PSSRU

A Comparison of Options for Indicators from the Adult Social Care Survey (ASCS)

Juliette Malley

PSSRU Discussion Paper 2736
August 2010
www.pssru.ac.uk

PERSONAL SOCIAL SERVICES RESEARCH UNIT

University of Kent

University of Kent
Cornwallis Building
Canterbury
Kent
CT2 7NF
Tel: 01227 823963/823862
PSSRU@kent.ac.uk

London School of Economics

London School of Economics
LSE Health & Social Care
Houghton Street
London
WC2A 2AE
Tel: 020 7955 6238
PSSRU@lse.ac.uk

University of Manchester

University of Manchester
Dover Street Building
Oxford Road
Manchester
M13 9PL
Tel: 0161 275 5250
PSSRU@manchester.ac.uk

www.PSSRU.ac.uk

Contents

Box

Figures

Tables

Acknowledgements

This is an independent report commissioned and funded by the Policy Research Programme at the Department of Health. The views expressed are those of the researchers and not necessarily those of the Department of Health.

I am grateful to Professor Ann Netten, the members of the Social Services User Survey Group and an anonymous reviewer for their guidance and comments on an earlier draft of this report. My thanks also go to the councils that took part in the pilot and to the participants in the survey.

Executive summary

The Department of Health (DH) commissioned the Personal Social Services Research Unit (PSSRU) at Kent to develop an 'outcomes-focused' indicator for social care service users based on data gathered via the Adult Social Care Survey (ASCS, formerly the National Adult Social Care User Experience Survey and the Putting People First Survey). This work was commissioned under the previous government. Although the new government is keen to reform the performance framework, it seems likely that an indicator which focuses on users' views of their outcomes from services will be of interest since it fits with the principles and ethos of the new outcomes framework, albeit for primarily acute health care, set out by the Coalition.

The approach taken here to developing a measure is to develop several indicator specifications and compare them. To choose between options for indicators, it is necessary to have a set of criteria for making choices. This is complicated for indicators since they generally serve a variety of functions, and the way the data are used and by whom have important consequences for the characteristics that make a good indicator. Despite these problems, it is possible to identify a number of desirable characteristics which can be summarised under five headings as follows: acceptability, feasibility, reliability, validity and sensitivity. These qualities are often in conflict, and choosing between specifications for indicators may involve making trade-offs within and between these five dimensions. The relative balance of these qualities is therefore a matter for debate, but proper debate depends on clarity over which of the uses and users are most important.

This report focuses on comparing empirically the validity, reliability and sensitivity of a set of proposed indicators. The dataset used to make these empirical comparisons is from the pilot of ASCS. The ASCS pilot was conducted following the same methods as the previous user experience surveys in social care: that is, the survey is conducted by councils with social services responsibilities (CASSRs) following guidance from the Information Centre for Health and Social Care (IC). Eighteen CASSRs took part in the pilot, which consisted of three elements: a pilot of the survey to all client groups and service types, a pilot of the residential care survey and a pilot of the survey for people with learning disabilities (LD). CASSRs could choose in which pilot to take part. Three participating CASSRs took part in more than one element. The pilot design and small size of the pilot samples (100 people) caused some problems for the analysis, which we discuss below.

Specification of the proposed indicators

Three types of indicators are developed and examined. All of the indicators are based on the Adult Social Care Outcomes Toolkit (ASCOT) questions, which have been designed by PSSRU over a number of years (see http://www.pssru.ac.uk/ascot/ for more information about ASCOT). The three indicators are summarised as a multi-item measure of social care-related quality of life (SCRQoL) developed using psychometric principles (referred to as the SCRQoL indicator), a preference-weighted multi-item SCRQoL measure developed using principles from economic theory and drawing on information about people's preferences for the various SCRQoL states measured in ASCOT (referred to as the PW-SCRQoL or preference-weighted SCRQoL indicator), and a measure of social care needs (referred to as the SCRQoL needs

indicator). All of these indicators are subjected to a series of analyses to investigate their validity, reliability and sensitivity.

The SCRQoL needs indicator has four different specifications which use different definitions of 'social care need'. The structure of the response options to the ASCOT items allows for several ways of defining need. The ASCOT items have four response options, designed to capture the outcome states of high-level needs, low-level needs, no needs where the person's aspirations are not met (mustn't grumble state), and no needs where the person's aspirations are met ('ideal' state). This leads to two potential ways of defining need according to severity: presence of high-level needs and presence of either high- or low-level needs. The range of needs can also be varied. The ASCOT measure has eight dimensions and of these the ASCOT team have generally considered the dimensions of personal care, food and nutrition, safety, and accommodation cleanliness and comfort as the 'basic' need dimensions. The range of SCRQoL needs can therefore be defined as a need in a basic SCRQoL dimension or a need in any SCRQoL dimension depending on the stakeholders' interests.

Analysing the qualities of the indicators

The three types of indicator were analysed to explore their reliability, validity and sensitivity. The following methods were applied:

- Validity: The indicators were correlated against a survey question asking the service user to report their overall quality of life on a seven-point scale and a question asking the user how satisfied they are with the care they receive from social services. Given the aim to develop a measure of social care outcomes, we can conclude that the measure that has the highest correlation with these two questions fits best with national priorities and therefore has the best validity.

- Reliability: This was examined using two different methods. The first method draws on generalisability theory. The analysis splits the observed variation in the indicators into that attributable to the CASSR and that attributable to the individual, generating what is known as a variance partition coefficient (VPC). The VPC is simply the proportion of variation due to the CASSR. From the VPC it is possible to calculate the number of respondents required to achieve an adequate level of reliability. The second method draws on the criterion used by the IC, which is that the confidence interval around the indicator should not be greater than plus or minus four per cent of the length of the scale. As with the generalisability method, this method can be used to calculate the number of respondents required to meet this criterion.

- Sensitivity: The sensitivity of the indicator can be determined by exploring whether the measure captures meaningful changes. What we mean by meaningful is difficult in this context as there is no yardstick. We therefore simply explore whether there are any differences between CASSRs on each of the indicators as an indication of their sensitivity and usefulness as indicators, the rationale being that if there is no variation the indicator is not useful.

Comparing the qualities of the indicators

Analyses conducted to develop the indicators demonstrated that the SCRQoL indicator had good psychometric properties. The preference weights used to weight the ASCOT items to generate the PW-SCRQoL indicator are illustrative weights from a pilot study conducted as part of the Outcomes of Social Care for Adults (OSCA) study. We cannot therefore be completely certain about the exact properties of this indicator, although it seems unlikely that the weights from the full study will be sufficiently different to change the conclusions drawn about this indicator in this report. One reason for this is that when comparing the SCRQoL and PW-SCRQoL measures (effectively equally-weighted and differentially-weighted versions of the ASCOT) as we do below, we note that the preference weights do not seem to greatly alter the qualities of the SCRQoL indicator.

Although both the SCRQoL and preference-weighted SCRQoL indicators were skewed the skew did not seem severe enough to cause problems for analysis. By contrast, all the SCRQoL need indicators were highly skewed and could not be analysed using methods for normally-distributed data. We therefore recalculated these indicators as binary variables. By this we mean that the new variables were generated from the counts of needs, which reflected the proportion of people with at least one social care need in any CASSR.

The interpretation of social care need in each of the SCRQoL needs indicators makes a very big difference to the proportion of people estimated to have at least one social care need. Only the measure with the broadest definition, i.e. a low- or high-level need in at least one SCRQoL domain, had good variability. All of the other measures had a very small proportion of people reporting at least one need and consequently they had poor variability. This was particularly so for the measure with the narrowest definition, i.e. a high-level need in at least one of the basic SCRQoL domains, which had on average across the whole sample only three per cent of people reporting an social care need. Although these results are good from the point of view of services since services should not be leaving people in a state of social care need, measures with poor variability are not good for making comparisons across groups as large sample sizes are required for accurate and reliable estimates, to enable accurate and reliable monitoring of variations over time and between CASSRs.

The evidence from the analysis of the validity, reliability and sensitivity of three different types of indicators found that the SCRQoL and preference-weighted SCRQoL indicators were very similar. Both require small numbers of people to attain the required level of accuracy for indicators (according to both the IC criterion and the generalisability theory method) and they both show similar relationships with other measures of outcome states, such as the general measure of quality of life and satisfaction with services. There were also observable differences between CASSRs on both of these measures. Further analysis showed that they also produce very similar rankings of CASSRs, which means that the choice between these two measures will not have significant consequences for the ordering of CASSRs. The choice between these two measures should therefore be determined by their acceptability to interest groups.

The SCRQoL needs indicators, by contrast, produce different rankings of CASSRs depending on how social care need is defined. The broadest measure of social care need produces a ranking of CASSRs that is most similar to the ranking obtained with the SCRQoL and preference-weighted SCRQoL indicators. All of the needs indicators also have weaker relationships with the general measure of quality of life and satisfaction with services than both the SCRQoL and PW-SCRQoL indicators, indicating that they are less valid measures of outcome states. This is unsurprising since the concept of need, although similar to the concept of outcome state, is conceptually distinct. Again it is the broadest measure of social care need that has the strongest correlation with the measures of quality of life and satisfaction with services. However, the broadest measure of social care need, because it has the greatest variability, requires a much larger number of respondents than any of the other measures to achieve the required level of accuracy according to the IC criterion. Although this was not the case when the generalisability method for estimating reliability was used, because of the limitations discussed below we suggest caution in interpreting this finding. On the positive side, the concept of SCRQoL need captured by these indicators is likely to resonate more with CASSRs since it is more closely related to the concerns of the service and the scores produced are meaningful: if they know that people in their care have needs in any of the SCRQoL domains then they should want to rectify this situation.

The limitations of the dataset, in particular the small number of participating CASSRs and the systematic differences between CASSRs in the way data have been gathered because of the choice of pilot, caused a number of problems for the analysis. We advise caution in interpreting the results of the reliability analysis based on generalisability theory and the sensitivity analysis. For the former it is unclear how estimates of reliability are affected as the two problems work in opposing directions. Since it was found that the scores on each of the indicators varied according to whether the person had help from a care worker and the versions of the questionnaire used (see below), it is possible that differences in the proportion of people responding via these methods between CASSRs explains the differences in the value of the indicators observed across CASSRs. Multivariate analysis would help to disentangle the effects of these factors, but it has not been possible to explore this here because of the small number of CASSRs participating in the pilot. It would be of value to repeat the reliability and sensitivity analyses on the full dataset to better understand these issues.

Despite these problems, with the limitations of using a pilot dataset we are able to conclude that the most suitable measures for an overarching national indicator would be either of the SCRQoL, the preference-weighted SCRQoL or the SCRQoL needs indicator, where need is defined as the proportion reporting at least one low- or high-level need in any of the SCRQoL domains. The reason for choosing this SCRQoL needs measure over the other needs measures is that the lack of variability in the other SCRQoL needs measures will make them difficult to use. The choice between these three indicators depends on their acceptability to users and feasibility issues associated with their use. It also depends on the balance to be sought between these two qualities and the qualities of validity, reliability, sensitivity.

Aggregating data gathered by different methods

An important issue requiring further investigation regards the effect of the way data are gathered on the value of the indicators. Data are gathered using a number of different methods and techniques. People can complete different versions of the questionnaire and may have help from another person, including their care worker. The analysis conducted here found that people who completed the LD version of the questionnaire and those who have help from care workers systematically report better outcomes and that these two variables (completion of LD questionnaire and help from a care worker) are also associated. Prior to combining responses gathered from different methods, it is important to be clear about the reason for these observed differences: for example, to be clear that the differences arise because of real differences in outcomes between these groups, rather than being caused by the method used to gather the data. To understand this fully, it is necessary to investigate these relationships in more detail using multivariate techniques when a suitable dataset becomes available, perhaps supported by qualitative work to aid interpretation. In the interim, a method for reweighting responses could be considered to ensure that CASSRs with a greater number of people completing LD questionnaires, for example, do not benefit until the effect of the method for gathering data on the indicator score is clear.

Other considerations for the future ASCS

Another consideration for the future is the suitability of the accuracy criterion used by the IC for determining the sample if either of the SCRQoL and PW-SCRQoL indicators are used. The IC criterion leads to very small sample sizes (fewer than 100 people), which may not be large enough to identify some real differences between CASSRs. Samples of this size would produce unsuitable sized samples for further intra-CASSR analysis. It is also worth noting that this criterion is not suitable for measures of rare or relatively infrequent events, as we have seen with the SCRQoL needs indicators. For the SCRQoL needs indicator with the narrowest definition of need, the confidence interval produced is much wider than the indicator point estimate. Alternative accuracy criteria could be considered, as well as alternative approaches to estimating a required sample size, such as power calculations for detecting differences between subgroups.

This analysis also uncovered some inconsistencies between the care setting and client group and the version of the questionnaire that was sent to the client. It is unclear what is the source of the error that produces these inconsistencies, but it is important to understand this as it could have an effect on response rates.

Response rates were very variable across CASSRs and have been so in all past surveys of this kind. Should the variations persist through to the full survey, it would be of value to investigate the factors explaining differences in response rates across CASSRs using multivariate techniques, as this could provide useful information on good process for future surveys.

Attributing outcomes to services

It is important to note that all of the measures suggested as indicators focus on *outcome states* rather than the *outcome* from services. The outcome state of a person is determined by many factors, including factors unrelated to the service. Because of the effect of other factors, it is difficult to attribute differences between CASSRs on these indicators to differences in the action of services across CASSRs and/or the policies of CASSRs; the observed differences may simply be a result of differences between CASSRs in these other factors that are unrelated to the action of services or the CASSR. If these data are to be published or if the aim of measurement is to capture the quality of publicly-funded services, it is important that the indicators are adjusted to reflect the outcome from such services, or the contribution that services make towards the outcome state of an individual. In addition, such analysis can help to identify potential inefficiencies. There is therefore a strong argument to be made for developing a measure that captures the outcome from services rather than the outcome state of individuals receiving services, or at least conducting the required analysis so people using the data are aware of its limitations for assessing differences between CASSRs.

In this report two approaches to isolating the effect of services have been suggested, which can be viewed as potential short- and long-term solutions to the problem of attributing outcome states to the action of social care. The potential short-term solution draws on the production of welfare (POW) approach. POW is an analytical framework that sets out the key factors influencing the process of producing social care outcomes for an individual. It can be used to build regression models to estimate the contribution that services make to the outcome state of an individual. Analysis using this framework is currently being conducted on the 2009 older people's home care user experience survey. Results from this analysis should be available in early 2011 and will provide useful insight into the feasibility of this approach.

The potential longer-term solution draws on the approaches developed to estimating outcomes from social care as part of ASCOT. ASCOT contains two parts: a set of questions asking about a person's current SCRQoL and a set of questions asking about the person's SCRQoL in the absence of services. The latter set of questions is labelled 'expected SCRQoL' in the absence of services. These questions can only be asked in an interview, but by mapping the 'expected SCRQoL' in the absence of services to other measures of need that are easier to measure in a self-completion survey, such as activities of daily living (ADLs) measures, for clients with specified characteristics it may be possible to estimate the service outcome for each individual and produce a 'value-added' measure for social care. The success of this approach depends on developing robust mapping equations between expected SCRQoL in the absence of services and other needs measures. A separate study would be required to produce the mapping equation and test it, and this would require additional resources and fieldwork.

Conclusions and recommendations

This report has demonstrated the validity, reliability and sensitivity of three different types of indicators that could be generated from the ASCS data and the ASCOT items, in particular. The evidence suggests that the SCRQoL and preference-weighted SCRQoL indicators will produce very similar rankings of CASSRs and have similar levels of validity, reliability and sensitivity, although, as noted above, the latter two attributes were hard to test given the limitations of the dataset.

The SCRQoL needs measures, by contrast, produce slightly different rankings of CASSRs depending on how need is defined. The broadest measure of SCRQoL need behaves similarly to the SCRQoL and preference-weighted SCRQoL indicators, but it is a less valid measure of social care outcome. The other three SCRQoL needs measures capture relatively infrequent events, which means that large sample sizes are likely to be needed to provide an accurate picture. We suggest therefore that these three measures are less suitable for overarching indicators.

As identified above, the small number of participating CASSRs and the systematic differences between CASSRs in the way data have been gathered, caused a number of problems for the analysis. Since this was a result of the small samples used to pilot the survey, the full survey is likely to provide a much more suitable dataset and it may be of value to repeat the reliability and sensitivity analyses on this dataset when it becomes available.

The analyses conducted here raise a number of questions about the compatibility of questions derived using different data collection methods, the suitability of the accuracy criterion used by the IC for these data and the effect of response rates on indicator values. I have suggested that further research and analysis are conducted to better understand the issues raised.

Finally, as noted above, all of the indicators capture outcome states, not outcomes from social care. Two approaches have been suggested, a short- and long-term solution, that could be used to generate an indicator from the ASCOT items that is more sensitive to the outcomes of social care. To generate and test such an indicator requires detailed analysis and further data collection. We plan to investigate the suitability of the short-term solution on the older people's home care survey and this analysis should be available in early 2011. This analysis is important since it will indicate the extent to which differences between CASSRs on the indicators can be attributed to the services and the policies of the CASSR. The long-term solution requires a separate data collection exercise and analysis of the dataset to generate and test a value-added measure of social care outcomes.

Introduction

This report compares options for a new indicator for social care service users based on data from a new survey of social care service users known as the Adult Social Care Survey (ASCS).[1] The work presented here was commissioned under the previous Labour government and its aim was to develop an ‘outcomes-focused’ performance indicator (PI) that could be used to populate the National Indicator Set (NIS) for the new spending period starting from 2011. Although the new government is keen to reform the performance framework, it seems likely that an indicator which focuses on users’ views of their outcomes from services will be of interest since it fits with the principles and ethos of the new outcomes framework set out by the Coalition for health care (Department of Health, 2010).

The focus of this paper is on comparing potential ‘outcomes-focused’ indicators. To this end we present empirical evidence for the validity, reliability and sensitivity of the proposed indicators, across the range of service areas and client groups surveyed. However, these are not the only important characteristics for an indicator to possess; acceptability of the indicators and feasibility are also important qualities. Indicators are one part of a wider regulatory and performance framework designed to, amongst other things, assure the quality of services for the people who use them. The place of indicators within this framework, how they are to be used and by whom are all important questions to address since the answers to these questions determine the balance between these five characteristics. The evidence presented here should therefore be considered in the light of the intended function of and audience for the indicator.

The report is organised as follows: In the next section, I provide some background to the analysis in this report, focusing in particular on the development and content of the ASCS and the options for indicators that could be generated from the ASCS data. A framework for choosing between the options is set out which describes the desirable characteristics of indicators. The analysis in this report draws on a pilot of the ASCS conducted by a number of volunteer councils with adult social services responsibilities (CASSRs) under guidance from the NHS Information Centre for Health and Social Care (IC). The data and its collection are described, followed by a section setting out the methods used to assess the validity, reliability and sensitivity of the indicators. The results of these analyses are then presented alongside some descriptive analyses to provide some context to the main analyses. This section of the report and the methods section are highly technical. Readers who are not interested in the technical detail can skip over these sections to the discussion where I summarise the findings and their implications for the choice of indicator, paying attention to the extent to which each indicator achieves the desirable characteristics. I also discuss the pros and cons associated with each choice of indicator and consider how they could be judged against the criteria of acceptability and feasibility.

---

[1] Originally known as the Putting People First Survey, and subsequently the National Adult Social Care User Experience Survey).

Background

Since 2001, all councils with adult social services responsibilities (CASSRs) have been required to conduct surveys of users' experiences of social services (user experience surveys (UESs)). These are national surveys carried out by CASSRs following guidance from the IC and are regarded as an important part of the overall performance framework for social care. They provide councils with information about how they might improve services locally and central government and regulators with information to monitor and compare the performance of councils. CASSRs are required to submit their raw data to the IC for analysis centrally, and user surveys are now an established part of CASSRs' annual data returns to central government.

These surveys have made a valuable contribution towards putting the views of users at the heart of decision-making and are seen as an important source of data for PIs. However, a criticism of the past surveys and indicators based on these surveys is that they have focused on the service users' experiences of aspects of the process of service delivery rather than outcomes, and have captured different user groups each year, limiting the possibilities for monitoring improvements over time (Department of Health, 2009). In view of this criticism, the DH was keen to develop a survey for social care that included the views of all social care users, regardless of client group or care setting, and better reflected service users' outcomes, in line with the then Labour government's policy to realign the performance framework around the objective of improving outcomes for all (Office of the Deputy Prime Minister, 2005). One purpose of the survey would be to provide data which could be used to populate indicators in the performance framework, at that time the National Indicator Set (Communities and Local Government, 2007). With the transition to the new Coalition Government, the exact nature of the new performance framework is not clear. However, the general direction taken by Labour towards developing an outcomes-focused indicator which draws on the perspectives of users is echoed in the recent Coalition government proposals for a new outcomes framework (Department of Health, 2010).

Development of the Adult Social Care Survey (ASCS) questionnaire

To meet the needs of the performance framework, the DH commissioned the Personal Social Services Research Unit (PSSRU) and the Tizard Centre at the University of Kent to assist in developing a new Adult Social Care Survey (ASCS) for 2010/11. The aim of this work was to develop a survey that was outcomes-focused and applicable to all social care client groups regardless of care setting. From the outset, it was acknowledged that this was a challenging aim because of the heterogeneity of the social care user population and the services they use. Challenges were also imposed by the difficulties in obtaining a sampling frame, particularly for privately-funded social care users. (For a more detailed documentation of the challenges associated with this survey please refer to Malley and Netten, 2009). For this reason, it was decided that the survey would focus, in the first year at least, on publicly-funded clients for whom a reasonably accurate and complete sampling frame exists in the form of CASSR records.

In consultation with stakeholder groups, it was agreed that a set of questions from the Adult Social Care Outcomes Toolkit (ASCOT) should be used as the core questions for the survey. ASCOT comprises a set of instruments which have been developed by researchers at PSSRU over a number of years. For ASCS only the set of questions that capture current social care–related quality of life (SCRQoL) or current social care outcome states are used. ASCOT has been refined recently as part of the Outcomes of Social Care for Adults (OSCA) study, which is funded by the NIHR under the Health Technology Assessment programme. The body of work developing ASCOT has involved consultation with stakeholders and has drawn upon the substantial literature reporting social care service users' definitions of social care outcomes and their quality of life to develop the domains of outcome assessed in the measure (see e.g. Malley et al., 2006, Netten et al., 2010a, Netten et al., 2005, Netten et al., 2002). The OSCA study reviewed the data from these previous studies to refine the domains and questions, and further tested the questions via cognitive interviews with about 30 people(Netten et al., 2009). The aim of cognitive interviews is to ensure that the questions are expressed in language that makes sense to service users and that the questions are not offensive (Willis, 2005). Together these studies have involved testing of the domains and questions with over 90 service users. For more information about ASCOT, visit the ASCOT website at: http://www.pssru.ac.uk/ascot/.

Because the revised ASCOT questions had not been tested in the self-completion format in all the social care settings and with all the social care client groups, it was agreed that further development work should be carried out to ensure that the questions were suitable for the breadth of social care services and client groups surveyed as part of ASCS. The development work consisted of three research strands, examining the feasibility of conducting the self-completion survey in various settings. Projects examined the variety of help received by service users in completing the questionnaire and consequences for the validity of the data, the feasibility of the self-completion approach and the suitability of the questionnaire for people living in care homes, exploring the feasibility of asking advocates to help service users to complete the questionnaire and the consequences for the validity of the data. A fourth strand developed a version of the questionnaire suitable for people with learning disabilities (PWLD) and explored the feasibility of the self-completion approach with this group. As part of this research, the ASCOT questions were tested with another 40 people evenly split between people living in their own homes and in care homes (nursing and residential), and with a further 33 people with learning disabilities. One aim of this work was to make sure the phrasing of the questions was clear and that the questions were relevant to these clients in these settings. Another aim was to understand the implications for implementation arising from the care setting and client characteristics.

The findings were generally positive about the feasibility of conducting the survey across the care settings studied and clients included in the development work. However, it was recommended that people who were unable to give to consent to participate in the survey and would require a proxy should be excluded, awaiting a more detailed investigation into how valid data could be gathered from proxy respondents. As well as the standard questionnaire, two variants were produced: one for people with learning disabilities and one for people in residential care. These variants were necessary to ensure that the questions were understood clearly by these different user groups (Malley et al., 2010). The standard questionnaire is reproduced in Appendix 1. On the basis of the report's recommendations, the IC carried out a pilot of the ASCS with a sample of CASSRs in 2010. The pilot dataset is used for the analysis in this report.

Measuring performance: Qualities of an indicator

In selecting an indicator, a key question is how will it be used in the future and by whom. There are a variety of potential users of the data and a number of ways in which these same stakeholders could make use of it. For example, government departments and regulators may use the information to monitor performance nationally and make comparisons across different localities to identify ‘beacons’ or ‘coasting’ areas. In the past the government has chosen some indicators as targets and has attached financial rewards to achievement of the target. The general public, service users and carers could also make use of the statistics to assess the effectiveness of their local services, and depending on whether the data are available at the level of the individual service, to inform their purchasing choices between providers. CASSRs and third sector organisations might use the results to gauge the success of local initiatives for service users, and CASSRs may also use the data to inform their commissioning decisions.

Indicators can clearly serve a variety of functions and the way data are to be used and by whom has significant consequences for the desired characteristics of an indicator. It is possible, however, to identify general desirable qualities for an indicator, which are set out in Box 1. Many of these qualities are in conflict with each other, so when choosing an indicator it is important to consider the appropriate balance between these characteristics – an activity that is likely to depend on how the indicators are to be used and by whom. For example, simplicity is often in conflict with the principle that the indicator should not be able to be gamed since more simple measures are usually easier to game. Clearly, where financial incentives are attached to targets, or organisations are rated by their performance on the indicator, it is important that the indicator cannot be easily gamed. However, this goal would be less important if data were solely to be used for information since the stakes associated with poorer relative or absolute performance would be less high and there would be less incentive to game the indicator. The relative balance of these qualities is a matter for debate, but informed debate depends on clarity over which of the uses and users are most important, and recognition of the need to compromise.

Box 1: Desirable qualities of an indicator[2]

Acceptability: the extent to which the indicator is acceptable to those being assessed and those undertaking the assessment. This includes aspects such as relevance to current and future policy aims, whether the indicator is easy to understand, is well-defined and easy to calculate. It can also include aspects linked to uses, such as whether it can provide data that are comparable over time and across different locations; and whether it can be attributed to the action of CASSRs (or providers) who are accountable for the outcome. Another area of concern is around ensuring that the indicator cannot be gamed and does not induce perverse incentives.

Feasibility: the extent to which valid, reliable and consistent data are available for collection and reporting in a timely manner, so the data are useful.

Reliability: the indicator should have minimal measurement error, or at least be accurate enough for its purpose. In the case of survey-based indicators, where each service user is a rater of CASSR performance, inter-rater reliability should be high: that is, responses from raters should be similar. Where responses are variable, it is important that a good level of reliability can be achieved by increasing the number of raters.

Sensitivity: the indicator should have the capacity to detect changes in the unit of measurement, in this case variations in outcomes across CASSRs (and potentially providers) and variations in outcomes within CASSRs over time.

Validity: the extent to which the indicator accurately measures what it purports to measure. Key issues here are whether the indicator is underpinned by evidence and/or consensus, the extent to which the indicator has the ability to accurately predict outcomes, and whether the basis for scoring and combining responses is defensible.

The first two of the desirable qualities – acceptability and feasibility – depend in large part on the intended uses and audiences for the indicator. They cannot be assessed empirically, although it is possible to undertake analyses to inform these decisions. The latter three features – reliability, sensitivity and validity – can be assessed empirically (at least to some extent) prior to implementation and can reasonably be seen as part of the development and testing phase. In section 4 I set out the types of analyses that can be conducted to inform assessments of reliability, sensitivity and validity. In the next section I discuss the options for indicators based on questions within ASCS.

Outcomes-focused indicators: options for consideration

ASCS contains a question on overall quality of life and a couple of questions on aspects of health, including depression/anxiety and pain. The questions with specific relevance to social care outcomes in the questionnaire are the following:

- Control over daily life: Which of the following statements best describes how much control you have over your daily life?

---

[2] Adapted from a document produced by Nalyni Shanmugathasan, DH; Campbell et al. (Campbell et al., 2003); and Bird et al. (Bird et al., 2005).

