PSSRU

Using Survey Data to Measure Changes in the Quality of Home Care

Analysis of the Older People's User Experience Survey 2006

Juliette Malley, Ann Netten and Karen Jones

Discussion Paper 2417~2
November 2007

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
Cowdray House
Houghton Street
London WC2A 2AE
Tel: 0171 955 6238
PSSRU@lse.ac.uk

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

www.PSSRU.ac.uk

Contents

Index of Tables and Figures

Acknowledgements

This work was funded by the Department of Health. Our thanks are due to the Social Services User Survey Group, for their suggestions, help and support in getting the study underway and to the ADASS for their support. We are grateful to the participating councils, in particular our liaison contacts. We are also very grateful to Lesley Cox for all her help and support in co-ordinating the data collection.

The views expressed in this report are those of the authors and do not necessarily reflect the views of the Department of Health.

Executive summary

1. The development of the User Experience Survey (UES) and the inclusion of the user's perspective in the Performance Assessment Framework (PAF) were novel and important steps forward in the world of performance assessment and monitoring of social services. Commentators had previously criticised the lack of connectedness between the performance/quality and user involvement agendas (Beresford et al., 1997; Watson, 2002) and these developments represented real steps forward in meeting this goal. Analysis from the 2003 extension to the UES for older people has demonstrated that user experience measures can usefully discriminate between Local Authorities (LAs). While this new research raises some important questions, this type of approach is clearly an improvement on performance measurement based on processes and inputs.

2. Fifty authorities took part with representation from all Government Office Regions and LA types. Analysis of responses to the compulsory questions and comparison with responses to these questions nationally revealed very similar proportions responding to each category. We can assume from these findings that the sample of authorities is broadly representative of the views of service users across England.

3. Non-response to the survey was quite high, with response rates at on average 58 per cent. However, this figure is identical to the average across England as reported by the Information Centre (2007). Non-response did vary considerably by LA and anecdotal evidence suggests that this may be partly explained by differences in practices across authorities, in terms of the degree of over-sampling and chasing. Procedural differences may affect comparability of the results between authorities and also their collation to provide a national picture. Future surveys may wish to consider different approaches to sampling or data collection that remove these problems.

4. Gender, age, ethnicity, provider type, number of providers, and planned hours of home care were all found to vary with non-response. This meant that the respondent population was composed of a disproportionately high number of females, people aged 75 and over, people with a White ethnic background, people receiving in-house provision, provision from one provider and fewer than 5 hours of home care per week according to Local Authority (LA) records. These variables were found in the 2003 extension study to vary with quality, but they did explain a very small amount of the variance (Netten et al,. 2004). It is hoped that this will mean the non-respondents will have very little effect on biasing estimates, but this is not necessarily the case. Further research will be conducted

to explore the impact of non-response on estimates of satisfaction and quality using data for which this analysis is possible.

5. Non-response to questionnaire items for those who had responded ('item non-response') was acceptable for most variables, at around 5 per cent, but was quite high for several variables at over 10 per cent. Those in the latter category included some of the 'auxiliary' variables collected using LA records (provider type, planned hours and number of providers), questions presented in a grid format (the questions on care worker attitudes and outcomes), questions asking for either a 'yes' or 'no' response (questions about ADLs), the question on direct payments and the question asking service users how many hours of home care they receive.

6. The high item non-response indicates different problems with the questions depending on the type of question. If auxiliary variables were missing, they were generally missing for whole LAs at a time; LAs reported in these circumstances that they were unable to provide these data. A relationship was found between satisfaction and missing items: cases where the auxiliary variables were missing were more likely to be less satisfied with their services, suggesting that an authority's ability to access information is an important factor in ensuring service users are satisfied with their services.

7. The relatively high non-response to some questions in the questionnaire indicates that the questions potentially posed some difficulties to users when they tried to answer them. This could be because users were not able to remember or did not know the answer, for example to what services they received, or they felt the questions were intrusive, for example the questions on disability. It could also reflect problems with the phrasing of the questions. For all questions with high item non-response, cognitive testing could be considered to try to understand what types of problems users may be having with the questions. Where cognitive testing indicates problems with phrasing, reformulation of the questions (followed by cognitive testing of the reformulated questions) could be considered to try to improve response rates.

8. Overall, people were more likely not to answer further questions if they were less satisfied with their services, but certain of the missing questions had interesting relationships with satisfaction beyond this more general relationship. For the grid questions it was found that negatively phrased statements were less likely to be answered by people with higher satisfaction, whereas positively phrased were less likely to be answered by people with lower satisfaction. This relationship is likely to bias estimates of

quality that are derived using these types of questions. For this reason we have restricted analysis to our measure of service quality which does not include any grid questions.

9. The question on hours of care also seemed to pose problems. Comparison, at the individual level, of planned hours with reported hours found large discrepancies for a large proportion of the data. Even excluding extreme outliers, which represented about 14 per cent of all cases in which we had data for both questions, the standard deviation was around 2hours and the range 11hours. This seems improbable and leads us to assume that the reporting of hours of home care by either of or indeed both service users and LAs is inaccurate. Reformulation of these questions may want to be considered for future surveys.

10. Analysis of responses to questions across the sample of participating authorities revealed large differences across all variables by authority type and by authority. London boroughs in general had the least positive responses, but within each authority type there was a wide variation in responses. This indicates that service users have very different experiences of the quality of services across authorities. Exploration of what accounts for the different experiences within each authority would be an interesting extension of this work and would help authorities to think about how they can improve their services for users. Some of these questions have already been considered by PSSRU studies looking at the relationship between provider characteristics (Netten et al., forthcoming) and quality and through the analysis of the 2003 extension study (Netten et al., 2004). Further analysis from this study will explore this question in more detail.

11. Comparison with 2003 indicated that, amongst LAs taking part in both years, there was an increase in the proportion of service users receiving provision from multiple providers and an increase in the proportion of service users with provision from independent providers. This is compatible with national data reporting a similar trend in increasing independent provision of services (Commission for Social Care Inspection 2006,b).

12. Comparison of service user views between 2003 and 2006 showed a complicated picture for the 2006 compulsory questions. Satisfaction with services as measured by the performance indicator (PI) remained roughly the same; although the number reporting they were extremely satisfied fell, but the number reporting they were very satisfied increased. Views associated with the flexibility of the service worsened and views associated with outcomes (safety, social participation and control over daily life) improved. This pattern was not repeated for non-compulsory items. For these, all views associated with outcomes worsened, including personal care and employment and

occupation outcomes. For the other items, views associated with the reliability and continuity of the service also worsened. The picture was more complex for views associated with staff attitudes and staff skills and knowledge, with some improvement and some decline.

13. Validation of the quality measures derived from the 2003 extension study was only carried out on the service quality measure for reasons identified above. Reliability and validity as measured by the relationship with satisfaction and service quality in 2003 were good. However, content validity of the 2006 measure was compromised by the lack of a question covering the domains of communication, staff skills and knowledge and outcomes. The original service quality indicator did include a question on communication but it had to be dropped as the question had changed for 2006. Analysis of reliability and validity including this item indicated that it was important and should be a part of the measure. Leaving the communication item aside, overall the outcomes, carer quality and service quality measures capture the majority of the content domains of quality, but some areas are missing, such as food and nutrition and accommodation cleanliness, order and accessibility. Questions to cover these domains may want to be considered for future surveys, so quality, particularly with respect to outcomes, can be assessed more fully.

14. Validation of the weighted satisfaction measure was conducted using the service quality measure for the reasons identified above. It was felt that this should give a good indication of the performance of the overall quality measure as the majority of the questions composing the service quality variable were used for overall quality. The analysis looked at whether the weighting structure, meaning whether there were still four different levels of quality associated with the different responses to satisfaction, was the same for both years. Although the structure was the same, the values for the weights were different depending on whether they were calculated using the 2003 or 2006 dataset. Each significantly different level of satisfaction was associated with a lower average value of service quality for 2006 compared with 2003. This implies that people are as satisfied with their services despite reporting receiving lower quality services along the dimensions of reliability, continuity, flexibility and staff attitudes. This finding may present problems to the use of the satisfaction item for measuring quality change. We explored this questions raised by this measure, by examining change in quality, as measured by both the service quality indicator and satisfaction.

15. Examination of change in quality from 2003 to 2006 was conducted using the satisfaction and service quality measures. No significant change in satisfaction between 2003 and 2006 was found, whether using the PI or the mean value for satisfaction. When the

analysis was repeated using cases that had answered both of the service quality and satisfaction questions, average satisfaction was found to improve between 2003 and 2006, but the PI showed no change. In contrast, service quality was found to decrease over the same period. These results indicate that the PI is not as sensitive as the mean satisfaction value to changes, which has some importance for its value as a PI. If the measure is not sensitive to change in user experiences of services then it is perhaps not an adequate measure of quality. The findings also suggest that the service quality and satisfaction measures are not measuring the same things or that something is interfering in their relationship as was implied by the analysis of the weighted satisfaction measure. In this respect these two analyses are in agreement.

16. It is unclear from the data why this relationship between average satisfaction and service quality is observed, although we have three hypotheses:

    i. It is possible that the service quality measure does not capture aspects of quality that have improved over the period. If we assume that satisfaction measures quality and *nothing else* then this could explain the findings as we have already noted that certain aspects of quality, for example outcomes and staff skills and knowledge, are not included in the service quality measure. However, the evidence from the analysis of individual questions asking about these missing areas does not support the hypothesis that there has been overwhelming improvement in other areas. It is also possible that the missing areas (that have improved) are more important to people, hence they have a stronger influence on their satisfaction, but evidence from previous work, combined with evidence about which areas have improved, again does not seem to support this proposition (Francis and Netten, 2004).

    ii. It is also possible that the satisfaction measure captures things other than quality and that it is too 'noisy' an indicator to use to measure quality. Satisfaction is the user's evaluation of their experience and not a report of their experience as the measures that compose the service quality measure are. This makes the satisfaction measure more likely to reflect things other than just experience, expectations, and more personal preferences about services and what they should do could influence a persons' evaluation of the service and these can be influenced by variables well beyond the control of councils. The literature seems to support this view of satisfaction as being a 'noisy' measure and if we reject hypothesis one this seems to be a reasonable explanation of the finding.

iii. Another possible reason is that there is gaming associated with the satisfaction indicator. Gaming associated with the reporting and collection of statistical data is found where the data are used to control the behaviour of organisations. We have described three types of effects widely reported in the literature and known as: ratchet effects, threshold effects and output distortions and have discussed that the proportion responding 'extremely' or 'very' satisfied to the question about satisfaction with services acts like a target, with the target informed by both the bandings and last years' performance. We suggest that if misrepresentation of data exists there would be a larger peak at 'very' satisfied compared with 2003 and this is indeed what is observed when the 2006 distribution of responses to the question is compared with the 2003 distribution.

17. We have discussed three possible explanations for the lack of consistency in the relationship between satisfaction and service quality between 2003 and 2006. It is likely that each of these explanations contributes to a greater or lesser extent to solution. However, we have argued that the second (satisfaction is a 'noisy' indicator) and the third (gaming) are more probable. We have suggested, however, that our confidence in this conclusion could be improved by further analysis investigating the relationship between satisfaction and the other derived variables from 2003. However, before attempting this analysis we need to try to adjust for non-response.

18. The problems we have identified have consequences for the use of measures. First, in their use of measures of quality change, gaming and the 'noisiness' of the measure make its use difficult. The former because it affects the validity of the measure and the latter because it makes movements in the indicator difficult to interpret. We have also discussed that the measure appears to be insensitive to changes in user experience, which makes it inappropriate for use as a measure of quality change. Second in its use as a PI, the same problems of interpretation and validity are also important but the latter because of its detrimental effect on public trust. Satisfaction is currently viewed in the performance assessment framework (PAF) as a measure of quality and this work challenges this view. Whilst satisfaction does reflect quality, it does not measure only quality and its position in the framework could be considered. (Boyne, 2002) describes satisfaction rather as a measure of responsiveness to citizens and consumers, which could be viewed as an aspect of quality, and this is perhaps a better description of what it measures.

19. Solutions to the problem of gaming have been discussed by a number of authors in the literature and most authors are clear that losing targets and indicators altogether is not a solution as the incentives attached to them can and do have positive effects. Their

solutions are usually to try to reduce or lessen the impact of the negative effects. So called 'anti-gaming strategies' could include introducing greater uncertainty into the specification of PIs or in this situation collection of the data by an independent third party would reduce the possibility of misrepresentation and opportunities for gaming.

