data -- a pattern of missingness known as not missing at random (NMAR) (Little and Rubin 1987). In this report we explore the patterns of missingness and therefore the potential for bias due to nonresponse within a postal survey of adult social care clients in England, known as the Adult Social Care Survey (ASCS).

It is important to understand the effects of missingness in the ASCS because of how the survey estimates are used. A number of the questions within the ASCS are used to populate the Adult Social Care Outcomes Framework (ASCOF), which is a national tool designed to capture information on outcomes for publicly-funded adult social care users within each of the 152 councils with adult social services responsibilities (CASSRs) in England. The ASCOF is not a performance management tool; nevertheless, it is used nationally to monitor outcomes for service users within CASSRs, support national policy development and support the central government's responsibility to account for public expenditure to the public and Parliament. Locally, it is expected that CASSRs will use ASCOF for 'benchmarking' between areas, to support local policymaking and drive service improvement (Department of Health 2011b). Each year the ASCOF indicators are reported for each CASSR in England in a way that allows for comparison between areas. If the ASCS-based ASCOF indicators are biased due to nonresponse then this will affect not only the estimates of absolute performance but also the relative performance of CASSRs against each other. Significantly biased indicators could cause national and local policymakers to take potentially inappropriate policy decisions. It is therefore important to not only understand the extent of bias and try to estimate its effects on the survey estimates, but to develop methods to minimise the effects of bias.

One feature of the ASCS that makes it particularly interesting from the perspective of nonresponse is the way the survey is managed. Each CASSR is in charge of running the survey within their area. This means that many of the aspects of the survey process, such as sampling, following up nonrespondents and targeting hard-to-reach groups, are within the control of CASSRs. Although CASSRs must follow centrally-set guidance which covers these issues, they are allowed a degree of flexibility. There does appear to be quite a substantial degree of variation in response rates across CASSRs. Part of our interest in conducting this work was therefore to explore the degree of local variation in the management of the survey to analyse whether differences in the survey management may be important in determining response rates or whether individual characteristics are more important determinants of CASSR response rates. The intention is that such analysis would not only generate useful recommendations for improving the response to future ASCSs, but would also provide insight into the effects on nonresponse in a multi-site study, where the management of the survey is delegated to the individual sites.

The objective of this report is twofold. First to explore the extent and implications of missing data in the measurement of social care outcomes based on the ASCS survey. Specifically, we examine strategies that could control for any possible deleterious effects of missing data on survey estimates, so minimising the potential for misinterpretation of the data. Second we explore the factors that are important in determining both CASSR-level response rates and whether an individual is a respondent to the survey. This analysis should generate useful recommendations for future ASCSs and provide an insight into the effects of nonresponse in a survey managed similarly to the ASCS. To inform our analytical approach, we begin by reviewing the evidence concerning nonresponse in surveys and the potential mechanisms for nonresponse. We then describe the ASCS in some detail, paying particular attention to the ways in which CASSRs may deviate from the guidance. The results from the modelling of nonresponse and the exploration of bias are presented, followed by some discussion of the implications of our findings for future ASCSs, the ASCOF and research in the area.

Background

There are a variety of theories that seek to describe conceptually the reasons behind survey nonresponse, such as social exchange theory and leverage-saliency theory (for a summary of these theories see Dillman et al. 2009). Theories of survey participation broadly posit that non-participation is not a constant attribute of a person; rather different sets of influences act on the sample members to determine their likelihood of participation. Influences may include features of the survey design, individual characteristics, household characteristics, interviewer attributes, and the social environment. Since the aim of this research is not to test these theories but to use insights from them to analyse response propensity to the ASCS, we are therefore most interested in the types of factors these theories suggest are important. The survey attribute, individual characteristics and social environment categories are most relevant to a postal survey like the ASCS.

In postal surveys the causes of nonresponse can be divided into three categories: failure to receive the survey request (for example, because of non-delivery, or interception by another and failure to forward onto the intended recipient), refusal to participate, and an inability to respond (for example, due to illiteracy in English, or physical or cognitive impairments). Due to the nature of postal surveys, it is rarely possible to determine within which of these three categories nonrespondents fall. Nevertheless, it is helpful to use this categorisation to identify potential factors influencing response propensity within the ASCS. This conceptual framework is all the more important since empirical research identifying the types of factors that influence response propensity to postal surveys of populations similar to those surveyed in the ASCS is extremely limited.

Social exchange theory in particular, suggests that a way of preventing people refusing to engage in research is to increase the benefits of participation. Aspects of the survey design, such as incentives, can be important in this respect. However, given the focus of the ASCS on adult social care users, a more important cause of nonresponse is likely to be that a large proportion of the sample are unable to easily respond. There is evidence that visual impairments (Rahi et al. 2004) and proximity to death (Kauppi et al. 2005) affect response to postal surveys. However, other conditions, which are highly prevalent in the population, such as cognitive impairments, physical impairments, other sensory impairments, intellectual disability, mental health problems and substance misuse problems are also likely to make it difficult for the person to respond. Such conditions may also make it more likely that the intended recipient will fail to receive the survey request due to hospitalisations, for example. Given the extent of variability in local eligibility policies in England, this factor could therefore lead to differences in the rates of nonresponse between authorities.

We might expect attributes of the survey that help users to overcome their cognitive, intellectual, physical or visual impairments to be important in determining response rates. Such attributes could include the use of alternative modes of administration, such as face-to-face interviews which may be easier for disabled people, and the use of alternative version formats, such as braille, EasyRead or large print. Additionally, features of the individual's social environment, such as the availability of someone to help the person, may also be important. However, the availability of someone the help is not necessarily a positive thing in the context of this survey. These people may also act as 'gatekeepers', who, out of a desire to protect the intended recipient of the survey, may intercept the survey and choose not to pass it on.

Since the ASCS is run independently by each of the CASSRs in England, we might also expect the behaviour of CASSRs to also influence response propensity. CASSRs have a degree of flexibility over how they manage the survey, within the boundaries of centrally-set guidance. However, as we discuss in more detail below, some CASSRs do not follow the guidance to the letter and others use the allowed flexibility in the guidance to vary aspects of the survey design, for instance by using monetary incentives. Since the surveys are a mandatory requirement and not something all CASSRs would necessarily choose to do, we might expect there to be variations in CASSRs' attitudes towards the survey. This may make gaming within some CASSRs more likely and may also impact on how they manage the survey. Negative attitudes to the survey, for example, may affect the extent to which CASSRs take steps to ensure good response rates (e.g. by making alternative formats available, or by taking care when extracting the sample and checking the data held about those people). There is anecdotal evidence that some CASSRs use their data systems to identify those who will need alternative formats (e.g. due to visual impairment or illiteracy in English), prior to sending out questionnaires whilst other CASSRs wait for service users to contact the authority to request an alternative format. The quality of data the CASSR holds about the potential respondent is likely to be important in determining which of these approaches is adopted. Data quality may influence response propensity through other mechanisms; for example, where data quality is poor, the likelihood of surveys being incorrectly addressed is likely to be higher, increasing the likelihood of failure to receive the survey request.

For the ASCS we therefore assume that response propensity *(R)* is a function of the characteristics of an individual and their environment *(C)*, the attributes of the survey *(S)*, the attitude of the CASSR *(A)* and interactions between these variables. This is illustrated in Figure 1 and can be represented mathematically as:

$$R = f(C, S, A).$$

Importantly, since the survey is managed by CASSRs, the attributes of the survey vary at the CASSR level. Models of response propensity will therefore include both variables at the individual- and CASR-level, potentially requiring multilevel modelling. Such models are rare in response propensity research, although their use is increasing (Johnson et al. 2006; Durrant and Steele 2009; Smith 2011; Steele and Durrant 2011). This study therefore provides some useful evidence about the value of multilevel modelling and area-level variables in the predicting response propensity

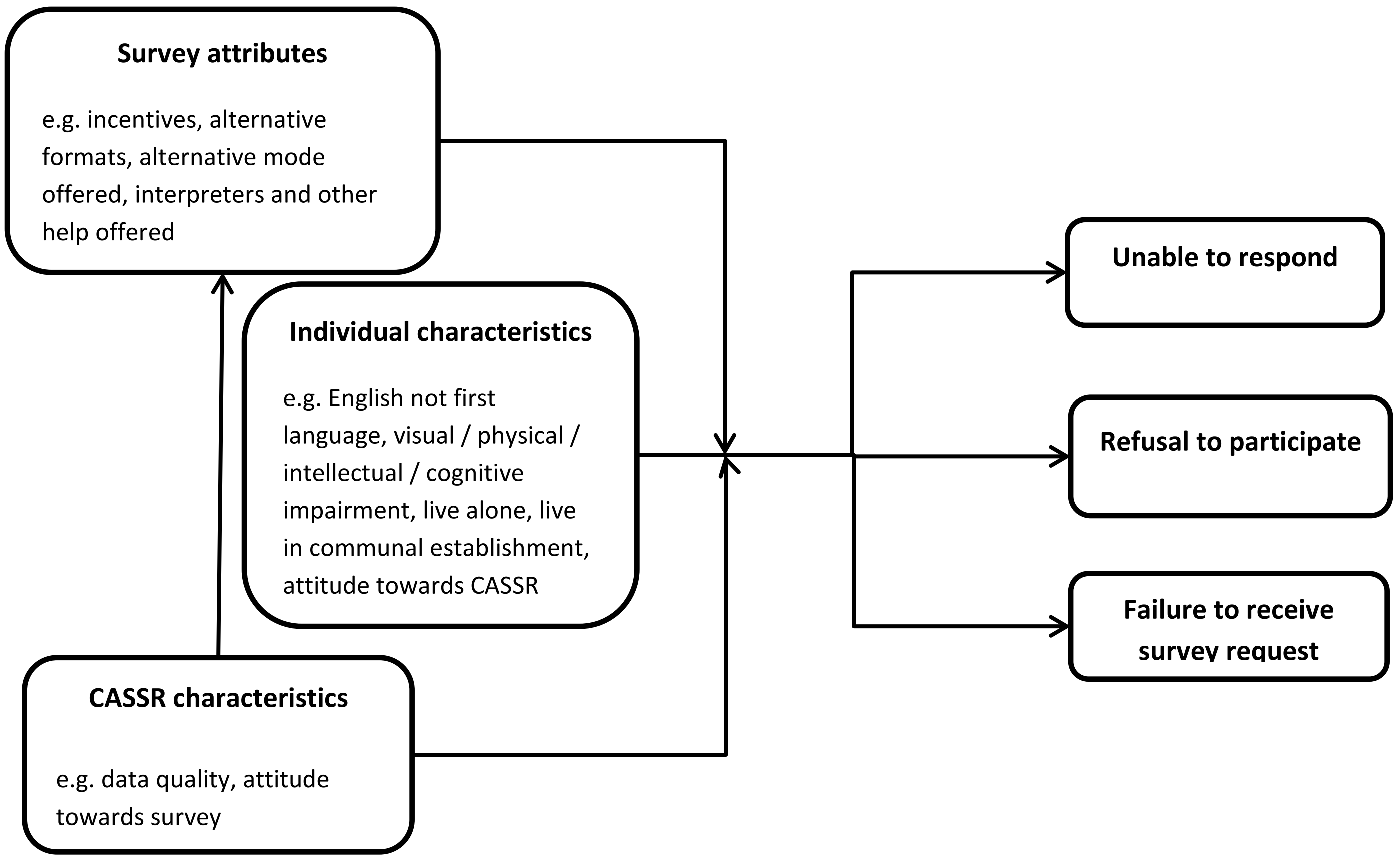

Figure 1: Illustration of the sets of factors influencing response propensity within the ASCS

The Adult Social Care Survey

Survey design

The ASCS is managed and run at a local level by each CASSR following centrally-set guidance (The Information Centre for Health and Social Care 2010) and is a mandatory requirement for all CASSRs in England. The guidance covers sampling, data collection and data management issues. More specifically, it requests CASSRs to select from their records a random sample of eligible service users aged 18 and over, where eligibility was defined as "a person receiving services[1] on 30 September who had the capacity to consent to take part in the survey". CASSRs are instructed to take care to remove people from the sample who have died or moved away. Guidance for calculating sample size is also provided and followed standard ASCS practice that councils should aim to achieve a margin of error of no more than ± 5 per cent for each question (The Information Centre for Health and Social Care 2010). All participating authorities also agree to use the same questionnaire (Malley et al. 2010).

Response to the survey

In 2011, 149 out of 152 CASSRs in England participated in the ASCS[2]. These 149 CASSRs sent out a total of 150,676 questionnaires[3] and of these, 16,294 were returned blank (indicating that the person did not wish to respond), and 61,105 were returned completed (or partially completed). This corresponds to a 41 per cent overall response rate for the sample. However, 74 questionnaires could not be linked back to the CASSR client records as the questionnaire identification number was missing, either because it had been removed by the service user (five cases) or through administrative error (in tota,l 69 cases across two councils). These cases were effectively treated as nonrespondents in the dataset to avoid duplication. Although this solution is not ideal, the number of cases was very small in relation to the size of the sample and re-estimation of models excluding the affected CASSRs did not substantially change the findings.

The average response rate for CASSRs was similar to the overall response rate at 42 per cent but, as we have found in previous surveys, the response rate varied significantly across CASSRs ranging from 21 to 82 per cent, as shown in Figure 2.

---

[1] The service user is receiving one or more services provided or commissioned by social services which are part of a care plan following a Community Care Assessment and the care being received is managed by the CASSR. Services provided or commissioned by an NHS health partner under section 75 arrangements are also included

[2] Two councils are exempt from the survey as the number of service users within their area who meet the survey eligibility criteria were too small to guarantee statistically robust results. A further council chose not to participate in this survey, despite it being mandatory.

[3] This excludes the eight questionnaires sent to people who should not have been included in the survey.

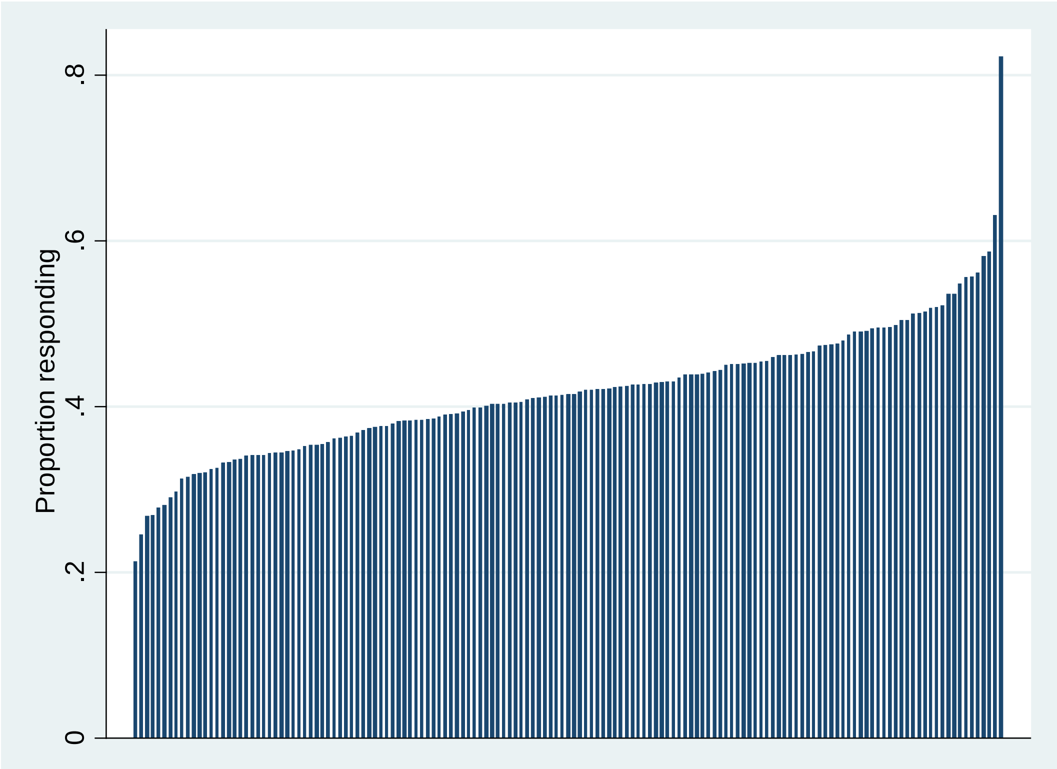

Figure 2: Distribution of unit response rates to the survey across CASSRs

Variations in the sampling, survey management and data collection processes

Although all CASSRs have to follow guidance on sampling and the management of the survey they have some freedom within the guidance to modify the process to suit their circumstances. To capture the degree of variation across CASSRs and to understand the effect this may have on nonresponse, CASSRs agreed to collect paradata (data on the management of the survey, such as the date at which questionnaires were sent out and returned) and provide a commentary of problems encountered.