- Personal care: Thinking about your personal care, by which we mean being clean and presentable in appearance, which of the following statements best describes your situation?
- Food and drink: Thinking about the food and drink you get, which of the following statements best describes your situation?
- Accommodation cleanliness and comfort: Which of the following statements best describes how clean and comfortable your home is?
- Personal safety: Which of the following statements best describes how safe you feel?
- Social participation: Thinking about how much contact you've had with people you like, which of the following statements best describes your social situation?
- Occupation: Which of the following statements best describes how you spend your time?
- Dignity: Thinking about the way you are helped and treated, and how that makes you think and feel about yourself, which of these statements best describes your situation?

As I have already described, these questions are taken from the ASCOT which has been developed by researchers at PSSRU at the University of Kent over several years. Although the questions are outcomes-focused, they do not capture social care outcomes *per se*; rather, they capture the *outcome state* of the person. The questions cover aspects of a person's quality of life that we can reasonably expect to be affected by social care services. I therefore refer to these questions as capturing social care-related quality of life (SCRQoL)(Netten et al., 2009). Capturing social care outcomes is particularly complex because of the special characteristics of social care, including the ongoing nature of services and the fact that their aim is usually to maintain the quality of life of the person (rather than improve or cure the person's underlying impairment) often in the face of an unavoidable decline in their health. PSSRU has developed techniques that use these SCRQoL questions alongside, for example, measures of need, to estimate social care outcomes. The methods and additional tools required to make these estimations are all part of the ASCOT.

Each SCRQoL question has four response options which are phrased to capture the outcome states of 'high-level needs', 'low-level needs' and 'no needs'. The 'no needs' category is divided into two options, representing a state where the person has no needs but their aspirations in the domain are not met ('mustn't grumble' state), and a state where the person has no needs and their aspirations in the domain are met ('ideal' state).[3] In discussion with the advisory group for this study, several ways of developing an outcomes-focused indicator from the ASCS data were identified. These are outlined below.

---

[3] The dignity domain is structured slightly differently to the other domains because of the nature of the attribute. Although the two bottom response options represent high- and low-level needs states, the two top response options do not represent 'mustn't grumble' and 'ideal' states, and in some circumstances could be considered equivalent. For example, evidence from the cognitive interviews suggests that the second response option is particularly common amongst people receiving relatively low levels of service input. They feel that 'services do not affect the way they think or feel about themselves' because services do not have enough of an impact on their everyday lives.

Summary score across all *dimensions* of SCRQoL

An indicator could be specified as a summary score across all the SCRQoL items, where each response level takes a different value (e.g. zero for high-level needs and three for the 'ideal' state). An indicator developed this way would be based in psychometric theory, where multiple items representing a single concept are frequently summed together to produce a single score, provided the items meet the required psychometric criteria. This method of summing across items assumes that each response has equal weight. As an indicator for the CASSR, this could be presented as the average SCRQoL score for individuals within a CASSR. This measure is referred to as the SCRQoL indicator.

Summary score across all dimensions of SCRQoL, weighted by preferences

This is very similar to the measure in 0 in that it produces a summary score across multiple SCRQoL items, which as an indicator could be reported as an average for each CASSR. However, unlike the measure in 0 where each response option and dimension is given an equal weight in constructing the score, in this measure each response option and dimension would have its own weight determined by preferences for that state. For example, if people in general prefer a state of low-level needs in the occupation domain to low-level needs in the personal care domain, then this preference ordering (and the magnitude of the preference) will be reflected in the weights. Whereas the measure specified in 0reflects the social care outcome state, this measure would reflect the value of the social care outcome state. In this document this measure is referred to as the PW-SCRQoL indicator.

As part of work for the development of ASCOT, PSSRU is currently conducting a study to gather preferences for the outcome states in the SCRQoL measure. Preferences are being gathered from service users and a general population sample. It should be noted that until these data are available (early in 2011) we do not know whether service users have different preferences to the general population, although work in other areas of policy suggests that there do tend to be variations between those people with experience of services and the general population. It should also be noted that the service users are not a representative sample and that, given the relatively high level of cognitive functioning required for the preference task, certain users will always be precluded from participating (e.g. those with moderate to severe dementia). This means that it will never be possible to attain a set of preference weights based on a representative sample of users. The preference weights will be available early in 2011, but provisional preference weights produced from a general population pilot study are currently available (Burge et al., 2010).

Presence of SCRQoL needs

One of the benefits of the approach to item construction in ASCOT is that the response options are designed to be meaningful in the sense that they describe states of no, low- and high-level needs. We can make use of this structure to generate an indicator, which captures the extent to which SCRQoL needs are being met across CASSRs.

There are many different ways that the ASCOT items could be combined to generate a measure of SCRQoL needs. These include:

- A count of the number of SCRQoL domains within which the person has either a low- or high-level need.
- A count of the number 'basic' SCRQoL domains within which the person has either a low- or high-level. The difference between this measure and the one above is that this would only be formed from the basic SCRQoL domains, i.e. the personal care, safety, accommodation cleanliness and comfort, and food and drink domains.

Further, it may be considered more appropriate to capture the number of high-level needs rather than low- or high-level needs. The following indicators could be formed on this basis:

- A count of the number SCRQoL domains in which the person has a high-level need.
- A count of the number of basic SCRQoL domains in which the person has a high-level need. Again, the difference between this measure and the one above is that this would only be formed from the basic SCRQoL domains.

As an indicator, these measures could be reported either as the average number of domains in which people in a CASSR have needs, or as the proportion of people in a CASSR experiencing needs in at least, for example, one SCRQoL domain. The choice between these two options may be partly determined by the distributional characteristics of the measure, but also by the role of the indicator. This measure is referred to in the remainder of this report as the SCRQoL needs indicator.

Data and data collection

The pilot ASCS was carried out by a small sample of CASSRs that had volunteered to take part. In order to ensure enough data was gathered about the feasibility of conducting the survey in residential homes and with people with learning disability, the pilot had three strands:

- A survey of those living in residential care
- A survey of those with learning disabilities
- A survey across all social care users.

CASSRs could volunteer to take part in one or several strands of the pilot. In total, 18 councils volunteered to take part in at least one element of the pilot survey, and three CASSRs participated in more than one element of the pilot.

Guidance on sampling, data collection and management was sent to all participating CASSRs to ensure the data collected are comparable across CASSRs (NHS Information for Health and Social Care, 2010a). The guidance requested CASSRs to select a random sample of eligible service users from their records, where eligibility was defined as anyone aged 18 or over who was in receipt of some form of social care on a specified date. This includes all people in receipt of residential care, equipment, day centres and home care, but excludes people who pay privately for social care. For the pilot, the IC guidance was that the sample size should be at least 100, on the basis that a sample of this size should be large enough to test the feasibility of the survey and explore the distributional characteristics of the questionnaire items.

All CASSRs used the three questionnaires supplied by the IC which had been developed by PSSRU (Malley et al., 2010). CASSRs were instructed to send the questionnaire for people with learning disabilities to anyone with a learning disability irrespective of care setting, to send the residential care questionnaire to anyone in a residential care setting (unless the person had a learning disability), and to send the standard questionnaire to all other social care clients.

Methods

To provide background to the analysis, I first describe the characteristics of the sample respondents and their responses to the ASCOT questions. In particular, I describe the patterns of help received by respondents to answer the questionnaire and explore whether there are any variations in the responses to ASCOT questions according to the type of help received and the type of questionnaire completed. Formal testing of differences is conducted with chi-squared tests of association.

Development of the indicators

Summary score across all dimensions of SCRQoL

In developing an indicator which is a summary score across all dimensions of SCRQoL, it is necessary to establish the validity of the basis for combining responses. In this proposal the value of the responses to each of the SCRQoL items are summed. The scale so produced from the multiple items represents the SCRQoL of an individual, but such a scale is only valid if the items are similar enough to justify adding the responses to each of the items together. I use well-established psychometric techniques to demonstrate the validity of this approach.

Within the framework of classical test theory, factor models are commonly employed to develop multi-item scales from a pool of items since the aim of the model is to explain the common variance of the items through a smaller number of latent variables, known as factors (De Vellis, 2003). The structural relationship between the items is investigated first through examination of inter-item correlations. Polychoric correlations[4] are used because of the ordinal nature of the measurement scale for each item (Olsson, 1979a). Exploratory factor analysis is carried out on the polychoric correlation matrix, as research has shown that Pearson correlations can lead to incorrect conclusions where the items are ordinal (Holgado–Tello et al., 2010, Olsson, 1979b). I use the maximum likelihood method for factor extraction primarily because it allows for statistical evaluation of the factor solution (Fabrigar et al., 1999). Strong loadings of the items on the latent variable and low values for the unique variance of the items indicate that the latent variable explains the items well and the items can be summed together as a scale.

If CASSRs' social care policies and the services they provide have an impact on service users' SCRQoL, I would expect service users experiencing the same set of policies and services to have more similar SCRQoL (all other things being equal) than service users experiencing a

---

[4] Polychoric correlations are calculated in STATA version 11 using the user-written polychoric programme.

different set of policies and services. It is likely that service users within the same CASSR will have more similar SCRQoL than service users in different CASSRs, giving the dataset a clustered structure. Clustering of responses, in this case by CASSR, can have an effect on the factor solution and it is generally recommended that where data are clustered steps are taken to adjust for the clustering by using a multilevel factor analytic approach (Steele and Goldstein, 2006). It is possible to gauge whether the multilevel factor analysis will substantially affect the results by exploring using a variance-components model the proportion of variance that is attributable to the CASSR for each of the items (Grilli and Rampichini, 2003, Muthén, 1994). Since multilevel factor analysis is highly resource-intensive, requiring significant computing power, I only explore the potential effects by modelling the CASSR-level variability in each SCRQoL item using a cumulative logistic model for ordinal responses.[5]

To reflect the fact that the probability of endorsement of a response option varies across the quality of life (QOL) items in a graduated way, for this multi-item measure I also used a less familiar technique known as Mokken scaling, or non-parametric item response theory (IRT), which is a more appropriate model than the factor model for developing scales when items are hierarchical (Sijtsma and Molenaar, 2002, Sijtsma et al., 2008, Moorer et al., 2001). If items are hierarchical, then it is possible to determine a person's response to any of the items in the set based on their response to one item, as endorsement of one item response option necessarily entails endorsement of other item response options that are less difficult: i.e. there is a transitive relationship between the item response options. IRT models the hierarchical nature of items probabilistically; in an IRT model the probability that a person positively responds to a certain item response option is a function of both the position of the person along the latent continuum and the position of the item response option along the latent continuum. I would expect some of the items, such as those representing QOL, to show this hierarchical pattern of response because the QOL attributes are not equivalent in the sense that some states are less desirable than others. The states that are less desirable should be endorsed less frequently than those that are more desirable (c.f. Moorer et al., 2001).

Mokken scaling is a non-parametric version of the more familiar Rasch model and makes fewer restrictions about the function relating the probability of response to the location parameters for items and persons. Mokken scaling only requires that the function increases with the latent trait value (monotone homogeneity). Items may obey the requirement of monotone homogeneity yet not be useful scale items. Loevinger's coefficient of homogeneity (H), which is the ratio of observed Guttman errors (errors in the aforementioned transitivity relationship) to total errors under the null hypothesis that items are totally unrelated, is used to assess scalability. It can be constructed for an item, item pairs and the set of items (scale). Higher values of H for the scale indicate a more reliable ordering of items and persons, with values of $H \geq 0.3$ indicating a weak scale, values ≥ 0.4 a medium scale and values ≥ 0.5 a strong scale (Mokken, 1971, Sijtsma and Molenaar, 2002). As a rule of thumb, items with H values < 0.3 are considered to be unscalable since they do not provide a reliable ordering of persons (Sijtsma and Molenaar, 2002).

---

[5] This is conducted using the user-defined gllamm procedure in STATA version 10 (Rabe-Hesketh et al., 2002).

Like factor analysis, Mokken scaling can be used to investigate the dimensionality of the QOL items and identify those that scale strongly and less strongly together. I use a variant of the Mokken procedure proposed by Hemker et al. (1995), applied by Moorer et al. (2001) and available in the MSP (version 5) software (Molenaar and Sijtsma, 2000). Hemker et al. note that a multi-dimensional item bank will often appear to form one scale at low values of *c*, where *c* is equal to H for the scale and is the acceptance criteria for the scale. As *c* is increased to values up to 0.6, if the scale is multi-dimensional it will often break up into several homogeneous subscales, while unidimensional scales will remain intact. Since I expect the QOL items to form a unidimensional scale we would therefore expect the scale to break up into individual items as *c* is raised and not to form several scales.

Preference-weighted summary score across all dimensions of SCRQoL

Much of the development work for this indicator has been conducted as part of the ongoing PSSRU programme of work to develop a measure of adult social care outcomes. Preference studies to generate preference weights for ASCOT have been conducted, although at this time only illustrative weights are available from the pilot study (Burge et al., 2010). Researchers at PSSRU have also conducted extensive validation work for the ASCOT descriptive system, evaluating the extent to which each of the ASCOT conceptual dimensions is captured by the corresponding item and the extent to which each dimension is conceptually distinct from the other dimensions (Netten et al., 2010b). In this report, therefore, I simply explore the distributional characteristics of the preference-weighted measure within this population, using illustrative preference weights from the pilot study.

SCRQoL needs measures

Four different types of measures have been suggested as potential indicators that capture SCRQoL needs, using different definitions of ‘need’. All of these measures are designed to be counts of the number of SCRQoL domains within which the individual experiences a need (defined as either high-level or both low- and high-level). I have also distinguished between the basic domains in which I would expect social care services to meet needs and the non-basic domains where the extent to which CASSRs are able to meet needs is likely to be much more variable.

Count data are frequently characterised by highly skewed distributions and are commonly modelled using a Poisson, rather than a normal, distribution since the latter requires a symmetric distribution. Poisson distributions have only one parameter, λ, which is equal to both the population mean and variance. Because the mean and variance are equal, the distribution is said to be equidispersed, and different methods are required for analysis and the generation of confidence intervals compared to those used for normally-distributed data. It is, however, not necessarily the case that these data will be better modelled using a Poisson distribution. A Poisson distribution can be approximated to a normal distribution when λ is large. It is therefore important to explore the distributional characteristics of these data to determine the extent to which methods for normally-distributed data are applicable.

For analytical purposes, count data are generally expressed as rates. In this example the data can be expressed as the number of needs per person, leading to an indicator defined as the average number of needs per person, or as the incidence of needs per person domain since people are ‘exposed’ to a specified number of domains and the number of needs per person is dependent on the number of domains to which they are ‘exposed’. This would produce an indicator specification of the average number of needs per person domain. Given the lack of experience within social care of using indicators like this, we suggest that if count data cannot be approximated by the normal distribution, it would be better to generate an indicator based on the proportion of people in a CASSR with, for example, at least one need across any of the domains. Such a measure could also be examined with more familiar methods for binomially-distributed data, i.e. methods for proportions.

Assessing validity

Validity assesses the extent to which the instrument measures what it is intended to represent. In Messick’s words, it is ‘an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment’ (Messick, 1990: 1). Here the measure is intended as a performance indicator; it should be able to provide information for a variety of stakeholders to help them make decisions about commissioning services and to assess the extent to which services are providing good outcomes to their users. Validating the indicator is therefore a matter of identifying the extent to which the instrument captures its intended focus, i.e. social care outcomes.

Establishing validity tends to be approached by examining whether the social care outcome measure behaves as expected with other variables. Detailed work along these lines has been conducted as part of the Outcomes of Social Care for Adults (OSCA) study, which has examined the relationship between each of the individual items comprising the ASCOT measure and other variables measuring the concepts represented by each of the items for a group of older people receiving publicly-funded home care and other community-based services, including Direct Payments, meals and day care (Netten et al., 2009, Netten et al., 2010b). The main focus of this work will therefore be to establish the relationship of the indicator with other variables that capture social care outcomes for this particular population. The analysis reported here is therefore limited to (i) the correlation between the indicator and the general quality of life item, question 2 (in Appendix 1), which we would expect to be related to a measure of social care outcome state; and (ii) the correlation between the indicator and the general satisfaction with services item, question 1, which we would expect to be related to a measure of social care outcomes. Importantly from an indicator perspective, this analysis can also establish whether simpler measures, such as the satisfaction and quality of life items, can capture social care outcomes adequately.

Where the indicators are dichotomous a different form of correlation coefficient is required, which is known as the polychoric correlation coefficient and takes into account the binary nature of the indicator. However, this correlation coefficient is based on the assumption that the data are drawn from an underlying normal distribution, which, given the data come from a Poisson distribution, is not true. For this reason we also explore the relationship between the indicators and the general satisfaction and quality of life measures using a non-parametric test of association. In this instance we use Kendall’s tau-b, which is a test of association between two ordinal variables. This statistic takes values from -1 to 1, where 1(-1) is evidence of perfect positive (negative) association and implies that the value of the independent variable perfectly predicts the value of the dependent variable.[6]

A third aspect of validity that is relevant here is how valid it is to combine responses to questions obtained in different ways. In particular, we are concerned with whether it is possible to combine responses gained from the LD questionnaire with those gained from the standard or care home questionnaire. Given the wording of the questions varies in the LD compared to the standard and care home questionnaire, we need to be clear that any differences in the distribution of the indicators are due to real variations in social care outcomes rather than the different ways of asking the questions on which the indicators are based. In addition, a large number of the questionnaires, particularly of those people in care homes and people with LD, are likely to be completed with the help of somebody else. Likewise, it is important to know that any differences in the distribution of indicators according to who helped and the type of help given are a consequence of real differences in social care outcomes rather than due to the additional perspective of the helper. To explore differences in the indicator scores across these subgroups, we conduct a one-way between-subjects analysis of variance (ANOVA).[7] For binary data we explore differences across subgroups using chi-squared tests of association.

Assessing reliability

Measurement is always subject to error, arising from random error associated with misunderstandings for example, and systematic error, for example, associated with the measurement instrument, which may not be a perfect measure of the concept. Measures of reliability aim to quantify how sensitive measurement is to error. Reliability is related to the standard error of measurement (SEM), which is commonly used to construct confidence intervals around estimates, through the formula:

$$SEM = \sigma_X \sqrt{(1-r)} \quad ,$$

where r is the reliability and σ_X is the standard deviation of the test (X).

---

[6] This test is used in preference to Goodman and Kruskal's gamma test, since it tends to be more stable under different categorisations of the variable. By contrast, gamma tends to be larger when a variable is measured with fewer categories (Agresti and Finlay, 1997). This is an important consideration as each of the indicators has a different number of levels.

[7] Bartlett's test for homogeneity of variance was conducted. When this test is violated, the results of the ANOVA may not be trustworthy, namely that the reported p-value from the significance test may be too liberal (yielding a higher than expected type I error rate – reject the null hypothesis when it is actually true) or too conservative (yielding a lower than expected type I error rate). Where Bartlett's test is violated we perform simulations to estimate the degree to which the test is liberal or conservative and examine the extent to which alternative tests (F star and W test) are more reliable, using the simanova programme in STATA (Wilcox et al., 1986).

Cronbach's alpha is commonly used to assess the reliability of multi-item measures. It is a measure of the internal consistency of the items and as such provides an indication of how reliably the instrument or test captures a concept, for example the concept of SCRQoL, for each individual. Since each item is assumed to measure the concept with some degree of error, the more items in the test the more information there is for each individual and the more accurate the measurement for each individual. Therefore Cronbach's alpha increases as the number of items in the test increases.

Whilst Cronbach's alpha provides a good indication of how reliably the multi-item scale measures an individual service user's SCRQoL, it does not indicate the reliability of the measure as an indicator. This is because when the multi-item measure or preference-weighted scale is used as an indicator, each service user effectively becomes a rater of the performance of the CASSR, and each service user rates the performance of the CASSR with error. Drawing on generalisability theory, in this situation systematic error not only arises from the instrument but also from systematic variation in the ratings of service users within one CASSR; the true score can be thought of as the between-CASSR variation in the indicator score (Shavelson and Wenbb, 1991, Brennan, 2000, Cronbach et al., 1997). Reliability of the indicator is therefore given by the following formula:

$$r = \frac{VAR_B}{VAR_B + \frac{VAR_W}{n}} ,$$

where VAR_B is the between-CASSR variance, VAR_W is the within-CASSR variance, and *n* is the number of respondents to the test. Reliability is therefore a function of the number of raters and like Cronbach's alpha reliability will increase as the number of respondents increases. Where *n* is one the formula reduces to the variance partition coefficient (VPC) or intra-class correlation coefficient (ICC).

To estimate reliability for each indicator we estimate a variance-components model to obtain an estimate of the VPC via restricted maximum likelihood.[8] For dichotomous variables we also estimate a variance-components model, but with a logit link, to reflect the binary nature of the variables.[9] Simulation studies have shown that these methods can underestimate the variance components when the number of groups is less than 50 and the errors are not normally distributed. These problems are compounded when the VPC is low at around 0.1 (Maas and Hox, 2004). This dataset is not ideal for estimating the VPC since only 18 CASSRs are participating in this study, and they are not all surveying the same groups. We therefore advise

---

[8] A variance-components model is used instead of analysis of variance since it is more appropriate where designs are not balanced, as is the case here (Rasbash and Goldstein, 1994). Restricted maximum likelihood estimation rather than maximum likelihood estimation is used since it is more accurate when the aim is to provide estimates of the variance components (Hox and Maas, 2006).

[9] Alternative methods for estimating the intra-class correlation coefficient have been suggested for dichotomous variables, which are not based on the latent scale (Ridout et al., 1999).

that these results should be treated as indicative of the reliability of each indicator rather than conclusive.

In the past the IC has used the criterion that the 95 per cent confidence interval for the mean should have a width no greater than eight per cent of the scale (or plus or minus four per cent of the scale length) to determine the required attained sample size for each CASSR. The formula for estimating sample size, n, can be written as:

$$n = \frac{z_{\alpha/2}^2 \sigma^2}{d^2},$$

where d is half the width of the required confidence interval, z is the value of the normal deviate for a two-sided 95% confidence interval (i.e. the value at p=0.025, which is approximately 1.96) and σ^2 is the population variance. For binary data, the population variance is given by $p(1-p)$, where p is the probability of success (in this case success is rather awkward terminology since success is defined as the probability of unmet need). We use this formula to estimate the required sample size according to the IC criterion for the accuracy of the indicator estimate.

Assessing sensitivity

The sensitivity of an instrument can be assessed by examining whether the instrument is able to detect meaningful changes in the unit of measurement. In this instance sensitivity is measured in terms of whether the indicator is able to detect meaningful differences in the performance of CASSRs at one point in time. (It is not possible to examine the ability to detect meaningful changes over time as the data are cross-sectional.) Detecting *meaningful* differences is the key issue at stake, and it is therefore important to define what we mean by meaningful. For health measures it is common to find meaningful being equated with therapeutically significant. However, there is no obvious parallel for service users, and anyway we are interested in detecting meaningful differences among CASSRs, not individual service users. Instead, here I focus on whether there are any differences across CASSRs on the proposed indicators using tests based on ANOVA[10] and comment on the significance of these findings, by examining standard errors and absolute differences in the scores across CASSRs. The rationale is that a measure that identifies differences across CASSRs is a useful indicator; one that did not identify differences across CASSRs would not be a valuable measure.

For dichotomous variables we analyse differences across CASSRs using a random-effects variance-components model with a logit link to reflect the binary nature of the variables. Differences between CASSRs are tested by examining the significance of the VPC, where a significant VPC indicates that there are differences between CASSRs. As I have discussed above the small number of participating CASSRs may mean that the variance components are underestimated. I therefore also examine the significance using a logistic regression and entering the CASSRs as dummy variables. Differences between CASSRs are formally tested using an omnibus Wald test.

---

[10] As above, we conduct checks to ensure Bartlett's test of homogeneity of variances is not violated.

Results

As identified above, 18 councils took part in at least one element of the pilot survey. Three councils participated in more than one element of the pilot. The participating councils included all authority types (inner and outer London boroughs, unitary authorities, shire counties and metropolitan boroughs). Councils covered seven of the nine Government Office Regions, with only the South East and Yorkshire and the Humber not represented. The main method used to administer the questionnaire was a self-completion postal questionnaire. Just over three per cent of questionnaires were conducted as face-to-face interviews and less than one per cent as telephone interviews. One council made use of an interpreter for four clients. Two councils used advocacy services: one for only four clients and the second, which only participated in the LD pilot, used the advocacy services to collect data for the entire sample of 80 respondents.

In total, 3,453 questionnaire were sent out and 1,364 were returned completed, an overall unit response rate of 40 per cent. The IC (2010b) reports that response rates varied across CASSRs, but that response rates were higher for those CASSRs that chased non-respondents, where the average response rate was 51 per cent compared to 31 per cent for those that did not chase non-respondents. We would therefore expect response rates to be higher in the national survey.

Data collected about both the respondents and non-respondents can be used to examine whether the respondent population differs systematically in any way from the sample population. A number of variables were collected for both respondents and non-respondents, including sex, age, ethnicity, sexual orientation, religion, client group (primary and secondary), and service receipt information. We have not analysed whether there are any differences in the characteristics of the respondent and non-respondent samples here, although from Table 23 in Appendix 2 it can be seen that the are some differences. Formal testing for the significance of observed differences would be valuable. However, as there were significant variations in the methods applied by CASSRs in the pilot – e.g. the pilot they participated in, the length of time allowed for those sampled to respond, and whether follow-up questionnaires were sent – and since these factors are also likely to affect response rates, any analysis would need to take the interaction between these CASSR-level variations and the individual-level variations into account. This is likely to be very difficult given the small number of CASSRs participating, so this dataset is not ideal for investigating the factors affecting response rates.

Councils were supplied with three different questionnaires: a standard questionnaire for those clients based in their own home without a learning disability, one for clients based in a residential care home and one for adults with a learning disability. For those clients based in residential care that have a learning disability the guidance was that the LD questionnaire should be used. Analysis conducted by the IC comparing the pilot element to the questionnaire sent confirms that councils used the questionnaires as expected: see Table 1, reproduced from the IC (2010b)).

Table 1: The type of questionnaire sent out by councils by pilot element conducted

	Pilot Element			
Questionnaire Sent	All Service Users	Residential Care Homes	Adults with Learning Difficulty	More than one element
Standard	865	0	0	170
Residential Care Home	480	650	0	315
Adults with Learning Difficulty	325	55	280	210
Total	1,670	700	280	695

Reproduced from IC (2010b) with permission.

However, a comparison between the data on service receipt and client group provided by councils and the questionnaire sent revealed a number of inconsistencies. For example, in Table 2, which shows the distribution of questionnaires sent by primary client group, it can be seen that 236 people identified on CASSR systems as having LD as their primary client group were not sent an LD questionnaire. Conversely, 215 people who are not recorded in the LD client group were sent an LD questionnaire. It is possible that some of these people who were sent an LD questionnaire but were not recorded as having LD as their primary client group may have been recorded as having LD as their secondary client group. However, as Table 3 shows only 101 people were recorded as having LD as a secondary client group and only 89 of those were sent an LD questionnaire.

Table 2: The type of questionnaire sent by primary client group

	Questionnaire sent				
Primary client group	Standard	Residential Care Home	Adults with Learning Difficulty	Missing	**Total**
Physical or sensory impairment	762	994	69	66	1,891
Mental health	94	238	139	28	499
Learning disability	60	176	652	14	902
Substance abuse	6	3	1	0	10
Vulnerable people	112	31	4	0	147
Missing	0	1	2	1	4
Total	1,034	1,443	867	109	3,453

Table 3: The type of questionnaire sent by secondary client group

Secondary client group	Questionnaire sent Standard	Residential Care Home	Adults with Learning Difficulty	Missing	Total
Physical or sensory impairment	89	107	107	0	303
Mental health	26	24	26	1	77
Learning disability	2	10	89	0	101
Substance abuse	2	1	2	0	5
Vulnerable people	10	13	8	0	31
Unknown	400	549	357	0	1,306
Missing	505	739	278	108	1,630
Total	1,034	1,443	867	109	3,453

Similar inconsistencies are observed when council records of service receipt are cross-tabulated with the type of questionnaire sent, as is shown in Table 4, although there is a lower degree of error. This table shows that 53 people in institutional care were not sent the residential care home questionnaire and 13 people not in institutional care were sent the care home questionnaire.

Table 4: The type of questionnaire sent according to whether the person is in institutional care

Institutional care category	Questionnaire sent Standard	Residential Care Home	Adults with Learning Difficulty	Missing	Total
In institutional care	53	1,430	312	0	1,795
Not in institutional care	863	13	555	0	1,431
Missing	118	0	0	109	227
Total	1,034	1,443	867	109	3,453

Characteristics of users in the sample

Table 5 show the characteristics of sample respondents. They are more likely to be female than male and the sample is dominated by older people, with roughly 60 per cent of respondents over the age of 65. Only around 25 per cent of respondents are under the age of 50. The respondent sample is predominately white. Information on sexual orientation and religion was completed poorly by councils. Sexual orientation was missing for nearly 99 per cent of the sample and religion for nearly 50 per cent. For respondents where religion was reported, the majority are identified as Christian (Table 23 in Appendix 2).