20. Solutions to the problem of 'noisiness' are more complicated as they depend on what it is intended to measure. For the work on outputs and for the PI, based on the domains described in the PAF, a measure of quality is required. We have suggested that a composite measure based on the service quality measure could be a good alternative measure. Its benefits over the satisfaction measure are:
    i. It is easier to interpret since the items that constitute it are reports of experience of what happened rather than evaluations of what happened so are less susceptible to influence by attitudes or expectations towards the service.

    ii. It is more reliable than a single item measure as summation across several items all measuring quality provides a more rounded picture.

    iii. It is capable of finer discrimination as it is a composite measure making it a better measure for capturing change.

    iv. It is more complicated so potentially less open to gaming.

21. Although the current service quality measure is a good start and, we feel, the direction for future measures of quality, we have identified some weaknesses with it due to the fact that it was developed from a questionnaire not designed for the purpose of scale construction. These problems with the questionnaire that have been discussed and the problems with the scale that are consequent to them would need to be overcome through redevelopment of the questionnaire and further development and testing of the scale. As we have suggested there are certain assumptions underlying the construction of a composite measure and these would need to be rigorously tested.

22. We have also identified multiple and in some cases competing interests in a measure of quality. The various end-users – Local Government performance officers, central Government, CSCI, ONS, evaluators of policy, and so on – have in some instances different requirements for the measure and balancing these will be essential. The development of any measure of quality will benefit from engagement with and input from these end users.

1. Background

The mandate for conducting surveys of user's experiences and satisfaction with services was first given in the white paper *Modern Local Government: in Touch with the People* (Department for Transport Local Government and the Regions, 1998). For some time now councils with social service responsibilities (CSSRs) have been required to conduct a user experience survey (UES) for at least one client group per year. In 2002 the Office of National Statistics (ONS) and SPRU developed a set of questionnaires for this purpose (Qureshi and Rowlands, 2004). Subsequent national surveys have drawn on this work to identify a set of compulsory questions for each round.

The surveys were seen as part of the new Performance Assessment Framework (PAF), with the aim being to publish the compulsory questions so that authorities could compare their performance and central Government could monitor progress nationally (Department of Health, 1998). In line with this aim, certain questions have been chosen within each survey as performance indicators (PIs) and these have been reported nationally within the PAF and as part of the Best Value initiative. In 2003, the first year of the older people's survey, four items were chosen as PIs, with two being Best Value PIs. The items chosen were:

- Overall how satisfied are you with the help from Social Services that you receive in your own home?
- If you ask for changes in the help you are given, are those changes made?
- Does anyone contact you from Social Services to check you are satisfied with the home care that you receive?
- Do care workers come at times that suit you?

The first two of these questions, on satisfaction and changes, were the Best Value PIs.

The PAF was split into several parts to capture the various aspects of performance. These were identified as: national priorities and strategic objectives, cost and efficiency, effectiveness of service delivery and outcomes, quality of services for users and carers, and fair access (Department of Health, 1998). Under quality, the Government provided examples of potential PIs, including delayed discharge from hospital, proportion of residents provided with single rooms, and user and carer satisfaction surveys. These items and the PIs chosen from the user satisfaction surveys need to accurately reflect the quality of services, in order that they add something new to the assessment of performance.

The inclusion of the user perspective in the PAF was a novel addition to the framework under the current administration; it is also central to policy. Indeed, commentators have been quick

to criticise the lack of connectedness between the performance/quality and user involvement agendas (Beresford, et al., 1997; Watson, 2002). Assessments from the user's perspective create the right kind of incentives for local authorities to improve their services and are also the closest measures to measures of outcomes, since users include in their evaluations of quality, amongst other considerations, their perceptions of their quality of life or the actual outcome of the service (Qureshi and Rowlands, 2004).

Given that the aim of the modernisation agenda is to improve the quality of services, the question of whether services are actually improving is of interest. Combined information from performance information and inspection reports in the form of the star ratings show that services are improving and have been for the past four years (Commission for Social Care Inspection, 2006a). However, a recent report by CSCI was more critical stating that although surveys showed widespread satisfaction, more detailed interviews and group discussion elicited more critical feedback especially concerning: care workers being rushed, shortness of visits, timing of visits and reliability (associated with care workers rushing between visits and turning up late). They also found that the experience of staff members was that services were short staffed. CSCI suggested that satisfaction ratings indicate people's gratitude for the service and the value they place on the relationship with care workers (Commission for Social Care Inspection, 2006b). The evidence for quality improvement seems to be inconclusive and the repetition of the UES for older people in 2006 offers us a unique opportunity to gain a fresh perspective on this question since it covers not just satisfaction but questions designed to capture the exact aspects of quality alluded to in the CSCI report.

1.1 The 2003 UES extension study

The questionnaire used for the 2003 extension study was based on a questionnaire, developed by the Office for National Statistics (ONS) with support from the Social Policy Research Unit (SPRU) at York University (Qureshi and Rowlands, 2004), with some amendments introduced by PSSRU as a result of the findings from a study to establish the key aspects of quality of home care (Francis and Netten, 2004). SPRU provided the ONS with briefings based on research into user experience and satisfaction derived from various studies (Qureshi, et al., 1998; Qureshi and Rowlands, 2004). These studies had identified various domains of quality and outcomes as assessed by service users and their carers and this informed the coverage of the questionnaire; cognitive testing was used to shape the exact nature of the individual items and the construction of the questionnaire itself. The results of the PSSRU study led to the introduction of additional items that covered domains of meaning not captured in the ONS/SPRU questionnaire. For example a question on communication was added (Francis and Netten, 2004; Netten et al., 2004). Thirty-four authorities took part in the 2003 study.

The 2003 extension study led to the development of a number of indicators of quality derived using a factor analytic procedure on the questionnaire items. These factors were an overall quality measure (the one factor solution), a service quality measure, two carer quality measures and an outcomes quality measure (the four factor solution). The derived variables, ostensibly measures of the various aspects of quality, were then used to validate the various PIs. The method of validation used sought to confirm the construct validity of the PIs by comparing the PI measure, itself supposed to be an indicator of quality, to the derived quality measures. The study found evidence to support the use of the satisfaction PI on the basis of a statistically strong relationship between the PI and the various quality indicators (Netten et al., 2004).

Although the primary use of the quality measures in this study was to validate the PIs, developments in other areas, particularly around the measurement of outputs for National Accounts, means that there is an increasing interest in a measure that can reflect the quality of social services at a national level (Atkinson, 2005). Such a measure needs to draw on current and ongoing statistical collections and it was proposed by PSSRU at Kent that for home care for older people (and perhaps other services and client groups should the UES become a regular collection) the UES could be used for this purpose (Netten et al., 2006). A weighted satisfaction measure was proposed that weights the proportion of the English population responding to each significantly different level of satisfaction (extremely, very, quite and neutral/dissatisfied) by an index of quality. The index of quality is constructed from the value of the overall quality measure at each level of satisfaction. Since the quality items are not collected regularly, for this measure to work, the relationship between overall quality and satisfaction, which determines the weights, needs to remain constant from year to year. The 2006 extension study was seen as an opportunity to validate the weightings and the other measures of quality.

The 2003 extension study also uncovered several factors associated with perceptions of quality, as measured by the derived quality factors. Individual and service characteristics included: gender, age, ethnicity, receipt of practical help, reported intensity of home care provision, number of different providers, and type of provider (in-house, independent or both) (Netten, et al., 2004). The relationship between quality and type of provider was particularly interesting and a separate study of providers was conducted to investigate this in more detail (Netten, et al., forthcoming). Area level associations were also uncovered showing that better experiences were associated with lower average weekly expenditure for home care per person, being resident in metropolitan areas, higher hourly cost for home care, and lower employment and local wage rates. However, the authors note: ‘Although significantly associated with users’ experiences, these factors explained a very low proportion

of the variation in the indicators of home care quality.' (Netten et al., 2004). They posit that this could be a result of non-inclusion of many important likely predictors of people's experiences; it could also demonstrate the importance of quality as something that is unique and helps to shape people's experiences of services.

Functional ability and levels of morale were suggested as possible predictors as the 2003 study found that more positive perceptions of home care were significantly associated with receiving 10 hours or less of home care. Although service use is determined by need, it is only a proxy to need since other factors intervene, in particular receipt of informal care and ability to pay for services (Arber et al., 1988; Evandrou, 2005; Pickard et al., 2000; Wittenberg et al., 2006), so a measure of functional ability is needed to assess this hypothesis. It was acknowledged in the 2003 study that many of the individual characteristics that were associated with perceptions of quality were likely to be related to high dependency on services and low functional ability with associated low morale. The 2006 extension study was also seen as an opportunity to test this theory.

2. Aims and objectives

The principal aims of the research are to:

- add value to the UES for a sample of participating local authorities by enhancing comparability across dimensions of quality not included in the compulsory items and for authorities taking part for the second year, comparability across time is also enhanced
- validate the measures of quality for wider use, including the use of the results for measuring the value of personal social services (PSS) outputs and potentially efficiency gain
- explain variation in reported quality within the 2006 study
- investigate any variation in reported quality between 2003 and 2006.

Individual reports have been provided for participating councils to facilitate the first aim.

This report focuses on the evaluation of the conduct of the survey, the validity of the derived quality measures and the weighted satisfaction measure, variations in quality between 2003 and 2006.

3. Method

In 2005, all LAs were approached to see if they would like to participate in an extension to the 2006 survey, following initial contact made by the Information Centre for Health and Social Care (IC) on behalf of the DH. Participating authorities all agreed to use the same questionnaire, which was principally the questionnaire used for extended survey in 2003, with minor amendments reflecting developments in our thinking. The data collection process followed the guidance set by the IC, which included requirement for collection of some auxiliary data from LA data systems (2005). LAs were also responsible for entering and returning the anonymous data in a common format (Excel worksheets) to PSSRU for analysis.

3.1 Additions and changes to the questionnaire for 2006

The 2003 survey identified several problems with the questionnaire. The positively and negatively phrased statements of attitudes, presented in grids, had quite high (item) non-response in the range of 11 to 31%, with higher rates of non-response for the negatively phrased questions. The advisory group decided to rephrase the negatively phrased statement that was included in the compulsory questions to be positively phrased on this basis. However, this approach to all negatively phrased statements was rejected as an option by the PSSRU on methodological grounds. Attitudinal statements need to have a balance of both positive and negative statements to ensure that people are answering and reading the questions correctly (Nunnally, 1967). It would also not solve the high non-response rates to the positively phrased statements. The PSSRU considered dropping these questions altogether; but, given the overwhelming contribution they make towards the quality variables, it was not possible to exclude them without finding replacements and time was not available to undertake the testing required to do so rigorously.

Another issue identified in the 2003 study was that value could be added to the analysis by introducing questions that provided more information about user characteristics, in particular information about disability level in terms of functional dependency and level of morale. Work undertaken by the PSSRU for Kent County Council in 2005 indicated that questions on ability to perform a selected number of activities of daily living (ADLs) and instrumental activities of daily living (IADLs) and a question on perceived health would help in this regard. Another question on length of time in receipt of services was introduced on the basis of literature identifying a potential relationship between this and satisfaction (Calnan et al., 2003).

The perceived health question was based on the five point scale suggested by Robine and colleagues as part of a European project on health indicators (Robine et al., 2002). The ADLs

and IADL questions were modelled on the 3-item Barthel questionnaire (Ellul et al., 1998). However, the continence question was dropped as it was felt that this was inappropriate for a self-completion questionnaire. A replacement question was introduced that was based on the GHS item 'wash face and hands'. Analysis of distributions of respondents to a survey of over 300 home care service users conducted in 2005 (Darton et al., 2006) implied that this worked well as a replacement. An IADL question was also introduced to capture lower level needs as the Barthel index tends to not capture lower levels that may be more prevalent in a population living in the community. The IADL was chosen on the basis of analysis of the Kent home care survey indicating that it had the best distribution amongst home care recipients (Jones et al., 2005).

LAs also had some concerns in 2003 about the length of the questionnaire and its effect on response rates. Although analysis demonstrated that these concerns were unfounded, it was decided to try to reduce the length of the questionnaire. Certain questions that were not found to be associated with satisfaction or part of the quality variables were dropped (for example, a question asking about the respondent's knowledge of their care plan). It was also decided to try to reduce the number of the items in the grid format to respond to concerns about these questions. Items (my care workers are miserable, understanding and gentle) were excluded on the basis that they had the least variability across respondents. In addition, there were several changes to the ordering of questions to accommodate the concerns of the advisory group.