The central guidance encouraged some variations in the process used for implementing the survey. Other variations were allowed but discouraged, whilst some were not allowed and were considered to be deviations from the guidance. The types of variations are shown in Table 1, according to their status within the guidance.

Table 1: Variations in survey management and their status in the guidance

Variation in survey management	Status in guidance
Incentives	Allowed but not recommended
Different delivery mode (e.g. telephone, face-to-face, interpreter, supported completion)	Main mode required to be postal to avoid response bias Alternative modes allowed and encouraged for chasing nonrespondents and engaging hard-to-reach groups
Different format or version of questionnaire (e.g. braille, audiotape, translated, large font)	Allowed and encouraged to engage hard-to-reach groups
Chasing nonrespondents	Alternative modes and formats/versions of the questionnaire allowed Required to send at least one reminder and allowed to send a maximum of two reminders Required to send reminders to whole nonrespondent population
Sampling frame	Required to exclude clients who lacked capacity to consent due to ethical requirements Allowed to remove people who had recently taken part in surveys Allowed to remove people who had requested not to take part in future surveys Allowed to exclude people in active dispute with the CASSR No other exclusions allowed
Supplementary sample	Not recommended in the guidance but allowed later for CASSRs that failed to meet target for respondents
Date questionnaires were sent out	Recommended mid to late January in order to complete fieldwork before the census date (end March), deviation discouraged but allowed
Addition of questions	Allowed, but neither encouraged nor discouraged
Modification of questions	Not allowed

Many CASSRs took advantage of opportunities to modify the management of the survey and others deviated from the guidance. The extent of variation in the use of incentives, methods for chasing nonrespondents and engaging minority groups, exclusions, timing of the survey and adding questions is shown in Table 2. Two CASSRs (1%) also reported modifying questions[4], but these modifications were minor formatting changes. In addition, twenty-five CASSRs (17%) reported 'other' deviations from the guidance, which included surveying a second supplementary sample in order to meet the target sample (6, 4%), using telephone interviews as the preferred mode of delivery (1, 1%), sending out questionnaires in batches rather than on one day (1, 1%), chasing a sample of nonrespondents rather than all nonrespondents (2, 1%), and exclusion of mental health clients from the sample due to problems generating a sampling frame for this group (3, 2%). A number of authorities also reported that they had not chased nonrespondents but a closer inspection of the commentary provided by CASSRs made it clear that others had also deviated in this way but not reported it as a deviation. In total 16 (11%) of CASSRs did not chase nonrespondents.

Table 2: Extent of variation in survey management and deviation from the guidance across CASSRs (N=149)

Type of variation	**Frequency**	**Percentage**
Used incentives	12	8%
How nonrespondents were chased		
Post	127	85%
Email	5	3%
Phone	17	11%
Interview	5	3%
no chasing took place*	16	11%

[4] One authority modified the yellow highlighting to make it specific to its situation and the other authority modified the formatting of the numbers slightly for its form recognition software. These are not major alterations and there is no point exploring any further with this variable.

Type of variation	Frequency	Percentage
How engaged minority groups		
IC translated questionnaires	33	22%
locally translated questionnaires	8	5%
interpreter via phone	21	14%
interpreter via face-to-face interview	12	8%
friend/family member provided interpretation	30	20%
no engagement*	78	52%
Excluded people who participated in recent survey	13	9%
Sent out questionnaires		
early January	4	3%
late January	38	26%
early February	73	49%
late February	39	26%
early March	22	15%
late March	7	5%
Added questions	27	18%

*These categories were generated from responses to the questions and comments. The question was not directly asked of CASSRs.

A further issue identified when examining the commentary was that some CASSRs had carried out their own surveys or consultations very close to or over the period of the ASCS. In total, 10 CASSRs (seven per cent) reported carrying out their own survey close to the period over which the ASCS data was collected.

A key aspect of the sampling, required by the ethics committee, was for CASSRs to exclude all people lacking the capacity to consent. Due to differences in the availability of appropriate information and resources across CASSRs, there were some concerns about the ability of CASSRs to execute this task with the same degree of thoroughness. CASSRs were asked to report the percentage of people excluded from the initial sample on the basis that they lacked the capacity to consent to participate in the survey. The variability across CASSRs is shown in Table 3.

Table 3: Distributional statistics for CASSR-level data on removal of people lacking the capacity to consent to participate (N=146)

	Mean	SD	Median	Max	Min
Proportion of sample removed due to lacking capacity	11.45	13.44	7.20	58.80	0.00

CASSRs also collected information about the way data was inputted into the data return, a factor which could influence the quality of the data. The different methods used for inputting questionnaire and client record data are shown in Table 4.

Table 4: Variation across CASSRs in the way survey and client record data is input into the data return (N=149) *

	Frequency	Percentage
Information from client records is input		
Manually	51	34%
Using a database	85	57%
Other	23	15%
Information from questionnaire is input		
Manually	119	80%
Using software	17	11%
Other	16	11%

*Percentage does not sum to 100 as some CASSRs used multiple approaches

Auxiliary information on service users

CASSRs agreed to collect a number of auxiliary variables from their databases for each sample member. This information can be used in order to assess the representativeness of the sample achieved. The choice of variables available from CASSR databases was limited, with key variables, like disability and receipt of informal care, not available. Some CASSRs were not able to provide all of the requested variables and the quality of some of the data was often poor with data missing for many cases or, in the case of the budget data, appearing to have a number of errors making it unreliable[5].

Table 5 shows the extent of missingness of the auxiliary data across the CASSRs as well as the overall completion rate for each variable within the sample. As is clear from this table, only the variant of the questionnaire delivered (translated version and care home, learning disability or standard) was fully observed across the CASSRs. Some of the variables that are partially observed across several CASSRs, such as sex, age group, ethnicity and several services, actually have high completion rates indicating that, although cases are missing for a number of CASSRs, the number of missing cases is not high. It is only the budget, religion, sexual identity and secondary client group data that are poorly completed, being missing for entire CASSRs.

Table 5: Extent of missingness of auxiliary data across the CASSRs and within the overall sample

	Number of CASSRs			Sample completion rate
Variable	**Fully observed**	**Partially observed**	**Completely missing**	
Questionnaire variant	149	0	0	100.0%
Sex	135	14	0	100.0%
Age group	127	22	0	99.9%
Ethnicity	17	132	0	98.3%

[5] CASSRs were asked to provide the gross budget but in many cases it appeared that the budget net of user contributions was provided. It is not clear to what extent this has occurred across CASSRs and it is difficult to identify erroneous budgets for exclusion. There are also differences across CASSRs in the composition of the budget, as not all CASSRs were able to provide the budget for non-chargeable services, in-house services and services that are frequently shared with the NHS, such as equipment and mental health services. For this reason we have not been able to use this data.

	Number of CASSRs			Sample completion rate
Variable	**fully observed**	**partially observed**	**Completely missing**	
Sexual identity	0	21	128	1.1%
Religion	4	128	17	52.2%
Primary client group	117	32	0	99.6%
Secondary client group	49	44	56	42.8%
Residential care	143	6	0	100.0%
Nursing home	141	8	0	99.1%
Home care	139	10	0	99.0%
Day care	139	10	0	98.3%
Meals	136	7	6	94.6%
Short-term residential care	137	11	1	97.5%
Direct Payments	141	8	0	97.7%
Personal Budgets	133	9	7	92.2%
Professional support	134	8	7	93.8%
Equipment	138	8	3	96.3%
Other services	136	10	3	95.6%
Total budget	13	70	66	41.0%

The frequency distribution for each auxiliary variable, except the budget data, is shown in Table 6. It is clear from the table that some groups are small (particularly when crossed with response status). Those with mixed ethnicity, people who report their religion as Buddhist, Jewish or Sikh, those who are reported to have substance misuse problems as their primary client group, and most secondary client groups are all quite small. The lack of information about secondary client group and religion across CASSRs means that we are unable to use these variables in the multivariate analysis as we would lose too many degrees of freedom. We also considered these variables inappropriate for imputation as they were missing in at least 50 per cent of cases (Rubin 2003).

Table 6: Distribution of responses to the auxiliary data, and relationship of each auxiliary variable with ASCS response status

		Percentage returning…		
Variable	**Frequency**	**Completed questionnaire**	**Blank questionnaire**	**Questionnaire not returned**
Questionnaire variant (N=150,676)				
Standard community-based	106,028	40.4%	11.6%	48.0%
Standard care home based	25,148	37.9%	11.6%	50.5%
LD community-based	14,213	42.1%	5.8%	52.2%
LD care home based	5,287	51.2%	5.2%	43.6%
Sex (N=150,605)				
Male	55,971	39.6%	10.1%	50.3%
Female	94,634	41.0%	11.2%	47.7%
Age group (N=150,519)				
18-24	5,176	30.0%	7.9%	62.0%
25-30	7,889	32.2%	8.6%	59.1%
31-39	10,983	35.4%	8.8%	55.8%
40-49	14,057	39.2%	8.5%	52.3%
50-64	14,982	42.7%	8.9%	48.5%
65-74	19,060	43.0%	10.6%	46.4%
75-84	35,474	42.1%	11.8%	46.1%
85 and over	42,898	41.8%	12.8%	45.4%

Variable	Frequency	Percentage returning… Completed questionnaire	Blank questionnaire	Questionnaire not returned
Ethnicity (N=148,142)				
White	133,237	41.3%	10.8%	47.9%
Mixed	977	33.6%	9.5%	56.9%
Asian	6,695	34.7%	12.9%	52.4%
Black	5,627	34.8%	7.5%	57.7%
Other	1,586	35.1%	13.3%	51.6%
Religion (N=78,719)				
None	8,999	39.6%	11.0%	49.4%
Christian	58,754	42.7%	10.7%	46.6%
Buddhist	144	38.2%	11.1%	50.7%
Hindu	1,287	42.2%	13.8%	44.0%
Jewish	1,016	45.7%	7.9%	46.5%
Muslim	3,092	32.5%	9.3%	58.2%
Sikh	748	33.3%	25.3%	41.4%
Other	4,679	39.6%	13.6%	46.9%
Primary client group (N=150,040)				
Physically disabled	102,186	42.6%	11.2%	46.2%
Mental health problem	23,666	28.8%	13.1%	58.1%
Learning disability	20,056	44.3%	5.7%	50.1%
Substance misuse problem	633	21.5%	10.1%	68.4%
Vulnerable person	3,499	41.6%	13.3%	45.1%

Variable	Frequency	Percentage returning… Completed questionnaire	Blank questionnaire	Questionnaire not returned
Secondary client group (N=64,437)				
Physically disabled	3,365	37.3%	10.8%	51.9%
Mental health problem	2,066	33.2%	12.5%	54.3%
Learning disability	371	41.5%	11.9%	46.6%
Substance misuse problem	276	23.9%	12.3%	63.8%
Vulnerable person	1,471	41.9%	16.5%	41.6%
None	56,888	42.2%	8.6%	49.1%
Personal care only home (n=150,608)				
Resident	23,167	42.2%	9.5%	48.4%
not a resident	127,441	40.2%	11.1%	48.7%
Nursing home (N=149,382)				
Resident	7,732	33.9%	13.4%	52.7%
not a resident	141,650	40.9%	10.7%	48.5%
Home care (N=149,128)				
Recipient	49,819	43.9%	9.0%	47.2%
not a recipient	99,309	38.9%	11.8%	49.4%
Day care (N=148,148)				
Recipient	20,296	42.4%	9.2%	48.4%
not a recipient	127,852	40.3%	11.1%	48.7%
Meals (N=142,605)				
Recipient	7,834	39.3%	12.2%	48.5%
not a recipient	134,771	40.7%	11.0%	48.3%

Variable	Frequency	Percentage returning… Completed questionnaire	Blank questionnaire	Questionnaire not returned
Short-term residential care (N=146,944)				
Recipient	2,205	36.5%	13.3%	50.2%
not a recipient	144,739	40.6%	10.8%	48.6%
Direct Payments (N=147,188)				
Recipient	12,314	43.8%	7.8%	48.5%
not a recipient	134,874	40.3%	11.1%	48.7%
Personal Budgets (N=138,970)				
Recipient	19,593	43.1%	9.6%	47.3%
not a recipient	119,377	40.2%	11.2%	48.7%
Professional support (N=141,272)				
Recipient	23,822	34.5%	14.9%	50.6%
not a recipient	117,450	41.9%	10.1%	48.0%
Equipment (N=145,129)				
Recipient	30,936	43.4%	11.8%	44.8%
not a recipient	114,193	39.9%	10.7%	49.4%
Other services (N=144,037)				
Recipient	11,621	42.1%	11.5%	46.5%
not a recipient	132,416	40.4%	10.6%	49.0%
All services (N=128,056)†				
recipient of at least one service	125,324	40.9%	11.1%	47.9%
does not receive any services	2,732	30.7%	9.0%	60.3%

†Derived from the service receipt variables.

The extent of completion of the auxiliary data was used to generate further indicators of CASSR data quality. Since these data are extracted from the local client record database, they can be used to proxy coverage and quality of the CASSR client records. We generated four variables: a count of auxiliary data variables that were fully observed, a count of the number partially observed, a count of the number missing, and the overall percentage of auxiliary data that was missing from the sample. The distribution of these data across CASSRs is shown in Table 7.

Table 7: Distributional of CASSR-level quality proxy indicators (N=149)

	Mean	SD	Median	Max	Min
Count of fully observed auxiliary variables	18.63	2.94	19	24	5
Count of partially observed auxiliary variables	4.46	3.04	4	18	0
Count of missing auxiliary variables	2.91	1.82	3	7	0
Proportion of auxiliary data missing	0.16	0.07	0.16	0.58	0.04

Additional CASSR-level data

In addition to the CASSR-level paradata, we downloaded data that reflected local conditions and features of the SSD that could potentially explain patterns of missingness. These included data on population density (Office for National Statistics 2001), which has been fairly consistently identified as a predictor of response propensity in general population surveys (Groves 2006) and area indices of multiple deprivation (McLennan et al. 2011), which have also been shown to affect response propensity (Johnson et al. 2006). The distributional characteristics of these indicators are shown in Table 8.