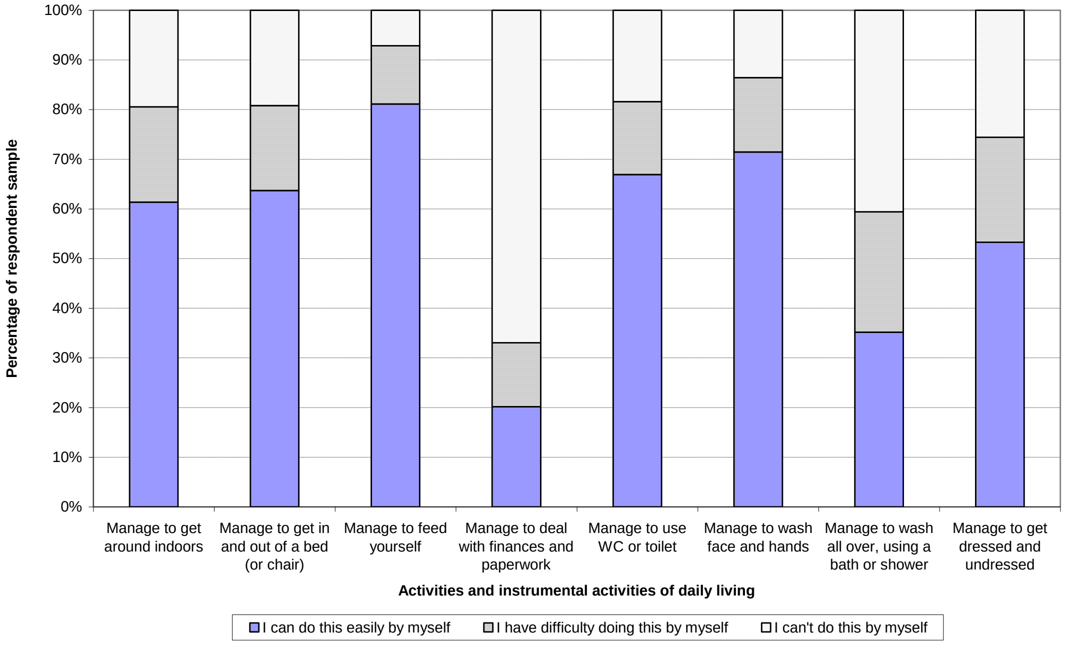
Table 5: Characteristics of the sample respondents

Characteristic	Frequency	Valid Percentage	Sample Percentage
Gender			
Male	503	38.0%	36.9%
Female	821	62.0%	60.2%
Valid total	1,324	100%	97.1%
Missing	40		2.9%
Total	1,364		100%
Age group			
18-24	51	3.9%	3.7%
25-30	51	3.9%	3.7%
31-39	89	6.7%	6.5%
40-49	118	8.9%	8.7%
50-64	197	14.9%	14.4%
65-74	139	10.5%	10.2%
75-84	256	19.4%	18.8%
85 and over	421	31.8%	30.9%
Valid total	1,322	100%	96.9%
Missing	42		3.1%
Total	1,364		100%
Ethnicity			
White	1,232	93.8%	91.0%
Mixed	3	0.2%	0.2%
Asian	29	2.2%	2.1%
Black	42	3.2%	3.1%
Other	7	0.5%	0.5%
Valid total	1,313	100%	97.0%
Missing	41		3.0%
Total	1,364		100%

Over 50 per cent of sample respondents have a physical and sensory impairment. The next largest client group is people with a learning disability (26 per cent), then people with mental health problems (15 per cent). Less than five per cent of respondents have substance misuse problems or are considered vulnerable people. Councils were not able to fully complete secondary client group information – around 50 per cent of sample cases had missing data. However, of those completed, physical and sensory impairment was the largest category for secondary client group (Table 23 in Appendix 2).

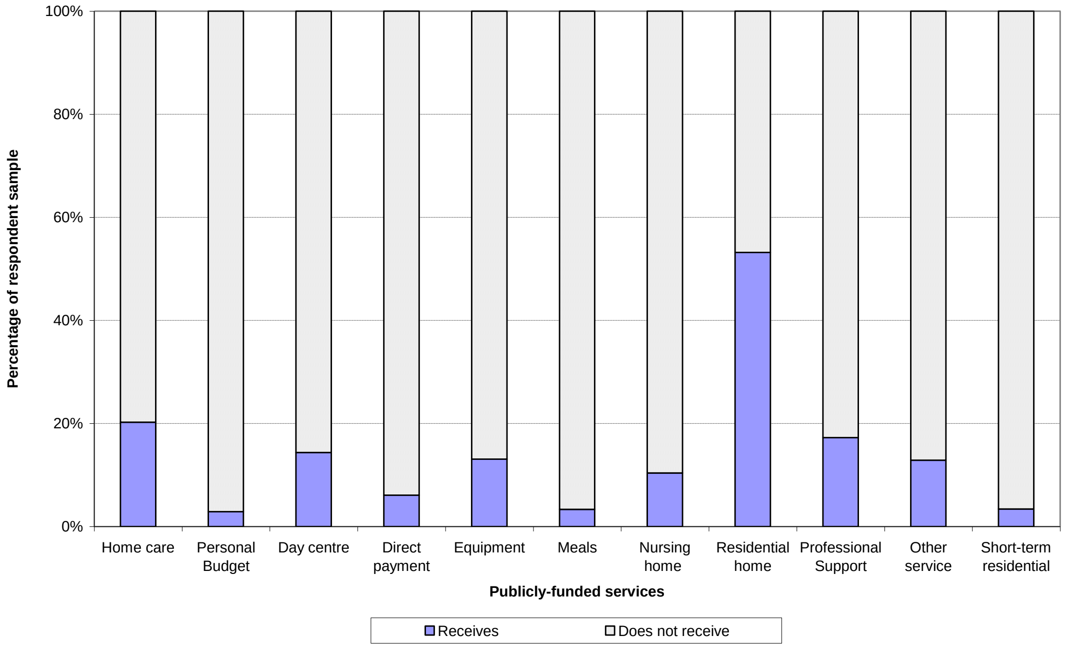
Need for help with activities of daily living (ADLs) and instrumental activities of daily living (IADLs) is a good indicator of need for social care services. Eight questions about ADLs and IADLs were asked in this questionnaire, and the distribution of the respondent sample's need for each of these ADLs is shown in Figure 1. There is variation in the type and degree of need across the sample. Whilst only 20 per cent of the sample report being able to manage their finances and paperwork easily on their own, over 80 per cent of the sample report being able to feed themselves on their own. A large proportion, around 65 per cent, report having difficulty with or not being able to wash all over in a bath or shower.

Figure 1: Distribution of need for help with activities of daily living and instrumental activities of daily living in respondent sample



CASSRs provided details of the publicly-funded social care services received by sample respondents (and non-respondents). The distribution of social care services amongst this population is shown in Figure 2. Residential care is the most used service amongst the respondent sample, with more than 50 per cent of the sample in residential care. The less used services amongst the respondent sample are personal budgets, meals and short-term residential care, which are only used by around three per cent of the respondent sample.

Figure 2: Distribution of receipt of publicly-funded services in respondent sample



Data on receipt of informal care and private help were also collected from the questionnaire. Roughly 20 per cent of the respondent sample report not receiving any regular practical help from family and friends. About a third of the sample reports receiving regular practical help from someone living inside the household and a similar percentage report receiving regular practical help from someone living outside the household. Less than ten per cent report receiving help from both someone living inside and outside the household. Compared to practical help, a much smaller proportion of the sample reported purchasing additional care or support privately, or paying more to top up their care and support. The majority of the sample (over 70 per cent) reported not purchasing additional care and support privately. Around 20 per cent reported buying additional care and support with their own money, and less than ten per cent reported that their family paid for additional care and support. A very small percentage (less than two per cent) reported both that their family bought extra care and support and they paid for extra out of their own money.

Table 6 shows the percentage of people receiving help to complete the questionnaire and the source of the help. A very large number of respondents to the sample, just over three-quarters, reported that they had help to fill in the questionnaire. Of the people reporting they had help, the largest proportion, at nearly 50 per cent, reported that they had help from someone living outside their household, but a substantial proportion at just over 35 per cent reported that they had help from a care worker.

Table 6: Percentage of people receiving help to complete the questionnaire and source of help

Source of help to complete questionnaire	Frequency	Valid percentage	Percentage
Responded by self	320	24.5%	23.5%
Had help from a care worker	360	27.6%	26.4%

Had help from someone living in my household	160	12.3%	11.7%
Had help from someone outside my household	466	35.7%	34.2%
Missing	58		4.3%
Total	1,364	100%	100%

The help given varied considerably. Approximately half the respondent sample reported having the questions read to them and a similar proportion reported that someone else wrote down the answers for them. Around 30 per cent report that someone translated the questions for them and another 30 per cent that they talked through the questions with someone else. A smaller number, at around eight per cent, reported that someone filled in the questionnaire for them without asking them the questions, i.e. the respondent was a proxy.

Distribution of ASCOT items

Item response rates to ASCOT items are all good at over 90 per cent. The dignity item has the lowest item response rate at 93 per cent; all other items have response rates at or over 95 per cent. The distribution of responses to the ASCOT items is shown in Table 7. For the items capturing the more basic SCRQoL domains (i.e. personal care, safety, accommodation cleanliness and comfort and food and nutrition) the majority of respondents report that they have no needs and that their needs are met in the 'ideal' way. However, for the SCRQoL domains capturing the less basic aspects of quality of life, such as social participation, occupation and control over daily life, a much larger proportion report having needs. The pattern of responses to the dignity domain is similar to the basic domains in that the majority report that the way they are helped and treated makes them think and feel better about themselves.

Examination of the relationship between the SCRQoL items and the type of questionnaire received revealed highly significant ($p<0.005$) associations between each of the items and questionnaire type. The exact pattern of association varied by item, although generally people responding to the standard questionnaire were more likely to report worse outcomes compared to those responding to the care home or LD questionnaire. For example, to the question about control over daily life, people responding to the LD questionnaire were more likely to report no needs and less likely to report needs; people responding to the care home questionnaire were more likely to report both the ideal state and the high-level needs state and less likely to report the mustn't grumble and low-level needs states; and people responding to the standard questionnaire were more likely to report low-level needs and less likely to report the ideal or mustn't grumble state. By contrast, to the social participation and involvement question, people responding to the LD questionnaire were more likely to report the mustn't grumble state and less likely to report the ideal or low-level needs state; people responding to the care home questionnaire were more likely to report the ideal state and less likely to report all of the other states; and people responding to the standard questionnaire were more likely to report needs (low and high) and less likely to report both of the no needs states.

Highly significant associations ($p<0.001$) were also observed between all the SCRQoL items and the question about the source of help to answer the questionnaire, except for the dignity item where the association was only significant at the 10 per cent level ($X^2(6)=11.38$, $p=0.077$). People who had help from a care worker consistently reported the best outcomes, being more likely to report the ideal state for all SCRQoL items and less likely than expected (or equally likely as expected) to report all other outcome states. For the other sources of help, the responses varied by question. For example, people who reported no help to answer the questionnaire were more likely to report the ideal state to the control over daily life question and less likely to report all others states; people with help from someone living inside the household were more likely to report needs (low and high) and less likely to report no needs; and people with help from someone living outside the household showed the same pattern as those with help from someone inside the household. For the accommodation item, people who reported not having any help were more likely to report the mustn't grumble state or low-level needs and less likely to report the ideal state; people who reported having help from someone inside the household were not more or less likely than expected to report any of the outcome states, whereas people with help from someone living outside the household were more likely to report the mustn't grumble state and less likely to report the ideal state.

Table 7: Response to the ASCOT items of social-care-related quality of life

Questions	**Frequency**	**Percent**
Could you tell me which of the following statements best describes how much control you have over your daily life?		
I have as much control over my daily life as I want	466	35.5
I have adequate control over my daily life	578	44.1
I have some control over my daily life but not enough	183	14.0
I have no control over my daily life	84	6.4
Total	1,311	100
Thinking about your personal care, by which we mean being clean and presentable in appearance, which of the following statements best describes your situation?		
I feel clean and am able to present myself the way I like	830	63.3
I feel adequately clean and presentable	429	32.7
I feel less than adequately clean or presentable	44	3.4
I don't feel at all clean or presentable	9	0.7
Total	1,312	100
Thinking about the food and drink you have, which of the following statements best describes your situation?		
I get all the food and drink I like when I want	844	64.8
I get adequate food and drink at OK times	393	30.2
I don't always get adequate or timely food and drink	55	4.2
I don't always get adequate or timely food and drink, and I think there is a risk to my health	10	0.8
Total	1,302	100
Could you tell me which of the following statements best describes how clean and comfortable your home is?		
My home is as clean and comfortable as I want	915	70.3
My home is adequately clean and comfortable	341	26.2
My home is not quite clean or comfortable enough	39	3.0
My home is not at all clean or comfortable	6	0.5
Total	1,301	100
Could you tell me which of the following statements best describes how safe you feel?		
I feel as safe as I want	936	71.8
Generally I feel adequately safe, but not as safe as I would like	304	23.3
I feel less than adequately safe	46	3.5
I don't feel at all safe	17	1.3
Total	1,303	100
Thinking about how much contact you've had with people you like, which of the following statements best describes your social situation?		

I have as much social contact as I want with people I like	647	49.8
I have adequate social contact with people	436	33.6
I have some social contact with people, but not enough	175	13.5
I have little social contact with people and feel socially isolated	41	3.2
Total	1,299	100
Could you tell me which of the following statements best describes how you spend your time?		
I'm able to spend my time as I want, doing things I value or enjoy	516	40.0
I'm able do enough of the things I value or enjoy with my time	475	36.8
I do some of the things I value or enjoy with my time but not enough	230	17.8
I don't do anything I value or enjoy with my time	70	5.4
Total	1,291	100
Thinking about the way you are helped and treated, and how that makes you think and feel about yourself, which of these statements best describes your situation?		
The way I'm helped and treated makes me think and feel better about myself	785	62.1
The way I'm helped and treated does not affect the way I think or feel about myself	370	29.2
The way I'm helped and treated sometimes undermines the way I think and feel about myself	96	7.6
The way I'm helped and treated completely undermines the way I think and feel about myself	14	1.1
Total	1,265	100

Development of the SCRQoL indicator

The eight items relating to SCRQoL had good response rates and were suitable for further testing to establish the feasibility of creating a multi-item measure (referred to in this report as the SCRQoL indicator). Table 8 shows the inter-item polychoric correlations for each of the SCRQoL items. The correlations are mostly moderate to large, between 0.3 and 0.7 (Cohen, 1988). Dignity has the lowest correlations with the other items, with most of the correlations around 0.3. The implication from this analysis is that the items are sufficiently strongly correlated to be considered for scaling.

Table 8: Matrix of polychoric correlations for SCRQoL items

	Control	P. care	Food	Accom.	Safety	Social	Occup.	Dignity
Control	1.0000							
Personal care	0.5722	1.0000						
Food	0.5209	0.5425	1.0000					
Accommodation	0.4132	0.5721	0.5630	1.0000				
Safety	0.3432	0.4004	0.3613	0.5345	1.0000			

Social	0.4559	0.4698	0.4079	0.5036	0.5439	1.0000		
Occupation	0.6274	0.5557	0.4952	0.4627	0.4834	0.5941	1.0000	
Dignity	0.3793	0.3584	0.3462	0.4030	0.3592	0.3921	0.4348	1.0000

Factor analysis

Tests for the suitability of the factor analysis procedure were all good. The Kaiser-Meyer-Olkin (KMO) test of sampling adequacy for this dataset was 0.87, which is considered ‘meritorious’. Bartlett’s test for sphericity rejected the null hypothesis that the variables are not inter-correlated ($X^2(28)=2529.482$, p=<0.001). Maximum likelihood (ML) factoring extraction was performed through STATA factormat on the polychoric correlation matrix of the eight SCRQoL items.[11] One factor was extracted with an Eigenvalue 3.78. The likelihood ratio test of independence against the saturated model is significant ($X^2(28)=4037.31$, p=<0.001) indicating that the factor analysis is meaningful and the items are inter-correlated.

The one-factor model is shown inTable 9. All the variables load onto the factor with a loading greater than 0.4. The factor explains the majority of the variance of most items; only the safety and dignity items have a unique variance greater than 0.6, which indicates that the factor does not explain these variables very well. Cronbach’s alpha for this scale is 0.81, which is considered to be very good and it is not increased by dropping any items from the scale. The strong loadings of all the items and the positive results from Bartlett’s test of sphericity and the KMO suggest that the items are similar enough to be summed together into a SCRQoL scale.

Table 9: Pattern matrix for the one-factor solution

Variable	Factor	Uniqueness
Control	0.7078	0.4990
Personal care	0.7385	0.4547
Food	0.6801	0.5374
Accommodation	0.7097	0.4964
Safety	0.6205	0.6150
Social	0.7017	0.5076
Occupation	0.7756	0.3985

---

[11] The maximum likelihood factoring extraction method assumes that the items are multivariate normal, an assumption which is not met with these data: Mardia’s test for skewness = 14.599, X^2 (120) =2899.529, p < 0.001; Mardia’s test for kurtosis = 113.060, X^2 (1) =2028.757, p < 0.001; Henze-Zirkler = 27.193, X^2 (1) = 50787.323, p < 0.001; Doornik-Hansen X^2 (16)= 3115.814, p < 0.001. We therefore repeated the analysis using principal axis factoring which is recommended when the assumption of multivariate normality is violated, but the same solution was found (Fabrigar et al., 1999).

Dignity	0.5407	0.7076

Items with unique variance > 0.6 shown in bold.

Analysis to examine whether there is clustering of responses by CASSR for the SCRQoL items reveals that for most of the items very little of the variance is attributable to systematic variation at the CASSR level as shown in Table 10. However, given the small number of CASSRs it is possible that the VPCs are underestimated. Assuming that the VPCs are all underestimated to the same extent, three of the items appear to have a larger proportion of their variance attributable to the CASSR. The accommodation cleanliness and comfort and occupation items have VPCs of 5.0 per cent and 5.9 per cent respectively. The safety item has a larger VPC at 8.8 per cent. Given that all of the items have low VPCs at less than 10 per cent, even if the VPCs are underestimated it seems unlikely that a multilevel factor analysis would alter the results. In addition, as we discuss below, there is an argument that these results are affected by differences in the composition of the samples for some of the CASSRs, which is likely to have the effect of overestimating the VPC.

Table 10: Variance Partition Coefficient (VPC) for each SCRQoL item

Item	**VPC**
Control over daily life	2.2%
Personal care	2.1%
Food and drink	1.4%
Accommodation cleanliness and comfort	5.0%
Safety	8.8%
Social participation	2.2%
Occupation	5.9%
Dignity	3.0%

Mokken scaling analysis

The results of the procedure to investigate the dimensionality of the SCRQoL items are shown in Table 11 using lowerbounds from 0.3 to 0.6 with steps of 0.1.[12] At 0.3 all items form one scale, but as the lowerbound is increased items drop out of the scale. Dignity and safety drop out first as the lowerbound is increased to 0.4. This fits with the findings from the factor analysis where safety and dignity were identified as the two items with the most unique variance. All of the items drop out of the scale at 0.6. Although safety forms a secondary scale with social participation and involvement, we do not interpret this finding as indicating that

---

[12] All cases with missing data for at least one of the SCRQOL items are excluded from the analysis, leaving a sample of size 1,188.

the items form multi-dimensional scales. Conceptually the items seem distinct and whilst it is possible that people who feel safer may feel safer because they have contact with others and vice-versa, previous analysis with other samples has not found these two items to form a separate subscale (Netten et al, 2009) . For this reason we interpret the secondary scale as a quirk of the sample. Overall this analysis seems to support the factor analysis where one factor was extracted.

Table 11: Scales determined using Mokken item selection procedure (N=1,188)

c	Scale 1	Scale 2	Excluded
0.3	1, 2, 3, 4, 5, 6, 7, 8		
0.4	1, 2, 3, 4, 6, 7		5, 8
0.5	1, 2, 7	5, 6	3, 4, 8
0.6			1, 2, 3, 4, 5, 6, 7, 8

1 Control over daily life, 2 Personal care, 3 Food and drink, 4 Accommodation cleanliness and comfort, 5 Safety, 6 Social participation, 7 Occupation, 8 Dignity

Analysis of the fit of the MH model to the SCRQoL items shows that there are no significant violations of the model.[13] Loevinger's homogeneity (H) coefficient, which is a proportional reduction in error statistic, forms the basis of the assessment of scalability. It is a summary measure of the goodness of fit of the Mokken model to the observed data and varies between zero and one, with a value of one meaning perfect fit and a value of zero indicating no improvement over the null model. The H coefficient for the scale is 0.43, which indicates that the items form a moderate scale and the respondents can be ordered by means of the set of items.

The scalability parameters and statistics are summarised in Table 12. The item H-coefficient gives an indication of the power of the items to discriminate between persons. The H for the all the items, except safety and dignity, is greater than 0.4, which is taken to indicate that the items discriminate well between individuals. The safety and dignity items discriminate adequately enough between people to be considered for the scale since each has an item H coefficient greater than 0.3.

Table 12: Scalability parameters and item characteristics (N=1,188)

	Mean	H	No needs		Low Needs	High needs
Items			**Ideal**	**Mustn't grumble**		
Control over daily life	1.90	0.46	36.11%	44.02%	14.06%	5.8%
Personal care	1.41	0.44	63.55%	32.32%	3.54%	0.6%

---

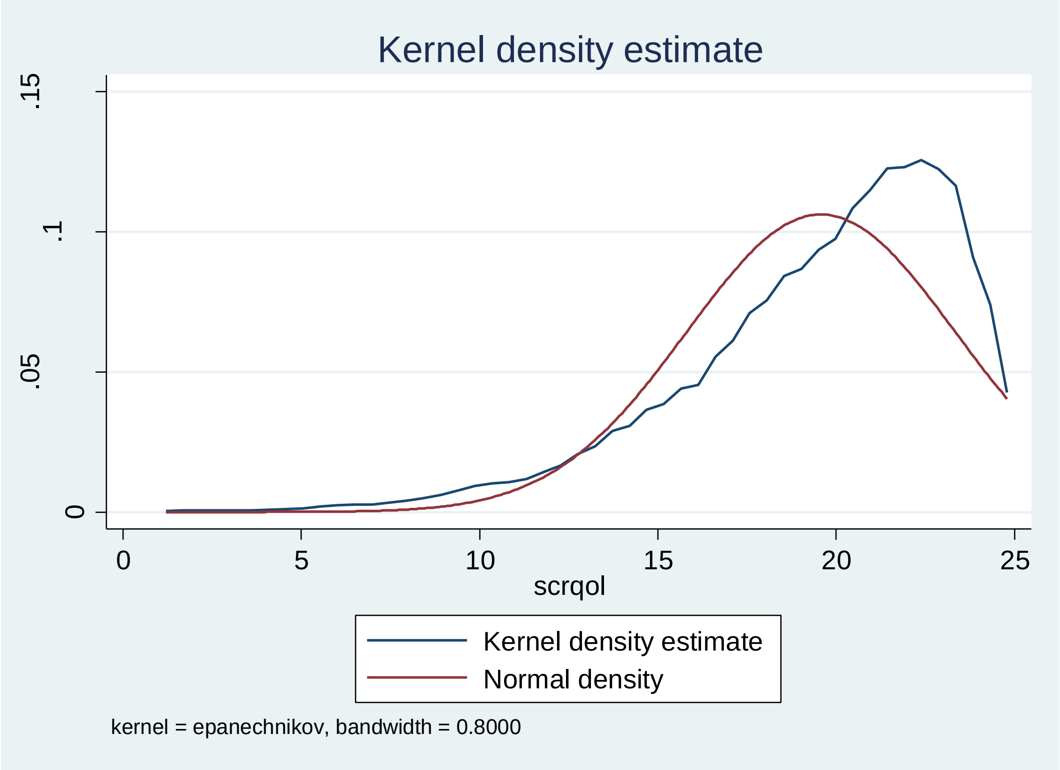
[13] Tests were conducted with groups of size 118, minimum violation of 0.03 and 5% significance level. Altering these parameters did not substantially change the conclusions.

Food and drink	1.41	0.41	65.07%	29.71%	4.38%	0.8%
Accommodation cleanliness and comfort	1.34	0.43	70.37%	26.26%	2.86%	0.5%
Safety	1.34	0.38	72.47%	22.90%	3.28%	1.3%
Social participation	1.69	0.45	50.25%	33.50%	13.05%	3.2%
Occupation	1.88	0.50	40.15%	37.04%	17.59%	5.2%
Dignity	1.47	0.35	62.63%	28.96%	7.32%	1.1%

Distribution of the SCRQoL indicator

The results for the eight items were combined by following a scoring calculation (shown in Appendix 3). The distribution of the SCRQoL indicator in this sample is shown in Figure 3. The mean score for the SCRQoL indicator is 19.57 (SD=3.76, n=1,188), although the scale is significantly skewed (skew=-1.12, p<0.001) and leptokurtic (kurtosis=4.30, p<0.001). The maximum observed score for the scale is 24 and the minimum is two.

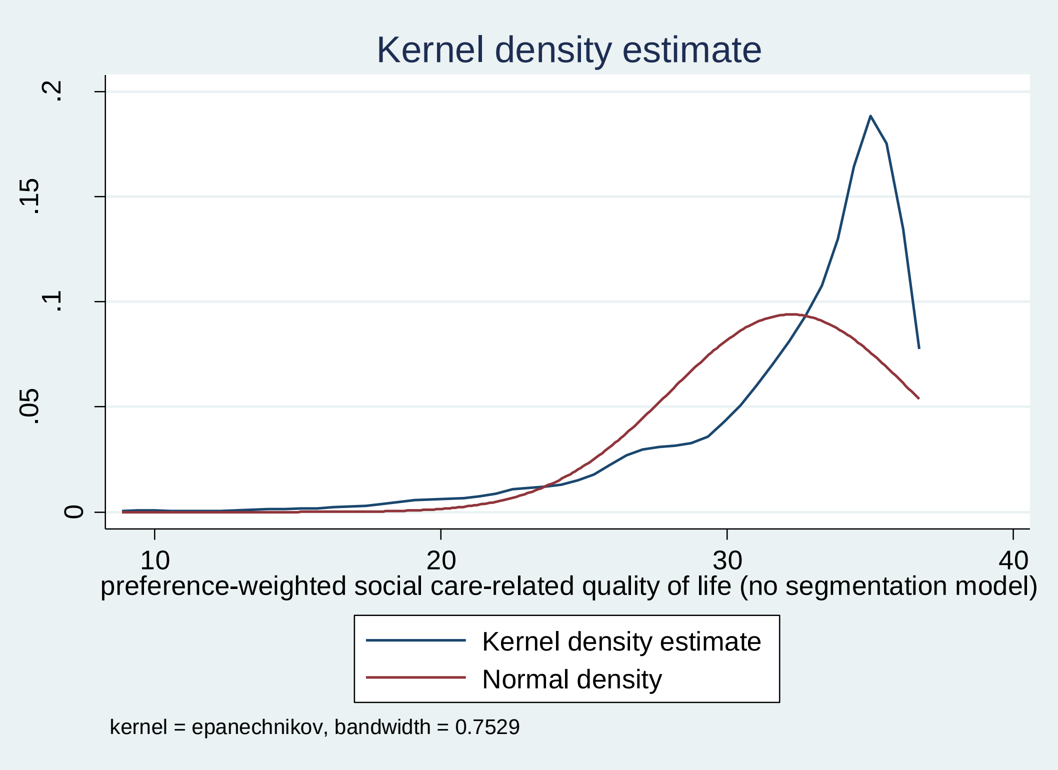
Figure 3: Distribution of the SCRQoL indicator



Development of the preference-weighted SCRQoL indicator (PW-SCRQoL)

To calculate the preference-weighted indicator we have used the illustrative preference weights from a pilot study (Burge et al., 2010). The calculation is set out in Appendix 4. The distribution of the preference-weighted SCRQoL scale in this sample is shown in Figure 4. The mean score for the SCRQoL scale is 32.24 (SD=4.24, n=1,188), although the scale is significantly skewed (skew=-1.81, p<0.001) and leptokurtic (kurtosis=6.80, P<0.001). The maximum observed score for the scale is 35.97 and the minimum is 9.61.