There is a wide literature reporting how changes in phrasing and ordering of questions in questionnaires can affect the way people respond (Kalton and Schuman, 1982; Schuman and Presser, 1996) and it is possible that some of these re-wordings and re-orderings could affect the way people respond to questions and hence the reliability of the derived quality factors. Important changes affecting the quality scales were the following:

Dropped items:

- The help I get from social services has made me more independent than I was – included in overall quality factor
- My care workers are gentle – included in overall and carer quality factors
- My care workers are understanding – included in overall and carer quality factors
- My care workers are miserable – included in overall and carer quality factors.

Changed items:

- The item "I don't feel safe in my home" was changed to "I feel safe in my home" – included in overall and outcomes quality factors
- The item "I always feel clean" was changed to "I always feel as clean as I want to be" – included in overall and outcomes quality factors

- The item "Are you kept informed, by your home care service, about changes in your care? (e.g. your visit will be late or you'll have a different carer)" was changed to incorporate four responses in 2006 compared with three for 2003. – included in the service quality factor.

The full questionnaire for the 2006 study is shown in Appendix A.

Since some of the factors identified in 2003 included items not in the 2006 questionnaire, the carer, service quality and overall quality factors had to be revised using only those questions present in both years. Analysis of the 2003 dataset indicated that loss of these factors did not impact too significantly on the reliability of the quality scales. The outcomes scale was, however, not shortened to exclude the safe item as it was felt that this change should not have an effect on the integrity of the scale, although it is acknowledged that it might have an effect on the reliability through changes in the way people respond to the new statement. The items that could be included in the scales compared with those that could have been in 2003 are summarised in Appendix B.

3.2 Analyses

Several different analyses were planned to investigate each of the aims. These are described under the following headings:

- Validation of the derived quality measures
- Change in quality from 2003 to 2006
- Local variations in quality.

3.2.1 Validation of the derived quality measures

Validity is a difficult concept since it is defined in terms of the operational theory, in other words a valid measure is one that does what you intend it to do. Various types of validity have been described for measures and (Nunnally, 1967) summarises these under three headings: content, predictive and construct validity, stating that the type of validity that is most important depends on the function of the measure[1]. The quality measure required for use in the work measuring outputs is one that captures fully all the aspects of quality that would be associated with social services; it is important that certain aspects of quality are not emphasised over other aspects, unless there is a good reason for doing so. It is also important that the measure of quality distinguishes between poor and high quality services and is sensitive to changes in quality. From this perspective, establishing content validity is most

[1] Predictive validity is determined through examination of the relationship between what is measured and the criterion, or what is predicted from the measure; content validity is about the adequacy of the domain of content that is sampled; and construct validity is applicable when the measure is capturing something abstract and is about the extent to which the measure adequately represents the abstract variable of interest.

important. Beyond that, establishing construct validity, that quality as measured through the questionnaire items is related to other constructs, such as satisfaction, as expected, is also important. Predictive validity is not as important as the point of the measure is to describe quality within local authorities not predict future quality or some other characteristic.

Content validity can be assessed by examining the extent to which the items in the measure of quality capture all the domains of quality identified in the literature. These are identified as falling into two areas: quality associated with the process or act of providing care and quality associated with outcomes from that care. Aspects of quality identified from the literature and explored in in-depth interviews with older people (Francis and Netten, 2004) were:

- Continuity
- Flexibility
- Reliability
- Communication
- Staff attitudes
- Staff skills and knowledge.

Work on social care outcomes by Netten and colleagues (Netten et al., 2006; Netten et al., 2005) has identified eight domains of outcome applicable across all client groups:

- Personal safety
- Personal cleanliness and comfort
- Social participation and involvement
- Control over daily life
- Employment and occupation
- Meals and nutrition
- Accommodation cleanliness, order and accessibility
- Role support.

Construct validity is also important and to assess this validity a clear theoretical model is required, since we are really trying to determine whether the measure fits with other measures in the way we expect according to our best theory[2]. If construct validity is about identifying whether expected relationships are observable, then examination of construct validity is limited by the other items in the questionnaire or other items that can be collected. We might expect quality to be related to a number of variables that were not collected, for example the

[2] There is a circularity to this argument which has not been clearly resolved. If measures are judged by their ability to fit with theory, then they will only ever be as good as the best theory. Measures will need to develop as theory changes.

unit cost of the service, various other economic conditions and the state of the service in the previous study. We would also expect that people receiving good quality services would be more likely to be satisfied with their services, but this may depend on their expectations of the service or their previous experiences. Complex multilevel models could be built to try to validate the construct of quality, but these will be limited by the data we have collected and those that are available. At the very least we would be able to examine whether the quality measures correlate with satisfaction and the 2003 measures.

We would also want to check the reliability of the various measures. Although reliability does not prove validity, it is a necessary condition for validity and can be investigated using Cronbach's alpha as a measure of internal consistency (Nunnally 1967). We would expect the measures obtained from 2006 data to have high and similar alpha coefficients to those obtained from the 2003 dataset.

3.2.2 Change in quality from 2003 to 2006

The UES offers an opportunity to explore change in quality through comparison of LAs by PIs that capture quality and comparison of LAs using other variables that are not designated as PIs, namely the derived quality variables. A simple comparison of difference between the 2003 and 2006 sample can be conducted using a t-test.

In addition to comparison of the difference between ratings for 2003 and 2006 along each of these variables, it would also be helpful to be able to explain the variation. For example, do changes in the characteristics of clients, such as, an increase in more disabled clients explain the difference observed? This type of explanation requires multilevel modelling techniques, to account for three levels of analysis: time, area effects and individual characteristics.

3.2.3 Local variations in quality

Evidence from the 2003 study and from the 2006 national study (Information Centre, 2007) indicates that there is a great deal of variation in quality between authorities. Explaining these variations would help to understand how quality might be improved. It would also help to interpret PIs and ensure that variation is not due to, for example, variations in case-mix across authorities (Bird et al., 2005). Variations in quality between authorities and individuals can be examined using regression techniques and, where there are several levels of analysis, multilevel models.

However, explaining variation, especially where there are multiple levels of analysis, requires a good theoretical base, as this is needed to guide the collection of data and the construction of a model. It is possible to think about variations in quality from a number of different perspectives, for example differences may result from differences in experience, in the

organisational structures, or in the political economy. A theory linking these different areas needs to be developed.

4. Results

All 150 councils were invited to participate in the 2006 extended UES and in total 50 councils volunteered to take part. This compared favourably with 2003 when 34 authorities took part. Of those 34 authorities 21 took part in both the 2003 and 2006 studies.

Prior to analysis all data was checked through various SPSS programmes for the accuracy of data entry, missing values, and the fit between the distributions of the variables and the assumptions of multivariate analysis. Several problems were identified with the accuracy of the data, including:

- Values out of range or coded incorrectly. These were checked with the LA where possible or treated as non-response
- Inconsistencies in responses to statements in the grid (i.e. ticking all the way down one side). Cases where individuals had responded in this way were treated as a non-respondent for these items. 810 cases had to be treated as non-respondents for question 6 and 13 cases for question 14
- Inconsistencies between planned hours and reported hours. We report this analysis in more detail in Appendix C. These variables were not included in analyses reported here as they appeared to be unreliable
- Inconsistencies in reporting practical help. Respondents with inconsistencies in their responses to each of the statements, such as ticking all three boxes implying that they receive practical help and also do not receive practical help, were treated as non-respondents to the practical help items. 5995 cases had to be treated as non-respondents
- Inconsistencies between satisfaction and responses to the other survey questions. This analysis is reported in Appendix D. It was unclear whether this represented errors in recording or genuine differences in the importance of various domains of quality to forming individual's satisfaction with services. Some of these responses though are likely to contribute to multivariate outliers where satisfaction and the derived quality variables are included in the same analysis.

4.1 Response rates

Information was collated from 28,840 individual respondents out of a possible 49,542 – an overall (unit) response rate of 58 per cent. Although the average response rate to the survey was 58 per cent, it varied widely between authorities from 32 to 82 per cent. However, the

response rate for the extension study is the same as that reported for the national UES[3] (Information Centre, 2007), but is slightly lower than that recorded for 2003 extension study of 65 per cent (Netten et al., 2004).

It is hard to conclude much from the variation in response rates between authorities and from 2003 to 2006 as much depends on the sampling methodology, chasing and data collection procedures. Although the IC produced guidance (2005), anecdotal evidence suggests that all of these strategies varied between authorities. The guidance suggests a simple random sampling method to provide a confidence interval of +/- 0.4 per cent around the satisfaction PI. This would produce samples roughly proportionate to the size of the authority. The guidance also suggests that authorities over-sample for non-response. However, the percentage to over-sample is not specified and it varied widely between authorities. Some authorities carried out full population surveys despite being medium to large sized authorities. Some of those authorities that over-sampled also did not chase non-respondents as they had received all the responses required to produce the necessary confidence interval around the PI after the initial posting of the questionnaire. Other authorities carried out more chases to try to reach the target confidence interval. These procedural differences are significant as those people responding to later chases may have different characteristics that those responding to earlier requests for information. Not chasing and grossly over-sampling can also lead to low response rates. Both of these may introduce bias into estimates derived from the samples and may negatively impact on comparability. It may also affect the legitimacy of the estimates for England as certain authorities may be over- or under-represented.

Of those responding to the questionnaire, there is another source of non-response known as item non-response. This arises where people only answer certain questions or only answer some questions correctly. The item response rate also varied widely by question as is shown in Table 18, Appendix E and although it is quite low for most variables only 5,067 cases had complete data for all variables, which represents fewer than 20 per cent of respondents. A significant reason for this was the lack of complete cases for the auxiliary data where entire variables were missing for 22 authorities. However, a significant number of cases were also incomplete because respondents did not complete the questionnaires. Those items with the lowest responses were those in the grids – questions 6a-f and 14a-m, particularly those phrased negatively – and those questions where respondents were required to tick 'yes' or 'no' – question 19a-d asking about services and question 27a-d asking about functional

[3] It should be noted that in the national sample respondents responding to only one question were also classed as non-respondents. However, the proportion responding to only one of the compulsory questions in the PSSRU study was very small at 0.1% of all questionnaires returned. This would have an insignificant effect on the value of the response rate.

(dis)ability. In addition the first question in the grid always had the highest response rate – a rate equivalent to the rest of the questions in the questionnaire – indicating that some of the problem with the grid questions is navigational. Other questions with lower response rates were questions 17 and 20 asking about direct payments and hours of home care respectively. There may be several reasons for non-response. In some cases it is likely that non-response is a result of difficulties answering the questions. However, in the case of disability questions people may just not want to respond, perhaps seeing the question as intrusive. In the case of services it may be they are not sure or feel they are being checked up on. However, where there are problems, it is worthwhile investigating these questions and perhaps revising for future surveys.

Non-response in both its forms – unit and item – can cause bias in the estimates where missing data are missing systematically according to characteristics of the service users. This is not easy to deal with, although some methods are available. Unit non-response can be adjusted for by weighting the estimates to take account of sections of the population that are missing from the sample that responded. However, this is only necessary should the missing population have characteristics that explain variation in the estimates of interest, for example, if gender explained satisfaction and men were less likely to respond than women we might want to weight for non-response by gender. We examine the potential for bias introduced by unit non-response in Appendix F.

Item non-response can cause bias where items are not missing completely at random (MCAR), i.e. where there is some association between the missing items and either the characteristics of the respondents and non-respondents or the value of the missing item itself. In the situation where items are not MCAR, most texts advise using imputation procedures such as multiple imputation (Fichman and Cummings, 2003; Little and Schenker, 1994; Schafer and Graham, 2002). Little's test (1988) can be employed to examine for MCAR, although the conclusions of this test should not be regarded as definitive (Little and Schenker, 1994). Given that the methods for adjusting for non-response are quite complex and require adequate time for testing, it is important to determine whether there is any need for adjustment. We examine the need for multiple imputation in Appendix F.

In summary, the results of the non-response analysis had the following findings:

Unit non-response

There is the potential for some bias in the estimates for satisfaction and service quality as a result of unit non-response. The following variables, summarised in Table 1 were associated

with non response. Since these variables were found in the 2003 study to explain a very small amount of the variation in quality (Netten et al., 2004) we would expect the bias to be small. However, a weighting procedure could be used to establish the extent of this bias, when reporting estimates.