Table 8: Distributional statistics for CASSR area characteristics

	Mean	SD	Median	Max	Min
Density (people/hectare) (N=148)	24.02	26.47	15.105	131.02	0.61
Deprivation (average IMD2010 score) (N=149)	23.16	8.62	23.10	43.40	5.40

Social Care Outcome Indicators

Five outcome indicators are specified in the Adult Social Care Outcomes Framework (ASCOF) from the ASCS:

- the ASCOT measure of self-reported current SCRQoL (Netten et al. 2011; Potoglou et al. 2011; Malley et al. 2012; Netten et al. 2012)
- the safety domain from ASCOT (safety QI)
- the control over daily life domain from ASCOT (control QI)
- the satisfaction with services question (satisfaction QI)
- the access to information question (information QI).

In the *ASCOF Handbook of Definitions* the attributes in the SCRQoL measure are combined assuming equal weights for each domain (Department of Health 2011a). This decision was based on the analysis of the data from the pilot ASCS but was also justified by the lack of alternatives at that stage (Malley et al. 2010). Since then a set of utility weights (anchored to the dead state, using time-trade off, TTO) for the domains has been developed (Netten et al. 2011; Potoglou et al. 2011; Netten et al. 2012). We use both the equally-weighted and utility (TTO)-weighted versions of the ASCOT SCRQoL measure as outcome indicators in the analysis.

The distributional and missingness statistics for the QIs are shown in Table 9. Missingness is much higher for the ASCOT measures, since these are composed of eight separate items from the questionnaire. In addition, the information QI has a much lower number of observations than the other items because it has a not applicable (N/A) response option to allow people to report that they had never tried to seek out any information. The variation in these QIs across the 149 participating CASSRs is shown in Figure 3, along with approximate 95 per cent confidence intervals (adjusted with the finite population correction factor). The graphs show that there are likely to be some differences across CASSRs in the QIs[6].

[6] The confidence interval provides an indication of the uncertainty associated with the estimate of the mean for each CASSR. Importantly, and contrary to popular opinion, differences are not observed between CASSRs where the errors bars do not overlap. The error bar required to represent this is much more difficult to construct. However, the correct error bar to achieve five per cent significance can be approximated by an error bar of width 1.39 times the standard error of the mean (Goldstein and Healy 1995). It should be noted that this approximation is for a single comparison, not multiple comparisons.

Table 9: Distributional and missingness statistics for the QIs for the whole sample

	SCRQoL-TTO (n=54,350)	SCRQoL (n=54,350)	Satisfaction QI (n=57,929)	Control QI (n=59,478)	Safety QI (n=59,688)	Information QI (n=42,884)
Mean	0.80	18.62	0.62	0.75	0.95	0.74
SD	0.19	3.93	0.49	0.43	0.22	0.44
Max	1.00	24	1	1	1	1
Min	-0.17	0	0	0	0	0
% Missing	12.3%	12.3%	5.3%	2.6%	2.2%	8.0%
% N/A	n/a	n/a	n/a	n/a	n/a	34.3%

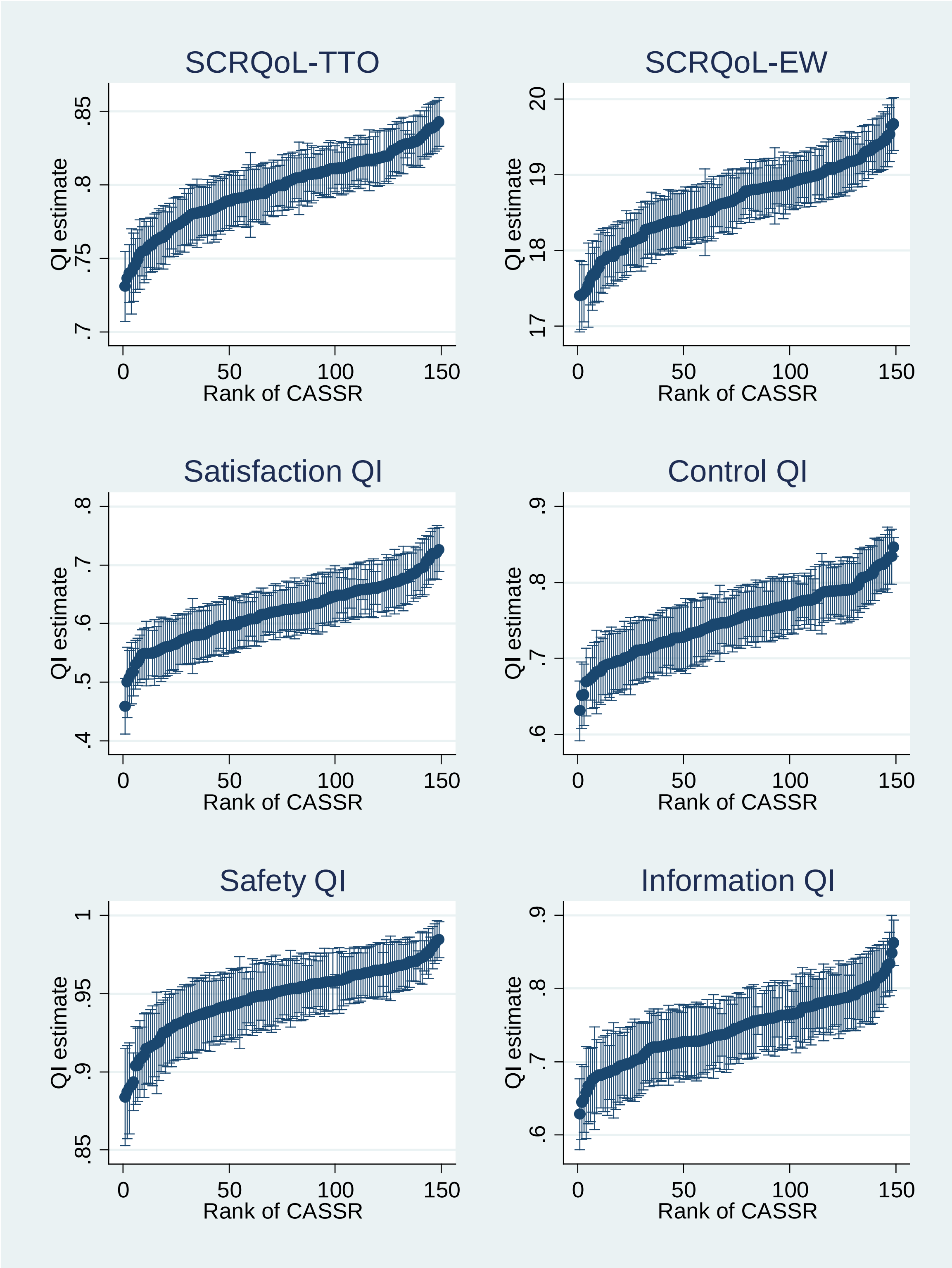

Figure 3: QI estimates for CASSRs with approximate 95 per cent confidence intervals, for each of the QIs

Statistical analysis

The aim of the statistical analysis is twofold:

- to understand the patterns of missingness in the ASCS and identify the key factors influencing missingness;
- to assess the likely bias resulting from nonresponse and to identify ways to correct for any potential bias.

Our strategy is to develop a statistical model to predict response propensity using the auxiliary data collected about the characteristics of individuals within the sample and the characteristics of CASSRs and the management of the survey within CASSRs. We then use the predicted probabilities from this model to develop weights to rebalance the respondent sample for nonresponse. However, we must first deal with missingness within the auxiliary data. We attempt to recover the missing auxiliary data using multiple imputation, as we describe in more detail below.

Regarding the response propensity model, response status, *R*, is recorded as respondent or nonrespondent, where the nonrespondent category is composed of two subgroups: nonrespondents who returned blank forms (blank form) and nonrespondents who did not respond to the survey request (no response). Those responding with blank forms are treated as nonrespondents as the survey cover letter, in line with ethics committee requirements, instructed survey recipients to return a blank form if they did not wish to participate in the survey.. To accommodate this trichotomous nominal outcome, we employ a multinomial logistic (MNL) regression model (Long and Freese 2006). A Wald test to determine whether the three alternatives are distinguishable was significant for each pair, and of importance the "blank form-no response" pair ($X^2(26)=860.72$, $p<.001$), indicating the appropriateness of the MNL model. Nevertheless, for illustrative purposes we also present results from a binomial logistic (BNL) regression model on the dichotomous "response-nonresponse" outcome. To account for the effect of clustering of responses by CASSR on standard errors we use the sandwich variance estimator (Rogers 1993).

We also apply a multi-level random-effects regression model (Rabe-Hesketh and Skrondal 2008) to model the effect of CASSR-level clustering of responses. The value of using such a model is that it computes separate intercepts for each CASSR and directly estimates between-CASSR variance with the model parameter, ζ. The parameter, ζ, therefore captures any unobserved systematic heterogeneity between CASSRs not captured by the observed CASSR-level variables. It also allows us to test for differences in response propensity between CASSRs and apportion the variation in response propensity between the individual- and CASSR-level. This is important because differences in the extent and patterns of missingness between CASSRs may be due to differences in the processes they use (not captured by CASSR-level variables) and differences in attitudes towards the survey.

The choice of explanatory variables in the models is driven by the conceptual framework, the availability of data and their ability to predict response propensity. Main and interaction effects were explored, although no interaction effects were retained due to a lack of stability in the estimates. Only those variables for which the β-coefficient was found to be significant ($p<.10$) were retained in the models, unless they were considered to be theoretically important, in which case they were retained despite their lack of significance. The final set of variables included is shown in Table 10 with a rationale for their inclusion in the model.

Table 10: Variables included and retained in the response propensity models

Variable	Rationale
Age group	Important for sample representativeness. Much research has found young and very old less likely to respond (Herzog and Rodgers 1988; Kaldenberg et al. 1994; Elliott et al. 2005).
Ethnicity	Important for sample representativeness. May be a weak indicator of proficiency with English.
Client group (physical disability[7], learning disability, mental health problem, substance misuse problem)	Client groups have different problems that make it difficult to answer the questionnaire. Some groups, such as mental health users and people with substance misuse problems, may have less contact with social services and may see the survey as less relevant.
Resident in nursing home	Residency in a nursing home may act as an indicator of disability severity, since people in nursing homes are more likely to be cognitively impaired or severely physically impaired. Such people will be most unable to answer the questionnaire without help and, although staff are able to help, the extent to which they do may be dependent on the home's acceptance of the survey.

[7] Vulnerable people are also included in the physically disabled group since they do not differ from them in the regression.

Variable	Rationale
Resident in residential home	Residency in a residential home may act as an indicator of disability severity since people in residential homes are more likely to be cognitively impaired or severely physically impaired compared to people living in their own homes (although not as impaired as people in nursing homes). Such people will be most unable to answer the questionnaire without help and, although staff are able to help, the extent to which they do may be dependent on the home's acceptance of the survey.
In receipt of Direct Payments	May affect response rate as recipients may see the survey as checking up on their ability to manage the cash benefit appropriately.
In receipt of low-level services only	Low-level services include meals, day care and other services, such as transport. These services are generally less frequent and people receiving only these services may be less likely to respond as they do not think the survey is relevant to them.
In receipt of short-term residential care	Short-term residential care is a one-off service and people who have received this service may be less likely to respond as they do not think the survey is relevant to them.
In receipt of equipment	People who receive equipment will have little contact with social services and may not think the survey is relevant to them.
CASSR did not chase nonrespondents	Not chasing respondents will mean that people have fewer opportunities to respond and that there are fewer opportunities to overcome barriers to respond in place due to gatekeepers, forgetfulness or periods away from the home (e.g. because of hospitalisations)
CASSR did not report using any strategies to engage hard-to-reach groups	CASSRs that use multiple strategies to engage hard-to-reach groups are likely to have better response rates from such groups and therefore better overall response rates

Variable	Rationale
Proportion of people CASSR removed due to lacking capacity	CASSRs that take steps to remove people lacking capacity are more likely to have better response rates as they will have fewer surveys returned due to the person being unable to answer the survey due to severe cognitive impairment.
CASSR added further questions	There is evidence that longer questionnaires have lower response rates (Groves 2006)
Number of auxiliary items fully observed for a given CASSR	Indicator of the quality of the CASSR's data. Poor quality data may make it less likely for someone to receive the survey request and it may make it more difficult for the CASSR to target help or alternative modes/questionnaires effectively to people who need help to respond.
CASSR used incentives	There is evidence that incentives improve response rates (Dillman et al. 2009).
Deprivation score for CASSR	Other studies have found that deprivation of the area affects response rates (Johnson et al. 2006).

Assessing the potential bias in outcome estimates due to nonresponse is complicated by the lack of complete data on the characteristics of the full sample or the population more generally. Indeed, the analysis can only estimate the degree of representativeness of the sample and the potential for correcting any related bias, with respect to the indicators collected in auxiliary and paradata.

Whilst the correlation between the propensity score and the outcome indicators gives an indication of the likelihood of bias (Little 1986), it does not give an indication of the extent of the bias and, therefore, its importance. We attempt to estimate the extent of bias due to the presence of missing information by looking at the difference between a weighted version of QI, where the weight is designed to adjust the sample for the probability of responding, and the 'raw' unweighted QI (Elliott et al. 2005; Höfler et al. 2005; Groves 2006). Specifically, the weights are generated from the response propensity models already described and are inversely proportional to the estimated probability of being a respondent. They are applied such that missing or incomplete units in the sample are ignored and the complete data are inflated to reflect the probability of response for that unit[8].

[8] There are many different methods for weighting data for nonresponse (for a review see Kalton and Flores-Cervantes 2003), but the most appropriate weighting method for this dataset is one that uses

Due to missingness in the auxiliary and paradata, we are only able to construct a weight for those cases with full information (i.e. 124,072 cases out of 150,676 (82 per cent) and 128 of the 149 authorities (86 per cent)). As well as being an impractical method for correcting for nonresponse, this practice could itself introduce bias should the cases excluded from the propensity models, because they have partially observed data, be different from those with fully observed data. Since we find this to be the case for these data (see Table 11), we therefore attempt to recover the missing data on the characteristics of the full sample using chained equations multiple imputation (van Buuren 2007; White et al. 2011). We use this method since it allows greater flexibility in the specification of imputation equations for the missing variables – important for these data given the preponderance of nominal and ordinal variables. We also use additional CASSR-level data on service receipt and social care expenditure to allow for shifts in service receipt by CASSR (The Information Centre for Health and Social Care 2012c, a).

The magnitude of the estimated bias, and thus the relative importance of controlling for it, should be considered with respect to sampling error. We therefore compare the bias to the standard error (SE) of the unweighted estimate. Where bias is smaller than the SE, sampling variation dominates bias and the degree of bias could be considered to be quite small since it would only marginally shift the confidence interval for the estimate and most weighted estimates would fall within the confidence interval of the unweighted measure. Where bias is greater than the SE, however, bias dominates the sampling variation and could be considered to be quite large.