Figure 4: Distribution of the PW-SCRQoL indicator



Development of the SCRQoL needs indicator

We have calculated four different versions of a SCRQoL needs indicator, which are all counts of the number of SCRQoL domains in which the individual has a need. The calculation for each of the four versions is set out in Appendix 5. The distributional characteristics of the four versions are shown in Table 13.

Table 13: Distributional characteristics of the SCRQoL needs measures

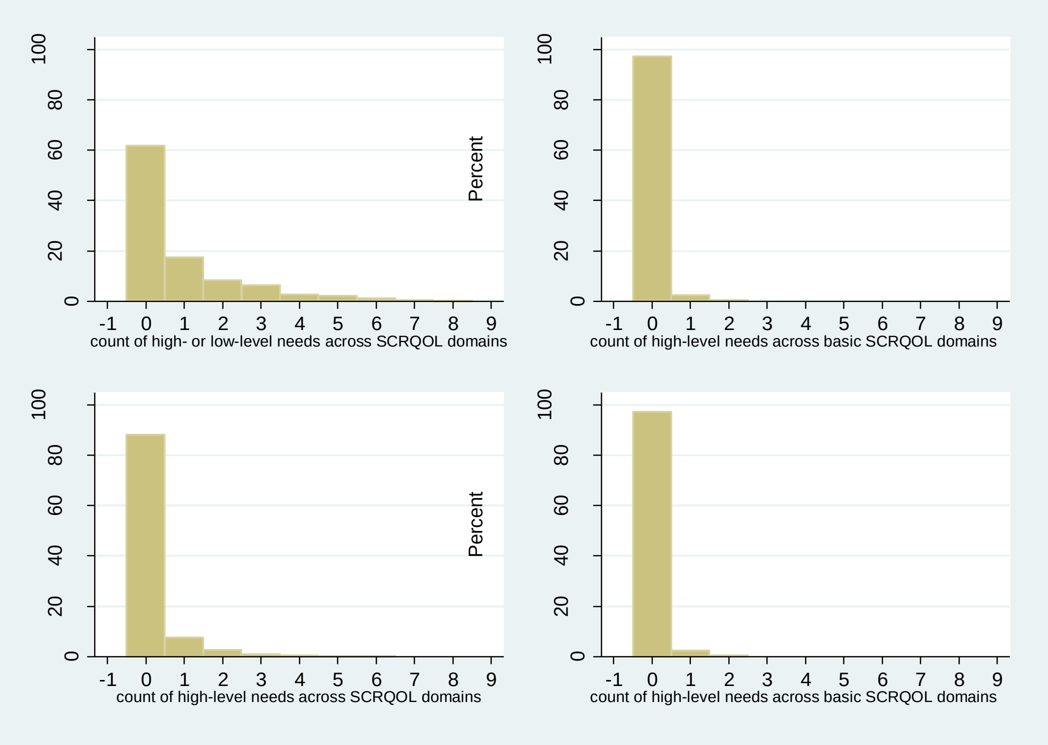
	Mean (VAR)	**Max**	**Min**	**Skew**	**Kurtosis**
Count of number of domains with high- or low-level need (n=1,188)	0.85 (2.01)	8	0	2.04***	7.16***
Count of number of basic domains with high- or low-level need (n=1,272)	0.17 (0.29)	4	0	3.97***	21.31***
Count of number of domains with high-level need (n=1,188)	0.19 (0.38)	6	0	4.65***	30.58***
Count of number of basic domains with high-level need (n=1,272)	0.03 (0.04)	2	0	6.80***	53.32***

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Compared with the SCRQoL and PW-SCRQoL indicators, the distribution of these indicators is much more highly skewed. From Table 13 it can be seen that the mean (interpreted as the average number of domains with needs per person) and variance for all of the variables are relatively similar, with the variance being around 1.3 to 2.4 times larger than the mean. This indicates that the Poisson distribution is likely to be a good fit to these data, although there is some evidence of overdispersion, since in all cases the variance is larger than the mean. As identified above, a Poisson distribution can be approximated to a normal distribution when λ is large, usually considered to be greater than 10, which is not the case with these data. Confidence intervals based on standard methods for normally-distributed data are therefore unlikely to be appropriate; neither are methods for the analysis of normally-distributed data. We therefore do not consider these indicators any further.

To convert the data to proportions, it is necessary to choose a cut point for the number of needs, i.e. the proportion with at least one need in any domain. For a measure with good distributional properties, this cut point should be determined by the point at which there is maximal variability, i.e. where p=0.5. The distribution of each variable is shown graphically in Figure 5.

Figure 5: Distributional characteristics of the four SCRQoL needs measures



As can be seen from the charts, the most numerous response for each of the variables is zero. Only the measure which captures the number of high- or low-level needs across all the SCRQoL domains has a good amount of variability. This measure has close to 40 per cent of people reporting a high- or low-level need in at least one SCRQoL domain. The other three measures are very severely skewed. The measure of the number of high-level needs in any of the basic domains has only three per cent of people reporting at least one need, which is quite a rare event. The other two measures (of high-level needs in all the SCRQoL domains and high- or low-level needs in the basic SCRQoL domains) have roughly 12 per cent of people reporting at least one need. We recommend that the cut point is chosen as at least one need for all of the measures (see Appendix 5 for scoring).

Although the events are not truly rare by epidemiological standards, they are uncommon for most of the measures, and therefore do not have much variability. Table 14 shows the sample mean and variance for each of the SCRQoL need measures, confirming that only the measure of the proportion of respondents with low- or high-level needs in at least one SCRQoL domain has good variability.

Table 14: Distributional characteristics of the SCRQoL needs measures, defined as proportion with at least one SCRQoL 'need' (with need defined differently for each measure)

	Mean	**Variance**
Proportion of respondents with a low- or high-level need in at least one SCRQoL domain (n=1,188)	0.38	0.24
Proportion of respondents with a low- or high-level need in at least one basic SCRQoL domain (n=1,272)	0.12	0.11
Proportion of respondents with a high-level need in at least one SCRQoL domain (n=1,188)	0.12	0.10
Proportion of respondents with a high-level need in at least one basic SCRQoL domain (n=1,272)	0.03	0.03

The distribution of these variables is likely to be problematic for the analyses that examine differences across CASSRs, i.e. the reliability and sensitivity analysis. The pilot only required CASSRs to sample 100 people, and some CASSRs returned very small sample sizes; consequently, the expected number of respondents likely to report at least one need is likely to be very low for some CASSRs. This is demonstrated in Table 15 and Table 16, which show the expected number of people likely to report at least one need based on the population mean for each indicator and the size of their sample. For the measure based on the proportion of people reporting at least one high-level need in a basic SCRQoL domain, the expected number is just one for six of the CASSRs. For the measures based on the proportion of people with a low- or high-level need in at least one of the basic domains and on the proportion of people with at least one high-level in one of the SCRQoL domains, it is common for the expected number of observations with a need to be fewer than ten – only five CASSRs have an expected number of ten or more. When the probability of an event is low and the sample size for the CASSR is small, there is the possibility that variation within CASSRs is due to a quirk of the sample rather than real variation. It is therefore important to be cautious when interpreting the results that examine differences between CASSRs, i.e. the results of the reliability and sensitivity analyses.

Table 15: Expected number of people likely to report at least one need in any of the SCRQoL domains, under two definitions of need

	Sample respondents	Expected number of respondents (0 d.p.)	
CASSR	All SCRQoL domains	Low- or high-level need	High-level need only
A	67	26	8
B	23	9	3
C	57	22	7
D	84	32	10
E	29	11	3
F	28	11	3
G	71	27	8
H	82	31	10
I	69	26	8
J	127	49	15
K	61	23	7
L	55	21	6
M	37	14	4
N	129	50	15
O	34	13	4
P	44	17	5
Q	124	48	15
R	67	26	8

Table 16: Expected number of people likely to report at least one need in the basic SCRQoL domains, based on two definitions of need

	Sample respondents	Expected number of respondents (0 d.p.)	
CASSR	Basic SCRQoL domains	Low- or high-level need	High-level need only
A	69	8	2
B	24	3	1
C	62	7	2
D	89	11	3
E	30	4	1

F	32	4	1
G	73	9	2
H	85	10	2
I	74	9	2
J	140	17	4
K	65	8	2
L	60	7	2
M	38	5	1
N	139	17	4
O	35	4	1
P	48	6	1
Q	134	16	4
R	75	9	2

Validity

The potential indicators are all designed to be outcomes-focused measures. We would therefore expect them to be correlated with the general quality of life item and the overall satisfaction with care and support item. Both of these items have seven response options for the standard and care home questionnaires and only five response options for the LD questionnaire. We have therefore explored the correlations separately for the LD questionnaire group. Because the dichotomous variables do not come from a normal distribution, violating the assumptions underlying the polychoric correlation, we also report the non-parametric statistic Kendall's tau-b. The correlations[14] and associations between the indicators and the quality of life and satisfaction items are shown in Table 17.

Table 17: Correlation between indicators and measures of quality of life and satisfaction with care and support

Indicator	**Satisfaction (n)**		**Overall quality of life (n)**	
	Corr	**Tau-b**	**Corr**	**Tau-b**

---

[14] This analysis was conducted using the Pearson correlation coefficient but was repeated using polyserial correlation coefficients to reflect the categorisation of the satisfaction and overall quality of life variables. Rho calculated via the polyserial correlation was very similar to the Pearson correlation coefficient, so we have reported only the Pearson correlation coefficient here. The correlation between the binary indicator of need and the two variables was calculated using a polychoric correlation.

Standard and care home	SCRQoL	-0.563***	-0.434***	-0.672***	-0.535***
		(838)	(838)	(846)	(838)
	PW-SCRQoL	-0.545***	-0.407***	-0.663***	-0.512***
		(838)	(838)	(846)	(838)
	Proportion of respondents with low- or high-level need in at least one SCRQoL domain	0.513***	0.358***	0.686***	0.485***
		(838)	(838)	(846)	(846)
	Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	0.567***	0.314***	0.560***	0.312***
		(891)	(891)	(903)	(903)
	Proportion of respondents with high-level need in at least one SCRQoL domain	0.448***	0.241***	0.694***	0.371***
		(838)	(838)	(846)	(846)
	Proportion of respondents with high-level need in at least one basic SCRQoL domain	0.598***	0.200***	0.666***	0.224***
		(891)	(891)	(903)	(903)
Learning disability	SCRQoL	-0.442***	-0.293***	-0.545***	-0.413***
		(300)	(300)	(304)	(304)
	PW-SCRQoL	-0.468***	-0.282***	-0.534***	-0.378***
		(300)	(300)	(304)	(304)
	Proportion of respondents with low- or high-level need in at least one SCRQoL domain	0.397***	0.232***	0.461***	0.300***
		(300)	(300)	(304)	(304)
	Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	0.572***	0.274***	0.508***	0.225***
		(321)	(321)	(324)	(324)
	Proportion of respondents with high-level need in at least one SCRQoL domain	0.436***	0.181***	0.498***	0.199***
		(300)	(300)	(304)	(304)
	Proportion of respondents with high-level need in at least one basic SCRQoL domain	0.435	0.059	0.701***	0.118***
		(321)	(321)	(324)	(324)

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

The SCRQoL and the PW-SCRQoL measures have strong correlations, generally greater than 0.5, with both of the satisfaction and quality of life measures (Cohen, 1988). The correlations are negative as a low score on both the SCRQoL and PW-SCRQoL indicators is associated with poor quality of life; on the satisfaction and overall quality of life items a low score is associated with high satisfaction and good quality of life. For both measures the correlation with the overall quality of life measure is stronger than the satisfaction measure. This is to be expected since conceptually the two measures are closer to overall quality of life than satisfaction with the service. This lends some support to the validity of these measures as outcomes-focused.

In general, the correlations between all the indicators are stronger amongst those who answered the standard or care home questionnaire, compared to those who answered the LD questionnaire. This is also true for tau-b, where the value of tau-b is always higher for those answering the standard or care home questionnaire. The differences may be due to differences in the wording of the ASCOT items or due to the fact that both the satisfaction and overall quality of life items have fewer categories in the LD questionnaire.

I have not compared the correlations for the SCRQoL needs measures to the correlations for the SCRQoL and PW-SCRQoL measures as the correlations are calculated using different methods. It is more instructive to compare tau-b scores. For these, the SCRQoL measure has consistently the highest tau-b, the PW-SCRQoL measure the second highest tau-b, and the unmet need measure capturing the proportion of respondents with high-level needs in at least one basic domain has consistently the worst tau-b. In general, the SCRQoL need measure capturing the proportion of respondents with low- or high-level needs in at least SCRQoL domain performs the best out of all the SCRQoL need measures.

Respondents completed different questionnaires and also a large number of respondents had help from different sources (care worker, person within household and outside of their household) to complete the questionnaire. Given the associations already identified between the ASCOT items and the type of questionnaire and source of help, it seems likely that these two variables may be associated with the scores for the indicators. Here we explore whether there are any significant differences in the indicators across these subgroups of the population. The variations in the mean for all the indicators across these two subgroups are shown separately in Table 18 and Table 19.

Table 18: Mean (S.D.) for indicators by type of questionnaire received

Indicator	Standard	Residential Care Home	Adults with Learning Difficulty
SCRQoL	17.83	19.99	20.53
	(4.24)	(3.53)	(2.99)
PW-SCRQoL	30.18	32.76	33.34
	(5.03)	(3.90)	(3.13)
Proportion of respondents with low- or high-level need in at least one SCRQoL domain	0.61	0.31	0.29
	(0.49)	(0.46)	(0.46)

Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	0.23	0.09	0.07
	(0.42)	(0.28)	(0.26)
Proportion of respondents with high-level need in at least one SCRQoL domain	0.18	0.12	0.06
	(0.39)	(0.32)	(0.24)
Proportion of respondents with high-level need in at least one basic SCRQoL domain	0.07	0.02	0.01
	(0.26)	(0.13)	(0.08)

Across all of the indicators, people responding to the standard questionnaire have the worst scores on the indicators and the people completing the LD questionnaire the best scores. The differences in the means are significant for the SCRQoL indicator (F(2,1149)=51.25, p<0.001, n=1,152)[15] and for the PW-SCRQoL indicator (F(2,1149)=56.31, p<0.001, n=1,152).[16] The differences in the indicators across the SCRQoL needs measures were examined using a chi-squared test.[17] A highly significant association (p<0.001) was observed between the questionnaire type and each of the SCRQoL need measures, indicating that the indicators vary by questionnaire type.

Table 19: Mean (S.D.) for indicators by source of help to complete the questionnaire

Indicator	By self, no help	Help from care worker	Help from someone inside household	Help from someone outside household
SCRQoL	19.80	20.71	19.30	18.52
	(3.89)	(3.02)	(3.71)	(3.96)
PW-SCRQoL	32.46	33.50	31.86	31.14
	(4.30)	(3.12)	(4.35)	(4.70)
Proportion of respondents with low- or high-level need in at least	0.34	0.25	0.44	0.51

---

[15] Bartlett's test for equality of variance rejects the null hypothesis that variances across the groups are equal ($X^2(2)$=38.27, p=<0.001). Simulation results confirmed that the ANOVA test is too liberal but demonstrated that this could be resolve using the W test and F star test. Both tests found that the differences in means were highly significant (Fstar(3, 348.26) = 37.64, p<0.001, and WStat(3, 162.41) = 30.60, p<0.001).

[16] Bartlett's test for equality of variance rejects the null hypothesis that variances across the groups are equal ($X^2(2)$=70.37, p=<0.001). Simulation results confirmed that the ANOVA test is too liberal. The problem was not resolved by using the W test or F star test, since both tests were also too liberal. However, the difference between the simulated p-value and the nominal p-value was not large and, given how small the observed p-value is, we have confidence in the result that there are differences between subgroups.

[17] Logistic regressions were also used to examine whether the mean varied across subgroups. These results agreed with the results of the chi-squared analysis.

one SCRQoL domain	(0.47)	(0.43)	(0.50)	(0.50)
Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	0.15	0.05	0.13	0.15
	(0.36)	(0.23)	(0.34)	(0.35)
Proportion of respondents with high-level need in at least one SCRQoL domain	0.07	0.05	0.15	0.20
	(0.25)	(0.21)	(0.36)	(0.40)
Proportion of respondents with high-level need in at least one basic SCRQoL domain	0.02	0.01	0.05	0.04
	(0.14)	(0.11)	(0.22)	(0.20)

Across all of the indicators, the best scores on the indicators are always recorded by those people who had help from a care worker. In general, those who do not report having any help report the second-best scores, those who have help from someone inside the household the third-best scores and those who have help from someone outside the household the worst scores. However, this pattern does vary for the two indicators that capture needs in the basic SCRQoL domains. The ANOVA test for differences between the means of the groups is highly significant for the SCRQoL indicator (F(3,1163)=22.73, p<0.001, n=1,166)[18] and for the PW-SCRQoL indicator (F(3,1163)=20.54, p<0.001, n=1,166).[19] Again, differences in the SCRQoL needs indicators by the source of help were explored using a chi-squared test. Highly significant associations (p<0.001) were observed between the SCRQoL need indicators and the source of help to complete the questionnaire, except for the SCRQoL need indicator capturing the proportion of respondents who report at least one high-level need in a basic domain. A significant association was still found between this indicator and the source of help, but the association was only significant at the five per cent level (X^2(3)=9.83, p=0.02).

Reliability

The decomposition of the variance into its between- and within-CASSR components is summarised in the variance partition coefficient (VPC) which is an estimate of the proportion of the variance attributable to the CASSR. The VPC for each of the indicators is shown in Table 20.

Table 20: Variance partition coefficient for each indicator

---

[18] Bartlett's test for equality of variance rejects the null hypothesis that variances across the groups are equal (X^2(3)=29.83, p=<0.001). Simulation results confirmed that the ANOVA test is too liberal but demonstrated that this could be resolve using the W test and F star test. Both tests found that the differences in means were highly significant (Fstar(4, 311.73) = 17.72, p<0.001, and WStat(4, 136.95) = 18.77, p<0.001).

[19] Bartlett's test for equality of variance rejects the null hypothesis that variances across the groups are equal (X^2(3)=60.37, p=<0.001). Simulation results confirmed that the ANOVA test is too liberal. The problem was not resolved by using the W test or F star test, since both tests were also too liberal. However, the difference between the simulated p-value and the nominal p-value was not large and given how small the observed p-value is we have confidence in the result that there are differences between subgroups.

Indicator	VPC
SCRQoL	7.33%
PW-SCRQoL	8.40%
Proportion of respondents with low- or high-level need in at least one SCRQoL domain	5.65%
Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	9.41%
Proportion of respondents with high-level need in at least one SCRQoL domain	2.81%
Proportion of respondents with high-level need in at least one basic SCRQoL domain	19.73%

The VPC is not large for any of the measures. The majority of the variation in the indicator is explained by rater variance, which is also confounded with random error. However, the VPCs reported here should be treated with some caution given the small number of CASSRs and the violation of the assumption of normally-distributed level two errors, which both tend to lead to underestimations of the variance components. However, working in the opposite direction is the effect of the differences in the composition of the CASSR samples, which would tend to produce overestimations of the VPC. Assuming that the degree of underestimation is equal for all indicators, which is not possible to verify, two stand out: the proportion of respondents with high-level need in at least one SCRQoL domain, which has a much lower VPC than the other indicators and the proportion of respondents with a high-level need in at least one basic domain, which has a much higher VPC than the other indicators at close to 20 per cent.

Figure 6 shows how the reliability of each of the indicators increases with the number of raters (service users). The chart can be used to provide an indication of how large the valid sample needs to be to achieve a reliable estimate for each indicator. Eighty per cent is generally considered to be the level of reliability required for a widely-used measure. The number of respondents required to meet this level of reliability is similar for most of the indicators at around 50 respondents, although the PW-SCRQoL indicator is marginally better than the other measures. Two of the measures stand out from this general picture. The SCRQoL needs measure of the proportion with at least one high-level need in one of the basic domains achieves a reliability of 80 per cent at roughly 20 respondents, which is a very small sample size. In contrast, the SCRQoL needs measure of the proportion with at least one high-level need in any of the SCRQoL domains does not meet the 80 per cent level of reliability until there are around 150 respondents. However, these results should be treated with some caution given the problems in estimating the variance components on which these calculations are based.

Figure 6: Reliability of indicators as a function of achieved sample size

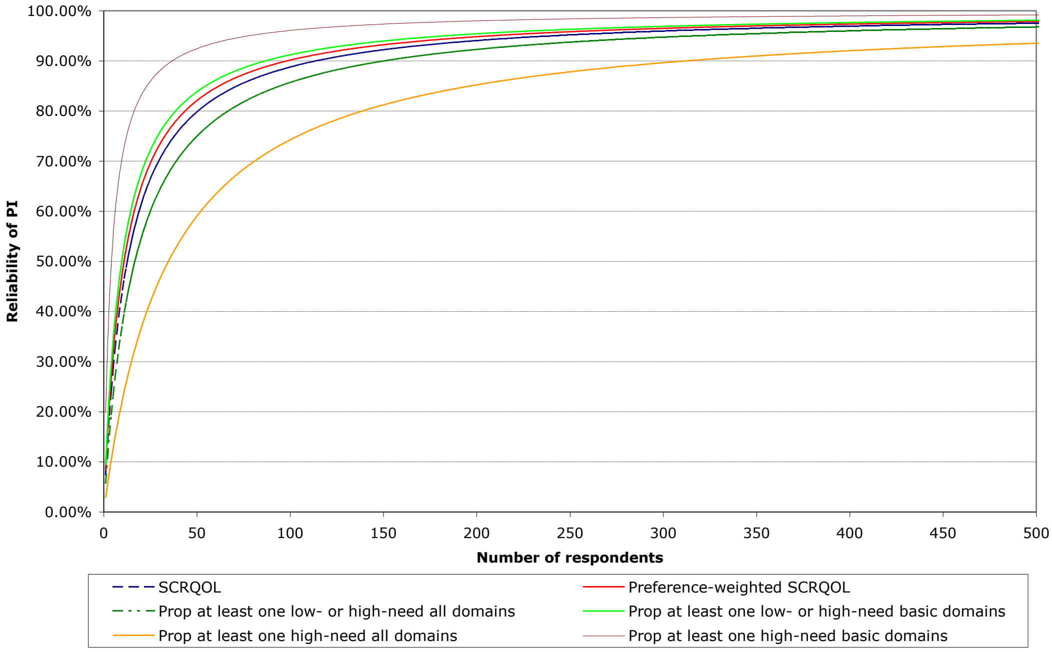


Table 21 shows the required attained sample size using the criterion that the 95 per cent confidence interval for the mean within each CASSR should have a width no greater than eight per cent of the scale (or plus or minus four per cent around the mean). According to this formula, the largest sample size is required for SCRQoL need measure of the proportion of respondents with low- or high-level need in at least one SCRQoL domain and the smallest for the PW-SCRQoL measure. For the proportion measures, we observe that the smaller the mean (and consequently variance), the smaller the required sample size.

Table 21: Required attained sample size using Information Centre criterion

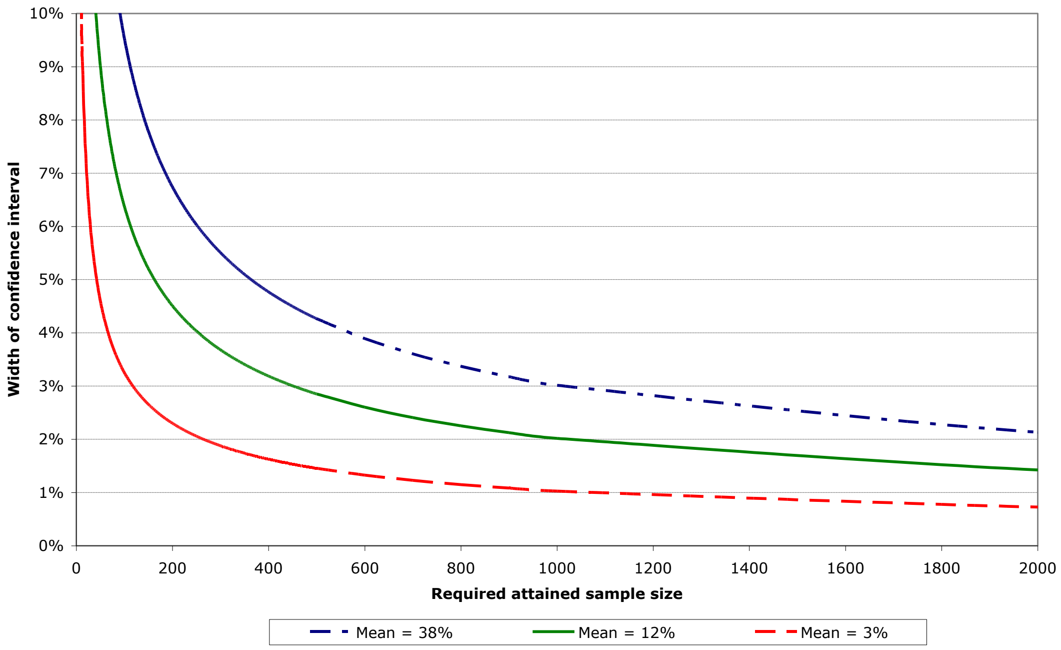
Indicator	Mean	Variance	+/- 4% of scale length (scale units)	Required N	% of mean
SCRQoL	19.57	14.13	0.96	59	9.8
PW-SCRQoL	32.24	18.01	1.11	56	6.9
Proportion of respondents with low- or high-level need in at least one SCRQoL domain	0.38	0.24	0.04	568	20.8
Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	0.12	0.11	0.04	254	66.5
Proportion of respondents with high-level need in at least one SCRQoL domain	0.12	0.10	0.04	250	67.9
Proportion of respondents with high-level need in at least one	0.03	0.03	0.04	66	282.7

basic SCRQoL domain

The results for the SCRQoL needs measures seem counter-intuitive since one would expect to need a very large sample size to detect a rare event accurately. Indeed, there is a problem with applying this formula to rare events, since as the mean (*p*) decreases the size of the confidence interval in proportion to the mean becomes much larger. For the proportion of respondents with high-level need in at least one basic SCRQoL domain, the confidence interval is in fact 280 per cent of the size of the mean compared to only seven per cent for the preference-weighted SCRQoL measure. Choosing a required confidence interval width of eight per cent when the real population mean is only three per cent means that the estimate of the population mean from a sample of the required size could be over twice the size of the real population mean. This is clearly not a very good estimate for the population mean, and it may make more sense for measures of rare or less frequent events to choose a confidence interval that is proportionate to the mean.

The chart in Figure 7 shows how the confidence interval width varies according to sample size for the means observed for these indicators. This shows that for the proportion of respondents with a high-level need in at least one basic SCRQoL domain, in order to achieve a confidence interval of width roughly equivalent to the mean i.e. 3 per cent, just over 100 respondents would be required; to achieve a confidence interval that is half the width of the mean, around 470 respondents are required; and to achieve a confidence interval that is about ten per cent of the size of the mean (similar to the SCRQoL and PW-SCRQoL measures) well over 10,000 respondents would be required. A sample size of greater than 2,000 is required to achieve a confidence interval that is about ten per cent the size of the mean for the indicators with means of 0.12.