Table 1: Relationship between response and characteristics of the user and provider

Auxiliary variable	*More likely to respond*	*Less likely to respond*
Gender	Female	Male
Age	75 to 84 85 and over	65 to 74
Ethnicity	White	Asian/Asian British Black/Black British Chinese
Provider type	In-house	Independent
Number of providers	One provider	More than one provider
Planned hours of home care	2 to fewer than 5hrs 0 to fewer than 2hrs	10 hrs and over

There are two problems with applying a weighting procedure. First not all authorities were able to provide the auxiliary data with which to proceed with weighting and secondly, the 2003 study did not collect auxiliary data so we are not able to weight this dataset. We are therefore not able to attempt to correct for any bias in any comparative work between authorities or between 2003 and 2006 or indeed any analyses using the whole dataset. We do intend, however, as a follow-up from this work to weight for unit non-response within those LAs who were able to provide all the auxiliary data to examine the potential impact on the accuracy of the estimates of quality.

Item non-response

Analysis of item non-response indicated that the items were not MCAR. This means that multiple imputation should be considered before conducting any kind of analysis. In addition several patterns of non-response and relationships between the values of the missing data and

the valid cases were uncovered, as summarised below (detailed analysis is provided in Appendix F):

- The average value for satisfaction when either provider type or number of providers is missing is significantly lower than in cases where the values for these variables are found. Given that these two variables are missing for whole authorities rather than individuals, this implies that authorities that were unable to provide auxiliary data had less satisfied service users. This has extremely interesting implications for the value of knowledge and data systems in contributing towards high quality services.
- The average value for satisfaction where the derived quality variables are missing, i.e. service quality, carer quality, outcomes quality and overall quality, is lower than in cases where the derived quality variables are present. This implies that less satisfied people are less likely to answer the questions. Therefore estimates of quality derived using these variables are likely to overestimate quality.
- When statements are positively phrased, the average value for satisfaction for all the missing cases is lower. This indicates that less satisfied people are less likely to answer positively phrased statements.
- Conversely, when statements are negatively phrased, the average value for satisfaction for all the missing cases is higher. This indicates that more satisfied people are less likely to answer negatively phrased statements.

Most of these problems are specific to the derived quality variables that include the grid format statements, although all quality variables were affected as people who were less satisfied also had a decreased propensity to answer any question about quality. This finding could affect the relationship between the quality variable and satisfaction.

Since multiple imputation is time consuming, we analysed the conditions under which valid case or complete case analysis might be robust. The following decisions were made:

- To not conduct any further analysis with the overall quality, outcomes quality and carer quality variables due to the pattern of non-response.
- To perform validation on only the service quality indicator and to use this measure for the weighted satisfaction measure.
- To perform analysis of quality change using the service quality and satisfaction variables.
- To not conduct any multilevel analyses due to lack of efficiency of the procedures.

Therefore we have limited the analyses reported here to validation of the service quality indicator and an investigation into change in quality. We will conduct further work, following

multiple imputation to validate the other derived quality measures, investigate changes over time in these measures and investigate local variations.

4.2 Description of participating authorities

In 2006 the sample included 10 London Boroughs, compared with only one in the 2003 sample – a result of another extended version of the UES having been commissioned by the London ADSS Benchmarking group in 2003 (Starfish Consulting, 2003). As Figure 1 shows the sample is relatively representative of the various authority types, although it is slightly biased in favour of shire authorities and against metropolitan districts. The participating councils were also spread quite evenly across the nine Government Office Regions (see Figure 2), although the North East was over-represented and the West Midlands were slightly under-represented.

Figure 1: Percentage of councils within each authority classification taking part in the 2006 study

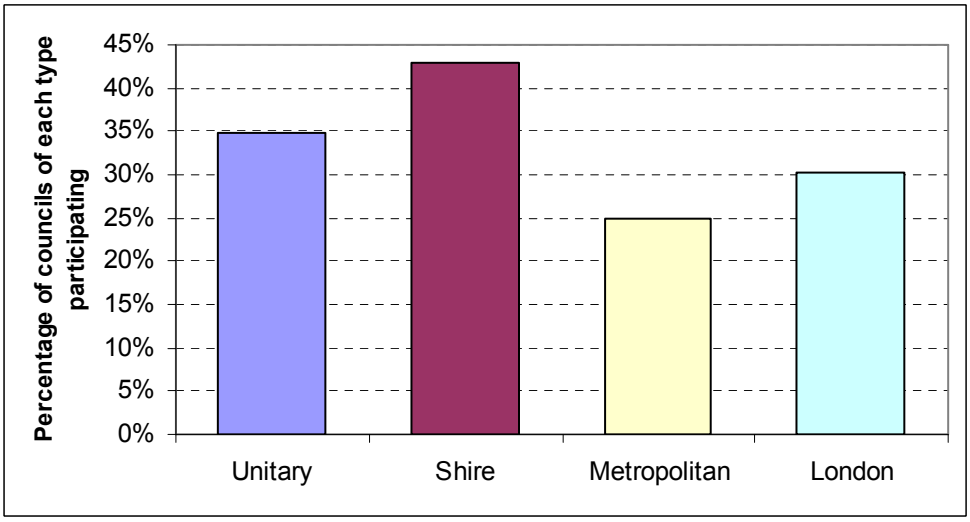
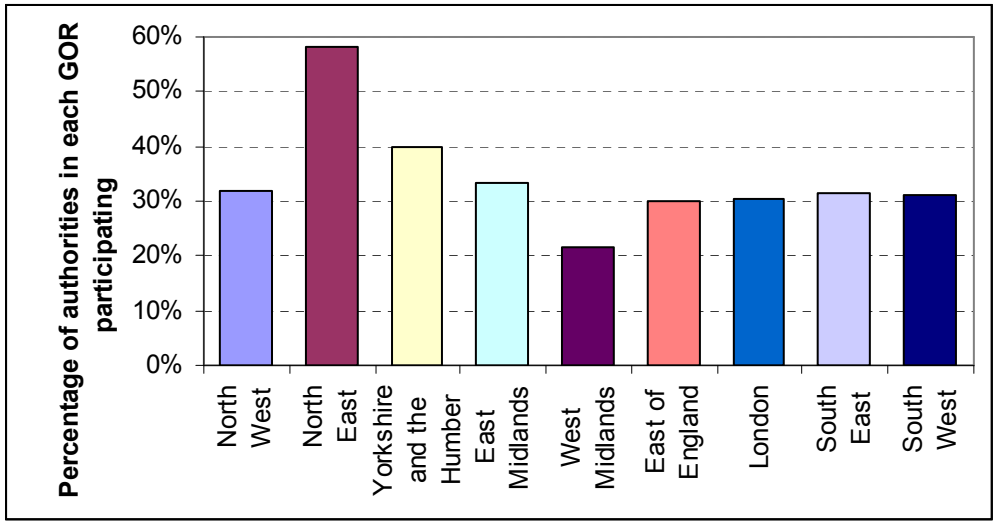


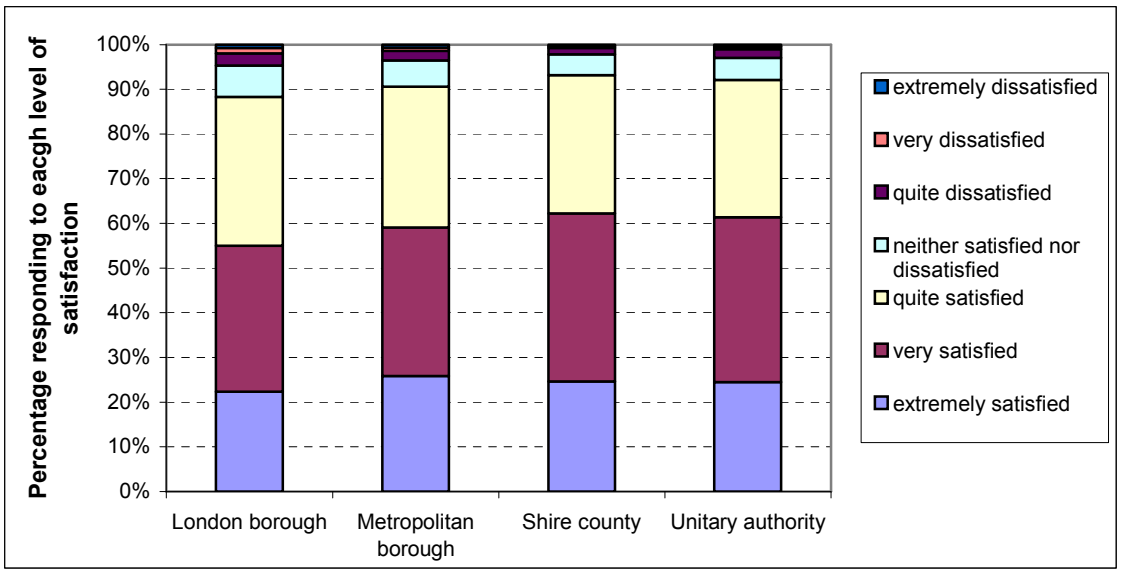
Figure 2: Percentage of councils within each GOR taking part in the 2006 study



4.2.1 Variations between authorities

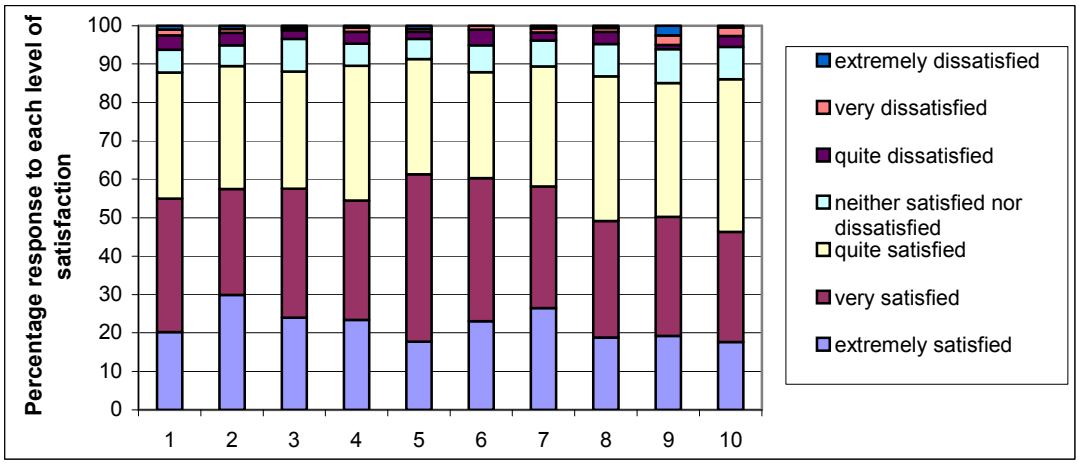
The analysis that was used to provide all participating authorities with short reports detailing their performance relative to other authorities showed a great deal of variation between localities and also between types of local authorities in terms of both the characteristics of service users, the characteristics of the home care service they were receiving and the views of the service users. We have illustrated in Figure 3 the variation by authority type using the satisfaction item, with London boroughs performing less well than other authority types on satisfaction. However, the same degree of variation was observed for all items (see Appendix G).

Figure 3. Responses to satisfaction item by local authority type, for authorities participating in extension study



There was also a large degree of variation within authority type categorisations. Figure 4 below shows the variation on the satisfaction item for London authorities in our sample.

Figure 4. Responses to satisfaction item for London boroughs participating in extension study



When comparing 2003 and 2006, it is important to try and control for this degree of variation at the authority level. The 2006 sample includes a larger proportion of London boroughs than the 2003 sample and this could potentially affect comparisons between the years as the London boroughs performed on average worse than all other authorities. Therefore for comparisons between years we include only those 21 authorities that took part in both years.

4.3 Characteristics of service users

The vast majority of respondents, 97 per cent, were contacted through postal questionnaires, 2 per cent via face-to-face interviews and 1 per cent via telephone interviews. However, these proportions varied between authorities, with one authority choosing to deliver all the questionnaires by face-to-face interview.

Table 2 shows the demographic make-up of the service users in the sample, citing all valid cases.

Table 2: Demographic characteristics of the sample

Characteristics	*n*	*%*
Gender		
Male	7,336	25.7
Female	21,175	74.3
Total	28,511	100.0
Age group		
65-74	4,350	15.2
75-84	11,722	41.1
85+	12,469	43.7
Total	28,541	100.0
Ethnic group		
White	27,155	96.6
Mixed	74	0.3
Asian/Asian British	269	1.0
Black/Black British	475	1.7
Chinese	43	0.2
Other	99	0.4
Total	28,115	100.0

The vast majority (85 per cent) were 75 or over and about a quarter were men. This is a very similar demographic profile to that found in 2003 and is the type of profile we would expect in this group. Generally people from black and minority ethnic (BME) populations are under-represented among older service users. The majority in this sample, 97 per cent, of respondents described themselves as white (British, Irish or any other white background). The largest single other group described themselves as Black/Black British and there was also a significant minority who described themselves as Asian/Asian British.