In the past, performance against each indicator for the survey-based indicators has been presented graphically in the form of bar charts and caterpillar plots which show the relative rank of each CASSR and also the error associated with each estimate (see e.g. (Department of Health 2003b, a; Commission for Social Care Inspection 2004; Department of Health 2004; Commission for Social Care Inspection 2007; The Information Centre for Health and Social Care 2007). We therefore explore how sensitive the ranking of the QIs is to non-

---

the sample auxiliary and paradata to identify respondents who are similar to nonrespondents and then increases the weight of respondents so that they represent similar nonrespondents. Sorting respondents and nonrespondents into adjustment cells on the basis of the auxiliary information available and then weighting the respondents in each cell by the inverse of the response rate in each cell is a straightforward method to achieve this (Little 1986; Kalton and Flores-Cervantes 2003). However, this method cannot manage continuous variables (these must be categorised) and can produce very large weights that significantly inflate the variances of survey estimates. This is particularly a problem where a large amount of auxiliary information is available as the sample sizes in adjustment cells can become very small and small sample sizes can lead to instability in the adjustments (Kalton and Flores-Cervantes 2003; Höfler et al. 2005). There are various solutions to these problems but we choose here to use response propensity score weighting largely because the response propensity model has value from a policy point of view since it can help identify factors that are important in driving nonresponse (Lepkowski et al. 1989; Mihelic and Crimmins 1997; Höfler et al. 2005).

response bias. To judge this we look at the Spearman's Rank Correlation Coefficient for the raw and weighted QIs. We also compute a Kendall's tau (τ) correlation coefficient for each weighted QI against the unweighted QI. A Kendall's τ of one indicates that CASSRs were ranked in exactly the same order, both before and after weighting. The tau coefficient allows us to quantify the proportion of possible CASSR pairs that changed order after the adjustment. Following Johnson et al (Johnson et al. 2010), we interpret a Kendall's τ correlation coefficient of 0.8 to mean that 10 per cent [(1–0.8)/2] of all possible CASSR pairings changed order after weighting.

All analysis is conducted in STATA version 12, using standard programmes for logistic regressions and multiple imputation, and the user-written programmes SPost2, to interpret the results from the multinomial logistic regression model (Long and Freese 2006), and gllamm for the random-effects multinomial logistic regression model (Rabe-Hesketh et al. 2004).

Table 11: Differences in the characteristics of cases with completely observed auxiliary information and at least one missing auxilliary data item

	Wald, Chi^2		**LR test**	
	Chi^2	P	Chi^2	p
Method of data collection (N=147,625)	96.91	0.000	88.41	0.000
Response status (N=150,672)	141.49	0.000	142.70	0.000
Sex (N=150,601)	0.68	0.409	0.68	0.409
Age group (N=150,515)	280.50	0.000	261.88	0.000
Ethnicity (N=148,119)	232.46	0.000	251.68	0.000
Religion (N=78,719)	1100.00	0.000	981.15	0.000
Primary client group (N=150,040)	221.04	0.000	231.90	0.000
Secondary client group (N=64,433)	254.70	0.000	292.47	0.000
Residential care home (N=150,608)	352.25	0.000	335.92	0.000
Nursing care home (N=149,382)	11.52	0.001	11.78	0.001
Home care (N=149,128)	3.73	0.053	3.73	0.054
Day care (N=148,148)	1.68	0.194	1.69	0.193

	Wald, Chi^2		LR test	
	Chi^2	P	Chi^2	p
Meals (N=142,605)	0.08	0.779	0.08	0.779
Short term residential care (N=146,944)	6.72	0.010	6.50	0.011
Direct payments (N=147,184)	0.31	0.581	0.30	0.581
Personal budgets (N=138,966)	31.45	0.000	30.67	0.000
Professional support (N=141,272)	410.74	0.000	384.92	0.000
Equipment (N=145,129)	50.26	0.000	51.19	0.000
Other services (N=144,037)	73.99	0.000	70.71	0.000
Advocate (N=150,672)	181.44	0.000	156.39	0.000
Interpreter (N=150,672)	8.22	0.004	10.38	0.001
Translated (N=150,672)	4400.00	0.000	3053.08	0.000
Type of questionnaire (N=150,672)	146.60	0.000	143.93	0.000
Replacement (N=150,672)	0.44	0.506	0.44	0.507

Results

The estimation results from the response propensity models are shown in Table 12 for the MNL specification and in Table 13 for the BNL specification. The estimations under both assumptions about the missing data mechanism for the covariates, i.e. casewise deletion, which assumes MCAR, and multiple imputation, which assumes MAR are shown in the tables. Due to the extent of missing data on the covariates, the results estimated on the casewise deletion sample have many fewer observations (124,072 out of a possible 150,672 cases) and fewer CASSRs (128 out of 149) due to certain variables being completely missing for CASSRs (see Table 2, Table 3, Table 4, Table 5, Table 7 and Table 8). Despite the differences in the samples used, it is still instructive to compare the results of the MNL and BNL specifications across the different assumptions about the missing data mechanism.

Looking first at the MNL specification, the model estimated on the multiply imputed dataset does not differ that much from the model estimated on the casewise deleted dataset. The factors that are most important (as determined by the significance of the odds ratio) in predicting the type of nonresponse (blank form versus nonrespondent), as compared to the base category of being a respondent, are very similar across the two MNL models. There are more differences for the “blank form” outcome, but given that this outcome represented only 11 per cent of the sample, some of these differences could be explained by the lower precision of these estimates. Although the models are well-specified, the pseudo-R^2 is very low and the proportion of the estimated probabilities correctly classified is extremely low at around 50 per cent, implying a rate of classification no better than chance. Therefore, despite there being some highly significant covariates, the overall ability of the model to explain response propensity is poor.

The random-effects (RE) specification of the MNL model, is also shown in Table 12 (MNL, RE) for the casewise deletion sample only since the software used does not have the capacity to estimate a RE model on multiply-imputed data. Interestingly, despite the fact that there appear to be quite large differences in response rates across CASSRs, rho for the model is very low at 0.026, but significant (likelihood ratio test, X^2=2065.48, p<0.001). This means that approximately three per cent of the variation in response propensity is due to systematic differences between CASSRs, after controlling for individual-level and CASSR-level factors. The estimate of rho for the variance components model (i.e. the model without covariates) is slightly higher at 0.039, but still low, indicating that it is overwhelmingly individual-level variation that is driving response propensity. This does not necessarily imply the weighting to correct for non-response bias is unnecessary: there is still the need for the reweighted sample to be representative for unbiased estimates about the population to be inferred.

Despite the limited contribution of systematic CASSR-level variation to response propensity, the RE model shows some differences in the estimation results, affecting the significance of some of the odds ratios, again particularly for the “blank form” outcome, but also for the service receipt and CASSR-level variables. However, all of the estimated odds ratios are broadly within the same area, and there are no reversals in the direction of effects. The differences therefore seem to be driven primarily by differences in the ways SEs are estimated in the two models and are not due to a flaw in the assumptions of the fixed effects model.

Table 12: Multinomial logistic regression models of response propensity, with fixed and random effects, under two assumptions regarding the missing data mechanism

	Casewise deletion				**Multiple imputed** (m=20)	
	MNL (N=124,072)		**MNL, RE** (N=124,072)		**MNL** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	SE
Fixed part						
Blank form†						
Mental health‡	1.602***	0.11	1.581***	0.047	1.613***	0.101
Learning disability‡	0.534***	0.037	0.528***	0.023	0.502***	0.034
Substance misuse‡	1.507	0.349	1.465*	0.245	1.605*	0.361
Age: 18-24§	1.039	0.166	1.027	0.075	1.145	0.158
Age: 25-30§	0.901	0.118	0.891*	0.051	0.907	0.112
Age: 31-39§	0.785	0.102	0.780***	0.038	0.774*	0.099
Age: 40-49§	0.690***	0.064	0.685***	0.03	0.683***	0.065
Age: 50-64§	0.658***	0.034	0.652***	0.026	0.661***	0.035
Age: 65-74§	0.801***	0.032	0.802***	0.027	0.790***	0.030

	Casewise deletion				Multiple imputed (m=20)	
	MNL (N=124,072)		**MNL, RE** (N=124,072)		**MNL** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	SE
Age: 75-84§	0.903**	0.033	0.897***	0.024	0.894***	0.028
White	0.971	0.18	0.979	0.034	0.932	0.216
Count of service types	0.865	0.103	0.931	0.042	0.838	0.085
Count of service types – sq	1.025	0.021	1.01	0.01	1.033	0.018
Nursing Home	1.715***	0.131	1.729***	0.077	1.636***	0.109
Residential Home	1.075	0.081	1.096**	0.036	1.020	0.064
Low-level services	1.741***	0.097	1.769***	0.052	1.686***	0.087
Direct Payment	0.895	0.062	0.913*	0.039	0.876*	0.059
Short-Term Residential	1.726***	0.223	1.850***	0.147	1.660***	0.185
Equipment	1.401***	0.11	1.422***	0.04	1.306**	0.112
No chase	0.717*	0.111	0.665***	0.065	0.738*	0.101
Deprivation: avscore	0.985	0.011	0.987***	0.003	0.987	0.009
Data quality: number fo	0.885**	0.039	0.877***	0.01	0.912*	0.040

	Casewise deletion				**Multiple imputed** (m=20)	
	MNL (N=124,072)		**MNL, RE** (N=124,072)		**MNL** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	SE
Remove lackcap	0.98	0.01	0.981***	0.002	0.983*	0.008
Add questions	1.022	0.216	0.994	0.091	0.884	0.165
Incentives	1.405	0.502	1.412**	0.169	1.561	0.617
No engagement	1.102	0.178	1.134*	0.068	1.017	0.173
Constant	4.333	4.308	4.385***	0.981	2.617	2.765
Nonrespondent†						
Mental health‡	1.621***	0.047	1.610***	0.032	1.580***	0.044
Learning disability‡	0.777***	0.03	0.770***	0.018	0.774***	0.029
Substance misuse‡	2.211***	0.314	2.168***	0.239	2.223***	0.271
Age: 18-24§	2.168***	0.121	2.159***	0.091	2.245***	0.108
Age: 25-30§	1.764***	0.08	1.751***	0.06	1.750***	0.075
Age: 31-39§	1.456***	0.054	1.449***	0.042	1.432***	0.049

	Casewise deletion				**Multiple imputed** (m=20)	
	MNL (N=124,072)		**MNL, RE** (N=124,072)		**MNL** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	SE
Age: 40-49§	1.221***	0.039	1.217***	0.031	1.218***	0.038
Age: 50-64§	1.038	0.032	1.03	0.024	1.038	0.031
Age: 65-74§	0.975	0.025	0.976	0.021	0.974	0.023
Age: 75-84§	0.987	0.021	0.98	0.017	0.988	0.019
White	0.822***	0.046	0.828***	0.019	0.830***	0.049
Count of service types	0.789***	0.042	0.848***	0.024	0.806***	0.039
Count of service types – sq	1.046***	0.011	1.030***	0.007	1.043***	0.010
Nursing Home	1.372***	0.064	1.377***	0.041	1.346***	0.058
Residential Home	0.939	0.041	0.956*	0.019	0.976	0.042
Low-level services	1.046	0.055	1.061**	0.021	1.067	0.049
Direct Payment	0.867***	0.024	0.888***	0.022	0.869***	0.023
Short-Term Residential	1.129	0.082	1.217***	0.066	1.170*	0.077
Equipment	0.957	0.043	0.973	0.018	0.940	0.039

	Casewise deletion				Multiple imputed (m=20)	
	MNL (N=124,072)		**MNL, RE** (N=124,072)		**MNL** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	SE
No chase	1.623***	0.198	1.556***	0.142	1.553***	0.165
Deprivation: avscore	1.013**	0.004	1.015***	0.003	1.013***	0.004
Data quality: number fo	1.017	0.042	1.013	0.011	1.011	0.028
Remove lackcap	0.997	0.002	0.998	0.002	1.000	0.002
Add questions	1.002	0.087	0.978	0.086	0.983	0.079
Incentives	0.89	0.117	0.908	0.104	0.887	0.110
No engagement	1.126	0.09	1.138*	0.066	1.070	0.080
Constant	0.757	0.633	0.696	0.151	0.830	0.474
Random part						
σ_u	n/a		0.299	-0.012	n/a	
Model statistics						
Log likelihood	-114735.18		-113702.44		n/a	
AIC	229578.4		227514.9		n/a	

	Casewise deletion				**Multiple imputed** (m=20)	
	MNL (N=124,072)		**MNL, RE** (N=124,072)		**MNL** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	SE
Wald test (F test mi data)	2761.57***		7688.07***		67.04***	
McFadden's R^2	0.04		0.04		n/a	
Proportion correctly classified	49.2%		50.6%		51.2%	

†Base category: Respondent; ‡ Base category: Physically disabled or vulnerable person; § Base category: aged 85 and over

legend: * p<.05; ** p<.01; *** p<.001

Table 13: Binomial logistic regression models of response propensity, with fixed and random effects, under two assumptions regarding the missing data mechanism

	Casewise deletion				**Multiply imputed data** (m=20)			
	BNL (N=124,072)		**BNL, RE** (N=124,072)		**BNL** (N=150,672)		**BNL, RE** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	Robust SE	Odds Ratio	SE
Fixed part								
Mental health‡	0.620***	0.017	0.624***	0.012	0.633***	0.017	0.631***	0.011
Learning disability‡	1.353***	0.049	1.366***	0.031	1.366***	0.047	1.380***	0.028
Substance misuse‡	0.484***	0.055	0.494***	0.053	0.472***	0.049	0.488***	0.049
Age: 18-24§	0.523***	0.025	0.526***	0.022	0.499***	0.022	0.514***	0.019
Age: 25-30§	0.638***	0.026	0.644***	0.021	0.641***	0.025	0.643***	0.019
Age: 31-39§	0.768***	0.026	0.772***	0.022	0.779***	0.025	0.778***	0.020
Age: 40-49§	0.909**	0.027	0.914***	0.023	0.911***	0.027	0.910***	0.021
Age: 50-64§	1.050*	0.03	1.060*	0.024	1.048	0.029	1.050*	0.022
Age: 65-74§	1.067**	0.023	1.068**	0.022	1.071***	0.021	1.070***	0.020
Age: 75-84§	1.031	0.02	1.040*	0.017	1.033	0.018	1.038*	0.016

	Casewise deletion				Multiply imputed data (m=20)			
	BNL (N=124,072)		**BNL, RE** (N=124,072)		**BNL** (N=150,672)		**BNL, RE** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	Robust SE	Odds Ratio	SE
White	1.182***	0.045	1.168***	0.026	1.181***	0.042	1.169***	0.024
Count of service types	1.244***	0.056	1.162***	0.032	1.232***	0.051	1.170***	0.030
Count of service types – sq	0.960***	0.009	0.974***	0.006	0.961***	0.008	0.974***	0.006
Nursing Home	0.702***	0.031	0.697***	0.02	0.718***	0.030	0.715***	0.019
Residential Home	1.042	0.039	1.023	0.019	1.018	0.038	1.011	0.017
Low-level services	0.867***	0.03	0.857***	0.016	0.860***	0.027	0.849***	0.015
Direct Payment	1.141***	0.029	1.113***	0.026	1.142***	0.027	1.116***	0.024
Short-Term Residential	0.813**	0.056	0.756***	0.039	0.801***	0.048	0.741***	0.035
Equipment	0.979	0.037	0.960*	0.017	1.003	0.035	0.965*	0.016
No chase	0.699**	0.076	0.728***	0.065	0.721***	0.070	0.740***	0.060
Deprivation: avscore	0.992**	0.003	0.990**	0.003	0.991**	0.003	0.990***	0.003
Data quality: number fo	1.024*	0.01	1.028**	0.011	1.016	0.009	1.017	0.009
Remove lackcap	1.006**	0.002	1.005*	0.002	1.002	0.002	1.002	0.002

	Casewise deletion				Multiply imputed data (m=20)			
	BNL (N=124,072)		**BNL, RE** (N=124,072)		**BNL** (N=150,672)		**BNL, RE** (N=150,672)	
	Odds Ratio	Robust SE	Odds Ratio	SE	Odds Ratio	Robust SE	Odds Ratio	SE
Add questions	0.987	0.069	1.014	0.087	1.031	0.066	1.044	0.081
Incentives	1.014	0.089	0.994	0.111	1.005	0.086	0.992	0.108
No engagement	0.894*	0.054	0.882*	0.05	0.940	0.052	0.928	0.048
Constant	0.441***	0.07	0.478***	0.101	0.518***	0.079	0.576**	0.102
Random part								
σ_u	n/a		0.293	0.020	n/a		0.295	0.018
Model statistics								
Log likelihood	-82163.54		-81180.44		n/a		n/a	
AIC	164381.1		162416.9		n/a		n/a	
Wald test (F test mi data)	1221.03***		2303.12***		47.67***		104.12***	
McFadden's R^2	0.021		0.033		n/a		n/a	
Percentage correctly classified	59.8%		59.8%		59.8%		59.8%	

‡ Base category: Physically disabled or vulnerable person; § Base category: aged 85 and over legend: * p<.05; ** p<.01; *** p<.001

For estimation simplicity, we therefore use the fixed effects MNL model estimated on the casewise deletion sample to examine the effects of the covariates on response propensity. The effects are illustrated in Figure 4 which shows the marginal change in response propensity (which takes values from zero to one) for each of the model covariates (the discrete change is shown for dummy variables). The effect of the count of the number of services a person is receiving is not shown in this graph since it has a nonlinear relationship with response propensity. Rather, the effect of the indicator “countservices” is illustrated in Figure 5. This shows that the effect of the number of services received on response propensity is very small when individuals have few services but, as the number of services received increases to five and beyond, the likelihood of being a respondent drops significantly. This effect probably reflects the positive association between dependency levels and the number of services received.