Figure 7: Width of the confidence interval as a function of required attained sample size for various values of *p* (mean)



Sensitivity

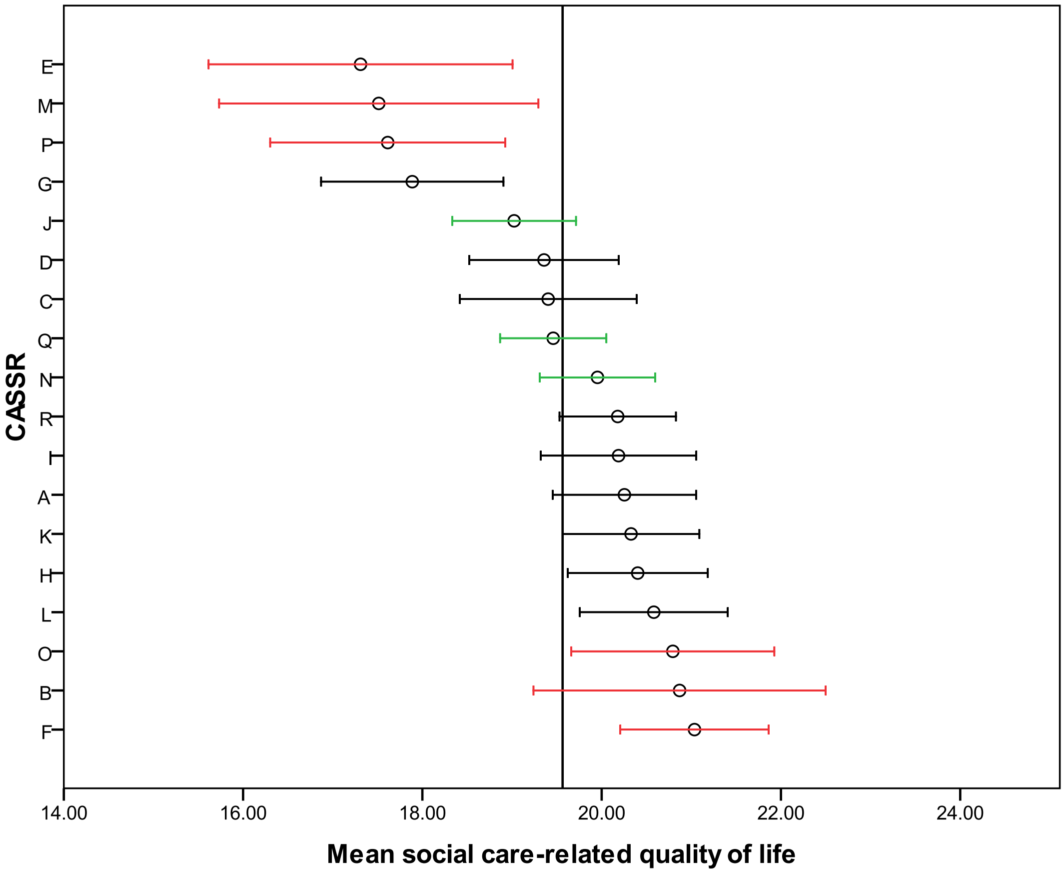
A one-way between-subjects ANOVA was conducted to compare the effect of living in different CASSRs on indicator scores. There are significant differences between CASSRs in SCRQoL scores (F (17, 1170)=4.92, p=<0.001 level).[20] The variation in SCRQoL scores is illustrated in Figure 8 which shows the mean SCRQoL scores for each CASSR with 95% confidence intervals.[21] The circles in Figure 8 represent the mean scores for each CASSR. The error bars around the mean for each CASSR represent the 95% confidence interval for the mean. The solid vertical black line represents the grand mean for the whole sample. Some of the confidence interval bars are coloured. The red bars indicate those CASSRs for whom the results do not reach the 80 per cent reliability level according to the analyses in section 5.7. The green bars indicate those CASSRs whose point estimates have a reliability of over 90 per cent.

Figure 8: Mean SCRQoL scores for CASSRs in sample

---

[20] Bartlett's test for equality of variance rejects the null hypothesis that variances across the groups are equal ($X^2(17)=58.40$, p=<0.001). Simulation results in STATA using simanova confirmed that the ANOVA test is too liberal. However, neither the W or F star test resolved the problem, although the W test was more reliable. The W test also found a highly significant difference between subgroup means (WStat(17, 327.27) = 4.28, p<0.001). Given this result, the finding that the difference between the simulated p-value and the nominal p-value was not large, and the small observed p-value, we have confidence in the result that there are differences between subgroups.

[21] The confidence intervals are calculated from standard errors not adjusted using the finite population correction. The finite population correction is commonly used by the IC when presenting these results and it has the effect of narrowing the confidence interval width since it takes account of the size of the sample in relation to the size of the population.

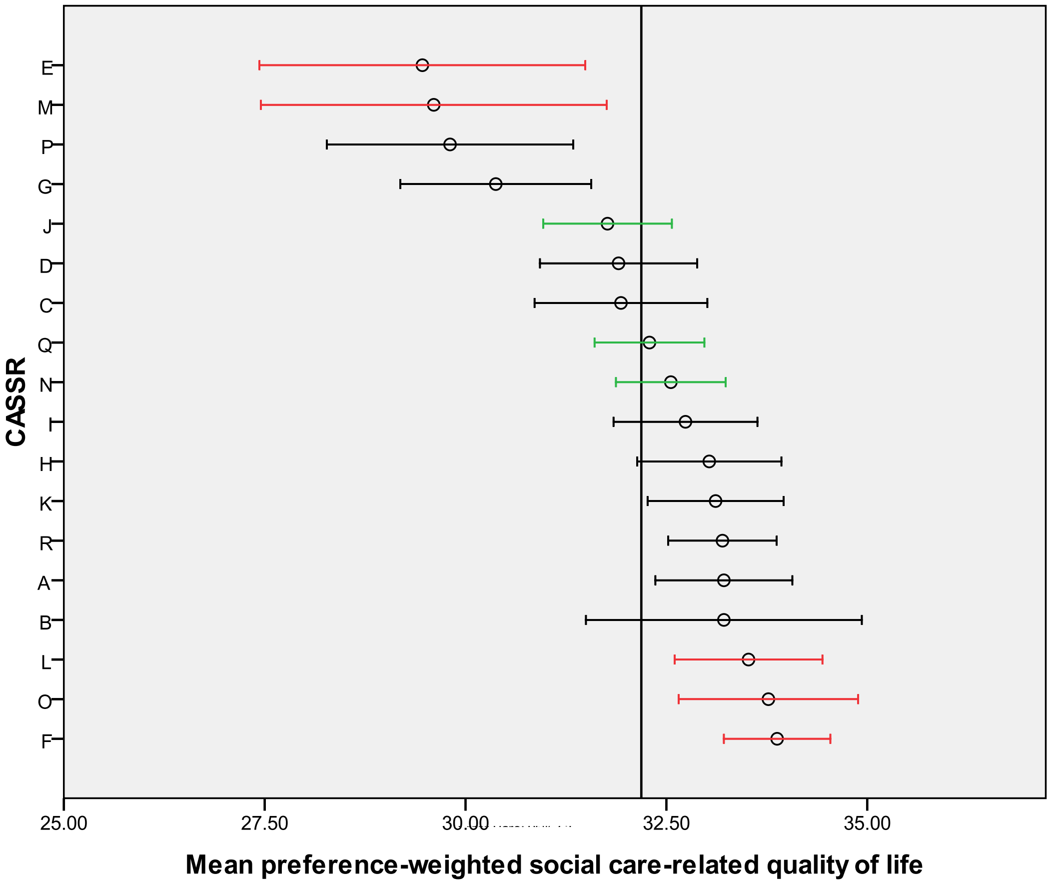


As for SCRQoL indicator we also find significant differences between CASSRs on the PW-SCRQoL indicator (F (17, 1170)=5.38, p=<0.001 level).[22] This is illustrated in Figure 9.

---

[22] Bartlett's test for equality of variance rejects the null hypothesis that variances across the groups are equal ($X^2(17)=98.19$, p=<0.001). Simulation results confirmed that the ANOVA test is too liberal. The W test almost completely resolves the problem. The W test also found a highly significant differences between subgroup means (WStat(17, 329.57) = 4.91, p<0.001). Given this result, the finding that the difference between the simulated p-value and the nominal p-value was not large, and the small observed p-value, we have confidence in the result that there are differences between subgroups.

Figure 9: Mean preference-weighted SCRQoL scores for CASSRs in sample



Of the SCRQoL needs indicators, all have significant differences across CASSRs except the measure of the proportion of respondents with high-level need in at least one SCRQoL domain (rho=0.028, X^2=1.45, p=0.114). The two measures of high and low-level needs have highly significant differences across CASSRs (for the measure covering all SCRQoL domains: rho=0.057, X^2=18.45, p=<0.001, and for the measure covering the basic SCRQoL domains: rho=0.094, X^2=15.97, p=<0.001). The measure capturing the proportion of respondents with high-level needs in at least one basic SCRQoL domain also shows significant differences across CASSRs (rho=0.197, X^2=9.08, p=0.001).[23] The indicator estimates for each CASSR are shown graphically in Figures 10 to 13, using the same colour coding as for the other two indicators.

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

---

[23] We repeated the analysis of the binary variables using a logistic regression and inputting the CASSRs as dummy variables. A global Wald test was used to test for differences between CASSRs and the results agreed with those reported above from the random-effects model.

approximation is for a single comparison, not multiple comparisons. A further more complicated adjustment is required for the latter purpose. Graphs based on this approximation for a single comparison are shown for all of the indicators considered in Appendix 6 (Figures 15 – 20).

Figure 10: Proportion with at least one need in any SCRQoL domain for CASSRs in sample

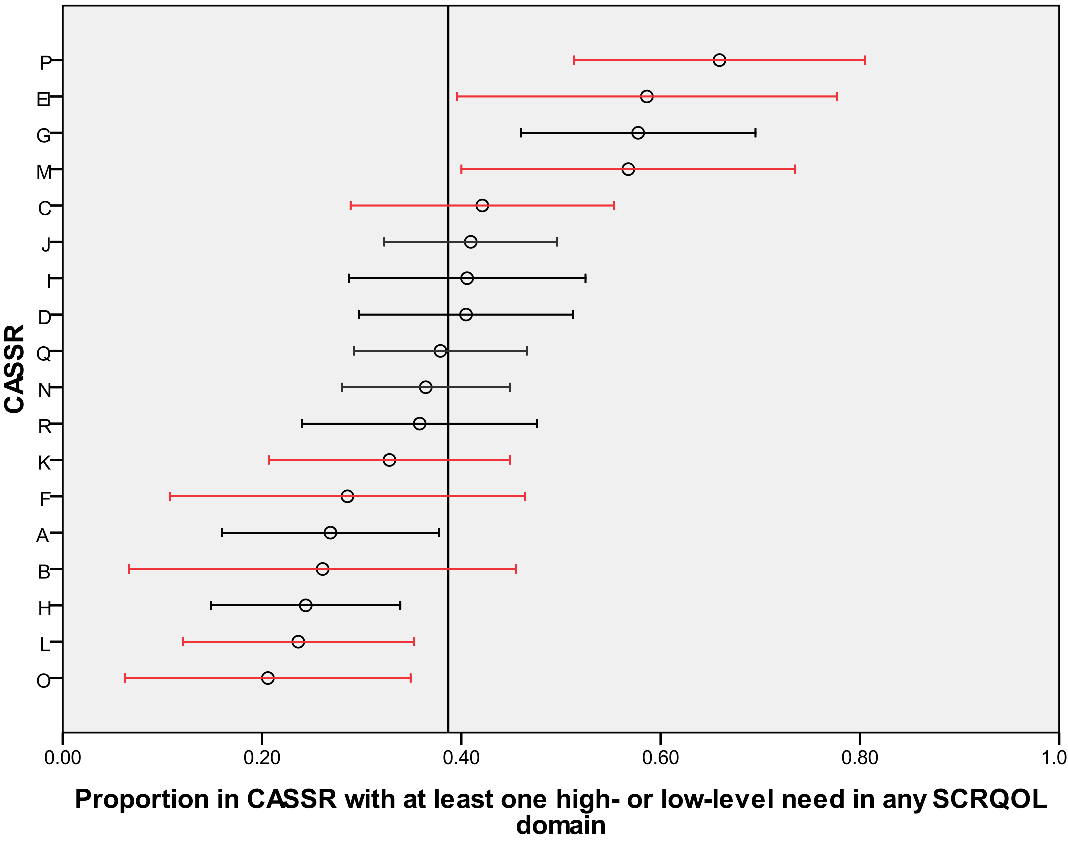


Figure 11: Proportion with at least one need in a basic SCRQoL domain for CASSRs in sample

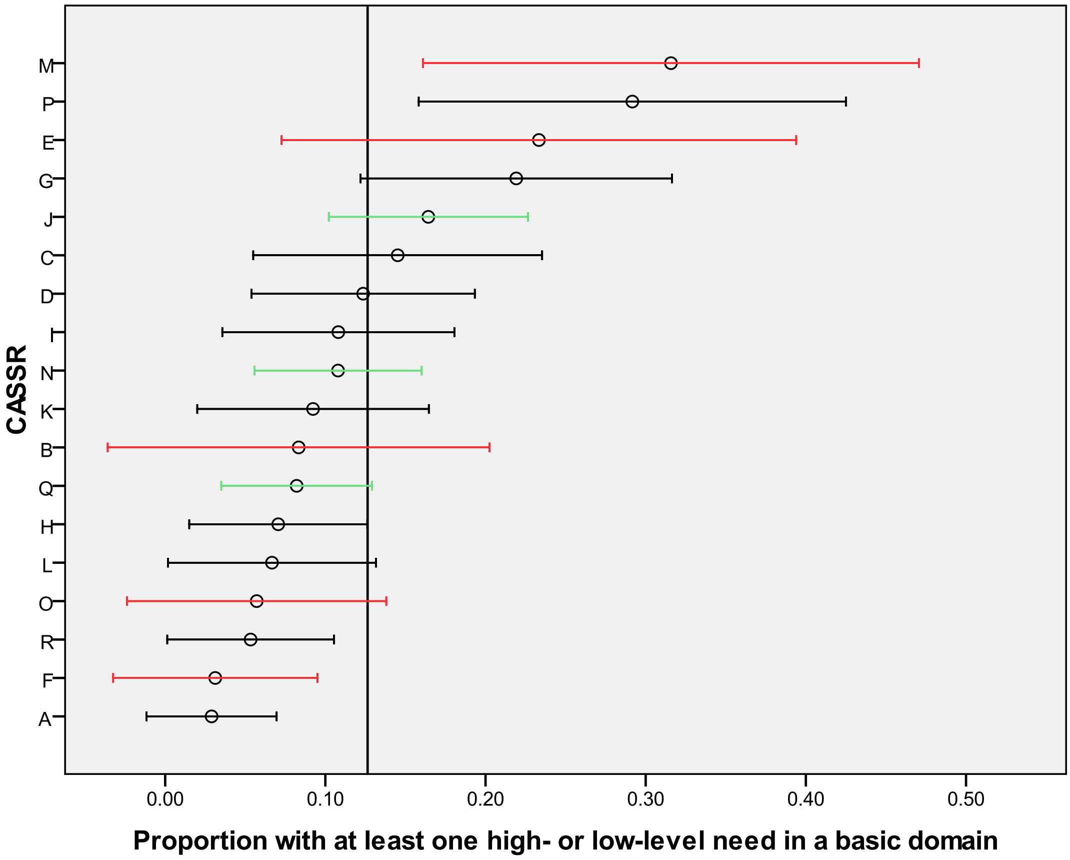

CASSR
M
P
E
G
J
C
D
I
N
K
B
Q
H
L
O
R
F
A
0.00
0.10
0.20
0.30
0.40
0.50
Proportion with at least one high- or low-level need in a basic domain

Figure 12: Proportion with at least one high-level need in any SCRQoL domain for CASSRs in sample

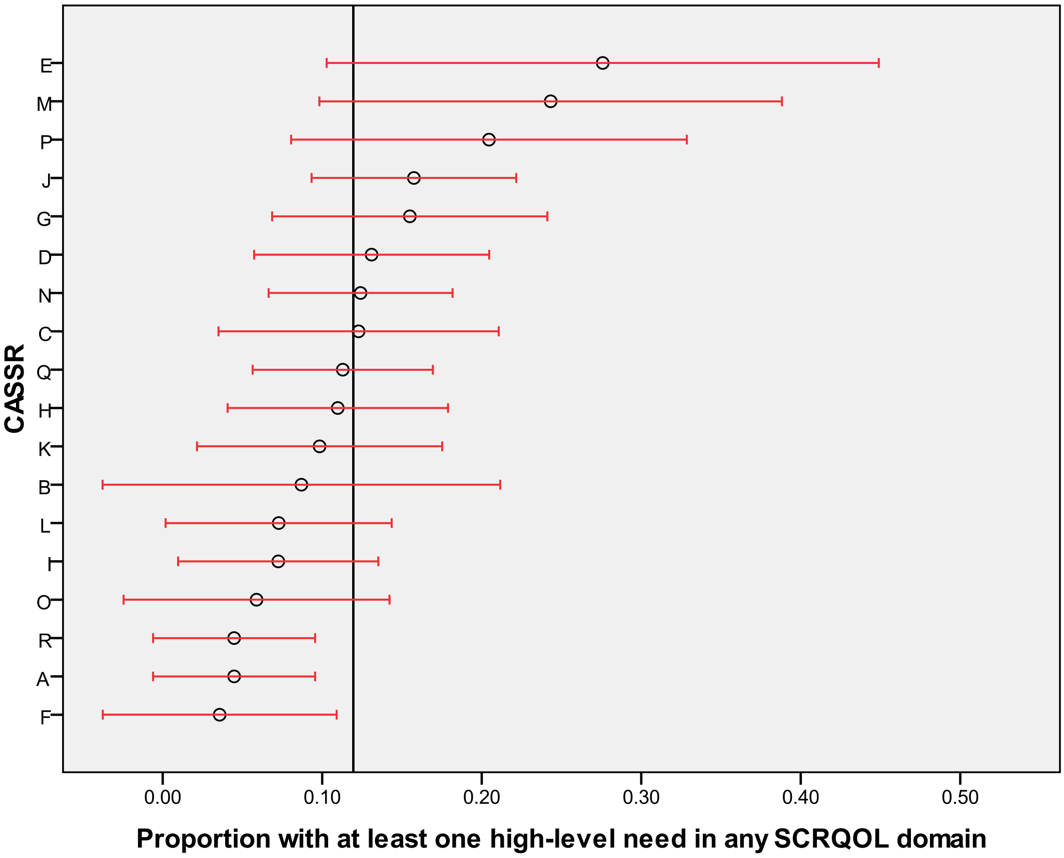
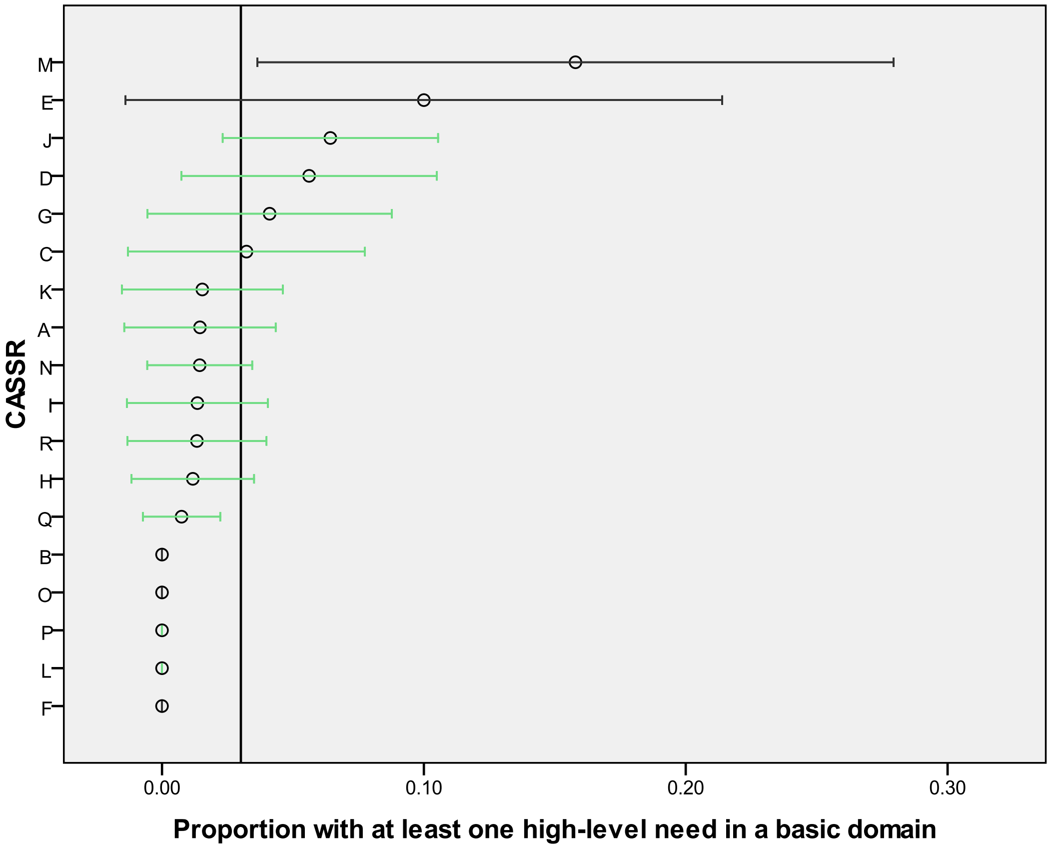


Figure 13: Proportion with at least one high-level need in a basic SCRQoL domain for CASSRs in sample



The results from the validity analysis demonstrated that there were differences in the value of the indicator across questionnaire types. Given that only ten of the CASSRs undertook the full pilot and gathered data using all three types of questionnaires, it was possible that the CASSRs that did not participate in the full pilot may be distorting the results of the sensitivity analysis. The funnel plots (Figures 21 to 26) shown in Appendix 7 illustrate the effect the CASSRs that did not participate in the full pilot may be having. The funnel plots show the CASSR's score on the indicator plotted against the number of observations on which the indicator score is based. The funnelling lines represent the confidence limits at two and three standard deviations away from the sample mean, which is represented by the thick horizontal black line. For each indicator, the majority of the better-performing CASSRs are those that did not take part in the full pilot, and the majority of the worse-performing CASSRs are those that did take part in the full pilot. It was possible that the observed differences between CASSRs on the indicators may be due to the type of pilot they participated in rather than the genuine differences.

I therefore repeated the analyses reported above excluding those CASSRs that did not participate in the full pilot. The ANOVA results for the SCRQoL and preference-weighted SCRQoL indicators still showed highly significant differences between the CASSRs ($p<0.001$). There were also highly significant differences ($p<0.001$) between CASSRs for the two SCRQoL needs measures where need is defined as both high and low-level needs. As I found above, there were no differences between CASSRs on the SCRQoL needs indicator capturing the proportion of respondents with high-level need in at least one SCRQoL domain. The SCRQoL needs indicator capturing the proportion of respondents with high-level needs in at least one basic SCRQoL domain did show significant differences across CASSRs as before, but only at the five per cent level of significance.[24]

Consequences of choice of indicator

The Spearman's rank correlation coefficients in Table 22 give an indication of how similar the rank position of the CASSRs is for each of the indicator pairs. The correlations for all pairs are strong, but there is some variation in ranks depending on the choice of indicator. The choice between the SCRQoL and the PW-SCRQoL indicator barely alters the ordering of the CASSRs, as can be seen by the extremely high correlation coefficient at 0.97. The SCRQoL needs indicator capturing the proportion of respondents with low- or high-level needs in at least one SCRQoL domain also leads to an ordering of CASSRs that is very similar to the SCRQoL and PW-SCRQoL indicators. However, choosing the SCRQoL needs indicator capturing the proportion of respondents with high-level needs in at least one basic SCRQoL domain, leads to some differences in ordering compared with that achieved with the other indicators. It is possible that some of the differences are due to the small sample size and the inaccuracy with which the proportion is estimated within some of the CASSRs.

---

[24] We repeated the analysis of the binary variables using a logistic regression and inputting the CASSRs as dummy variables. A global Wald test was used to test for differences between CASSRs, and the results agreed with those reported above from the random-effects model.

Table 22: Spearman's Rank Correlation Coefficient for ordering of CASSRs

	SCRQoL	PW- SCRQoL	low- or high-level need in any SCRQoL domain	low- or high-level need in basic SCRQoL domain	high-level need in any SCRQoL domain
PW-SCRQoL	0.97***				
Proportion of respondents with low- or high-level need in at least one SCRQoL domain	0.91***	0.91***			
Proportion of respondents with low- or high-level need in at least one basic SCRQoL domain	0.84***	0.91***	0.87***		
Proportion of respondents with high-level need in at least one SCRQoL domain	0.87***	0.93***	0.77***	0.92***	
Proportion of respondents with high-level need in at least one basic SCRQoL domain	0.72***	0.71***	0.61***	0.61***	0.62***

*** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level

Discussion

The purpose of this report is to provide empirical evidence on the qualities, in particular the validity, reliability and sensitivity, of indicators based on the ASCOT questions from the ASCS. We have considered here the qualities of three different types of measures that could be used as an indicator:

1. a summary score across all dimensions of SCRQoL (SCRQoL)
2. a summary score across all dimensions of SCRQoL weighted by preferences (preference-weighted SCRQoL)
3. a measure of need across the SCRQoL dimensions (SCRQoL needs)

The first two of these are measures of the social care outcome state of an individual. They differ in that (1) is based on psychometric principles for measure construction and is a measure of the social care outcome state or SCRQoL; whereas (2) is based on economic theory and is a measure of the value of the social care outcome state or SCRQoL state. Measure (3) is a slightly different measure capturing the degree to which a person has needs across the SCRQoL dimensions.

Four different ways measures of SCRQoL needs are suggested, which vary from each other according to how ‘need’ is defined using the ASCOT items. The structure of the response options to the ASCOT items allows for several ways of defining need. First, the ASCOT items have four response options, designed to capture the outcome states of high-level needs, low-level needs, no needs where the person’s aspirations are not met (mustn’t grumble state), and no needs where the person’s aspirations are met (‘ideal’ state). This leads to two potential ways of defining need according to severity: presence of high-level needs and presence of either high- or low-level needs. The range of needs can also be varied. The ASCOT measure has eight dimensions of which the ASCOT team have generally considered the dimensions of personal care, food and nutrition, safety, and accommodation cleanliness and comfort as the basic need dimensions. The range of SCRQoL needs can therefore be defined as a need in a basic SCRQoL dimension or a need in any SCRQoL dimension, depending on the stakeholders’ interests.

The results of the psychometric analysis supported the construction of the SCRQoL indicator from all the ASCOT items. The factor analysis and Mokken analysis (and supporting analyses) demonstrate that the variation in the ASCOT items can be explained by a single underlying latent factor, although we did note that the safety and dignity ASCOT items are not explained as well as the other items are by the latent factor. The results indicate, however, that the items are similar enough for the responses to each of the items to be summed together to form a single score for SCRQoL for each person (see Appendix 3 for scoring method).

The extent to which the ASCOT items meet the requirements for a measure of social care value has been studied in the Outcomes of Social Care for Adults (OSCA) study (Netten et al., 2010b). Using illustrative preference weights from a pilot study conducted within the OSCA project we generated the preference-weighted SCRQoL measure for each person in the sample (see Appendix 4 for scoring method).

Four measures of the number of SCRQoL needs per person were defined, and analysis of their distributions showed that they were severely skewed. From the analysis of the data I concluded that the measures were likely to be Poisson distributed, not normally distributed, and suggested that, given the lack of experience in social care organisations with such data, it would be better to develop indicators that express the extent of SCRQoL needs using proportions, where the proportion is specified as the proportion of people in a CASSR with at least one need. A binary measure (zero for no need and one for presence of at least one need) was generated for each of the specifications of SCRQoL needs, i.e. where need is defined as high- and low-level needs or only high-level needs and the range of needs is defined as all SCRQoL dimensions or only the basic SCRQoL dimensions. Analysis of these measures found that only the measure of the proportion of people with at least one high- or low-level need in any SCRQoL domain had good variation. The other measures had poor variability since within the sample the proportion of people with a need, as defined through the other three measures, was very low. The poor distributional characteristics of these three specifications of the SCRQoL needs indicator have consequences, in particular, for the reliability of the indicators and the sensitivity analysis.

Validity of the measures

The first set of analyses explored the validity of the three types of measures as outcomes-focused indicators. We have interpreted outcomes-focused as meaning that the indicator captures the *outcome state* of a person in the areas of outcome that can be expected to be affected by social care services. Since we have explored in some detail in development work for the ASCOT items the validity of the ASCOT items (Netten et al., 2009, Netten et al., 2010b), the focus of this work was on comparing the ability of each measure to capture the outcome state of a person. This was achieved by comparing the association of each measure with a single-item measure of general quality of life and a single-item measure of satisfaction with services. The rationale for choosing these items was that they are closely related to the concept of social care outcomes, at least as specified in government documents on outcomes from social care {Department of Health, 2006 #12}.

We found that all of the measures had significant associations with both the satisfaction and general quality of life items. However, the SCRQoL indicator had the strongest association followed by the preference-weighted SCRQoL indicator, although there was not much difference between these two measures. The SCRQoL needs indicators did not have such strong associations, but of these the measure of the proportion of people with at least one high- or low-level need in any SCRQoL domain was the best. It was also observed that for all of the measures the association is stronger with the general quality of life item than the satisfaction item.

These observed relationships are not surprising since the general quality of life item captures the person's evaluation of their quality of life and is therefore more similar conceptually to the ASCOT items than the satisfaction measure which captures the person's perceptions of the effect of services. Although the latter may capture their perceptions of the effect of services on their outcome state, the measure may also capture their perceptions of the process of service delivery and may be influenced by other factors such as expectations. We would therefore expect the general quality of life item to have a stronger association with all the items. Similarly, given the SCRQoL measure captures the person's outcomes state, this is also more similar conceptually to the general quality of life measure than the preference-weighted SCRQoL indicator which captures value of that outcome state, and the SCRQoL needs measures which capture the extent of SCRQoL need. Although need is related to outcome state, since people with high needs in the presence of services are likely to have poor outcomes, conceptually the two are distinct. This analysis demonstrates that the best measure of the outcome state of the person is the SCRQoL indicator.