Receipt of informal care is an important predictor of receipt of services (Pickard et al., 2000) and in this survey a question asked service users whether they received any practical help. As Malley et al. (2006) describe elsewhere, this question may over-report receipt of informal care since the concept 'practical' conveys a sense of occasional help as well as regular help,

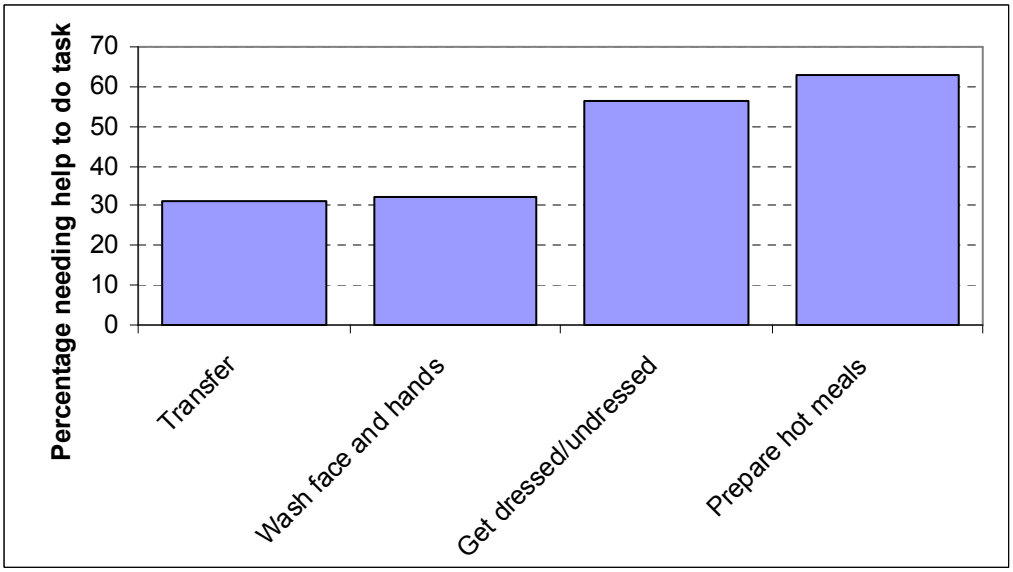
although this problem may not be apparent for older people. 17 per cent reported that they did not receive any practical help from anyone. The vast majority, 61 per cent, reported that they received help was from outside the household. 17 per cent reported receiving practical help from someone inside their household and only 5 per cent reported receiving practical help from both outside and inside the household.

In terms of perceived health, the distribution of responses to this question was fairly normally distributed. We found that the majority of people (56 per cent) reported fair health and very few people reported very good or very bad health, just 3 and 5 per cent respectively. 18 and 17 per cent respectively reported good and bad health.

Since the question reporting disability was a new addition, it is useful to analyse the distribution of the question. Figure 5 shows the difficulty rating of each of the items, meaning the ease with which respondents are able to endorse the question, in this case whether they need help to do the task. As can be seen, whilst the various questions appear to capture some variability, a different selection of questions might provide a more gradated pattern of response i.e. with similar distances between each of the 'steps'. For example, a slightly more difficult task might help to differentiate people better than the wash face and hands item, which has a very similar difficulty rating to the transfer question. In addition, a question that is more difficult to endorse than preparing hot meals, may help to differentiate low levels of disability.

A disability score composed of these items does seem to adequately distinguish between respondents, with 19 per cent reporting no need for help for any of the items, 14 per cent reporting need for help with one item, 16 per cent with two items, 23 per cent with three items and 29 per cent with all four items. The majority of the respondents report difficulty with three or four items, but it is complicated to provide a measure of the degree of disability of the population, as the score is dependent on the items within the measure and has no comparator.

Figure 5. Difficulty rating of the various ADL and IADL questions



We also included a measure of length of time in receipt of services. The sample showed a wide range of time in receipt of services. 10 per cent of the sample reported only having starting receiving services in the last six months. 17 per cent reporting having been receiving services for between 6 months and a year, 21 per cent between one and two years, 30 per cent between two and five years and 22 per cent for over five years.

In this study, there was also a question that asked about receipt of services other than home care. 21 per cent of the sample reporting receiving meals services, 22 per cent reported going to a day centre, 41 per cent reported receiving community nursing services and 8 per cent reported receiving another type of service. Of those people responding to these questions, 48 per cent received only home care services.

4.4 Characteristics of the home care service received by service users

Local authorities were asked to provide several pieces of information relating to the service users from their client records, as we have discussed already (see section on 4.1 Response *rates*). In particular, authorities were asked to provide information on the type of home care service that the service user received, the number of providers the service user received services from and the number of weekly hours of home care received. We identify in the section on non-response that these particular questions were answered poorly by a substantial number of authorities, with only 28 authorities able to provide data on all of the auxiliary variables. This means that any description of the sample based on these variables will exclude a number of authorities from the description and is therefore likely to give a biased picture of

the sample. Bearing in mind this caveat, we report figures for those for whom the data were available.

The overwhelming majority, 93 per cent, of service users received services from only one provider. Of those who received more than one provider, the majority (5 per cent) had two providers[4]. Comparing provision in 2006 with 2003, there is a slight increase from 5 to 7 per cent in provision from multiple providers and a corresponding decrease in provision from one provider only.

In 2006, the majority of the provision (69 per cent) was from independent providers only. 29 per cent received only in-house provision and just over 2 per cent were reported to be receiving both in-house and independent provision. Those authorities taking part in both years were less likely to use independent provision than other authorities in the sample, with only 57 per cent of service users receiving services from independent providers. However, in comparison with 2003, this represents a large increase in provision from this sector (from 42 per cent). These data imply a continued move towards use of independent provision and are in keeping with national data showing the same trend (e.g. Commission for Social Care Inspection, 2006b).

Two measures of hours of home care provision were recorded: planned as reported by the local authorities and actual hours as reported by service users. These variables provided a useful opportunity to explore whether home care users are receiving the hours of care that commissioners assume they are receiving. However, as we describe above, we identified several problems with these data, suggesting that any comparison is likely to be unreliable. The most limiting problem was the degree of mismatch between 'planned' and 'reported hours', with 'reported hours' often missing when 'planned hours' is present and vice-versa. Where comparison was possible, there were very large differences between hours reported and planned hours that could not be due to poor delivery by services. There seems to be a large degree of misreporting and it is not clear whether that comes from the LA or the user; both are likely sources. The issues are discussed in Appendix C.

4.5 Service User Views

This year, ten of the questions in the questionnaire were compulsory questions. The responses to these questions in the PSSRU sample are compared with those in the national sample in Table 3.

[4] The value for the sample including non-respondents is 92 per cent in receipt of services from one provider.

Table 3: Responses to compulsory questions in PSSRU and national samples

Question	*PSSRU extension*		*National figures*
	n	*%**	*%**
Overall, how satisfied are you with the help from Social Services that you receive in your own home?			
Extremely satisfied	6,876	24	24
Very satisfied	10,086	36	35
Quite satisfied	8,837	31	32
Neutral	1,518	5	6
Fairly dissatisfied	546	2	2
Very dissatisfied	176	1	1
Extremely dissatisfied	132	0	1
Total	28,171	100	100
Do your care workers come at times that suit you?			
Always	10,658	38	37
Usually	13,604	48	49
Sometimes	3,359	12	12
Never	462	2	2
Total	28,083	100	100
Are you kept informed, by your home care service, about changes in your care? (e.g. your visit will be late or you'll have a different carer)			
Always	9,579	35	34
Usually	11,349	41	41
Hardly ever	4,074	15	15
Never	2,757	10	10
Total	27,759	100	100
Do your care workers do the things that you want done?			
Always	18,382	66	65
Nearly always	7,445	27	27
Sometimes	1,887	7	7
Never	169	1	1
Total	27,883	100	100
Do you know how to make a complaint about the Home Care Service?			
Yes and I feel I could if I wanted to	20,867	76	76
Yes but I do not feel I could if I wanted to	2,164	8	8

No I do not know how to make a complaint	4,423	16	16
Total	27,454	100	100
I feel safe in my home			
Strongly agree	11,606	47	47
Agree	13,392	50	50
Disagree	692	3	3
Strongly disagree	168	1	1
Total	25,858	100	100
I have as much contact with other people as I want			
Strongly agree	7,790	34	35
Agree	13,544	54	53
Disagree	2,704	11	10
Strongly disagree	437	2	2
Total	24,475	100	100
I get up and go to bed at times that suit me			
Strongly agree	10,659	44	46
Agree	12,967	50	49
Disagree	1,103	4	4
Strongly disagree	242	1	1
Total	24,971	100	100
Which of the following statements best describes your present situation?			
I feel in control of my daily life	10,488	38	39
Services help me to feel in control of my daily life	12,655	46	46
I have some control over my daily life but not enough	3,133	11	11
I have no control over my daily life	1,057	4	4
Total	27,333	100	100
Has your social worker or care manager told you about direct payments?			
Yes	8,179	33	36
No	9,796	39	37
Don't know	6,903	28	27
Total	24,878	100	100

* Percentages may not add up to 100 per cent due to rounding

The percentages responding to each response category are very similar for both samples across all the compulsory questions, indicating that the PSSRU sample is relatively representative of the views expressed nationally.

One of the purposes of the extension survey was to provide a wider base for comparability between councils in terms of the quality of home care being delivered[5]. The other questions covered the areas of continuity, reliability, flexibility, communication, staff attitudes and skills and knowledge, identified as domains of quality that were important to service users in previous work (Francis and Netten, 2004).

In terms of continuity of the service, service users answered a question on whether they saw the same care workers. 34 per cent of respondents reported always seeing the same care worker, 60 per cent reported nearly always seeing the same care worker, five per cent reported hardly ever seeing the same care worker and only 1 per cent reported never seeing the same care worker.

Reliability of the service was assessed with three questions that all covered the timing of visits. These were whether care workers arrived on time, whether they were in a rush and whether they spent less time than they were supposed to. For these questions, 40 per cent of respondents said that their care workers always arrived on time, 40 per cent said that they often arrived on time, 14 per cent said that they sometimes arrived on time, 2 per cent that they never arrived on time, and 5 per cent that they never knew whether they were going to arrive. When asked whether care workers were in a rush, 49 per cent said they were never in a rush, 41 per cent that they were sometimes in a rush, 7 per cent that they were often in a rush and 4 per cent that they were always in a rush. In the same sample, 64 per cent said that their care workers never spent less time with them than they were supposed to, 28 per cent that they sometimes did, 6 per cent that they often did and 3 per cent that they always did.

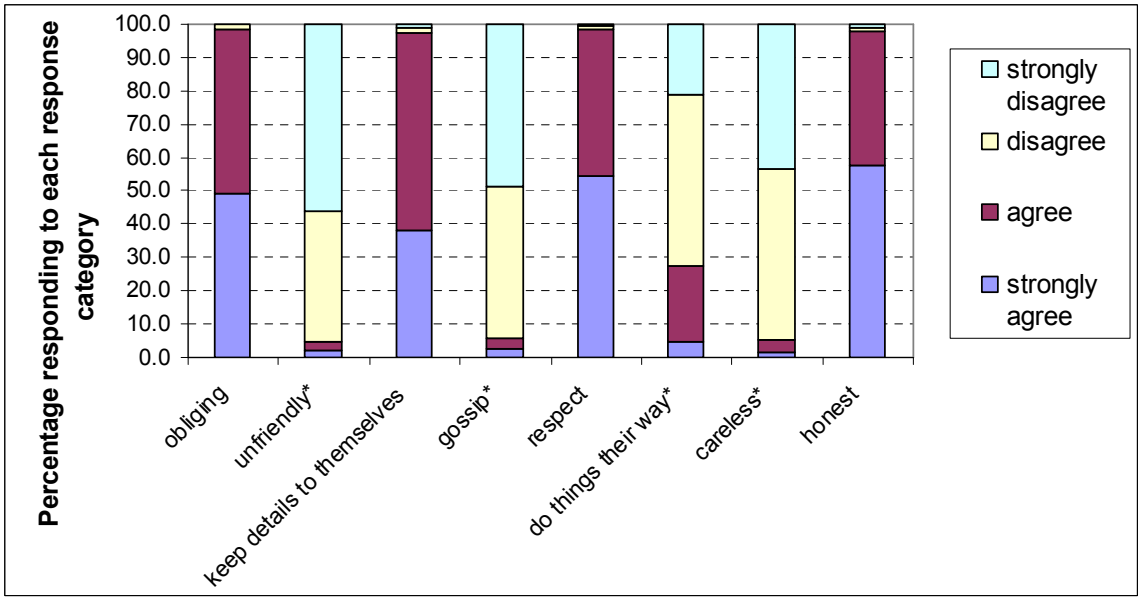
Flexibility was assessed in the following questions: care workers coming at times that suited, care workers doing the things you want done and having sufficient visits. The first two of these questions were part of the compulsory questionnaire and the responses are shown in Table 3. Ninety-one per cent of individuals responding to the sufficient visits question reported having sufficient visits, 8 per cent reported needing a few more and 1 per cent reported needing a lot more.