As Figure 4 shows, most of the covariates have very small effects on the probability of responding, of less than 0.05 points. Only the covariates relating to not chasing nonrespondents, being in a nursing home, being in any of the 18-24 or 25-30 or 31-39 age groups, and having substance misuse, learning disability or mental health problems change the probability of responding by more than 0.05 points. The covariate with the largest effect on the probability of responding is being in the substance misuse client group which is associated with a 0.17 reduction in probability of responding. However, this estimate has a large associated error and the confidence interval overlaps with estimates of the effect for the 18-24 age group, which also has a large effect (around a 0.15 reduction in probability of responding), the mental health client group, the 25-30 age group and not chasing nonrespondents. Interestingly, only one of these covariates, having a learning disability, has a positive effect on response propensity.

The effect of several of the continuous variables show a more considerable impact on the likelihood of response over the full range of values observed. These include the indicators of numbers of cases removed prior to sending out the questionnaire, due to the lack of capacity, and the indicator of deprivation level of the CASSR. The discrete effect on response propensity, from the minimum to maximum values of these variables is 0.084 and -0.069 respectively. Thus, the more people that CASSRs remove due to a lack of capacity prior to sending out the questionnaires the more likely people are to be respondents, and the more deprived an area the less likely people are the be respondents.

It is also useful to look at the elasticity of response propensity with respect to the covariates since it tells us how sensitive response propensity is to changes in the covariates. Specifically, the elasticity estimates the percentage change in response propensity linked to a one per cent change in a given covariate. When this ratio is below and above $|1|$, the response propensity is said to be inelastic and elastic to the variable, respectively. The elasticity of each of the covariates is shown in Figure 6. All values are less than $|1|$ indicating that response propensity is relatively inelastic to changes in any of the model covariates. Response propensity is most elastic, however, to the CASSR deprivation score, the number of services someone is receiving and the mental health status of the individual. It is also relatively elastic to the data quality (as estimated by the number of fully observed auxiliary data items); however, this estimate of elasticity has a very large SE so not much weight should be put on this finding.

To illustrate the differential effect of the covariates on individuals who returned a blank form and those who did not return the questionnaire, we illustrate the change in propensity for these outcomes associated with a discrete change in each of the covariates from the minimum to maximum values in Figure 7 and Figure 8 respectively[9]. Most of the covariates have either an unclear or very small effect on the probability of returning a blank form (less than 0.05 points). However, three CASSR-level covariates have a more significant impact. The more people a council removes from the sample due to lack of capacity, the indicator of data quality (in terms of auxiliary items fully observed) and the level of local deprivation are all associated with large decreases in the likelihood of returning blank forms over the range of values observed for these covariates.

The effects identified on the probability of not returning the form are quite different. Whilst many covariates have small effects, a number of variables are found to increase significantly the probability of not returning a form. These are covariates identified before as decreasing the likelihood of being a respondent: i.e. being in any of the 18-24 or 25-30 or 31-39 age groups; having substance misuse or mental health problems, which increase the probability of not responding by close to or more than 0.1 points; and the area deprivation, which decreases significantly the likelihood of returning a blank form. Not chasing nonrespondents to the first mail out is associated with a very large positive effect on the likelihood of not returning the questionnaire.

We also present the results from the BNL model since the interpretation of the effect of the covariates on response propensity is more straightforward. The findings are very similar to the MNL model, in terms of the factors that are important in predicting response propensity. For this reason we do not illustrate the effects of the covariates on response propensity for the BNL model. Similarly, despite being well-specified, the BNL models do not explain the observed variation in response propensity very well, with very low pseudo-R^2 and the percentage of predicted outcomes correctly classified approaching only 60 per cent.

The BNL models estimated on the multiply imputed dataset and the casewise deleted dataset are highly consistent. The main difference is for two CASSR-level variables (number of auxiliary items fully observed and proportion of the sample removing due to lack of capacity), which are found to be insignificant in the models estimated on the multiply imputed dataset. The results from the RE specification are also highly consistent with the

[9] Again countservices is not included. However, its effect on the “blank form” outcome is negligible and on the nonresponse outcome, the graph is the reflection (in the x-axis) of the response outcome.

fixed effects specification. Again rho is small at 0.025, although significant (likelihood ratio test, X^2=1966.22, $p<0.001$), and rho for the variance components model is only slightly larger at 0.033, confirming that the variation in response propensity is driven primarily by individual-level variation.

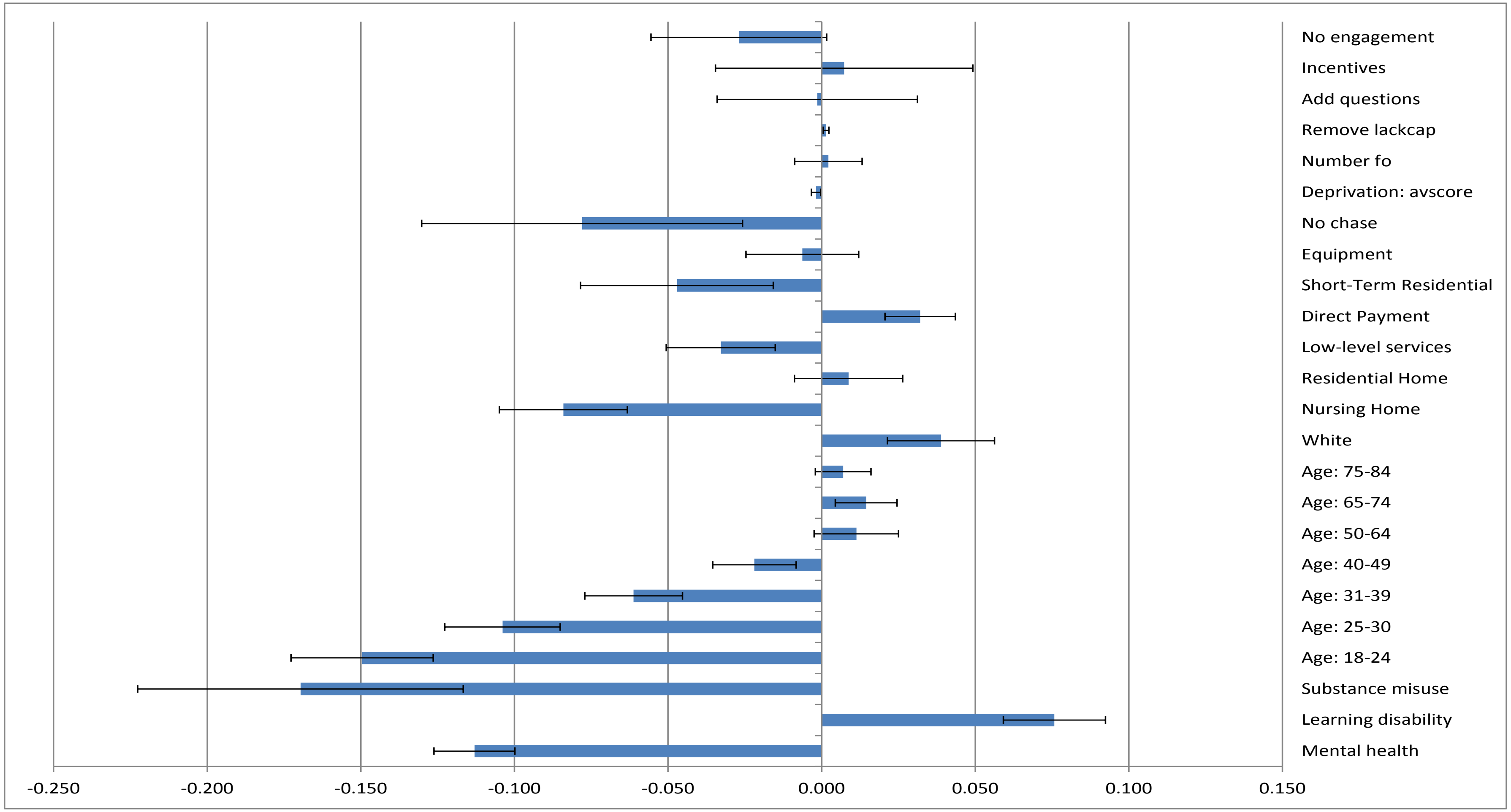

Figure 4: Marginal change in probability of responding, with approximate 95% confidence intervals, estimated using MNL model on the casewise deletion sample

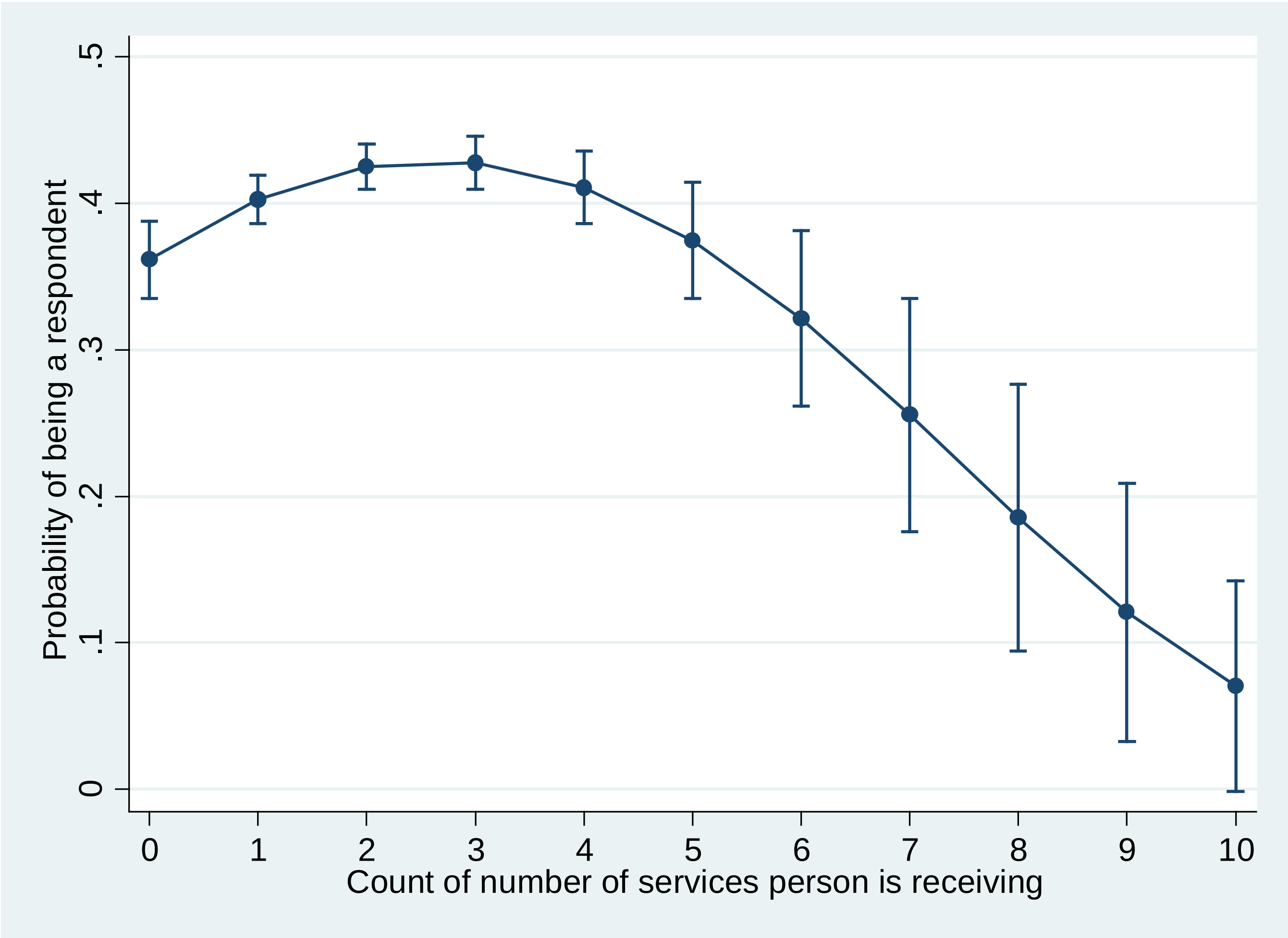

Figure 5: Variation in the effect of the covariate 'countservices' on response propensity, estimated by MNL model on the casewise deletion sample