Given the strong correlation (r=0.67 for the standard and care home questionnaire) between the SCRQoL indicator and the general quality of life item, it could be argued that the general quality of life item is a good indicator of the SCRQoL outcome state of an individual. Given the simplicity of the measure this may be a useful alternative to the more complex psychometric-based SCRQoL indicator. However, one of the benefits of the multi-item SCRQoL indicator is that the measure can be disaggregated to its constituent components. Another benefit is that it is clear which aspects of quality of life are included in each person's evaluation. This is not the case for the general quality of life item, where it is unclear what aspects of quality of life each person draws on to make their evaluation. It is possible that the types of factors included in evaluations will vary across subgroups of the population. For example, people with higher needs in the absence of services may focus more on aspects related to personal care; whereas those with fewer needs may focus on aspects related to their social well-being or even romantic life, an aspect which is beyond the control of social care services. Although there may be circumstances where the general quality of life item will function well as an indicator of outcome state, we would recommend using the SCRQoL measure to ensure clarity about what outcome states are being evaluated.

Combining responses across population subgroups

Social care services are provided to a heterogeneous population with very different abilities to complete a postal questionnaire. In order to gather responses from the variety of clients, different versions of the standard questionnaire were developed, including a version for people with LD and a version for people living in care homes. Instructions were also given with the questionnaire that where clients were not able to answer the questionnaire on their own they should seek help (although steps were taken to exclude proxy responses). Analysis of the returned data indicated that a large number of the responses were collected using the alternative versions of the questionnaire and that a large number of respondents had help from someone else to answer the questionnaire. An important question to be addressed when combining responses derived from different modes of inquiry is whether the mode of inquiry systematically influences the response.

We observed systematic differences in the indicators according to the source of help to complete the questionnaire. Not unexpectedly, clients who had help from their care workers had consistently better scores on all of the indicators compared to all other groups. People who completed the questionnaire on their own reported the next best scores and people who had help from someone living outside their household the worst. It is quite possible that this pattern (completed by self > help from inside household > help from outside household) is related to real differences in outcomes, as people who could complete the questionnaire without help are likely to be less disabled than those who need help. Those who have help from someone inside the household quite possibly regularly receive all kinds of practical help from this person, whilst those who had help from someone outside the home may have much reduced informal care networks and consequently worse social care outcomes. We would expect those who have help from care workers to be more similar to those who have help from someone inside or outside of the household, so this relationship is concerning, although it is possible that other factors explain why those who have help from care workers report better outcomes. Multivariate analysis is required to disentangle some of these relationships, but I would recommend that these results are interpreted with caution, particularly where there are variations in the proportion who have help from care workers across CASSRs. It may be judicious to consider weighting responses obtained from different groups before combining to ensure CASSRs with larger populations of people who had help from care workers do not benefit from this until it is clear what explains the differences observed across these groups.

The analysis also suggests that data gathered from the different questionnaire versions should be combined with some caution. There is evidence that all of the measures suggested as indicators vary systematically by the version of questionnaire used. Thus, people responding to the LD version have better results than those responding to the care home version who in turn have better results than those responding to the standard version. There are clearly many potential explanations for this finding. For example, the higher scores observed by those completing the LD questionnaire could be due to differences in the wording of the ASCOT items. The finding that the strength of the association between the general quality of life and satisfaction items and the indicators is different for people responding to the LD questionnaire compared to the care home and standard questionnaire, lends some support to this hypothesis. However, the strength of the association could also have been affected by the change in the wording and reduction of response categories for the quality of life and satisfaction items in the LD version rather than the wording of the ASCOT items. An alternative explanation could be that the differences are in fact driven by the higher proportion of responses from care workers in the LD (40 per cent) and care home (a third) versions compared to the standard questionnaire (seven per cent). Clearly, more detailed multivariate analysis is required to disentangle these effects.

The importance of these differences across versions should not be underestimated and are clear from the funnel plots shown in Appendix 7. From these plots it can be seen that for all the indicators the overwhelming majority of those CASSRs that participated in either the care home or LD pilot had better than average scores on all indicators, whereas those who participated in the full pilot has worse than average scores on all the indicators. Although we did not examine this in the funnel plots, it is obvious that if we excluded the care home and LD pilots the average score would be much lower for each measure and the conclusions about which CASSRs are better or worse than average would change substantially. It is therefore important that if there are variations in the proportion of responses gathered by the different versions across CASSRs, steps are taken to ensure any conclusions drawn about the achievements of CASSRs from these indicators are not influenced by the way the data have been collected. In the short term we would recommend reweighting of responses before combining responses gathered from different versions of the questionnaire to ensure CASSRs with larger care home and LD populations do not benefit unfairly, but would also advise that more detailed multivariate analysis is undertaken to understand what factors are most important in driving the variations observed, perhaps supported by qualitative work to aid interpretation.

Reliability of the measures

The results of the reliability analysis that draws on generalisability theory as an approach to estimating reliability should be interpreted with some caution for two reasons. First, the number of CASSRs is small and some of the model assumptions are not met, which both tend to lead to underestimation of the variance partition coefficient (VPC) and consequently the reliability of the measures. Second, only ten of the participating CASSRs conducted the full pilot and received completed questionnaires of all three versions. Since people responding to the care home and LD versions have better scores than those responding to the standard questionnaire, those CASSRs that took part in only the care home or LD pilot are likely to have much higher average scores for the indicator. This is likely to distort the VPC estimates since the CASSR effect is confounded with the questionnaire version in some CASSRs. However, we cannot be sure that the distortion arising from the pilot the CASSR participated in has the same effect on each indicator, which invalidates comparisons of reliability amongst the measures. We have not reanalysed the data excluding those CASSRs that did not participate in the full pilot since, although this would solve one problem, it would also further reduce the number of CASSRs creating more uncertainty over estimates of the VPC. We have concluded that to conduct this analysis it is better to use a larger sample of CASSRs, so the variance components are not underestimated and the effects of the questionnaire type can be adjusted for adequately. Such a sample could be gained from the full survey since this would not have the same restrictions on sample size as were in place for the pilot.

In the past the IC has used the Best Value criterion for the accuracy of an indicator of plus or minus four per cent. Our analysis shows that in order to achieve this degree of accuracy, a relatively small sample size of fewer than 60 respondents per CASSR would be required for both the SCRQoL and preference-weighted SCRQoL indicators. This sample size is very small and would make it very difficult for CASSRs to conduct any further analysis with their data: for example, analyses investigating associations between the ASCOT items and other variables would be very difficult due to small numbers in the cross-tabulation cells. This sample size is also likely to make it difficult to detect differences between CASSRs on these measures, although we have not conducted the necessary power tests here to determine what difference would be detected by a sample of this size. We suggest that if these measures are to be used, an alternative approach to determining the required sample size is used that is commensurate with the uses of the data. An example could be a power test for detecting differences between CASSRs or a power test for detecting differences between subgroups within CASSRs. Information from the pilot on the average scores on the measures across subgroups could be used for such analysis.

We also showed in the analysis that, for the SCRQoL needs measures, the lower the probability of observing a need (i.e. the smaller the mean) the smaller the required sample size. Thus, for the measure capturing the proportion of people with at least one high-level need in a basic SCRQoL domain, where the probability of observing at least one need is 0.03, the required sample size is 250 compared to 568 for the measure capturing the proportion of people with at least one high- or low-level need in any SCRQoL domain, where the probability of observing a need is 0.38. The reason for this is that the smaller the mean, the greater the width of the confidence interval in relation to the mean. Thus for the measure capturing the proportion of people with at least one high-level need in a basic SCRQoL domain, the confidence interval width is over twice the size of the mean. The concept of accuracy is therefore not constant across the scale since when the probability of observing a need is 0.03 and accuracy of plus or minus 0.04 is not adequate; it may be considered adequate, however, when the probability of observing a need is 0.38. We have therefore suggested that if any of these measures are chosen for indicators, careful consideration would need to be given to the criterion for accuracy. The criterion for accuracy has consequences for the size of the sample. For a probability of observing a need of 0.03, the sample size required to achieve a confidence interval of half the width of the probability (i.e. 0.015) is 470.

Sensitivity of the indicators

In this analysis we examined sensitivity by exploring whether there are any differences between the CASSRs on these measures. We are not able to look at differences over time as the dataset is cross-sectional, although this would be valuable. We found significant differences between CASSRs for all the measures except the measure of SCRQoL need, where SCRQoL need is defined as the proportion of people with at least one high-level need in any SCRQoL domain. However, it is important to recognise that some of the variation across

CASSRs is likely to be explained by the choice of pilot. We therefore repeated the analysis excluding the CASSRs that did not participate in the full pilot. Significant differences between CASSRs were still observed, again on all of the measures except the one capturing the proportion of people with at least one high-level need in any SCRQoL domain. We did, however, find that the differences between CASSRs on the measure of the proportion of people with at least one high-level need in a basic SCRQoL domain were only significant at the five per cent level. It is possible that differences across CASSRs in the proportion of people who had help from care workers and the proportion of people using the different versions of the questionnaire explain these differences. We therefore suggest that this analysis is treated with some caution until multivariate analysis is undertaken examining the effect of both questionnaire version and source of help on differences between CASSRs. The small number of CASSRs will make this analysis difficult to carry out on the pilot dataset.

Choosing between the indicators

To aid the decision-making about the appropriate indicator, we investigated whether and to what extent the ordering of CASSRs is altered by the choice of measure. We found that the Spearman Rank Correlation Coefficient (SPCC) was very high for the SCRQoL-preference-weighted SCRQoL pair, which indicated that these measures produce very similar orderings of CASSRs. The SCRQoL needs measures by contrast have much lower SPCCs with each other and with the SCRQoL and preference-weighted SCRQoL indicators. Out of these SCRQoL needs indicators, the measure of the proportion of people with a high- or low-level need in any SCRQoL domain has the highest SPCC with the SCRQoL and preference-weighted SCRQoL indicators. Therefore, whilst the choice between the SCRQoL and preference-weighted SCRQoL indicators does not have major consequences for the ranking of CASSRs, the choice between these two indicators and the SCRQoL needs indicators does have a consequence for the ranking of the CASSRs. It is particularly interesting that how need is defined in the SCRQoL needs indicators is important for the ranking of CASSRs. If any of the SCRQoL needs measures are used as an indicator, it is therefore important to think carefully about how need should be defined.

The measures of SCRQoL needs do not capture the outcome state of the individual as well as the SCRQoL and preference-weighted SCRQoL indicators, which has been demonstrated by the validity analysis, but the idea behind the SCRQoL needs indicators should resonate much more with what it is that services and CASSRs more generally are trying to do. If an individual to whom they are providing a service has a need, this should be of great concern to the service and should be something it wishes to rectify. By contrast, the SCRQoL and preference-weighted SCRQoL indicators are less meaningful, although population scores on these measures, which will be available on the ASCOT website, should help with benchmarking. All of the SCRQoL needs indicators, except the one that has the broadest interpretation of need, however, suffer from capturing rare or relatively rare events. This has consequences for the analysis of data and for the required sample size. Where events are rare, large datasets are required to identify the event with accuracy and analyse the data for differences across subgroups of the population. For these reasons we would recommend that the national indicator is based on a measure with good variation and suggest that a choice is made between the SCRQoL, preference-weighted SCRQoL and the SCRQoL needs indicator where need is defined as the proportion of people with at least one high- or low-level need in any SCRQoL domain.

Other considerations affecting the survey results

Analysis of the information CASSRs supplied on the services received by clients and client group revealed that there were some inconsistencies between the care setting and client group and the version of the questionnaire received. It is unclear where the source of the error is here. It could be that the wrong version of the questionnaire was sent out, the data on questionnaire type, service receipt or client group were input incorrectly into the return, or the data on service receipt and client group are recorded incorrectly on the database. It is important to resolve the reasons for these inconsistencies as this could affect response rates.

The IC (2010b) reports that response rates varied considerably across CASSRs and that those who chased respondents had higher response rates. There also seem to be some differences between the characteristics of the respondent and non-respondent population, although these have not been tested formally. There are clearly a number of factors that could explain variations in response rates across and within CASSRs, which it would be useful to explore to identify ways of improving response rates and also identifying whether steps need to be taken to adjust the sample for differences in the respondent and non-respondent population. This dataset is not ideally suited to this analysis because the number of participating CASSRs is small. However, should this finding persist through to the national survey, such analysis would be useful as it could inform the survey process and the interpretation of results.

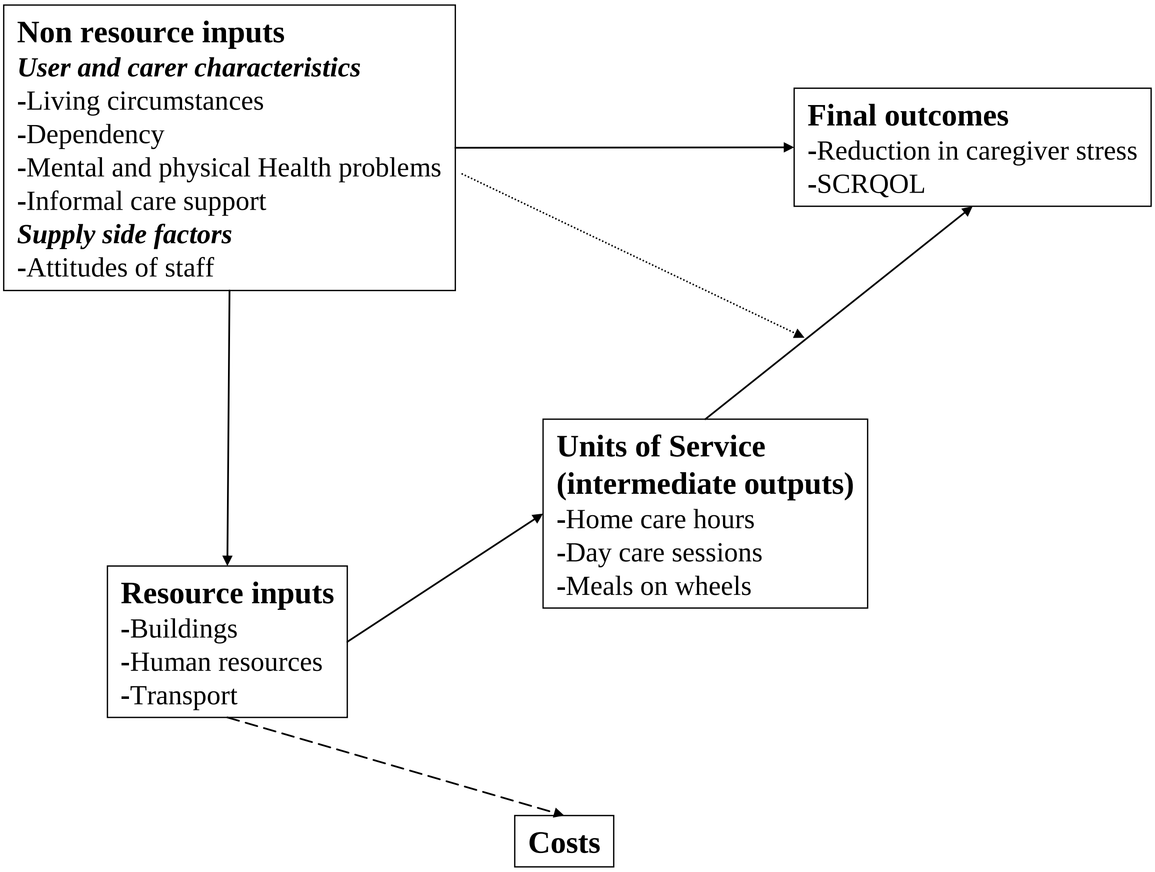
Attributing outcomes to services

It is important to note that all of the measures suggested as indicators focus on *outcome states* rather than the *outcome* from services. The outcome state of a person is determined by many factors including factors unrelated to the service. Because of the effect of other factors, it is difficult to attribute differences between CASSRs on these indicators to differences in the action of services across CASSRs and/or the policies of CASSRs; the observed differences may simply be a result of differences between CASSRs in these other factors that are unrelated to the action of services or the CASSR. If this data are to be published or if the aim of measurement is to capture the quality of publicly-funded services, it is important that the indicators are adjusted to reflect the outcome from such services, or the contribution that services make towards the outcome state of an individual. In addition, such analysis can help to identify potential inefficiencies. There is therefore a strong argument to be made for developing a measure that captures the outcome from services rather than the outcome state of individuals receiving services, or at least conducting the required analysis so people using the data are aware of its limitations for assessing differences between CASSRs.

As part of work aimed at investigating the productivity of social care services, PSSRU at Kent have devised a direct approach to capturing the service outcome (Netten et al., 2005, Netten et al., 2010a, Caiels et al., 2010). This involves asking users what their level of outcome on each SCRQoL dimension would be in the absence of services (expected level of outcome) and subtracting this from their level of outcome in the presence of services. This approach has currently been tested in face-to-face interviews where it seems to work well, although further validation work would be of value (Netten et al., 2010b). This approach has not been tested in a postal questionnaire format. The ASCOT group have reservations about applying it in a postal questionnaire because the questioning has relied heavily on computer-assisted personal interviewing technology to personalise the questions to individuals to reflect their individual patterns of service receipt. By mapping the expected level of outcome in the absence of services to other measures of need that are easier to collect, such as activities of daily living (ADLs) measures, for clients with specified characteristics it may be possible to estimate the service outcome for each individual and produce a 'value-added' measure for social care. The success of this approach depends on developing robust mapping equations between expected need in the absence of services and other needs measures that are easier to collect. A separate study would be required to produce the mapping equation and test it, and this would require additional resources and fieldwork. Data would not be available in the short term, but it may be a long-term solution to the problem.

An alternative way of approaching the problem of attributing the outcome state to the effect of services, which may produce a short-term solution, draws on the production of welfare (POW) approach. POW is an analytical framework that sets out the key factors influencing the process of producing social care outcomes for an individual and it can be used to estimate the contribution that services make to the outcome state of an individual. The POW framework is shown diagrammatically in Figure 14. As this framework makes clear, many factors other than the resources of the service can influence the final outcomes for the user. These are labelled within POW as 'non-resource inputs'. Important non-resource inputs are the needs characteristics of individuals, since these have an effect on both the amount of resources input and the outcome, for instance in the situation in which the person has very high-level needs it may be impossible for them to reach the same level of outcome as a person with low-level needs, even allowing for differences in the inputs from services. The issue with regard to attribution and producing a value-added measure of social care outcome is to separate the effect of the non-resource inputs on final outcomes from the effect of the 'resource inputs' (or 'intermediate outputs') (Malley and Fernández, in press). This is achieved using regression techniques, which partition the variation in the indicator into the two components (resource and non-resource inputs) identified. The success of such techniques, however, depends critically on the ability to control for all non-resource inputs, otherwise unobserved heterogeneity will lead to a positive bias in the estimates of the service contribution. We are currently applying this approach to the older people's home care survey data so evidence should be available shortly as to the success of such an approach.

Figure 14: The production of welfare framework



Non-resource inputs are not exclusively related to the needs characteristics of individuals. They can also include factors which, for example, affect the supply of care. This framework can therefore also be used to adjust the outcome state for contextual factors that may affect the quality of provision. Identifying relevant factors is complicated and they may not unambiguously be unrelated to need. An example of such a factor is the quality of the housing stock. This factor may create need, for example where the WC is on the top floor of the house and the person is unable to climb stairs without help. It can also impact on the ability of care workers to do their jobs well, for example where narrow corridors make it more difficult to carry out tasks. In the latter situation, an important question to address then is, is it fair that the effect of this factor, which may vary across CASSRs, is reflected in the indicator? In instances where all the recipients are in council housing stock an argument could be made for retaining this effect, as an incentive to councils to upgrade their housing. However, if recipients are in private housing stock it seems unfair to include this effect as there is much less the council can do to improve private housing. In such situations it is preferable to adjust the indicator to take account of such factors, by subtracting the effect of the quality of housing stock from the indicator. It is therefore the case that it may be of value to conduct such an analysis on the value-added version of the indicator, should it prove possible to generate such an indicator, to take account of contextual factors that also affect the outcome yet are beyond the control of the CASSR. The identification of such factors is often complex, as we have demonstrated with the housing stock example; another difficult issue would be whether the indicator should be adjusted for the effects of deprivation.

Conclusions and recommendations

This report has demonstrated the validity, reliability and sensitivity of three different types of indicators that could be generated from the ASCS data and the ASCOT items, in particular. The evidence suggests that the SCRQoL and preference-weighted SCRQoL indicators will produce very similar rankings of CASSRs and have similar levels of validity, reliability and sensitivity, although as I have noted the latter two attributes were hard to test given the limitations of the dataset.

The SCRQoL needs measures, by contrast, produce slightly different rankings of CASSRs depending on how need is defined. The broadest measure of SCRQoL need behaves similarly to the SCRQoL and preference-weighted SCRQoL indicators, but it is a less valid measure of social care outcome. The other three SCRQoL needs measures capture relatively infrequent events which means that large sample sizes are likely to be needed to provide an accurate picture. We suggest therefore that these three measures are less suitable for overarching indicators.

The limitations of the dataset, in particular the small number of participating CASSRs and the systematic differences between CASSRs in the way data have been gathered, caused a number of problems for the analysis. Since this was a result of the small samples used to pilot the survey, the full survey is likely to provide a much more suitable dataset and it may be of value to repeat the reliability and sensitivity analyses on this dataset when it becomes available.

Since data are gathered using a number of different methods and techniques, it is important to be clear about the effect of the method of gathering data on the scores reported by individuals before combining responses gathered from different methods. The analysis reported here finds that people who complete the LD questionnaire and those who have help from care workers systematically report better outcomes. Therefore, it is important to investigate using multivariate techniques the relationship between the score on an indicator for each CASSR and the method used for gathering data when the full survey dataset becomes available. In the interim a method for reweighting responses could be considered to ensure that CASSRs with a greater number of people completing LD questionnaires, for example, do not benefit until the effect of the method for gathering data on the indicator score is clear.

Another consideration for the future is the suitability of accuracy criterion used by the IC for this type of dataset. Very small sample sizes (fewer than 100 people) are required using this criterion for the SCRQoL and preference-weighted SCRQoL indicators, which may not be large enough to identify real differences between CASSRs and would anyway produce unsuitable sized samples for further intra-CASSR analysis. It is also not a suitable criterion for measures of rare events, as it would produce a confidence interval that is much wider than the indicator estimate. Alternative accuracy criteria could be considered, as well as alternative approaches to estimating a required sample size, such as power calculations for detecting differences between subgroups.

This analysis also uncovered some inconsistencies between the care setting and client group and the version of the questionnaire that was sent to the client. It is unclear what is the source of the error that produces these inconsistencies, but it is important to understand this as it could have an effect on response rates.

Response rates were very variable across CASSRs and have been so in all past surveys of this kind. Should the variations persist through to the full survey, it would be of value to investigate the factors explaining differences in response rates across CASSRs using multivariate techniques, as this could provide useful information on good process for future surveys.

Finally, we have noted that all of the indicators capture outcome states, not outcomes from social care. We have suggested two approaches: a short- and long-term solution, that could be used to generate an indicator from the ASCOT items that is more sensitive to the outcomes of social care. To generate and test such an indicator requires detailed analysis and further data collection. However, we plan to investigate the suitability of the short-term solution on the older people's home care survey, and this analysis should be available in early 2011. This analysis is important since it will indicate the extent to which differences between CASSRs on the indicators can be attributed to the services and the policies of the CASSR.

References

AGRESTI, A. & FINLAY, B. (1997) *Statistical Methods for the Social Sciences,* Upper Saddle River, New Jersey, Prentice-Hall.

BIRD, S. M., COX, D., FAREWELL, V. T., GOLDSTEIN, H., HOLT, T. & SMITH, P. C. (2005) Performance indicators: good, bad and ugly. *Journal of the Royal Statistical Society A,* 168**,** 1-27.

BRENNAN, R. L. (2000) Performance Assessments from the Perspective of Generalizability Theory. *Applied Psychological Measurement,* 24**,** 339-353.

BURGE, P., POTOGLOU, D., KIM, C. W. & HESS, S. (2010) How do the public value different outcomes of social care? Estimation of preference weights for ASCOT. WR-744-ONS. London, RAND.

CAIELS, J., FORDER, J., MALLEY, J., NETTEN, A. & WINDLE, K. (2010) Measuring the outcomes of low-level services: Final report. PSSRU discussion paper 2699. Canterbury, Kent, Personal Social Services Research Unit.

CAMPBELL, S. M., BRASPENNING, J., HUTCHINSON, A. & MARSHALL, M. N. (2003) Improving the quality of health care: Research methods used in developing and applying quality indicators in primary care. *BMJ,* 326**,** 816-819.

COHEN, J. (1988) *Statistical power analysis for the behavioral sciences,* Hillsdale, NJ, Erlbaum.

COMMUNITIES AND LOCAL GOVERNMENT (2007) The new performance framework for local authorities and local authority partnerships: single set of national indicators. IN COMMUNITIES AND LOCAL GOVERNMENT (Ed., TSO.

CRONBACH, L. J., LINN, R. L., BRENNAN, R. L. & HAERTEL, E. H. (1997) Generalizability Analysis for Performance Assessments of Student Achievement or School Effectiveness. *Educational and Psychological Measurement,* 57**,** 373-399.

DE VELLIS, R. (2003) *Scale Development : Theory and Applications,* London, Sage.

DEPARTMENT OF HEALTH (2009) Making Policy Count: Developing performance indicators for health and social care partnerships. Department of Health.

DEPARTMENT OF HEALTH (2010) Liberating the NHS: Transparency in outcomes – a framework for the NHS. A consultation on proposals. London, The Stationery Office.

FABRIGAR, L. R., WEGENER, D. T., MACCALLUM, R. C. & STRAHAN, E. J. (1999) Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods,* 4**,** 272-299.

GOLDSTEIN, H. & HEALY, M. J. R. (1995) The Graphical Presentation of a Collection of Means. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 158**,** 175-177.

GRILLI, L. & RAMPICHINI, C. (2003) Alternative Specifications of Multivariate Multilevel Probit Ordinal Response Models. *Journal of Educational and Behavioral Statistics,* 28**,** 31-44.

HEMKER, B. T., SIJTSMA, K. & MOLENAAR, I. W. (1995) Selection of Unidimensional Scales From a Multidimensional Item Bank in the Polytomous Mokken I RT Model. *Applied Psychological Measurement,* 19**,** 337-352.

HOLGADO–TELLO, F., CHACÓN–MOSCOSO, S., BARBERO–GARCÍA, I. & VILA–ABAD, E. (2010) Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity,* 44**,** 153-166.

HOX, J. & MAAS, C. (2006) Multilevel Models for Multimethod Measurements. IN EID, M. & DIENER, E. (Eds.) *Handbook of Multimethod Measurement in Psychology.* Washington, DC, American Psychological Association.

MAAS, C. J. M. & HOX, J. J. (2004) Robustness issues in multilevel regression analysis. *Statistica Neerlandica,* 58**,** 127-137.

MALLEY, J., CAIELS, J., FOX, D., MCCARTHY, M., SMITH, N., BEADLE-BROWN, J., NETTEN, A. & TOWERS, A.-M. (2010) A Report on the Development Studies for the National Adult Social Care User Experience Survey, Discussion Paper 2721, . Canterbury, Kent, Personal Social Services Research Unit, University of Kent.

MALLEY, J. & FERNÁNDEZ, J.-L. (in press) Measuring quality in social care services: theory and practice. *Annals of Public and Co-operative Economics*.

MALLEY, J. & NETTEN, A. (2009) Putting People First: Development of the Putting People First User Experience Survey. Discussion Paper 2637. Canterbury, Personal Social Services Research Unit.

MALLEY, J., SANDHU, S. & NETTEN, A. (2006) Younger Adults' Understanding of Questions for a Service User Experience Survey: A Report to the Information Centre for Health and Social Care. PSSRU Discussion Paper 2360. Canterbury, Personal Social Services Research Unit.

MESSICK, S. (1990) Validity of test interpretation and use. Research Report No. 90-11., Educational Testing Service.

MOKKEN, R. (1971) *A theory and procedure of scale analysis,* The Hague, De Gruyter.

MOLENAAR, I. W. & SIJTSMA, K. (2000) *Users manual MSP5 for windows. A program for Mokken scale analysis for polytomous items [Software manual],* Groningen, The Netherlands, iec ProGAMMA.