Communication was assessed through three questions that were all part of the compulsory questionnaire. These questions asked about direct payments, whether service users knew how to complain and whether they were kept informed about changes. The responses are shown in Table 3.

[5] A report was provided for individual authorities to compare their responses to the individual items with the responses for the whole sample, for each authority type (London borough, shire council etc) and with the whole sample from 2003. The template is presented in Appendix G.

Staff attitudes cover a number of questions, including the question asking service users whether they are happy with their overall treatment from care workers, and a number of the questions included in the grid. Seventy-four per cent of respondents reported always being happy with the way their care workers treated them, 23 per cent were usually happy, 4 per cent were sometimes happy and only a minority, less than 0.5 per cent, reported never being happy. Figure 6 presents the responses to the grid questions and shows that the proportion agreeing positively varies very little across items, except for the item asking about whether care workers do things their own way. The minority of respondents to this statement felt that care workers did things in their own way rather than as the service user would have preferred.

Figure 6. Responses to statements asking about care worker attitudes



* These statements were phrased negatively, so disagreement indicates positive sentiment

Staff skills and knowledge is assessed through two other items that were also part of the grid questions. These were: my care workers are excellent at what they do and my care workers are less thorough than I would like. Forty-three per cent strongly agreed that their care workers were excellent, 51 per cent agreed, 5 per cent disagreed and 1 per cent strongly disagreed with the statement. In response to the statement asking respondents to assess how thorough their care workers were, 3 per cent strongly agreed, 15 per cent agreed, 52 per cent disagreed and 31 per cent strongly disagreed with the statement.

There were also a number of questions that tried to uncover the degree to which the services were meeting the needs of the individuals in areas that service users assess as being important. Outcome areas identified as important to older people in earlier work were: food

and nutrition, personal care, social participation, safety, and control over daily living (Netten et al., 2002). This list was later revised to include the following domains: accommodation cleanliness, comfort and accessibility, employment and occupation, and role support (Netten et al., 2006; Netten et al., 2005). Not all of these domains are covered by the questions included in the survey.

Questions asking about safety, social participation and sense of control over daily life were all included in the compulsory questions and responses are shown in Table 3. These were for safety, "I feel safe in my home"; for social participation, "I have as much contact with other people as I want"; and for sense of control over daily life two questions were asked. These were: "I get up and go to bed at times that suit me" and the question asking which of the following statements about control over daily life best describes your situation. A question asking about employment and occupation was also included in the extension study, which stated "I spend too long with nothing interesting to do". Eleven per cent of respondents strongly agreed with this statement, 32 per cent agreed, 44 per cent disagreed and 13 per cent strongly disagreed[6].

Two questions on personal care were asked using the statement grid format. These were: "I am always as clean as I want to be" and "I always feel comfortable". Forty-six per cent of respondents strongly agreed that they were as clean as they wanted to be, 49 per cent agreed, 5 per cent disagreed and 1 per cent strongly disagreed. For the comfortable statement, 36 per cent strongly agreed, 55 per cent agreed, 9 per cent disagreed and 1 per cent strongly disagreed with the statement.

4.5.1 Comparison of responses to items between 2003 and 2006

Another purpose of the extension survey was to provide a wider base for comparability between 2003 and 2006. In 2003 only four questions were compulsory and of these two questions were dropped for the 2006 survey, leaving only two questions that could be potentially compared by authorities, the satisfaction question and the question about carers coming at suitable times. In the extension study, responses to six of the compulsory questions can be compared between 2003 and 2006 (see Table 4). Given the potential for any changes between 2003 and 2006 in the extension study sample to be the result of changes in the composition of the sample (e.g. the increased proportion of London Boroughs), we report changes between 2003 and 2006 just among LAs taking part in both years.

[6] A number of people appeared to answer this question incorrectly, not reading it as negatively phrased. This was picked up in the cleaning process when looking for respondents who had ticked all the way down one side. Although all respondents found ticking down one side were excluded, it is possible that some error still remains. The question will of course also greatly underestimate positive feelings for the service as more people ticked down the 'strongly agree' rather than 'strongly disagree' side, so more positive people had to be excluded.

Table 4: Responses to 2006 compulsory question in the PSSRU 2003 and 2006 extension studies for LAs taking part in both years

Question	*2003*	*2006*
	%	*%*
Overall, how satisfied are you with the help from Social Services that you receive in your own home?		
Extremely satisfied	26.5	25.3
Very satisfied	34.3	36.2
Quite satisfied	31.6	31.3
Neutral	4.5	4.8
Fairly dissatisfied	1.8	1.5
Very dissatisfied	0.8	0.5
Extremely dissatisfied	0.5	0.4
Total	100	100
Do your care workers come at times that suit you?		
Always	42.3	36.8
Usually	46.5	49.7
Sometimes	9.8	12.0
Never	1.4	1.4
Total	100	100
Do your care workers do the things that you want done?		
Always	75.1	67.4
Nearly always	20.0	26.0
Sometimes	4.5	6.1
Never	0.5	0.5
Total	100	100
I feel safe in my home		
Strongly agree	31.9	44.7
Agree	47.4	51.9
Disagree	14.2	2.8
Strongly disagree	6.5	0.6
Total	100	100
I have as much contact with other people as I want		
Strongly agree	26.0	31.7
Agree	57.2	55.9
Disagree	13.7	10.7
Strongly disagree	3.1	1.7
Total	100	100
I get up and go to bed at times that suit me		
Strongly agree	36.8	42.1
Agree	56.6	52.6
Disagree	4.9	4.4
Strongly disagree	1.7	0.8
Total	100	100

These data show some interesting and important patterns. Satisfaction with services, measured by the performance indicator cut-off point at 'very satisfied' that has appeared in

both UESs, appears to remain roughly the same between 2003 and 2006, although the distribution is more peaked around the very satisfied response. A positive value indicates that the values are more concentrated around the mean than they would be if the distribution were normal. Values for the 2003 and 2006 dataset both indicate that the distributions are more concentrated around the mean than would be expected, but the 2006 dataset is more peaked.[7] This indicates that over time responses have become more concentrated around the PI boundary of very satisfied.

The other comparable quality questions show what look like marked differences, which are highly significant (for times that suit: $F(1, 24388.615) = 75.989$, $p<0.001$; and for things want done: $F(1, 23853.308) = 152.184$, $p<0.001$)[8]. In particular, a smaller proportion reports that care workers always come at times that suit and always do things they want done in 2006 compared with 2003 (see Table 4), questions asking about the quality domain of flexibility. This is not offset, however, by the larger proportion in 2006 reporting that care workers nearly always or usually meet these aspects of flexibility. These findings indicate that the flexibility of the service has declined in the years between 2003 and 2006.

In contrast the outcomes questions show that an increasing proportion of people feel that the outcome areas of safety, social participation and control over daily life are being met. However, it should be noted that the marked increase in those responding "strongly agree" to the statement "I feel safe in my home" is potentially a result of the change in the phrasing of the statement from negative to positive. Research finds that it is easier to endorse a positively phrased statement than not endorse a negatively phrased one, which makes the direct comparison more difficult (Nunnally, 1967). The two other statements show a less marked increase in the proportion that are positive about these aspects of their life, which is probably a good guide to the effect of the change in wording on the response to the safety statement.

In addition to these items, the 2003 questionnaire also covered a number of other areas identified as domains of quality that mattered to older people using home care services and are present in the 2006 study. The only domain not covered on a comparable basis by both years of data was communication.

In terms of continuity, the question about whether service users saw the same care workers was repeated in 2006. In the sample of authorities participating in both years, the percentage

[7] The 'peakedness' of the distribution is referred to as kurtosis. The ratio between the standard error and the value of kurtosis is greater for 2006 (kurtosis for 2003 = 1.929, SE kurtosis = 0.034, ratio = 56.7; kurtosis for 2006 = 1.917, SE kurtosis = 0.029, ratio = 66.1). The mean for 2003 is 2.29 and for 2006 it is 2.28.

[8] Comparison between 2003 and 2006 was carried out using a one-way analysis of variance using the SPSS COMPARE MEANS ONE WAY ANOVA routine. Tests for homogeneity of variance indicated heterogeneous variance, so Welch/Brown-Forsythe robust tests of equality of means were carried out and reported.

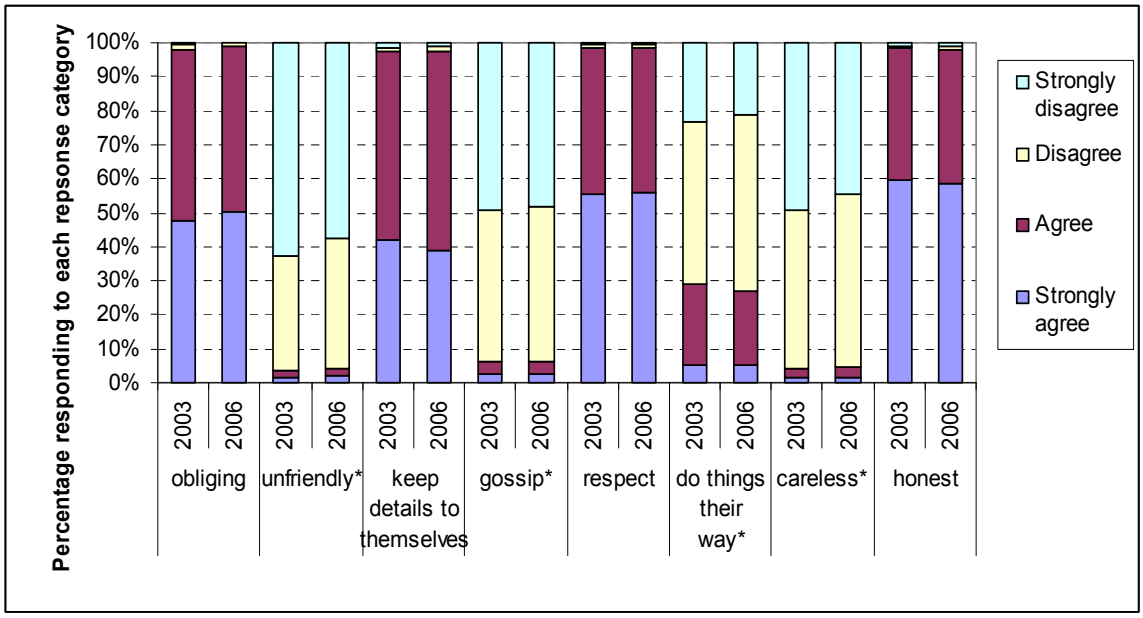
feeling positively about this area decreased. 36 per cent of respondents reported always seeing the same care worker in 2003, compared with 32 per cent in 2006. This was partly offset by an increase in numbers reporting that they nearly always saw the same care worker, from 59 per cent in 2003 to 63 per cent in 2006.

The three questions covering reliability of the service (whether care workers arrived on time, whether they were in a rush and whether they spent less time than they were supposed to) were all repeated in 2006. For each of these questions there was a decrease between 2003 and 2006 in the proportion responding that they had no problems (e.g. care workers always arrived on time), although again this was either completely balanced or partly balanced by those responding that there were some problems (e.g. care workers often arrived on time) (see Appendix G).

Questions on flexibility were also in both years of the survey and two of these questions are reported in Table 4. A third question asking whether service users had sufficient visits was also repeated for 2006. Between 2003 and 2006, slightly fewer service users reported not receiving sufficient visits from 10 per cent in 2003 to 9 per cent in 2006.