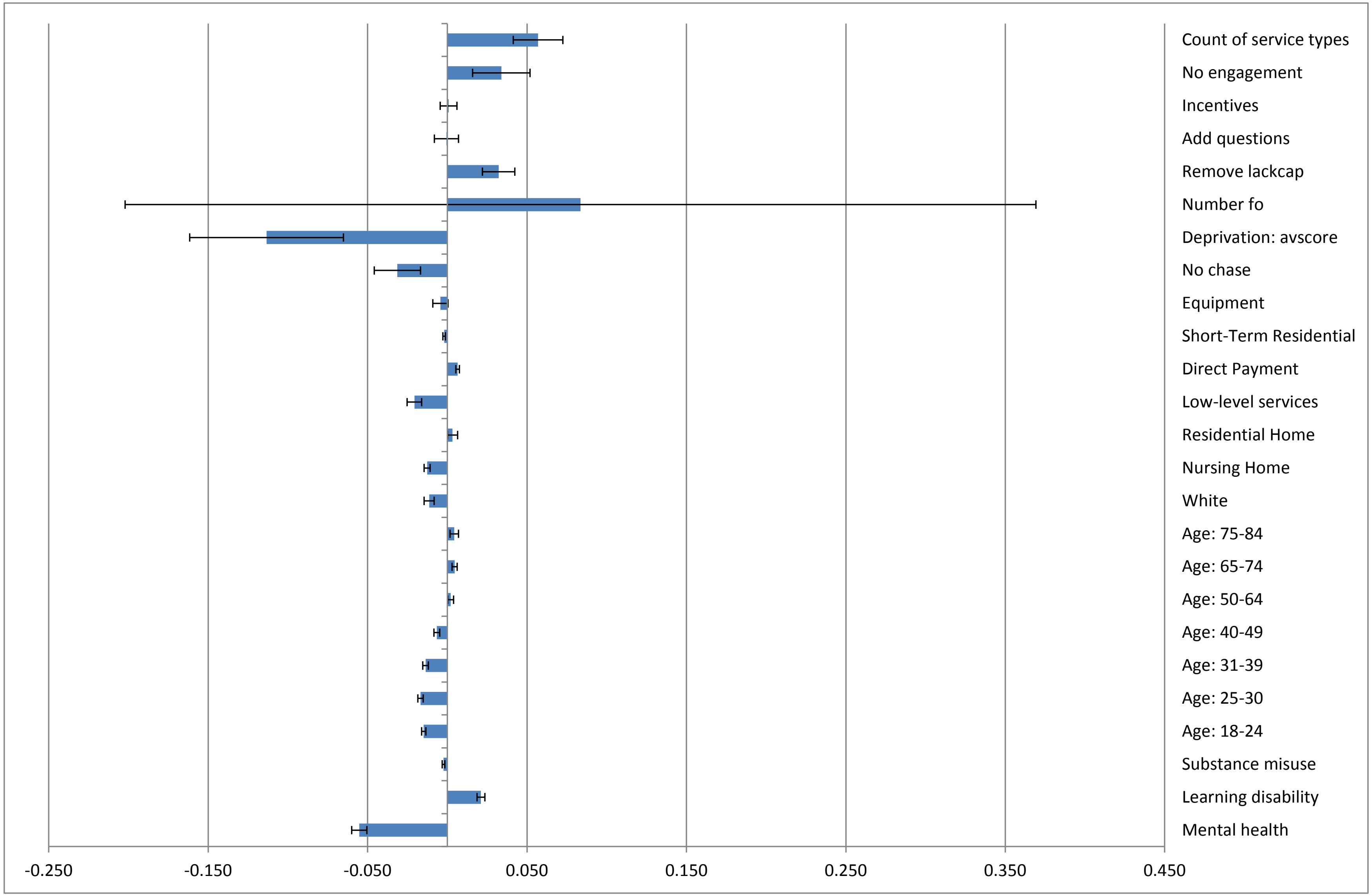

Figure 6: Elasticity of response propensity to each of the model covariates, estimates using MNL model on casewise deletion sample

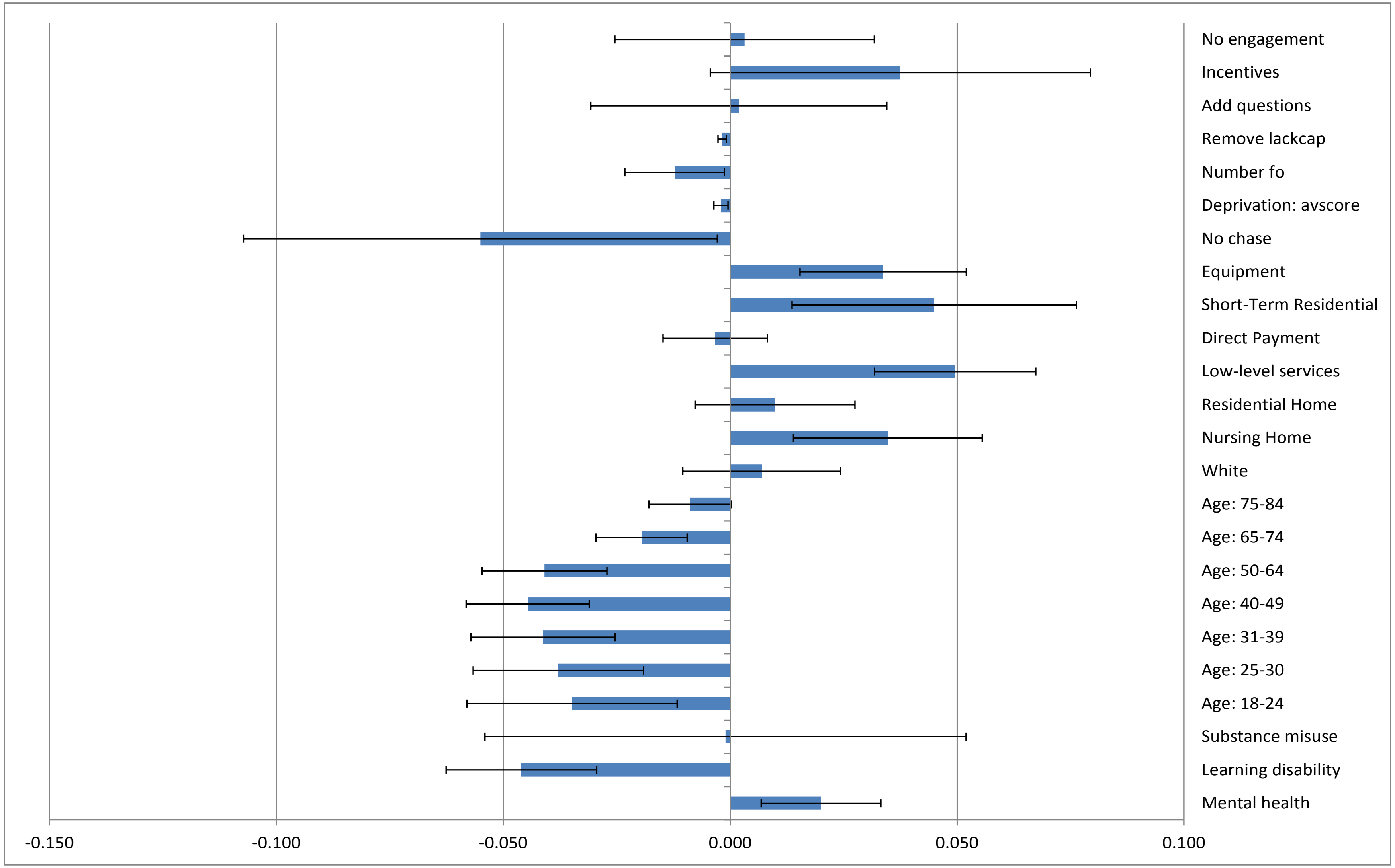

Figure 7: Marginal change in probability of sending back a blank form, with approximate 95% confidence intervals, estimated by MNL model on the casewise deletion sample

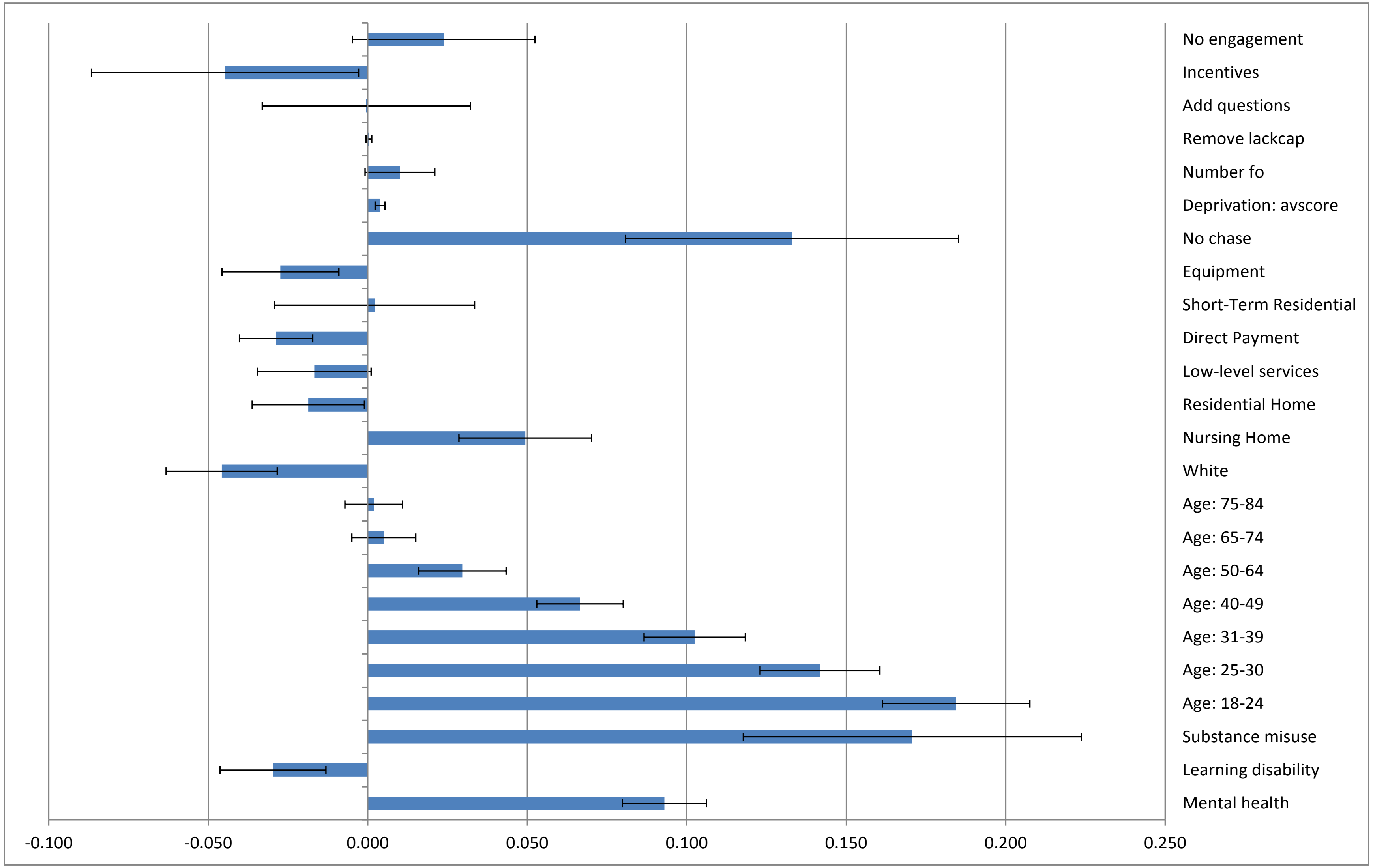

Figure 8: Marginal change in probability of not responding, with approximate 95% confidence intervals, estimated using MNL model on the casewise deletion sample

Following Little (1986), we regressed the QI on the predicted response propensity estimates using the respondent sample to test whether weighting is in order, this being the case if the coefficient of predicted response propensity is significantly different from zero. The β-coefficient for the response probability is shown in Table 14, for the casewise results, and Table 15, for the multiply imputed data. Response propensity, irrespective of the model form used, is strongly predictive of variation for all of the QIs. For all of the QIs, the respondents with the lowest propensity to respond experience the worst quality services. The model results indicate that all the QIs will benefit from weighting to reduce non-response bias and that weighting will on average lower the QI estimates.

Table 14: Significance of the response probability derived from the various response propensity models in predicting the QIs, estimated on the casewise deletion sample

	SCRQoL-TTO (n=45,120)	**SCRQoL** (n=45,120)	**Satisfaction QI** (n=48,143)	**Control QI** (n=49,449)	**Safety QI** (n=49,619)	**Information QI** (n=35,799)
BNL	0.259***	5.099***	0.976***	1.689***	4.572***	1.588***
	(0.033)	(0.642)	(0.247)	(0.255)	(0.408)	(0.195)
BNL, RE	0.267***	5.256***	0.958***	1.772***	4.632***	1.629***
	(0.033)	(0.643)	(0.252)	(0.254)	(0.407)	(0.195)
MNL	0.251***	4.927***	0.923***	1.627***	4.457***	1.563***
	(0.033)	(0.652)	(0.256)	(0.258)	(0.394)	(0.200)
MNL, RE	0.146***	2.850***	0.529**	1.017***	3.127***	1.034***
	(0.033)	(0.642)	(0.174)	(0.255)	(0.414)	(0.225)

legend: * p<.05; ** p<.01; *** p<.001; SEs in brackets

Table 15: Significance of the response probability derived from the various response propensity models in predicting the QIs, multiply imputed sample

	SCRQoL-TTO (n=54,350)	SCRQoL (n=54,350)	Satisfaction QI (n=57,929)	Control QI (n=59,478)	Safety QI (n=59,688)	Information QI (n=42,884)
BNL	0.272***	5.354***	1.053***	1.773***	4.851***	1.728***
	(0.03)	(0.591)	(0.227)	(0.234)	(0.366)	(0.178)
BNL, RE	0.291***	5.727***	1.074***	1.915***	4.936***	1.775***
	(0.03)	(0.579)	(0.228)	(0.228)	(0.363)	(0.179)
MNL	0.265***	5.200***	1.015***	1.709***	4.741***	1.696***
	(0.031)	(0.611)	(0.234)	(0.242)	(0.369)	(0.181)

legend: * p<.05; ** p<.01; *** p<.001; SEs in brackets

The distribution of the weights is shown in Table 16. The coefficient of variation is very similar across the models and there do not appear to be very large outlier values. For this reason we have simply used the inverse of the propensity score and not trimmed the weights or used adjustment cells (Little 1986; Kalton and Flores-Cervantes 2003). The inverse response propensity score weights are calibrated by multiplying by a post-stratification adjustment to ensure the population totals for each CASSR remain the same after weighting.

Table 16: Distribution of the non-response weights derived from the response propensity models

	Model	Mean	Std. Dev.	Min	Max	CV	DEFF
Casewise results (N=124,069)	**BNL**	1	0.231	0.548	4.964	0.231	1.053
	BNL, RE	1	0.223	0.569	2.896	0.223	1.050
	MNL	1	0.231	0.552	5.024	0.231	1.053
	MNL, RE	1	0.227	0.552	3.148	0.227	1.052
MI results (N=150,672)	**BNL**	1	0.229	0.540	4.839	0.229	1.052
	BNL, RE	1	0.222	0.544	3.028	0.222	1.049
	MNL	1	0.229	0.542	4.902	0.229	1.052

The extent of bias due to nonresponse is summarised in Table 17 and Table 18 for both the casewise deletion and multiply imputed samples, respectively. Since there are no substantial differences between the BNL and MNL models, we present only the results of the MNL model here, which is the most appropriate model for the data, despite having poor predictive power. The tables summarise the extent of bias found for each CASSR in the sample, showing the distribution of bias across the sample of CASSRs and the bias in the total sample. Several different estimates are presented: an estimate of bias calculated by subtracting the weighted indicator from the unweighted indicator; the absolute bias, which is simply the absolute value of the difference between the weighted and unweighted estimates; and the absolute bias as a percentage of the SE for the unweighted indicator, which gives an indication of the statistical significance of the bias. To get a better sense of the distribution of bias within the CASSRs and its seriousness, we also report the number of CASSRs where the bias is greater than one percentage point of the given scale, the number of CASSRs where the bias is negative, and the number of CASSRs where the bias is greater than 100 per cent of the unweighted indicator's SE. Since there are fewer CASSRs due to missingness within the auxiliary data in the casewise deletion sample (127[10] out of 149), to make it easier to compare across the multiply imputed and casewise deleted sample, we also report the frequencies of CASSRs in terms of the percentage of CASSRs within the particular sample (i.e. 127 for the casewise deletion sample and 149 for the multiply imputed sample).