MOORER, P., SUURMEIJER, T. P. B. M., FOETS, M. & MOLENAAR, I. W. (2001) Psychometric properties of the RAND-36 among three chronic disease (multiple sclerosis, rheumatic diseases and COPD) in the Netherlands. *Quality of Life Research,* 10**,** 637.

MUTHÉN, B. O. (1994) Multilevel Covariance Structure Analysis. *Sociological Methods & Research,* 22**,** 376-398.

NETTEN, A., BEADLE-BROWN, J., TRUKESCHITZ, B., TOWERS, A.-M., WELCH, E., FORDER, J., SMITH, J. & ALDEN, E. (2010a) Measuring the Outcomes of Care Homes, PSSRU Discussion Paper No. 2696/2. Canterbury, Personal Social Services Research Unit, University of Kent.

NETTEN, A., BURGE, P., MALLEY, J., POTOGLOU, D., BRAZIER, J., FLYNN, T. & FORDER, J. (2009) Outcomes of Social Care for Adults (OSCA). Interim findings. Discussion Paper 2648. Canterbury, Personal Social Services Research Unit.

NETTEN, A., BURGE, P., MALLEY, J., POTOGLOU, D., TOWERS, A.-M., BRAZIER, J., FLYNN, T. & FORDER, J. (2010b) Outcomes of Social Care for Adults (OSCA). Second Report of Interim Findings. Canterbury, Kent, Personal Social Services Research Unit.

NETTEN, A., MCDAID, D., FERNANDEZ, J.-L., FORDER, J., KNAPP, M., MATOSEVIC, T. & SHAPIRO, J. (2005) Measuring and understanding social services outputs. Canterbury, PSSRU.

NETTEN, A., RYAN, M., SMITH, P., SKATUN, D., HEALEY, A., KNAPP, M. & WYKES, T. (2002) The development of a measure of social care outcome for older people. PSSRU.

NHS INFORMATION FOR HEALTH AND SOCIAL CARE (2010a) Personal Social Services National Adult Social Care User Experience Survey. Guidance Document. Leeds, NHS Information for Health and Social Care,.

NHS INFORMATION FOR HEALTH AND SOCIAL CARE (2010b) Report for 2009-10 National Adult Social Care User Experience Pilot Survey (NASUEPS). Leeds, NHS Information for Health and Social Care,.

OFFICE OF THE DEPUTY PRIME MINISTER (2005) Securing better outcomes: developing a new performance framework. IN MINISTER, O. O. T. D. P. (Ed., The Stationery Office.

OLSSON, U. (1979a) Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika,* 44**,** 443-460.

OLSSON, U. (1979b) On The Robustness Of Factor Analysis Against Crude Classification Of The Observations. *Multivariate Behavioral Research,* 14**,** 485 - 500.

RABE-HESKETH, S., SKRONDAL, A. & PICKLES, A. (2002) Reliable estimation of generalized linear mixed models using adaptive quadrature. *Stata Journal,* 2**,** 1-21.

RASBASH, J. & GOLDSTEIN, H. (1994) Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model. *Journal of Educational and Behavioral Statistics,* 19**,** 337-350.

RIDOUT, M. S., DEMETRIO, C. G. B. & FIRTH, D. (1999) Estimating Intraclass Correlation for Binary Data. *Biometrics,* 55**,** 137-148.

SHAVELSON, R. J. & WENBB, N. M. (1991) *Generalisability Theory: A Primer,* London, Sage.

SIJTSMA, K., EMONS, W., BOUWMEESTER, S., NYKLÍČEK, I. & ROORDA, L. (2008) Nonparametric IRT analysis of Quality-of-Life Scales and its application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref). *Quality of Life Research,* 17**,** 275.

SIJTSMA, K. & MOLENAAR, I. W. (2002) *Introduction to nonparametric item response theory,* Thousand Oaks, CA, Sage.

STEELE, F. & GOLDSTEIN, H. (2006) A Multilevel Factor Model for Mixed Binary and Ordinal Indicators of Women's Status. *Sociological Methods & Research,* 35**,** 137-153.

WILCOX, R. R., CHARLIN, V. L. & THOMPSON, K. L. (1986) New monte carlo results on the robustness of the ANOVA F, W and F statistics. *Communications in Statistics - Simulation and Computation,* 15**,** 933 - 943.

WILLIS, G. (2005) *Cognitive Interviewing: A Tool for Improving Questionnaire Design,* London, Sage.

Appendix 1: Model standard questionnaire

The questionnaire below is the standard questionnaire. The versions used for care homes and people with LD can be viewed in the report by Malley et al. (2010).

Your Social Care and Support Services

Introduction

We are contacting you because you receive care and support services that are paid for (at least in part) by [your local Social Services Department].By care and support services we mean you may be living in a care home, receiving a Personal Budget, home care, equipment, meals-on-wheels, Direct Payments, or attending a day centre. We want to improve and develop our services so we want to get your views on the services you receive. In particular, we want to hear about your quality of life and how services have affected the quality of your life.

Why you were selected

You have been selected at random from [Social Services'] records of people who are receiving social care and support services.

What we would like you to do

We would like you to help us by taking about ten minutes to give us your views about the care and support services you receive. If you choose not to answer this questionnaire this will not affect the services you receive.

What to do if you need help to give your views

If you would like, you can ask a friend, relative or an advocate to help you complete the questionnaire. Please remember that it is your views and your experiences that are important to us, rather than the views of anyone that helps you. You can also get in touch with [the telephone assistance line]to ask for someone independent from social services and your care provider to help you to complete the questionnaire.[Councils should mention here any telephone help line/advocacy group they have contact with through which assistance in completing the survey can be arranged]Staff from [Social Services]or anyone you pay who have been involved in the provision of your care or support should not help you.

If you are unable to complete this questionnaire either on your own, or by verbally giving answers for someone else to record, such as a friend or relative, then please either discard it or if you are able, return it uncompleted in the envelope enclosed.

What to do if you have queries or would like to know how to obtain information on the results

If you, or your friend or relative, have questions you would like to ask about the survey please ring……………… on Monday to Friday between 10.00 am and 12.00 pm or between 2.00 pm and 4.00 pm[Councils can vary these hours or expand this sentence e.g. to say leave a message and someone will get back to you]

What will be done with the results of the survey

The results of the survey will be used by the Care Quality Commission, the Department of Health and your[Social services department] to see how happy people are with their care and support services and assess their experiences of local care services. The results will also be used for further research or analysis.

Confidentiality

Your answers will be treated as confidential: they will not be passed on to your social worker, care manager, care and support worker or anyone providing you with services. You will not be personally identified and your answers will not affect the services you receive.

The code found [enter position] on this form is used for administration purposes only to make sure that when you return this questionnaire we do not send you another one. However, if you say on the form at question 7 that you are being hurt or harmed by anybody or your safety or health at risk then we will use this code to identify you so that someone (but not your care and support worker) can contact you to talk about it. This is the only circumstance under which this code will be used to identify you.

Reminder Letters

If you do not return this questionnaire then you may be sent reminder letters. If you do not wish to receive reminders then please send back an uncompleted questionnaire in the envelope provided.

Sending back the completed questionnaire

Once you have completed the questionnaire please return it in the envelope provided by **[DATE]**. You don’t need to put a stamp on the envelope.

Thank you for helping us by completing this questionnaire.

Section 1: Overall satisfaction with your social care and support

1. Overall, how satisfied are you with the care and support services you receive?

By 'care and support services' we mean any care provided by staff who are paid to help you. The staff could be from [Social Services], an agency, a care home or bought by you using money from [Social Services]through a Direct Payment.

Please tick (✓) one box

I am extremely satisfied ☐

I am very satisfied ☐

I am quite satisfied ☐

I am neither satisfied nor dissatisfied ☐

I am quite dissatisfied ☐

I am very dissatisfied ☐

I am extremely dissatisfied ☐

Section 2: Your quality of life

When answering the following questions please think about the quality of your life as a whole, including the help you get from others as well as [Social Services].

2. Thinking about the good and bad things that make up your quality of life, how would you rate the quality of your life as a whole?

Please tick (✓) one box

So good, it could not be better ☐

Very good ☐

Good ☐

Alright ☐

Bad ☐

Very bad ☐

So bad, it could not be worse ☐

3. Which of the following statements best describes how much control you have over your daily life?

By 'control over daily life' we mean having the choice to do things or have things done for you as you like and when you want

Please tick (✓) one box

I have as much control over my daily life as I want ☐

I have adequate control over my daily life ☐

I have some control over my daily life but not enough ☐

I have no control over my daily life ☐

4. **Thinking about your personal care, by which we mean being clean and presentable in appearance, which of the following statements best describes your situation?**

Please tick (✓) one box

I feel clean and am able to present myself the way I like ☐

I feel adequately clean and presentable ☐

I feel less than adequately clean or presentable ☐

I don't feel at all clean or presentable ☐

5. **Thinking about the food and drink you get, which of the following statements best describes your situation?**

Please tick (✓) one box

I get all the food and drink I like when I want ☐

I get adequate food and drink at OK times ☐

I don't always get adequate or timely food and drink ☐

I don't always get adequate or timely food and drink, and I think there is a risk to my health ☐

6. **Which of the following statements best describes how clean and comfortable your home is?**

Please tick (✓) one box

My home is as clean and comfortable as I want ☐

My home is adequately clean and comfortable ☐

My home is not quite clean or comfortable enough ☐

My home is not at all clean or comfortable ☐

7. **Which of the following statements best describes how safe you feel?**

By feeling safe we mean feeling safe both inside and outside the home. This includes fear of abuse, falling or other physical harm and fear of being attacked or robbed

Please tick (✓) one box

I feel as safe as I want ☐

Generally I feel adequately safe, but not as safe as I would like ☐

I feel less than adequately safe ☐

I don't feel at all safe ☐

8. Thinking about how much contact you've had with people you like, which of the following statements best describes your social situation?

Please tick (✓) one box

I have as much social contact as I want with people I like ☐

I have adequate social contact with people ☐

I have some social contact with people, but not enough ☐

I have little social contact with people and feel socially isolated ☐

9. Which of the following statements best describes how you spend your time?

When you are thinking about how you spend your time, please include anything you value or enjoy including leisure activities, formal employment, voluntary or unpaid work and caring for others.

Please tick (✓) one box

I'm able to spend my time as I want, doing things I value or enjoy ☐

I'm able do enough of the things I value or enjoy with my time ☐

I do some of the things I value or enjoy with my time but not enough ☐

I don't do anything I value or enjoy with my time ☐

10. Which of these statements best describes how having help to do things makes you think and feel about yourself?

Please tick (✓) one box

Having help makes me think and feel better about myself ☐

Having help does not affect the way I think or feel about myself ☐

Having help sometimes undermines the way I think and feel about myself ☐

Having help completely undermines the way I think and feel about myself ☐

11. Thinking about the way you are helped and treated, and how that makes you think and feel about yourself, which of these statements best describes your situation?

Please tick (✓) one box

The way I'm helped and treated makes me think and feel better about myself ☐

The way I'm helped and treated does not affect the way I think or feel about myself ☐

The way I'm helped and treated sometimes undermines the way I think and feel about myself ☐

The way I'm helped and treated completely undermines the way I think and feel about myself ☐

Section 3: The impact of social care and support services on your quality of life

12. In what ways do care and support services help you?

By ‘care and support services’ we mean any care provided by staff who are paid to help you. The staff can be from [Social Services], an agency or bought by you using money you receive from [Social Services], using a Direct Payment.

Please tick (✓) <u>all that apply</u>

To have control over my daily life ☐

With personal care ☐

Meals ☐

Social contact with people I like ☐

Doing things I value and enjoy ☐

Feeling safe and secure ☐

Keeping my home clean and comfortable ☐

Other (please tell us what in the box below) ☐

Section 4: Knowledge and information

13. In the past year, have you found it easy or difficult to find information and advice about support, services or benefits? *Please include information from different sources, such as voluntary organisations, and private agencies as well as* **[Social Services]**.

Please tick (✓) one box

Very easy to find ☐

Fairly easy to find ☐

Fairly difficult to find ☐

Very difficult to find ☐

I've never tried to find information or advice ☐

14. Thinking about the care and support you receive, if you felt unsafe or were worried about something that had happened to you, who would you talk to?

Please tick (✓) <u>as many</u> boxes as apply

A member of your family ☐

A friend, neighbour or someone you work with ☐

Your keyworker, Personal Assistant or care worker ☐

The manager of your care home or day centre ☐

Your care manager or social worker ☐

Someone else (please tell us who they are – their job, not their name) ☐

No-one I wouldn't say anything ☐

I don't know ☐

Section 5: Your health

15. How is your health in general?

Please tick (✓) one box

Very Good	☐
Good	☐
Fair	☐
Bad	☐
Very Bad	☐

16. **By placing a tick in one box in each group below, please indicate which statementsbest describe your own health state today.**

a. Pain or discomfort

Please tick (✓) one box

I have no pain or discomfort ☐

I have moderate pain or discomfort ☐

I have extreme pain or discomfort ☐

b. Anxiety or depression

Please tick (✓) one box

I am not anxious or depressed ☐

I am moderately anxious or depressed ☐

I am extremely anxious or depressed ☐

17. Please place a tick (✓) in the box that best describes your abilities for each of the following questions labelled from a to d.

	I can do this easily by myself	I have difficulty doing this myself	I can't do this by myself
a. Do you usually manage to get around indoors (except steps) by yourself?	☐	☐	☐
b. Do you usually manage to get in and out of a bed (or chair) by yourself?	☐	☐	☐
c. Do you usually manage to feed yourself?	☐	☐	☐
d. Do you usually deal with finances and paperwork- for example, paying bills, writing letters – by yourself?	☐	☐	☐

18. Please place a tick (✓) in the box that best describes your abilities for each of the following questions labelled from a to d.

	I can do this easily by myself	I have difficulty doing this myself	I can’t do this by myself
a. Do you usually manage to use the WC/toilet by yourself?	☐	☐	☐
b. Do you usually manage to wash your face and hands by yourself?	☐	☐	☐
c. Do you usually manage to wash all over by yourself, using either a bath or shower?	☐	☐	☐
d. Do you usually manage to get dressed and undressed by yourself?	☐	☐	☐

Section 6: About your surroundings

19. How well do you think your home is designed to meet your needs?

Please tick (✓) <u>one</u> box

My home meets my needs very well ☐

My home meets most of my needs ☐

My home meets some of my needs ☐

My home is totally inappropriate for my needs ☐

20. Thinking about getting around outside of your home, which of the following statements best describes your present situation?

You can include getting around by yourself or with help from someone else

Please tick (✓) one box

I can get to all the places in my local area that I want ☐

At times I find it difficult to get to all the places in my local area that I want ☐

I am unable to get to all the places in my local area that I want ☐

Section 7: About yourself, the service user

The answers to the next group of questions will be used to get a picture of who took part in this survey. For example, we will use these questions to help us make sure that services are delivered equally to people with different backgrounds.

21. Are you male or female?

Please tick (✓) one box

Male ☐

Female ☐

22. Which age group do you belong to?

Please tick (✓) one box

18-24 ☐

25-30 ☐

31-39 ☐

40-49 ☐

50-64 ☐

65-74 ☐

75-84 ☐

85 or over ☐

23. To which of these groups do you consider you belong?

Please tick (✓) <u>one</u> box

White (British, Irish, Traveller of Irish Heritage, Gypsy/Roma, any other white background)	☐
Mixed (White and Black Caribbean, White and Black African, White and Asian, any other mixed background)	☐
Asian or Asian British (Indian, Pakistani, Bangladeshi, any other Asian background)	☐
Black or Black British (Caribbean, African or any other Black background)	☐
Chinese	☐
Any other ethnic group	☐

24. What is your religion?

Please tick (✓) <u>one</u> box

None ☐

Christian ☐

Buddhist ☐

Hindu ☐

Jewish ☐

Muslim ☐

Sikh ☐

Other (please tell us what in box below) ☐

25. Do you receive any practical help on a regular basis from your husband/wife, partner, friends, neighbours or family members?

Please tick (✓) as many boxes as apply

Yes, from someone living in my household ☐

Yes, from someone living in another household ☐

No ☐

26. Do you buy any additional care or support privately or pay more to 'top up' your care and support?

Please tick (✓) as many boxes as apply

Yes, I buy some more care and support with my own money ☐

Yes, my family pays for some more care and support for me ☐

No ☐

27. Did you write the answers to this questionnaire by yourself or did you have help from someone else?

Please tick (✓) one box

Yes, I wrote the answers myself ☐

No, I had help from a care worker ☐

No, I had help from someone living in my household ☐

No, I had help from someone living outside my household ☐

28. What type of help did you have?

Please tick (✓) as many boxes as apply

None, because I wrote the answers myself ☐

Someone else read the questions to me ☐

Someone else translated the questions for me ☐

Someone else wrote down the answers for me ☐

I talked through the questions with someone else ☐

Someone answered for me, without asking me the questions ☐

Thank you for helping us by filling in this questionnaire.

Please post it back to us in the envelope provided.

You don't need to put a stamp on the envelope.

For your views to count please return this form by DATE

Appendix 2: Characteristics of the respondent and non-respondent sample

Table 23: Characteristics of the respondent and non-respondent sample

	Respondent sample			Non-respondent sample		
	Frequency	Valid Percentage	Sample Percentage	Frequency	Valid Percentage	Sample Percentage
Gender						
Male	503	38.0%	36.9%	106	34.9%	34.2%
Female	821	62.0%	60.2%	198	65.1%	63.9%
Missing	40		2.9%	6		1.9%
Age group						
18-24	51	3.9%	3.7%	14	4.6%	4.5%
25-30	51	3.9%	3.7%	5	1.6%	1.6%
31-39	89	6.7%	6.5%	22	7.2%	7.1%
40-49	118	8.9%	8.7%	23	7.6%	7.4%
50-64	197	14.9%	14.4%	38	12.5%	12.3%
65-74	139	10.5%	10.2%	29	9.5%	9.4%
75-84	256	19.4%	18.8%	76	25.0%	24.5%
85 and over	421	31.8%	30.9%	97	31.9%	31.3%
Missing	42		3.1%	6		1.9%
Ethnicity						
White	1,232	93.8%	91.0%	291	94.5%	93.9%
Mixed	3	0.2%	0.2%	3	1.0%	1.0%
Asian	29	2.2%	2.1%	2	0.6%	0.6%
Black	42	3.2%	3.1%	5	1.6%	1.6%
Other	7	0.5%	0.5%	7	2.3%	2.3%
Missing	41		3.0%	2		0.6%
Sexual orientation						
Heterosexual	40	88.9%	2.9%	0	0.0%	0.0%
Bisexual	2	4.4%	0.1%	0	0.0%	0.0%
Other	3	6.7%	0.2%	0	0.0%	0.0%
Missing	1319		96.7%	310		100.0%

Religion						
None	180	23.0%	13.2%	12	10.2%	3.9%
Christian	499	63.7%	36.6%	98	83.1%	31.6%
Buddhist	2	0.3%	0.1%	0	0.0%	0.0%
Hindu	10	1.3%	0.7%	0	0.0%	0.0%
Jewish	11	1.4%	0.8%	3	2.5%	1.0%
Muslim	15	1.9%	1.1%	1	0.8%	0.3%
Sikh	2	0.3%	0.1%	1	0.8%	0.3%
Other	64	8.2%	4.7%	3	2.5%	1.0%
Missing	581		42.6%	192		61.9%
Primary client group						
Physical impairment or disability	765	56.3%	56.1%	151	48.7%	48.7%
Mental health	159	11.7%	11.7%	57	18.4%	18.4%
Learning disability	397	29.2%	29.1%	90	29.0%	29.0%
Substance misuse	2	0.1%	0.1%	0	0.0%	0.0%
Vulnerable people	37	2.7%	2.7%	12	3.9%	3.9%
Missing	4		0.3%	0		0.0%
Secondary client group						
Physical impairment or disability	116	16.4%	8.5%	60	30.2%	19.4%
Mental health	33	4.7%	2.4%	16	8.0%	5.2%
Learning disability	41	5.8%	3.0%	14	7.0%	4.5%
Substance misuse	0	0.0%	0.0%	0	0.0%	0.0%
Vulnerable people	11	1.6%	0.8%	6	3.0%	1.9%
Unknown	506	71.6%	37.1%	103	51.8%	33.2%
Missing	657		48.2%	111		35.8%

Appendix 3: Scoring algorithm for SCRQoL indicator

Each of the items relating to SCRQoL has four options. For example:

I have as much control over my daily life as I want	**3**
I have adequate control over my daily life	**2**
I have some control over my daily life but not enough	**1**
I have no control over my daily life	**0**

For each of the eight questions, assign a score of 3, 2, 1 or 0 following the example above to reflect the respondent’s answer to the question. To calculate SCRQoL, total the eight values together. The scale constructed this way takes values from 0 to 24, where someone who scores zero has high-level needs across all SCRQoL domains and someone scoring 24 has no needs across all SCRQoL domains and has their needs met in an ideal way.

Appendix 4: Scoring algorithm for preference-weighted SCRQoL indicator

Each of the items relating to SCRQoL has four response options. Each response option has a value derived from preference study. The values for each response option are set out in Burge et al. (2010, see Table 31, first column) and are summarised in Table 24. To calculate the score for each individual first assign the value from the table to the corresponding response option for each SCRQoL question. After assigning these values, total the eight values together for each individual. The scale constructed this way takes values from 8.13 to 35.97, where someone who scores 8.13 has high-level needs across all SCRQoL domains and someone scoring 35.97 has no needs across all SCRQoL domains and has their needs met in an ideal way.

Table 24: Preference weights for SCRQoL item response options (based on unsegmented model)

	Preference weight
Could you tell me which of the following statements best describes how much control you have over your daily life?	
I have as much control over my daily life as I want	5.11
I have adequate control over my daily life	4.71
I have some control over my daily life but not enough	2.91
I have no control over my daily life	0.00
Thinking about your personal care, by which we mean being clean and presentable in appearance, which of the following statements best describes your situation?	
I feel clean and am able to present myself the way I like	4.62
I feel adequately clean and presentable	3.93
I feel less than adequately clean or presentable	1.31
I don’t feel at all clean or presentable	1.05
Thinking about the food and drink you have, which of the following statements best describes your situation?	
I get all the food and drink I like when I want	4.15
I get enough adequate food and drink when I want	3.96
I don’t always get enough adequate food and drink when I want, but I don’t think there is a risk to my health	1.87
I don’t always get enough adequate food and drink when I want and I think there is a risk to my health	1.20
Could you tell me which of the following statements best describes how clean and comfortable your home is?	
My home is as clean and comfortable as I want	4.14
My home is adequately clean and comfortable	3.90
My home is not quite clean or comfortable enough	2.05
My home is not at all clean or comfortable	1.59
Could you tell me which of the following statements best describes how safe you feel?	

I feel as safe as I want	4.49
Generally I feel adequately safe, but not as safe as I would like	2.46
I feel less than adequately safe	1.58
I don’t feel at all safe	0.56
Thinking about how much contact you’ve had with people you like, which of the following statements best describes your social situation?	
I have as much social contact as I want with people I like	4.42
I have adequate social contact with people	3.86
I have some social contact with people, but not enough	2.67
I have little social contact with people and feel socially isolated	1.32
Could you tell me which of the following statements best describes how you spend your time?	
I’m able to spend my time as I want, doing things I value or enjoy	4.88
I’m able do enough of the things I value or enjoy with my time	4.67
I do some of the things I value or enjoy with my time but not enough	2.80
I don’t do anything I value or enjoy with my time	0.97
Thinking about the way you are helped and treated, and how that makes you think and feel about yourself, which of these statements best describes your situation?	
The way I’m helped and treated makes me think and feel better about myself	4.16
The way I’m helped and treated does not affect the way I think or feel about myself	3.16
The way I’m helped and treated sometimes undermines the way I think and feel about myself	1.74
The way I’m helped and treated sometimes undermines the way I think and feel about myself	1.44

Appendix 5: Scoring algorithm for the unmet needs measures

To generate the unmet needs measures the following scoring algorithm is required. Each of the ASCOT items has four options. For example:

	High- and low-level need	**High-level need**
I have as much control over my daily life as I want	**0**	**0**
I have adequate control over my daily life	**0**	**0**
I have some control over my daily life but not enough	**1**	**0**
I have no control over my daily life	**1**	**1**

There are four options and the scoring for each option is as follows:

For the unmet need measure where unmet need is defined as a *high- or low-level need in any SCRQoL domain*, assign a score of 1 or 0 following the middle column in the example above to reflect the respondent's answer to the question. To calculate the count of unmet SCRQoL needs sum across the eight domains, to generate a score that will vary from zero (no unmet needs) to eight (unmet needs in each SCRQoL domain). To convert this to the binary version of the measure retain values of zero as zero and re-score all values greater than or equal to one as one.

For the unmet need measures, where unmet need is defined as a *high-level need in any SCRQoL domain*, assign a score of zero or one to each ASCOT items following the example in the far right column. To calculate the count of unmet SCRQoL needs sum across the eight domains, to generate a score that will vary from zero (no unmet needs) to eight (unmet needs in each SCRQoL domain). To convert this to the binary version of the measure retain values of zero as zero and re-score all values greater than or equal to one as one.

For the unmet need measure where unmet need is defined as a *high- or low-level need in a basic SCRQoL domain*, assign a score of 1 or 0 to the following ASCOT items: personal care, accommodation cleanliness and comfort, safety and food and nutrition, following the middle column in the example above to reflect the respondent's answer to the question. To calculate the count of unmet SCRQoL needs sum across the four domains, to generate a score that will vary from zero (no unmet needs) to four (unmet needs in each SCRQoL domain). To convert this to the binary version of the measure retain values of zero as zero and re-score all values greater than or equal to one as one.

For the unmet need measures, where unmet need is defined as a *high-level need in any SCRQoL domain*, assign a score of zero or one to the following ASCOT items: personal care, accommodation cleanliness and comfort, safety and food and nutrition, following the example in the far right column. To calculate the count of unmet SCRQoL needs sum across the four domains, to generate a score that will vary from zero (no unmet needs) to four (unmet needs in each SCRQoL domain). To convert this to the binary version of the measure retain values of zero as zero and re-score all values greater than or equal to one as one.

Appendix 6: Differences in indicator scores between CASSRs

The following set of graphs show the differences between CASSRs for each of the indicators, using the adjustment that allows for better identification of differences between a single pair of pre-chosen CASSRs. (It is not the correct confidence interval for multiple comparisons). Each confidence interval below is equivalent to 1.39 times the standard error.