All, save three, questions on staff attitudes were repeated in 2006. Between 2003 and 2006, a greater proportion of service users reported that they were always happy with the way their care workers treated them (from 75 to 76 per cent), but fewer service users reported being usually happy (from 23 to 21 per cent) and more service users reported that they were some sometimes happy (from 4 to 3 per cent), indicating a quite complex picture with some improvement and also some worsening of individuals' situations. Figure 7 presents the responses to the grid questions for both years and shows a very slight change in the proportion agreeing positively, with some improvements and some worsening, for example more people report that care workers are obliging but fewer report that they keep details to themselves.

Figure 7: Comparison of responses to statements about carer attitudes between 2003 and 2006, for authorities participating in both years



The two questions on skills and knowledge were also repeated in 2006. There was very little change in the proportions in agreement with the statement that they thought their care workers were excellent, but some change in those disagreeing strongly with the statement that their care workers were less thorough than they would like, from 34 per cent in 2003 to 32 per cent in 2006. However, those disagreeing with the statement on the thoroughness of the care workers rose between 2003 and 2006 from 50 to 52 per cent, balancing out the decrease in the proportion disagreeing strongly with the statement.

Some of the outcomes questions have already been discussed, but three of the other outcomes questions, covering the domains of personal care and employment and occupation were also repeated in 2006. In contrast to the compulsory statements that all saw an increased proportion responding positively, the responses to these statements show an increase in the proportion responding negatively (see Table 5). It is possible that these proportions have been affected by the increased number of cases that had to be excluded in 2006 due to incorrect answering of the statements and that the 'true' difference is not so great. However, since this problem also affected the compulsory statements, it is likely that the response to these areas is indeed less positive.

Table 5: Responses to outcomes questions asked in both 2003 and 2006, for authorities taking part in both years

	Feel clean		*Feel comfortable*		*Nothing interesting to do**	
	2003	*2006*	*2003*	*2006*	*2003*	*2006*
Strongly agree	52.9%	46.6%	39.8%	35.7%	8.1%	10.9%
Agree	44.5%	49.2%	53.2%	55.0%	27.3%	31.9%
Disagree	2.4%	3.6%	6.4%	8.5%	46.3%	44.4%
Strongly disagree	0.3%	0.6%	0.5%	0.9%	18.2%	12.9%

* Phrased negatively, so agreement with the statement is a negative response

4.6 Validation of the derived quality variables

One purpose of the extension study was to validate the derived quality measures from 2003. As we have discussed, non-response issues mean that we only consider here the service quality measure and the weighted satisfaction measure, the latter using weightings based on the service quality variable[9]. Validation of the service quality scale was provided by four types of checks: internal reliability of the scale, check for content validity, correlation with satisfaction and correlation with 2003 value (both for construct validity).

Reliability of the service quality measure

The service quality scale was found to have high reliability (KR20 = 0.7793, n = 25236, k=7)[10]. This was comparable with the reliability for the 2003 scale, although slightly lower. This is probably due to the loss of one item, which can have the effect of reducing the reliability, since reliability is in part a function of the number of items in the scale (Nunnally 1967). When the communication item was added, the reliability of the scale increased considerably (KR20 = 0.8218, n =24972, k=8), implying that the communication item improves the internal consistency of the scale.

[9] As the overall quality variable is composed of items from the service quality variable, and some others, this should give a good indication of the relationship between overall quality and satisfaction.

[10] Reliability is assessed with the Kuder-Richardson 20 (KR20) reliability statistic, which is a special case of Cronbach's alpha for dichotomous items (Nunnally, J. C. 1967 *Psychometric theory*, New York: McGraw-Hill Book Company.), using STATA software and the kr20 command.

Validity of the service quality measure

Content validity

Content validity can be assessed by examining the extent to which the items in the measure of quality capture all the domains of quality identified in the literature. Table 6 below lists the various domains of process quality identified in the literature (Francis and Netten, 2004) and the domains of quality of life or outcomes identified from the literature (Netten et al., 2006; Netten, et al., 2005) and lists against them the questions included in the various measures of quality derived from the 2003 study. We have included all the quality measures for completeness.

The table shows that most domains are covered by at least one of the derived factors. However, certain domains dominate the items in the questionnaire, for example the domain about staff attitudes. Certain domains are not covered by the questions in the questionnaire and these are mostly the domains concerned with outcomes. The service quality, carer quality and outcomes measures in combination provide the most complete coverage of the conceptual space. However, with these taken together a few domains are still absent.

Table 6: Content validity of 2003 derived quality measures

Quality domain	**Question**	**Overall quality**	**Service quality**	**Carer quality**	**Outcome quality**
Continuity	Do you always see the same care workers?		Y		
Flexibility	Do your care workers come at times that suit you?	Y	Y		
	Do your care workers do the things that you want done?	Y	Y		
	Do you have as many visits from your care workers as you need?				
Reliability	Do your care workers arrive on time?	Y	Y		
	Are your care workers in a rush?	Y	Y		
	Do your care workers spend less time with you than they are supposed to?	Y	Y		
Communication	Are you kept informed, by your home care service, about changes in your care? (e.g. your visit will be late or you'll have a		Only in 2003		

	different carer)?				
Staff attitudes	Overall, how do you feel about the way your care workers treat you? (e.g. whether they are understanding and treat you with respect)	Y	Y		
	My care workers are understanding	Y		Y	
	My care workers are obliging	Only in 2003		Only in 2003	
	My care workers are miserable	Only in 2003		Only in 2003	
	My care workers are unfriendly	Y		Y	
	My care workers do things in their way rather than mine	Y		Y	
	My care workers are careless	Y		Y	
	My care workers are gentle with me	Only in 2003		Only in 2003	
	As far as I know, my care workers keep any personal details they know about me to themselves	Y		Y	
	My care workers gossip to me about other people they care for	Y		Y	
	My care workers treat me with respect	Y		Y	
	My care workers are honest	Y		Y	
Skills and knowledge	My care workers are excellent at what they do	Y		Y	
	My care workers are less thorough than I would like	Y		Y	
Personal safety	I don't feel safe in my home	Y			Y
Personal cleanliness and comfort	I am always clean	Y			Y
	I always feel comfortable	Y			Y
Social participation and involvement	I have as much contact with other people as I want	Y			Y
	I get up and go to bed at times that suit me	Y			Y

Control over daily life	The help I get from social services has made me more independent than I was	Only in 2003			
Employment and occupation	I spend too long with nothing interesting to do	Y			Y
Meals and nutrition	No question identified				
Accommodation cleanliness, order and accessibility	No question identified				
Role support	No question identified				

Considering service quality alone, it covers most of the process domains. Two domains are absent: communication and staff skills and knowledge. The former is a result of changes in the structure of that question between surveys making it impossible to compare to 2003. The most significant absence from the service quality measure was any questions relating to outcomes or quality of life.

Construct validity

Construct validity was tested by examining the relationship between satisfaction and our quality measure. A one-tailed Pearson's product moment correlation test was performed with satisfaction and the service quality measure[11]. Service quality correlated significantly with satisfaction ($r = -0.594$, $p<0.001$ and $n = 25108$). A linear relationship appeared to be a good approximation[12]. The high correlation indicates a strong relationship between satisfaction and perceptions of service quality which is as expected and indicates good construct validity. The relationship was improved with inclusion of the communication item into the service quality measure ($r = -0.608$, $p<0.001$ and $n = 24848$), which implies that the communication item should be part of the service quality measure. However, the communication item was not included in both sets of data and could not be included in the measure.

[11] The SPSS CORRELATE BIVARIATE routine was used for this purpose. As deviations from normality can affect the interpretation of correlation coefficients, the data were screened for normality using the SPSS Frequencies routine (Tabachnick and Fidell, 2001). Both service quality and satisfaction were found to have skewed distributions. The direction of the skew was opposing, which can attenuate the correlation. Therefore, satisfaction was transformed using a square root transformation and this improved the distributional characteristics. Transformation of the service quality variable did not improve the distribution and it was left untransformed. Cases where values were missing for both items were excluded. In this instance 13 per cent of cases were excluded.

[12] A small but significant amount of the variation was explicable by a non-linear component. However, eta-squared was very similar to r-squared indicating a linear relationship.

A further one-tailed Pearson's product moment correlation test was performed between the service quality measure for 2003 and that for 2006, at the level of the authority, using only those authorities that participated in both studies. The 2003 measure was recalculated to include only those variables present in both studies to ensure comparability. The results showed that the 2003 and 2006 measures correlated significantly with each other ($r = 0.861$, $p<0.001$ and $n = 21$), which was as expected. This indicates that the construct validity is good.

Validation of the weighting structure for the weighted satisfaction measure

One aim of this work was to validate the quality measure for wider use and for this purpose a weighted satisfaction measure was proposed. This measure identified four significantly different levels of satisfaction. We have tested the validity of this measure by examining the relationship between satisfaction and service quality, as above, but have extended the analysis to examine whether the four different levels identified using the 2003 data remain[13].

Service quality was found to vary significantly with reported level of satisfaction, as summarised in Table 7 ($F (6, 18255) = 1597.411$, $p<0.001$ for 2003; and $F (6, 25101) = 2390.232$, $p<0.001$ for 2006). The strength of the relationship between satisfaction and service quality was also good (eta-squared = 0.344 for 2003; and 0.364 for 2006). The marginal means show that highest reported service quality was associated with individuals reporting extreme satisfaction and four levels of satisfaction were identified in both the 2003 and 2006 datasets[14]. This is encouraging and implies that the relationship between satisfaction and service quality has the same structure.

[13] A one-way between subjects analysis of variance was performed for satisfaction on service quality to examine the relationship between satisfaction and service quality using the SPSS GLM UNIVARIATE procedure. The procedure was repeated for the 2003 and 2006 dataset to uncover whether the weighting structure varied between the two samples. Unequal sample sizes were adjusted for using SPSS method 1, where marginal means are weighted by the sample size in the cell from which they are computed (Tabachnick and Fidell, 2001). Results of evaluation of assumptions for normality of sampling distributions, linearity and homogeneity of variance were all satisfactory. No multivariate outliers were found using the $p<0.001$ criterion for Mahalanobis distance. Some cases had missing data for one or both of the items and in total 3,729 cases or 13 per cent of the sample had to be excluded from the 2006 analysis and 3088 cases or 14 per cent of the sample had to be excluded from the 2003 analysis.

[14] Post-hoc tests were carried out using Scheffé's test for multiple comparisons to identify the different levels of satisfaction.

Table 7: Marginal mean of service quality at each level of satisfaction for 2003 and 2006

Level of satisfaction	*Marginal mean of service quality (95 per cent confidence interval)*	
	2003	*2006*
Extremely satisfied	5.767 (5.719-5.815)	5.730 (5.697-5.772)
Very satisfied	4.624 (4.582-4.666)	4.438 (4.402-4.472)
Quite satisfied	3.231 (3.188-3.274)	2.908 (2.870-2.946)
Neither	1.843 (1.730-1.955)	1.597 (1.504-1.690)
Quite dissatisfied	1.537 (1.368-1.706)	1.475 (1.322-1.628)
Very dissatisfied	1.372 (1.081-1.663)	1.322 (1.048-1.596)
Extremely dissatisfied	1.750 (1.398-2.102)	1.454 (1.114-1.793)

Although structurally the relationship remains the same, the weightings derived from such a structure are not the same for each dataset, as Table 8 shows. The weights for all of the groups in the 2006 dataset are lower than those for the 2003 dataset. This is of course only significant if it has an effect on the value of the weighted satisfaction measure for 2003 and 2006. We explore whether this is the case below.

Table 8: Weights for weighted satisfaction measure derived using service quality in the 2003 and 2006 datasets

Level of satisfaction	*2003*		*2006*	
	Mean quality score	*Standardised quality weight*	*Mean quality score*	*Standardised quality weight*
Extremely	5.767	1	5.730	1
Very	4.624	0.802	4.438	0.775
Quite	3.231	0.560	2.908	0.508
Neither/dissatisfied	1.716	0.298	1.541	0.269

4.7 Change in quality from 2003 to 2006

Clearly, a fundamental reason for repeating the UES is to identify whether there have been changes in service quality. We examined this question in our sample using the overall satisfaction indicator and our service quality measure.

A comparison of means[15] was performed on three (dependent) variables: satisfaction, the satisfaction PI and service quality. Data were first aggregated at the level of the authority, using average scores for satisfaction and service quality, and the proportion reporting that they were extremely or very satisfied with services for the PI. Only authorities participating in both years were included. The aim of this was to control for the high degree of variation between local authorities, as this could mask the variation between years if a test of differences was conducted at the individual level (Agresti and Finlay, 1997).