Table 17: Estimates of the extent of bias for each QI, using weights derived from the MNL model on the casewise deletion sample

	SCRQoL-TTO (n=45,119)	**SCRQoL[1]** (n=45,119)	**Satisfaction QI** (n=48,142)	**Control QI** (n=49,448)	**Safety QI** (n=49,618)	**Info QI** (n=35,799)
Difference (unweighted – weighted)						
Total sample	0.003	0.067	0.003	0.004	0.004	0.003
Mean	0.003	0.067	0.003	0.003	0.004	0.003
Maximum	0.015	0.295	0.024	0.018	0.023	0.015
Minimum	-0.005	-0.121	-0.014	-0.011	-0.003	-0.024

[10] This is one fewer than reported above since we lost a further CASSR at this stage because only one respondent remained in the sample for one CASSR due to missingness on the auxiliary data.

	SCRQoL-TTO (n=45,119)	**SCRQoL[1]** (n=45,119)	**Satisfaction QI** (n=48,142)	**Control QI** (n=49,448)	**Safety QI** (n=49,618)	**Info QI** (n=35,799)
Difference (unweighted – weighted)						
Freq >1% point (% sample)	6 (5%)	3 (2%)	12 (9%)	17 (13%)	8 (6%)	12 (9%)
Freq negative (% sample)	14 (11%)	15 (12%)	36 (28%)	34 (27%)	20 (16%)	29 (23%)
Absolute Difference (|unweighted – weighted|)						
Total sample	0.003	0.067	0.003	0.004	0.004	0.003
Mean	0.004	0.074	0.005	0.005	0.004	0.005
Maximum	0.015	0.295	0.024	0.018	0.023	0.024
Minimum	0.000	0.001	0.000	0.000	0.000	0.000
Absolute difference as a percentage of raw SE						
Total sample	391%	379%	126%	194%	368%	157%
Mean	38%	37%	20%	24%	35%	20%
Maximum	227%	227%	131%	173%	160%	80%
Minimum	0%	0%	0%	0%	0%	0%
Freq > 100% (% sample)	8 (6%)	6 (5%)	1 (1%)	1 (1%)	5 (4%)	0 (0%)

Table 18: Estimates of the extent of bias for each QI, using weights derived from the MNL model, on the multiply imputed sample[11]

	SCRQoL-TTO (n=54,350)	SCRQoL (n=54,350)	Satisfaction QI (n=57,929)	Control QI (n=59,478)	Safety QI (n=59,688)	Info QI (n=42,884)
Difference (Raw – weighted)						
Total sample	0.003	0.062	0.002	0.003	0.003	0.004
Mean	0.003	0.063	0.003	0.003	0.003	0.004
Maximum	0.018	0.346	0.024	0.019	0.018	0.017
Minimum	-0.005	-0.110	-0.013	-0.009	-0.002	-0.023
Freq >1% point (% sample)	5 (3%)	4 (3%)	10 (7%)	14 (9%)	8 (5%)	16 (11%)
Freq negative (% sample)	16 (11%)	19 (13%)	43 (29%)	33 (22%)	25 (17%)	33 (22%)
Absolute Difference (|Raw – weighted|)						
Total sample	0.003	0.062	0.002	0.003	0.003	0.004
Mean	0.004	0.071	0.005	0.005	0.004	0.005
Maximum	0.018	0.346	0.024	0.019	0.018	0.023
Minimum	0.000	0.000	0.000	0.000	0.000	0.000
Absolute difference as a percentage of raw SE						
Total sample	402%	389%	130%	197%	383%	181%
Mean	36%	36%	19%	23%	33%	21%
Maximum	229%	229%	130%	179%	163%	79%
Minimum	0%	0%	0%	0%	0%	0%
Freq > 100% (% sample)	8 (5%)	5 (3%)	1 (1%)	2 (1%)	6 (4%)	0 (0%)

[11] Total sample size varies between indicators due to missingness in the QIs. The multiple imputation procedure was not used to impute any of the questionnaire items at the same time as imputation of the auxiliary data because the software would not allow for imputation of questionnaire items only for respondents.

There are very few differences between the multiply imputed and casewise deletion samples, so we discuss the results from both samples together. For all of the QIs, the general direction of the bias is positive, as we expected given the findings from the regressions of response propensity on the QI. This means that weighting generally lowered the value of the QI for CASSRs with the greatest proportion of missing data and for the sample overall. However, this is not consistently the case and there are some CASSRs for whom the value of the QI is increased by weighting. The number of CASSRs affected in a positive way by weighting varies by QI with, for example, some 28 to 29 per cent having a better score after weighting for the satisfaction QI and around 11 per cent have a better score after weighting for the utility-weighted ASCOT measure.

These effects should not be overstated. The absolute bias uncovered through weighting is very small; for all QIs, except control, the difference is less than one percentage point on the scale for over 90 per cent of CASSRs. Whilst the effect of weighting varies by QI, for the overwhelming majority of CASSRs, the weighted QI estimates are well within the 95% confidence interval of the unweighted QI. For the information QI, the bias is not greater than a SE for a single CASSR. The percentage of CASSRs with an estimate of bias that is greater than a SE rises to one per cent for the satisfaction and control QIs, three to four per cent for the safety QI, four to five per cent for the equally-weighted SCRQoL measure and five to six per cent for the utility-weighted ASCOT measure. For the last two measures, one CASSR's weighted QI estimate was in fact outside the 95% confidence interval for the unweighted estimate. Interestingly this same CASSR experienced the largest change in QI score post weighting for the safety, control and satisfaction QIs.

Given this CASSR has quite a large estimated bias, and therefore is affected by the weighting procedure, it is useful to analyse further the characteristics of the CASSR and service users within this CASSR. Interestingly, this CASSR does not have the largest estimated absolute bias, although it is one of the largest. Looking in more detail at the characteristics of this CASSR compared to the rest of the CASSRs, it has fewer white service users, fewer clients in residential and nursing homes, and fewer people receiving equipment. It also has more users with low level services and has worse data quality (as measured by the number of auxiliary variables fully observed) compared to the other CASSRs. In addition, it did not make much effort to remove people lacking the capacity to complete the questionnaire. A further contributing factor is that this CASSR has the smallest SE for its unweighted QI estimates.

A further point to note in Table 17 and Table 18, reinforcing the issue of the relationship between precision and the magnitude of bias in determining the effect of bias, is the effect of bias over the whole sample. Since the level of precision of the estimates is much higher for the whole sample, the effect of bias is very much more significant.

Discussion and recommendations

Despite the fact that our regression models are only able to predict a minority of the variability in the probability of response, they provide important evidence about the range of factors associated with whether someone returns a completed form, a blank form, or does not return a form at all. Consistent with findings from other studies, young people were found to be less likely to respond to the ASCS (Herzog and Rodgers 1988; Kaldenberg et al. 1994; Elliott et al. 2005). In particular, service users under 39 years of age were much less likely to respond. In addition, people with mental health problems or substance misuse problems were also much less likely to participate in the survey. This could be because these groups have less contact with social services and so perceive the survey as less salient, or that they are more difficult to reach over the period of data collection by local authorities conducting the survey. For the mental health group it could also be due to data quality issues and potentially a greater likelihood of the questionnaire going astray. Many councils share responsibility for these people with health services and do not have direct access to their records. Interestingly, although response propensity was fairly inelastic to changes in any of the covariates, mental health status was one of the covariates with the highest elasticity, indicating that reducing the numbers of people with mental health problems in the sample would improve response rates. By contrast, people with learning disabilities were very much more likely to respond to the questionnaire, so there may be lessons to learn from this group. Several factors could be at play here, including the targeting of EasyRead questionnaires to this group and the finding that support to complete the questionnaire was very high amongst this group (The Information Centre for Health and Social Care 2012b).

Although the nature of the services received was not found to have a large effect on response propensity, receiving more than five services was found to reduce significantly the propensity to respond. Since service packages tend to be greater for people who have no available informal carers and have high levels of disability, it may be that this indicator is picking up people with severe levels of disability and those who are less likely to receive assistance from informal carers to complete the survey. In addition, being a resident in a nursing home had a relatively large negative effect on the probability of responding. Interestingly, being a resident in a residential home had no discernible effect on response propensity, indicating that the institutional effect identified is particular to either the characteristics of nursing homes or the types of residents they house. Nursing home residents are more disabled, both cognitively and physically than people in residential care homes (Darton et al. 2006). It may be that nursing home staff have less buy-in to the survey because of the nature of their residents, perhaps perceiving the survey as futile given the very high proportion of residents that will be totally unable to complete the survey due to advanced dementia.

Several CASSR-level indicators also had a significant and relatively large effect on response propensity. Not chasing nonrespondents had a relatively large negative effect on the likelihood of being a respondent although, interestingly, it had a greater positive effect on the likelihood of being a nonrespondent (reflecting the differential effect of this variable on the two types of nonrespondents, which we discuss further below). The odds of a case being a nonrespondent are approximately 1.6 times greater in CASSRs that do not chase nonrespondents compared to those that do, although the precision of this estimate was quite poor due to the relatively small number of CASSRs that did not chase nonrespondents. Deprivation, as measured through the IMD 2010, was also found to reduce significantly response propensity. This is consistent with a study by Johnson et al (2006) who explain the effect in terms of trust, arguing that more deprived areas are more socially disorganised and that this fosters a lack of trust in authorities and therefore a reduced likelihood of responding. It may also be related to the allocation of resources in deprived areas, with central management perhaps preferring to direct resources to frontline services rather than to achieving higher response rates in surveys.

As expected, CASSRs that excluded the largest proportions of people on the grounds of lack of capacity to complete the questionnaire prior to sending out the survey were more likely to achieve better response rates. This was also reflected in a differential effect on the two indicators of nonresponse. The more people a CASSR removed prior to sending out the questionnaire, the less likely they were to have people returning blank forms. The effect of removing people who lacked capacity on the probability of not returning a form was, by contrast, minimal. This finding confirms that where the ethical committee's strategy to remove people who could not complete the survey prior to sending out the questionnaires was followed stringently, then people who had relatively high levels of dependency seem to have been excluded. However, the fact that CASSRs who followed the ethical committee's strategy less stringently had a higher rate of return of blank forms, suggests that people acting on behalf of the intended recipient are effective at self-selecting out of the survey if the intended recipient lacks the capacity to complete the questionnaire. When deciding how best to proceed with the survey for this group of people from an ethical point of view, the potential distress caused by receiving an inappropriate questionnaire clearly needs to be balanced against the extremely high costs for CASSRs of complying with the strategy of removing people who lack capacity to complete the questionnaire prior to sending out the survey and the potential risk of biases in the sample arising from the inappropriate removal of potential respondents (The Information Centre for Health and Social Care 2012). Either way, excluding people lacking the capacity to complete the questionnaire from the survey undermines its representativeness relative to the overall population of recipients of social care and strategies should be found, such as questionnaire for proxy respondents, to try to include the experiences of these people in future surveys.

As we have already highlighted, variables had a differential effect on the two types of nonrespondent outcomes (returning a blank form or not returning a form). In addition to the effect of removing people who lack capacity and not chasing nonrespondents, two CASSR-level variables were found to have different effects on the type of nonresponse.

First, area deprivation was significantly associated with the likelihood that a form might not be returned, but negatively associated with the return of a blank form. This provides further support to the theory that the effect is associated with differences in local levels of trust services by levels of deprivation. Second, data quality, as measured by the number of auxiliary data without missing data, was not a significant predictor of returning no form but did decrease the likelihood of returning a blank form. It also had a positive relationship with being a respondent, although this relationship did not always reach statistical significance in the models. It is not clear why the strongest effect of data quality would be to reduce the likelihood of sending back a blank form. This finding, however, is not inconsistent with the argument that better quality data increases the likelihood of people responding since people are more likely to receive the survey request, receive an appropriate survey, and might be chased in case of nonresponse. Additionally, a number of individual-level variables were significant in predicting that a person would not return a form but were not significant, or had very small effects, on the probability of returning a blank form. These variables were being under 39 years old, having a substance misuse problem and having a mental health problem. All of these factors significantly and relatively substantially increased the likelihood of not returning the questionnaire.

It is also worth highlighting some of the effects that are not significant or are relatively small. One such factor is ethnicity, whereby being white has a small positive effect on the likelihood of being a respondent. We suggested that ethnicity may be an indicator of proficiency in English, albeit a poor one. It is possible that this explanation lies behind the effect of ethnicity but clearly a better indicator proficiency in English would be needed to test this theory. The dummy variable indicating whether or not CASSRs reported taking steps to engage hard-to-reach groups has a similar effect to ethnicity but was not always significant in the models. This lack of significance is largely a consequence of the small number of CASSRs reporting not to take steps to engage hard-to-reach groups, and the fact that there are fewer degrees of freedom at the CASSR level. However, it could also be due to differences in the degree and extent to which CASSRs who reported taking steps to engage hard-to-reach groups actually went about engaging with these groups.

Two factors, the use of incentives and additional questions, which have been found in many previous studies to affect response propensity, were found to be insignificant in this study, even though their effects were in the expected direction. In the case of additional questions, the effect was very small and the SE very large. This is largely due to the small number of CASSRs that choose to add further questions. In addition, limitations in the specification of the indicator are likely to have undermined its sensitivity. We did not know, for instance, the number of questions added. Further research would be needed to establish the effect of adding further questions to the ASCS. The effect of using incentives was larger but, again, insignificant due to a large SE. Again this is, in part, because very few CASSRs reported using incentives, but it could also be due to the variability in the incentive used across CASSRs. Although all CASSRs used a form of prize draw, the prize varied. We would caution against interpreting these findings as evidence of the ineffectiveness of incentives within the ASCS.

A further interesting finding regarding the modelling was the limited additional explanatory power of the random-effects model over the fixed-effects model. Despite the variations observed in CASSR response rates and the differences in how CASSRs manage the survey, the percentage of variation in underlying individual response propensity, explained by systematic differences between CASSRs, is very small at around three per cent. This tells us that it is primarily variations between individual sample members that are driving the variations in observed response rates. However, this does not mean that there is no need to weight the sample to adjust for the effects of nonresponse, since this method still has value in ensuring the sample is representative and allowing us to make accurate comparisons with other datasets and over time.

RECOMMENDATIONS CONCERNING IMPROVING RESPONSE RATES & REPRESENTATIVENESS

1. Mental health and substance misuse client groups are underrepresented in the ASCS sample. Response rates and the representativeness of the survey could be improved by focusing on these groups and trying to better understand the reasons for nonresponse. It is possible that we may be able to learn from the learning disability group since this group had much higher response rates.
2. Younger people (under the age of 39, but particularly those in the 18-24 age group) are underrepresented in the ASCS sample. CASSRs should engage these groups in future surveys to improve response rates and representativeness.
3. Despite the limited effect of CASSR differences on response propensity it is nevertheless important to recognise that some of differences in strategies used by CASSRs did have an observable effect on response rates. Specifically, chasing nonrespondents and removing from the sample people who lacked capacity had clear positive effects on response propensity. A number of other factors, such as data quality, and reporting engaging with hard-to-reach groups, had effects in the right direction that bordered on significance. Given the fact that CASSRs do have some power to affect response rates and the representativeness of the survey it is therefore important that they follow best practice when carrying out the survey to ensure they do not unintentionally undermine its representativeness.
4. The strong effect of chasing nonrespondents on response propensity suggests that the requirement to make at least one follow-up contact should be much more strongly enforced.
5. More detailed research into the effect of adding questions and using incentives on response rates in the ASCS should be carried out. The lack of significance of their effect in this analysis should not be interpreted as an indication that they do not have an effect

on response rates, given the small numbers of observations and insensitivity of the dummy variables used to explore these effects.