Figure 15: Mean SCRQoL scores for CASSRs in sample (with errors bars to show differences between CASSRs)

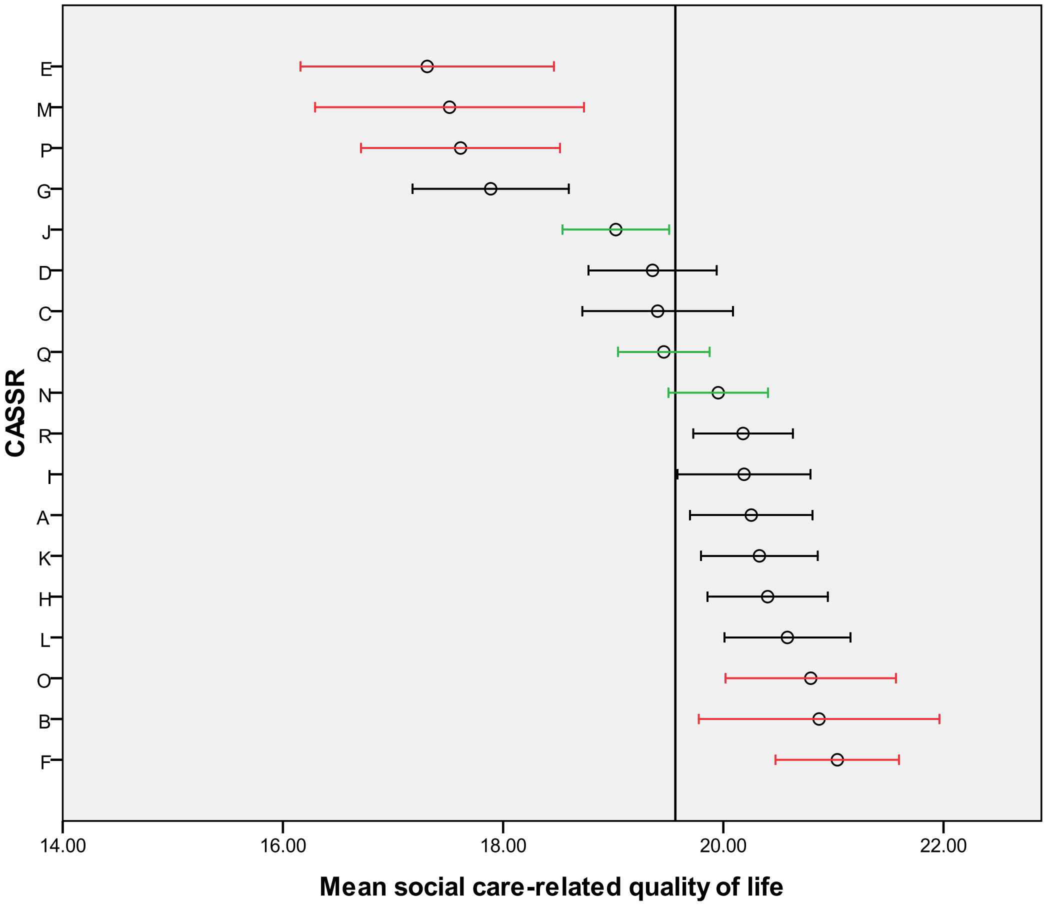


Figure 16: Mean Preference-weighted SCRQoL scores for CASSRs in sample (with errors bars to show differences between CASSRs)

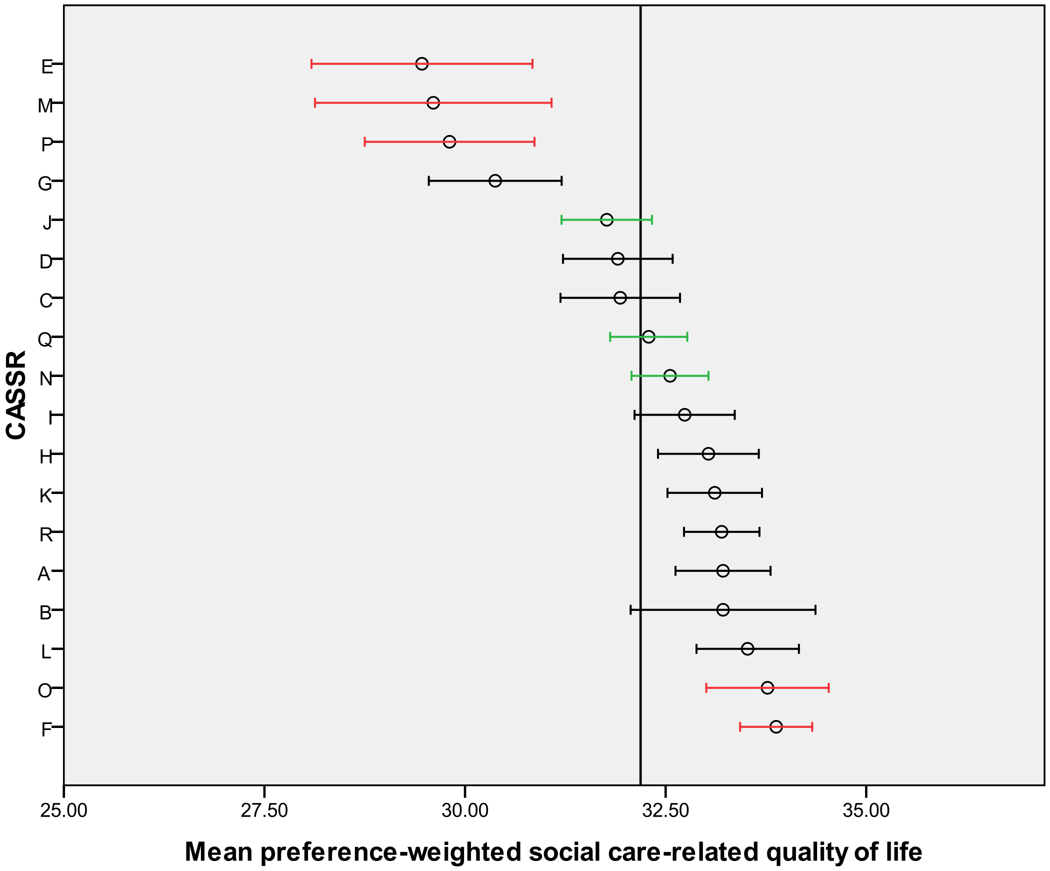

CASSR
E
M
P
G
J
D
C
Q
N
I
H
K
R
A
B
L
O
F
25.00
27.50
30.00
32.50
35.00
Mean preference-weighted social care-related quality of life

Figure 17: Proportion with at least one need in any SCRQoL domain for CASSRs in sample (with errors bars to show differences between CASSRs)

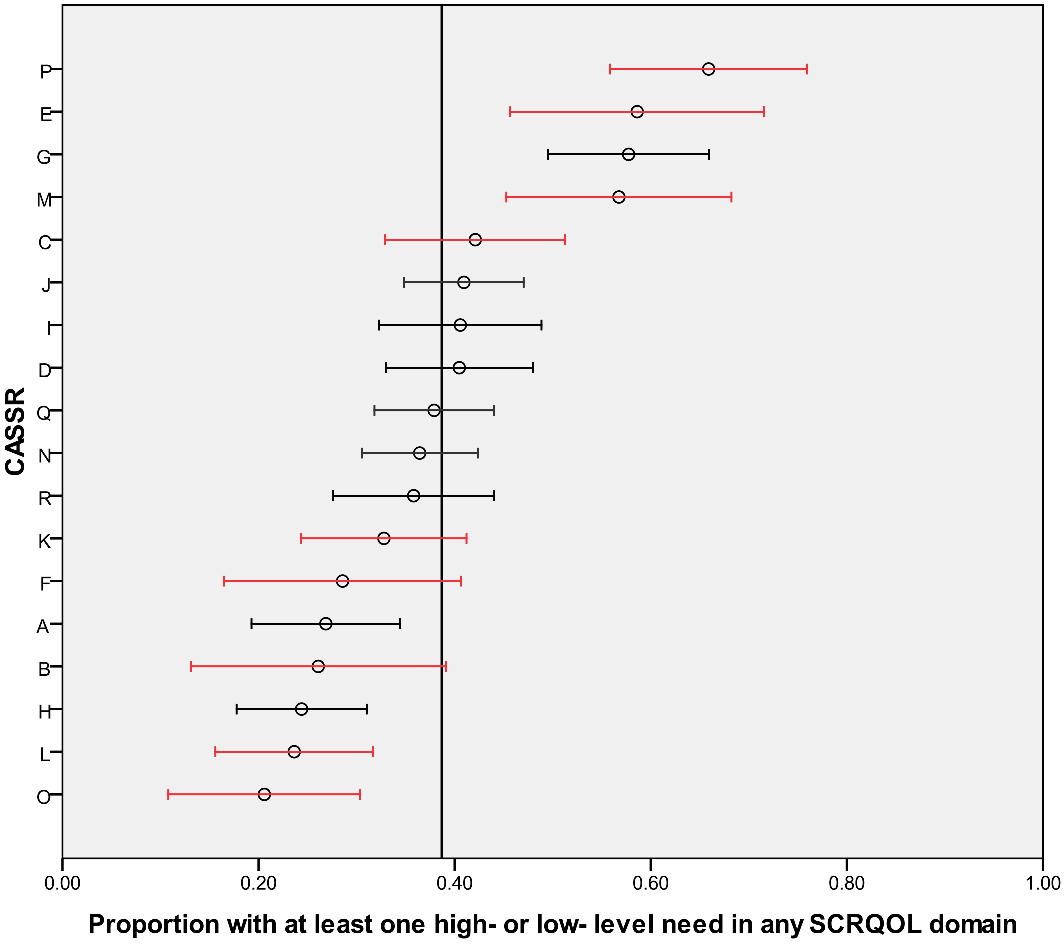


Figure 18: Proportion with at least one need in a basic SCRQoL domain for CASSRs in sample (with errors bars to show differences between CASSRs)

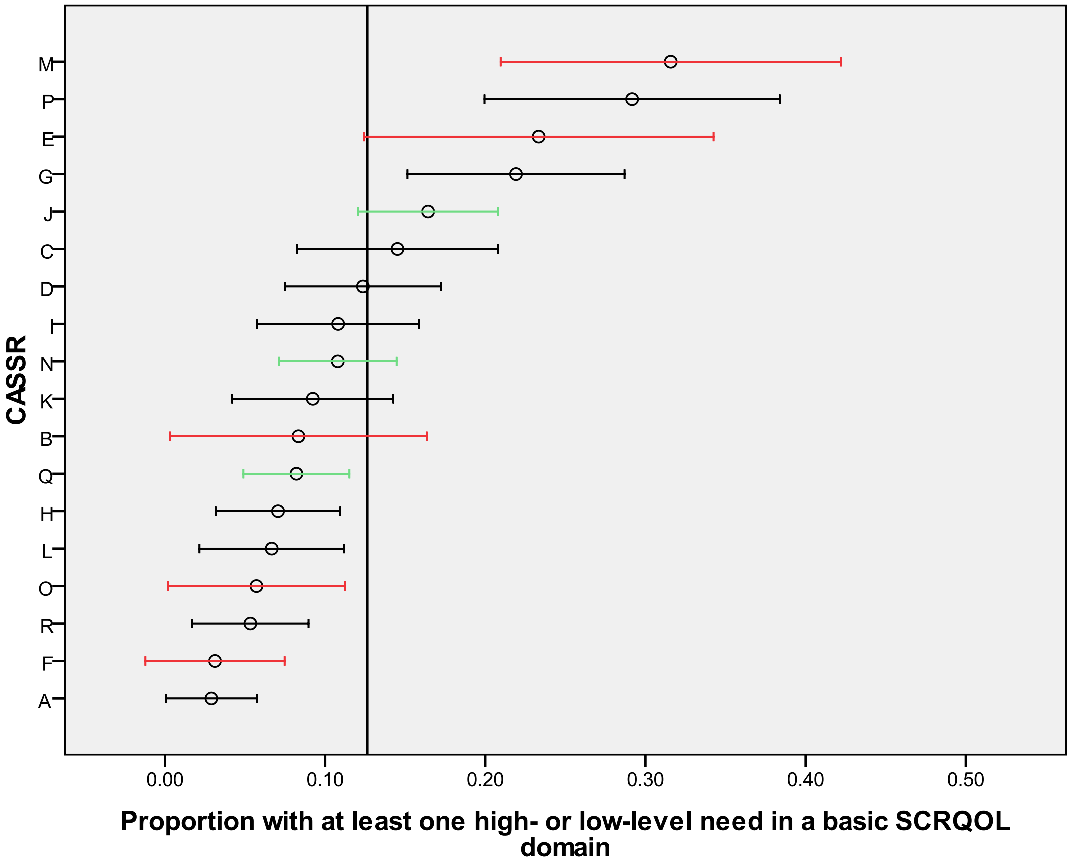


Figure 19: Proportion with at least one high-level need in any SCRQoL domain for CASSRs in sample (with errors bars to show differences between CASSRs)

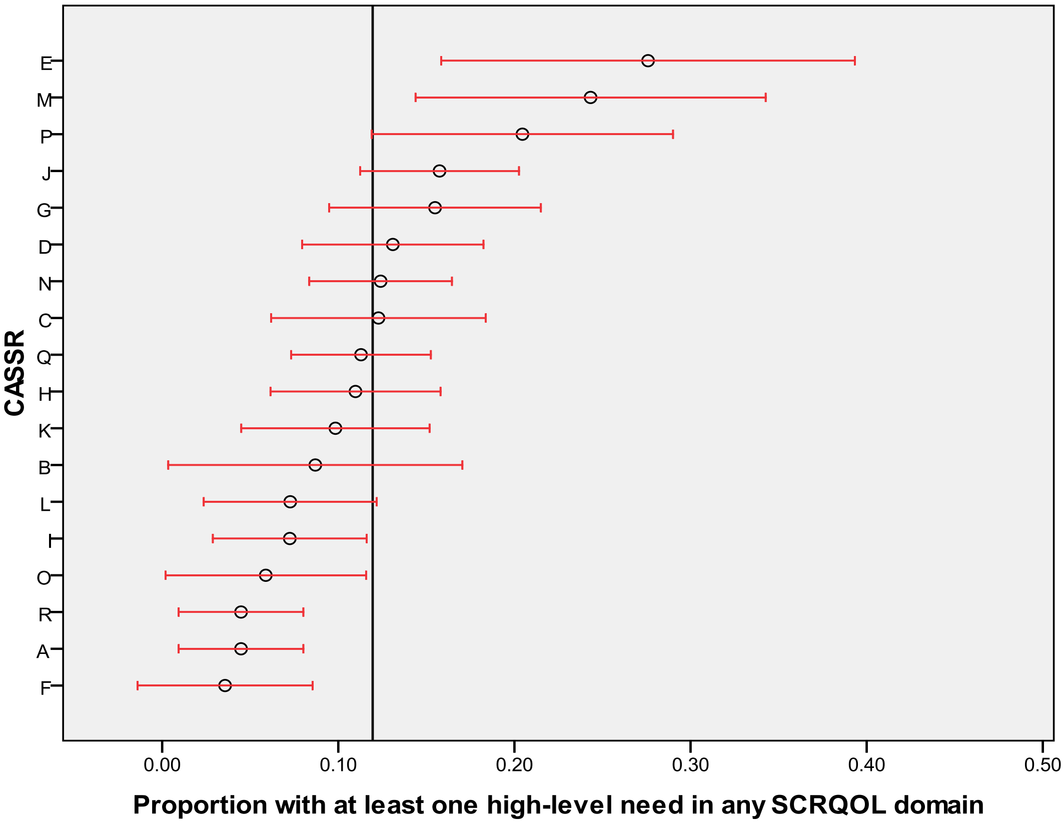
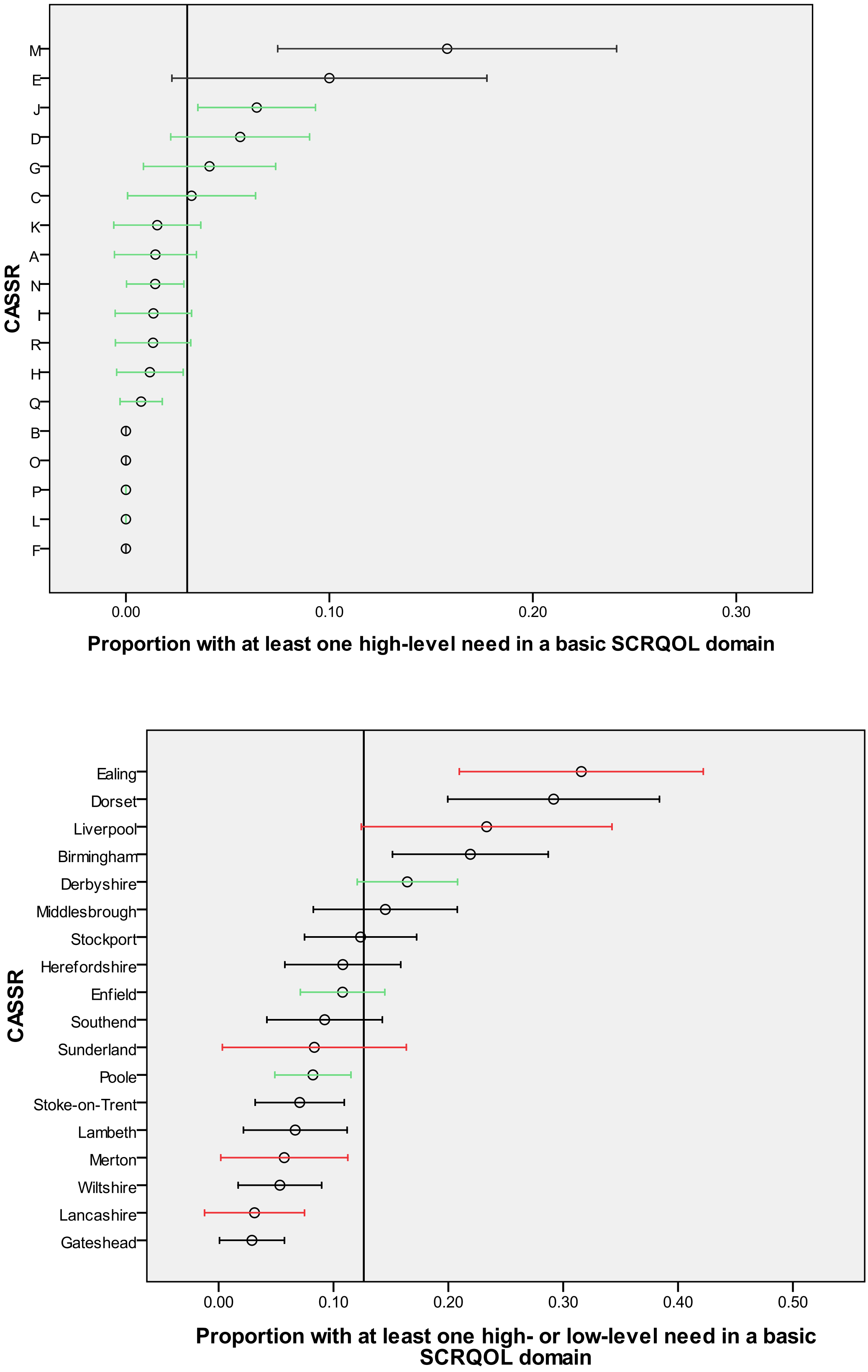


Figure 20: Proportion with at least one high-level need in a basic SCRQoL domain for CASSRs in sample (with errors bars to show differences between CASSRs)



Point estimates for CASSRs with red error bars have less than 80% reliability. Those with green error bars have greater than 90% reliability.

Appendix 7: Funnel plots for each indicator

The following set of graphs is a set of funnel plots showing the scores for each CASSR on each indicator. Separate markings are used to indicate the CASSRs that participated in the full pilot from those that participated in the care home or LD pilot.

Figure 21: Funnel plot showing score for each CASSR on the SCRQoL indicator

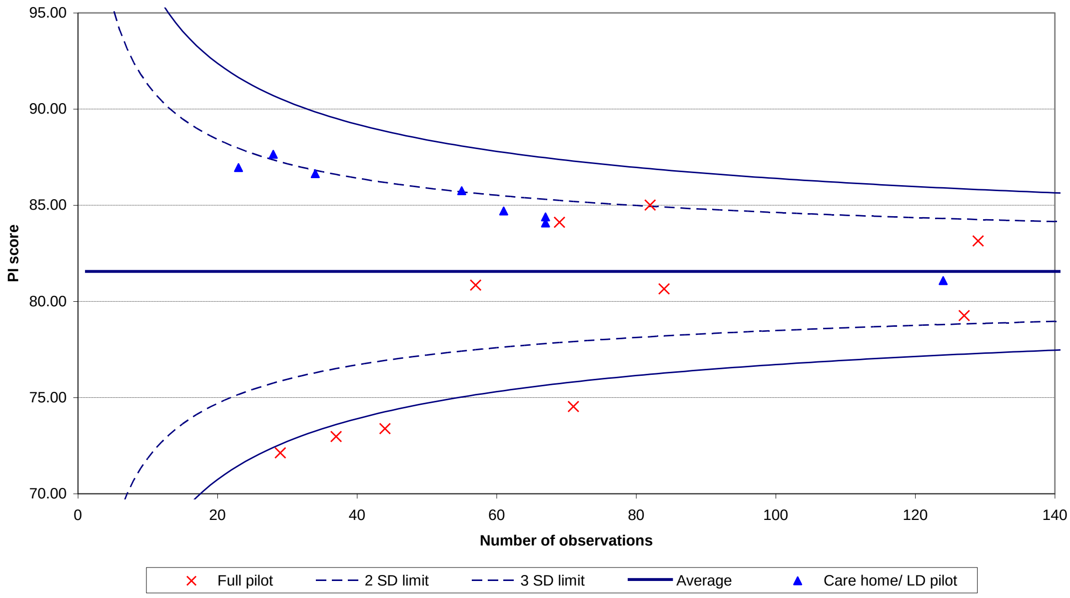


Figure 22: Funnel plot showing score for each CASSR on the PW-SCRQoL indicator

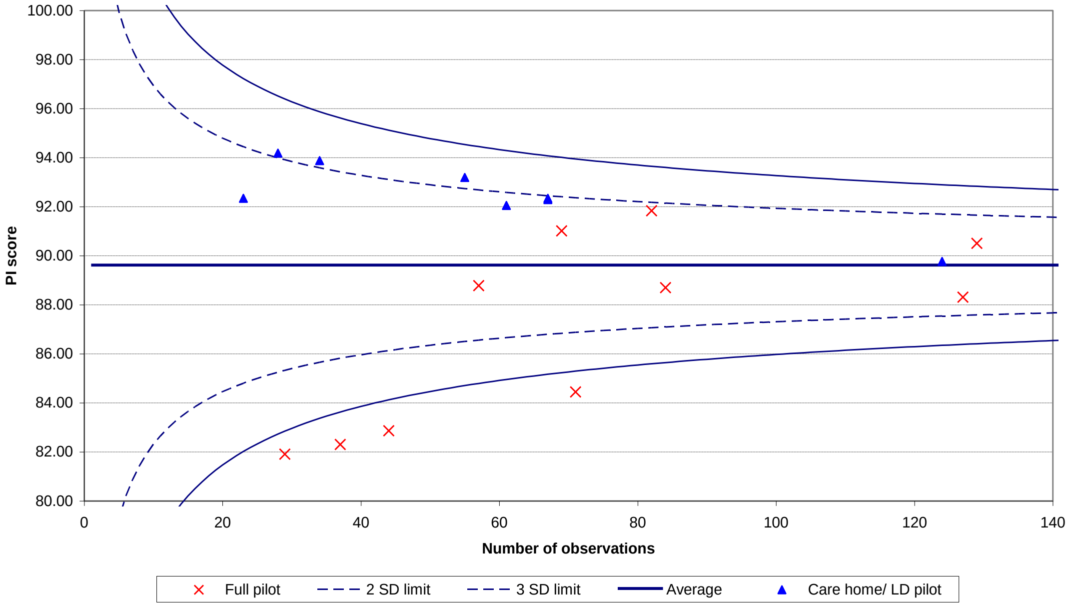


Figure 23: Funnel plot showing score for each CASSR on the proportion of people with at least one high- or low-level need in any SCRQoL domain indicator

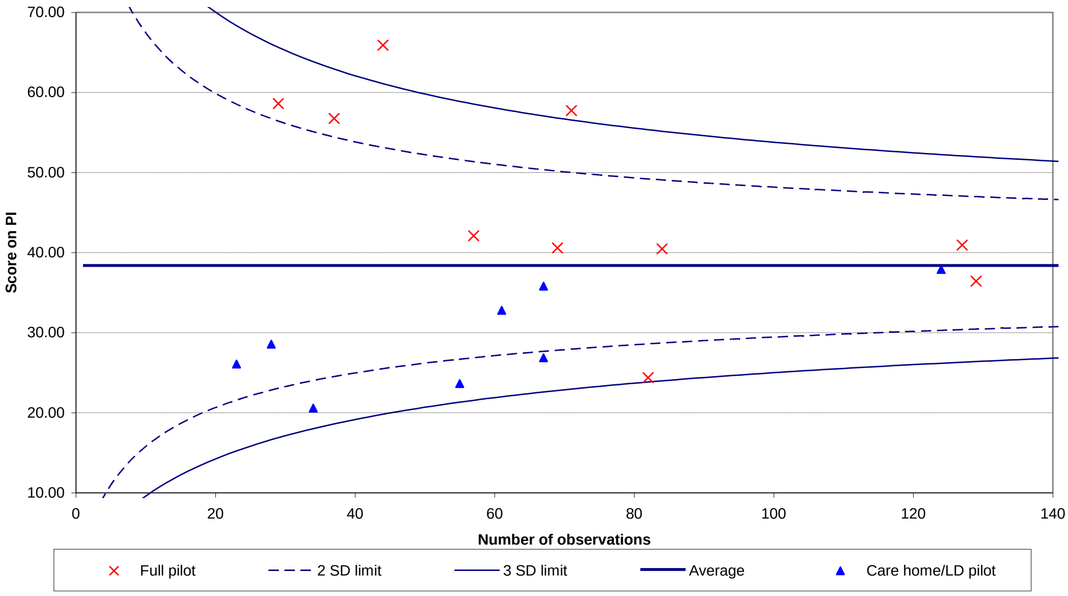


Figure 24: Funnel plot showing score for each CASSR on the proportion of people with at least one high- or low-level need in a basic SCRQoL domain indicator

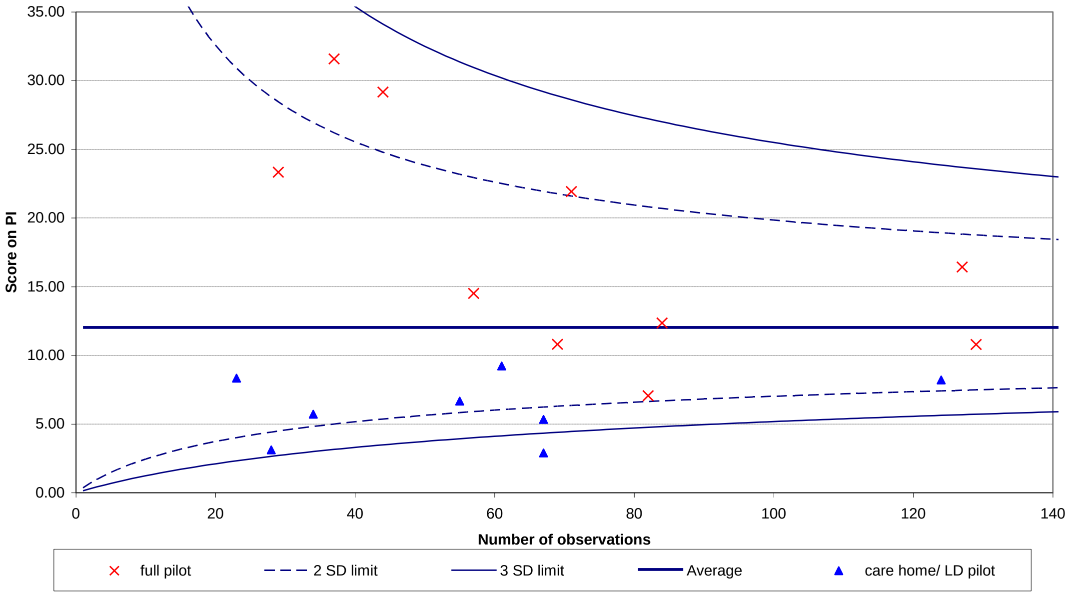


Figure 25: Funnel plot showing score for each CASSR on the proportion of people with at least one high-level need in any SCRQoL domain indicator

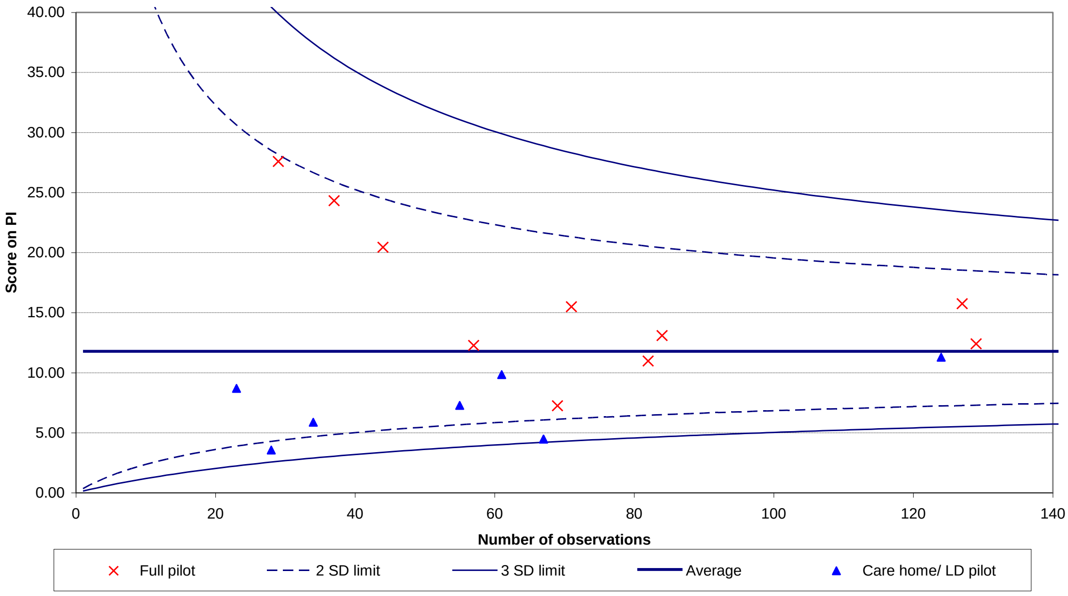


Figure 26: Funnel plot showing score for each CASSR on the proportion of people with at least one high-level need in a basic SCRQoL domain indicator