6. People with severe cognitive impairment are underrepresented in the ASCS. Steps should be taken to try and develop methods to capture the experiences of this group.

Although we identified a number of factors that have small or relatively modest effect on the probability of response and the type of nonresponse, the overall variation in the outcomes explained by the models was very low. Consequently, the reweighting procedure identified few large or statistically significant differences in LA ratings with respect to any of the QIs. We found only one CASSR with a statistically significant difference between weighted and unweighted estimates, although more CASSRs – for some QIs around 10 per cent – had differences of more than one percentage point.

Although we can be quite confident that the effect of weighting does not dramatically change the interpretation of QIs, it is not straightforward to judge its effect. The problem, alluded to above, is that we need to consider the meaningfulness (size) of differences as well as the statistical significance. Here we have used two rather arbitrary criteria for both of these factors: for the former, we equate a meaningful difference with a change exceeding one percentage point of a given QI scale; and for the latter, we consider a significant change as one which was outside of the 95 per cent confidence interval for the unweighted QI, since all survey estimates are reported with 95 per cent confidence intervals to illustrate the effect of sampling error. Using these criteria, we reach different conclusions about the number of CASSRs seriously affected by weighting, with the meaningful difference criterion producing more affected CASSRs than the statistical significance criterion. This issue is further illustrated when we consider the effect of weighting on the national-level estimates, since here differences of less than half a percentage point of the given scale are found to be statistically significant differences due to the greater degrees of freedom and therefore precision of the estimates. On purely statistical grounds these results would imply that it would be important to use weights for producing national-level, but not local-level, statistics. However, we feel this is not the correct conclusion to draw from these results.

Given there are fewer degrees of freedom at the local level, the statistical criterion used might be considered too restrictive. Whilst we think this approach is sensible given the use of 95 per cent confidence intervals to present the survey results, there may also be some value in varying the cut-offs and illustrating the proportion of effects significant at different statistical cut-off points. We also have very little evidence on which to make a judgement as to what is a meaningful difference in the QI scores. The value we have chosen does seem quite restrictive and differences of three or four per cent may be more meaningful from a policy point of view. However, lacking evidence it is difficult to judge this, particularly at this level of aggregation, where effects are averaged across very different user groups and service types. There may also be value in exploring different criteria for identifying meaningful differences. Taking both considerations into account, what these results seem to suggest is that weighting does not seem to greatly affect the QI estimates. However, it may be that weighting is considered important in any case to rebalance the sample to ensure its representativeness.

To overcome the problem of missing data in the auxiliary data used to develop the response propensity models underlying the calculation of the weights, we used a multiple imputation procedure. If it is not possible to improve reporting of these variables then such a process will be necessary for future ASCSs if weighting is to be considered. An important question then regards the degree to which we can have confidence in the imputed variables. The chains appeared to converge and the imputation method used has been shown to function well even in cases where the conditional specifications are incompatible (van Buuren et al. 2006). However, we have some reticence over the imputation of CASSR-level variables since some of the variability could be driven by individual-level variation. Although we excluded all individual-level variables from the imputation of all CASSR-level variables to guard against this possibility, one individual-level variable – response status -- had to be included since it was the outcome on which all of the covariates were regressed. An alternative programme is available to impute multilevel datasets, known as REALCOM (Goldstein 2009). Whilst we did not explore this here, because it seemed unnecessary given the small effects of many of the CASSR-level variables and the small differences between models estimated the imputed and casewise deleted datasets, it may be a more sensible option to use for any replications of the methods used here.

A further adaptation that may be investigated in future work is the effect of trimming the weights or stratifying them by forming adjustment cells. Little (1986) suggests stratification places less emphasis on the correct specification of the regression model, so it may be wiser to use this method for constructing the weights. However, given the small effects of the weighting procedure on the bias of estimates, we felt that this degree of manipulation was unnecessary.

Whilst our results imply that nonresponse does not substantially bias the QIs, we would caution against applying these findings without further consideration to all future ASCSs or drawing the conclusion that nonresponse to this survey does not produce any degree of bias. First, we have discussed at some length the difficulties inherent in trying to assess whether the estimated bias is substantial. It would seem important to consider what is substantial from a policy point of view and conduct a similar analysis for each ASCS before deciding that weighting is not in order. Second, in all surveys subsequent to the 2011 ASCS, CASSRs have not been required to follow such a demanding set of instructions to remove people who lack the capacity to complete the questionnaire. We would expect this to affect response rates and the return rate of blank forms. There may, in turn, be some effect on bias. Third, there are several factors that we thought were theoretically important but for which we were unable to find good indicators. These include availability of informal care, disability type and severity, and proficiency in English. We do not know whether, or to what extent, these factors would improve the models of response propensity. This largely depends on whether responding to a survey is a quasi-random process or whether there are important unobserved factors that determine response propensity. This extended quote summarises the position well:

> *"Generally, low correlations between available variables and survey non-response mean that propensity scores would predict little of the variance in responding. This would often be interpreted to mean that we lack strong variables to predict non-response and, as a result, we do not have a good propensity model. However, if responding was largely a quasi-random process (depending on transitory decisions and other idiosyncratic factors), then there would be few, if any, strong predictors and no propensity model would explain much of non-response. Showing that variables with a good theoretical basis for predicting non-response do not actually correlate with non-response might be seen as indicating that non-response is largely a quasi-random process and that there would therefore be relatively little non-response bias for most variables. Alternatively, if one had AD [auxiliary data] variables that were closely related to the substantive variables and showed that they did not vary across respondents and nonrespondents, this would also be an indicator that non-response bias might be either limited in magnitude or at least not related to the target variables."*

(Smith 2011: 395-396)

Whilst we have shown that some theoretically important variables have a limited effect on response propensity, we are missing a number of theoretically important variables and it would seem important to explore the possible effects of these before drawing any conclusions regarding the biasing or otherwise effects of nonresponse on QIs derived from the ASCS.

RECOMMENDATIONS CONCERNING BIAS

1. In this study we did not find a statistically significant biasing effect from nonresponse on CASSR-level QI estimates nor do we consider any of the differences observed to be meaningful from a policy point of view. Whilst we do not, therefore, recommend making adjustments for nonresponse at the CASSR level for this ASCS, we note that this study does not provide conclusive evidence as to the effect of nonresponse on QI estimates. More research is needed as further theoretically important variables become available and this analysis should be repeated for all future ASCS to ensure that there are not CASSRs with unfortunate patterns of missingness that mean the extent of bias in quality estimates is large.
2. We did find a significant biasing effect from nonresponse on QI estimates aggregated at the national level. However, we do not consider the size of this difference to be particularly meaningful from a policy point of view. Nevertheless, to ensure comparisons over time are not affected by nonresponse and ensure the representativeness of the sample, policymakers may consider that weighting to rebalance the sample is of value.

3. The method we have suggested to recover missing data in the auxiliary variables is complicated and time consuming. It would be better to enforce the completion of the auxiliary data, particularly those variables found to be significant predictors of response status.

References

Commission for Social Care Inspection (2004) Social services performance assessment framework indicators, 2003-2004, London, Commission for Social Care Inspection.

Commission for Social Care Inspection (2007) Social services performance assessment framework indicators, adults, 2005-2006, London, Commission for Social Care Inspection.

Darton R, Forder J, Bebbington A, Netten A, Towers A-M, Williams J (2006) Analysis to support the development of the relative needs formula for older people: Final report. PSSRU discussion paper no. 2265/3'. Canterbury, Kent, PSSRU.

de Leeuw E, de Heer W (2002) Trends in household survey nonresponse: A longitudinal and international comparison. *Survey nonresponse*. RM Groves, DA Dillman, JL Eltinge, RJA Little, New York, Wiley: 41-54.

Department of Health (2003a) Personal social services survey of home care users in England aged 65 or over: 2002-03. Bulletin 2003/26. Department of Health, London, The Stationery Office.

Department of Health (2003b) Social services performance assessment framework indicators 2002–2003. Department of Health, London.

Department of Health (2004) Personal social services survey of physically disabled and sensory impaired users in England aged 18-64: 2003-04, bulletin 2004/23. London, The Stationery Office.

Department of Health (2011a) The adult social care outcomes framework. Handbook of definitions. Version 2 (November 2011). Department of Health, London.

Department of Health (2011b) Transparency in outcomes: A framework for quality in adult social care. The 2011/12 adult social care outcomes framework. Department of Health, London.

Dillman DA, Smyth JD, Christian LM (2009) *Internet, mail and mixed-mode surveys: The tailored design method*. New York, John Wiley and sons.

Durrant GB, Steele F (2009) Multilevel modelling of refusal and non-contact in household surveys: Evidence from six UK government surveys, *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 172, 2, 361-381.

Elliott MN, Edwards C, Angeles J, Hambarsoomians K, Hays RD (2005) Patterns of unit and item nonresponse in the CAHPS® hospital survey, *Health Services Research,* 40, 6p2, 2096-2119.

Goldstein H (2009) REALCOM-impute: Multiple imputation using MLwiN, user guide'. Bristol, Centre for Multilevel Modelling, University of Bristol.

Goldstein H, Healy MJR (1995) The graphical presentation of a collection of means, *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 158, 1, 175-177.

Groves R (2006) Nonresponse rates and nonresponse bias in household surveys, *Public Opinion Quarterly,* 70, 5, 646-675.

Herzog AR, Rodgers WL (1988) Age and response rates to interview sample surveys, *Journal of Gerontology,* 43, 6, S200-S205.

Höfler M, Pfister H, Lieb R, Wittchen HU (2005) The use of weights to account for non-response and drop-out, *Social Psychiatry and Psychiatric Epidemiology,* 40, 4, 291-299.

Johnson ML, Rodriguez HP, Solorio MR (2010) Case-mix adjustment and the comparison of community health center performance on patient experience measures, *Health Services Research,* 45, 3, 670-690.

Johnson TP, Cho YI, Campbell RT, Holbrook AL (2006) Using community-level correlates to evaluate nonresponse effects in a telephone survey, *Public Opinion Quarterly,* 70, 5, 704-719.

Kaldenberg DO, Koenig H F, Becker BW (1994) Mail survey response rate patterns in a population of the elderly: Does response deteriorate with age?, *The Public Opinion Quarterly,* 58, 1, 68-76.

Kalton G, Flores-Cervantes I (2003) Weighting methods, *Journal of Official Statistics,* 19, 2, 81-97.

Kauppi M, Sokka T, Hannonen P (2005) Survey nonresponse is associated with increased mortality in patients with rheumatoid arthritis and in a community population, *The Journal of Rheumatology,* 32, 5, 807-810.

Lepkowski J, Kalton G, Kasprzyk D (1989) Weighting adjustments for partial nonresponse in the 1984 sipp panl, *Proceedings of the American Statistical Association, Section on Survey Research methods,* 296-301.

Little RJA (1986) Survey nonresponse adjustments for estimates of means, *International Statistical Review,* 54, 139-157.

Little RJA, Rubin DB (1987) *Statistical analysis with missing data*. New York, Wiley.

Long JS, Freese J (2006) *Regression models for categorical dependent variables*. College Station, Texas, Stata Press.

Malley J, Caiels J, Fox D, McCarthy M, Smith N, Beadle-Brown J, Netten A, Towers AM (2010) A report on the development studies for the national adult social care user experience survey, PSSRU discussion paper 2721. Canterbury, Personal Social Services Research Unit, University of Kent.

Malley J, Towers AM, Netten A, Brazier J, Forder J, Flynn T (2012) An assessment of the construct validity of the ASCOT measure of social care-related quality of life with older people, *Health and Quality of Life Outcomes* 10 DOI: 10.1186/1477-7525-10-21.

Martin J, Matheson J (1999) Responses to declining response rates on government surveys, *Survey Methodology Bulletin*, 45, 33-37.

McLennan D, Barnes H, Noble M, Davies J, Garratt E, Dibben C (2011) The English indices of deprivation 2010. London, Communities and Local Government.

Mihelic AH, Crimmins EM (1997) Loss to follow-up in a sample of americans 70 years of age and older: The lsoa 1984-1990, *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences,* 52B, 1, S37-S48.

Netten A, Beadle-Brown J, Caiels J, Forder J, Malley J, Smith N, Towers AM, Trukeschitz B, Welch E, Windle K (2011) Adult social care outcomes toolkit (ASCOT) main guidance v 2.1. PSSRU discussion paper 2716/3' Canterbury, Personal Social Services Research Unit, University of Kent.

Netten A, Burge P, Malley J, Potoglou D, Towers AM, Brazier J, Flynn T, Forder J (2012) Outcomes of social care for adults: Developing a preference-weighted measure, *Health Technology Assessment,* 16, 16.

Office for National Statistics (2001) Population density (uv02), from 2001 census. Office for National Statistics.

Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, Brazier JE (2011) Best–worst scaling vs. Discrete choice experiments: An empirical comparison using social care data, *Social Science & Medicine,* 72, 10, 1717-1727.

Rabe-Hesketh S, Skrondal A (2008) *Multilevel and longitudinal modelling using stata*. College Station, Texas, STATA press.

Rabe-Hesketh, S, Skrondal, A and Pickles, A (2004). 'Gllamm manual. U.C. Berkeley division of biostatistics working paper series, paper 160'. Berkeley, California, The Berkeley Electronic Press.

Rahi JS, Manaras I, Tuomainen H, Lewando Hundt G (2004) Engaging families in health services research on childhood visual impairment: Barriers to, and degree and nature of bias in, participation, *British Journal of Ophthalmology,* 88, 6, 782-787.

Rogers WH (1993) Regression standard errors in clustered samples, *Stata Technical Bulletin,* 13, 19-23.

Rubin DB (1976) Inference and missing data, *Biometrika,* 63, 3, 581-592.

Rubin DB (2003) Discussion on multiple imputation, *International Statistical Review,* 71, 619-625.

Smith TW (2011) The report of the international workshop on using multi-level data from sample frames, auxiliary databases, paradata and related sources to detect and adjust for nonresponse bias in surveys, *International Journal of Public Opinion Research,* 23, 3, 389-402.

Steele F, Durrant GB (2011) Alternative approaches to multilevel modelling of survey non-contact and refusal, *International Statistical Review,* 79, 1, 70-91.

The Information Centre for Health and Social Care (2012) Annex A: Adult social care survey 2010-11: Compliance costs analysis report. *Personal Social Services Adult Social Care Survey, England 2010-11 (Final Release)*. Leeds, The Health and Social Care Information Centre.

The Information Centre for Health and Social Care (2007)'Personal social services survey of home care users in England aged 65 and over, 2005-06. Leeds, The Health and Social Care Information Centre,.

The Information Centre for Health and Social Care (2010) Personal social services adult social care survey guidance document. Leeds, The Health and Social Care Information Centre.

The Information Centre for Health and Social Care (2012a) Community care statistics 2010-11: Social services activity report, England'. Leeds, Health and Social Care Information Centre,.

The Information Centre for Health and Social Care (2012b) Personal social services adult social care survey, England 2010-11 (final release). Leeds, The Health and Social Care Information Centre.

The Information Centre for Health and Social Care (2012c) Personal social services: Expenditure and unit costs -England 2010-11- final release. Leeds, The Health and Social Care Information Centre,.

van Buuren S (2007) Multiple imputation of discrete and continuous data by fully conditional specification, *Statistical Methods in Medical Research,* 16, 219-242.

van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB (2006) Fully conditional specification in multivariate imputation, *Journal of Statistical Computation and Simulation,* 76, 1049-1064.

White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: Issues and guidance for practice, *Statistics in Medicine,* 30, 377-399.