In order to try to control for differences in item non-response between the survey questions, averages were calculated using complete cases for both variables only, to ensure that the samples for each of the variables, satisfaction and service quality, in that year of the study contained the same people and were comparable. The tests were, however, repeated using all valid cases (those cases where responses to both questions are observed) to see what differences emerged, if any. It is important to recognise that by reducing the sample to only those authorities participating in both years, we are limiting the generalisability of the findings.

[15] A matched-pairs t-test was used to conduct this comparison of means.

Table 9 reports the findings for the sample of authorities taking part in both the 2003 and 2006 studies[16]. As is clear from the table the different indicators of quality provide a very different answer to the question of whether quality has changed. Several comparisons are revealing: First, comparing the mean score for satisfaction to the PI using the satisfaction variable, when all complete cases are used, both indicators report no statistically significant change in quality. However, it should be noted that although not significant, the change is in opposite directions, with the mean score indicating decline and the PI indicating improvement. The differences between these two indicators could be explained by the observed decline in the proportion responding that they were 'extremely satisfied' in 2006, compared with 2003, which is offset in the PI by the increased proportion responding that they are 'very satisfied' in 2006 compared with 2003. The 'extremely satisfied' scores would have more weight in a score based on an average.

Secondly, a comparison between the mean score for satisfaction, the satisfaction PI and the mean score for service quality, for all valid cases, give an equivocal picture. Both the mean score for satisfaction and the satisfaction PI indicate improvement, but only the mean satisfaction score is significant. By contrast the service quality indicator showed a statistically significant decline in quality. Whilst we cannot draw any conclusions about changes in quality country-wide, these findings are worrying for two reasons: they suggest that the scoring method for the PI loses a certain amount of the variation in people's responses, potentially allowing for invalid conclusions to be drawn. Secondly, they also suggest that the satisfaction indicator is measuring something different to the service quality indicator. This could have an impact on the weighted satisfaction measure.

[16] SPSS Paired samples t-test was used for the analysis.

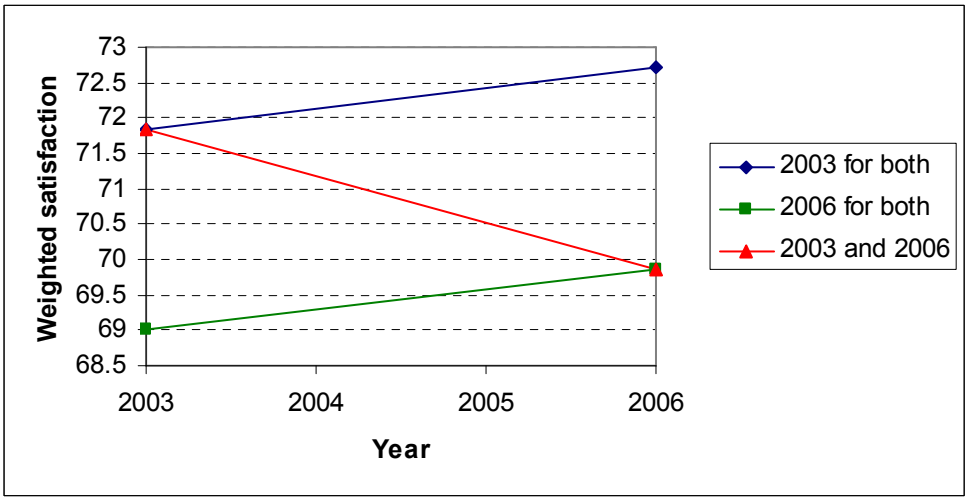
Table 9: Statistics showing change in quality between 2003 and 2006, using satisfaction and service quality measures

Measure of quality	*Complete cases* *t-statistic (2006-2003)*	*Valid cases* *t-statistic (2006-2003)*	*Df*
Mean score for satisfaction[+]	0.322	-4.802***	20
Proportion reporting extremely or very satisfied[~]	0.051	1.471	20
Mean score for service quality[~]	-4.770***	-4.850***	20

[+] Positive value for t-statistic means a decline in satisfaction
[~] Positive value for t-statistic means an increased proportion reporting extremely or very satisfied or an increase in service quality
***Indicates significance at the 1 per cent level; **Indicates significance at the 5 per cent level; Indicates significance at the 10 per cent level

Figure 8 shows how the different weightings affect the weighted satisfaction measure when different combinations of weightings are used. As can be seen from the figure, when the 2003 weightings are applied to the data from both years, an increase in quality as measured by the weighted satisfaction measure is found. When the 2006 weightings are applied, there is again an increase in quality, albeit from a slightly lower starting point. However, when the 2003 weights are used for the 2003 data and the 2006 weights are used for the 2006 data, a dramatic decline in quality is observed. This is likely to be explained by the change in the relationship between satisfaction and service quality.

Figure 8. Weighted satisfaction measure for 2003 and 2006 using different combinations of weightings



A comparison of means[17] was computed for each of the combinations of weightings, using data on proportions within each satisfaction level for each authority for 2003 and 2006 (data provided by IC). Table 10 reports the findings, showing that no significant change is found between 2003 and 2006 when the weights applied are not derived from that year, except when the 2003 weights are applied the difference is on the border of significance at the 10 per cent level. However, when the weights derived from the year are applied to that year a highly statistically significant difference is found. This of course raises the question of why. We discuss some potential explanations in the next section.

Table 10: Results of t-test for difference between 2003 and 2006 using the weighted satisfaction indicator

Weighting structure	*t-statistic (2003-2006)*	*Df*
2006 weights used for both years	-2.340	147
2003 weights uses for both years	-2.679*	147
2003 weights used for 2003 data and 2006 for 2006 data	7.985***	147

***Indicates significance at the 1 per cent level
**Indicates significance at the 5 per cent level
*Indicates significance at the 10 per cent level

5. Discussion

The aims of this study were to validate the measures of quality derived from the 2003 extension study for wider use, explore any variation in quality between individuals and authorities as well as between 2003 and 2006. In this report we have focused on evaluating the conduct of the survey (where possible), the service quality measure and change in quality using service quality for methodological reasons. The findings from these analyses have raised a number of questions, about the collection and usage of the statistics derived from the survey.

Participation in the 2006 study was good, with a third of all councils with social services responsibilities (CSSRs) taking part. This was better than the participation in the 2003 study. In addition the sample was more representative nationally, including a good number of

[17] A two-tailed matched pairs t-test for difference of means was used.

London boroughs and with participation from each Government Office Region, compared with 2003 where there was only one authority from London. Comparison of service user views on the compulsory questions with those from the national study reported by the IC (2007) was favourable and indicated that the sample was representative.

5.1 Non-response issues

As we have indicated non-response was a problem in the surveys and it meant that we were unable to analyse, in this first stage, the other quality variables derived from the 2003 study. We hope to complete this analysis following multiple imputation, a technique now widely accepted that can be used to adjust for non-response. The results of the non-response analysis did however suggest several problems with both the process of data collection and with some of the survey questions.

We found wide variation in non-response across authorities and anecdotal evidence suggests that the variation was not entirely due to differences in the propensity of service users to respond in each locality; rather differences in collection procedures seem to play an important role. In particular, authorities took different approaches to sampling and used different measures to combat non-response. In some LAs there was gross over-sampling and no follow-up, whereas in others there was limited over-sampling and energy was spent on following up non-respondents. Obviously in the former, response rates would be low and in the latter they would be high as a consequence of these different practices. These differences would not be a problem were it not possible that those respondents replying to the first round of response were different from those responding to later rounds, as has been suggested (Wood et al., 2006). This may affect comparability across LAs. The different sampling approaches can also affect the legitimacy of bringing the data together to provide a national picture as large authorities that have grossly over sampled may dominate the national picture. Hopefully the latter effect is small and insignificant, but given the wide degree of variation between authorities in terms of quality it cannot be dismissed. Future surveys may want to consider giving more explicit instructions to avoid these problems, perhaps monitoring the wave within which responses were received. Different approaches to sampling, such as a two-stage design, that are designed specifically to capture a national picture from multiple sites may also want to be considered.

We were also able to identify problems of non-response associated with certain types of questions, particularly those where respondents were required to answer ‘yes’ or ‘no’ or where the questions were set out as statement in grids. The question on direct payments and hours of home care were also poorly answered. There was also a great deal of variation between LA reported planned hours and user reported hours suggesting some problems with one or, more likely, both indicators. For some of the questions, such as services people may

not be responding as they are unsure or feel they are being checked up on. However, for a number of the questions, especially the grids, the non-response is likely to be a result of problems with the formulation of the question. Cognitive testing of all questions with high non-response may help to identify the root of these problems and reformulation for those subject to problems due to phrasing or structure may want to be considered before they are used in future surveys.

Non-response associated with the grid questions was particularly interesting as we uncovered a pattern whereby those reporting higher satisfaction less readily responded to negatively phrased statements and those reporting comparatively lower satisfaction less readily responded to positively phrased statements. In general these types of questions were poorly answered by older people and future surveys may want to consider replacing them with interrogation style questions rather than statements. Questions that could be used to replace the outcomes style questions have been developed by PSSRU for the younger adults' survey and are also being used in the Individual Budgets Pilot Evaluation. Should the response be better to these questions it would seem appropriate to use these questions in future surveys. The younger adults' survey and Individual Budgets Pilot also include some possible alternatives to questions asking about staff attitudes and knowledge that could be implemented again subject to the analysis of these surveys.

5.2 Changes between 2003 and 2006

The survey showed changes in the characteristics of the services provided that are in keeping with changes reported nationally (Commission for Social Care Inspection, 2006b). There had been a substantial increase in provision from independent providers over the period, with a corresponding decline in provision from in-house providers. There had also been an increase in provision from multiple providers.

The socio-demographic characteristics of service users were very similar between our samples from 2003 and 2006, with similar age and gender profiles. The ethnic composition had changed but this is probably a result of an increased participation of London boroughs. Information on health and disability was not collected in 2003, but collected in 2006 and indicated that the sample was in quite poor health, with the majority of the users reporting having fair health and being unable to carry out at least one of the tasks outlined in the section on disability. We have noted that a different set of ADLs and IADLs may be better able to distinguish levels of disability and this could be explored in future surveys.

Interestingly, the majority of service users reported receiving practical help from people living outside the household (56 per cent), a significant proportion at 17 per cent reported receiving no practical help and another 17 per cent reported receiving help from someone

they live with. This should be compared with data on the population at wide. Analysis of the 2001/2 GHS by the Long-Term Care Finance team at PSSRU at LSE, albeit from 2001/2, indicates that the majority of older people in fact receive informal help from someone living in the household (roughly 56 per cent), only 9 per cent receive no informal help and 35 per cent receive help from outside the home (authors own analysis). This comparison shows quite starkly the targeted nature of formal services towards people who have fewer informal care resources.

As in 2003, service users reported overwhelming satisfaction with services and the large majority reported positive or mostly positive experiences for all the domains of quality. However, comparison of 2006 responses to questions within each of the domains to those from 2003 revealed some interesting differences. Satisfaction, as measured by the PI, appeared to remain roughly the same between 2003 and 2006, with a marginally increased proportion reporting they were very or extremely satisfied in 2006. However, the distribution of responses to this question changed markedly, with an increasingly pronounced peak in the responses to 'very satisfied' in 2006 with fewer responses at the edge of the distribution.

Other compulsory questions showed different changes between 2003 and 2006. The questions asking about the flexibility of the service showed a marked decline in quality, and this was largely true of other areas of process quality, not assessed with compulsory questions. For example, the reliability and continuity of services appears to have worsened. Responses to questions assessing staff attitudes and staff skills and knowledge portrayed a mixed picture of changes in user experience.

Some compulsory questions focused on quality from the perspective of outcomes. The responses to these questions, which covered domains of safety, social participation and control over daily life, showed an improvement in outcomes in these areas, although the change in the wording of the safety questions makes comparison difficult. In contrast the non-compulsory questions that cover the outcome areas of personal care and employment and occupation showed a decline in quality for these areas. These findings are in contrast to reports from the regulators on compliance with national minimum standards (NMS) where standards for personal care (service user plan, privacy and dignity, autonomy and independence, and medication and health related activities) are reported as being most often met (Commission for Social Care Inspection, 2006b). This raises questions about the meaning that can be attributed to the standards and our measure and their comparability.

The decline in quality, as perceived by users, between 2006 and 2003, is largely due to movement between the top two levels of response to the four-point questions. There is very

little change at the bottom end of the distribution. When subjecting the 2003 dataset to factor analysis, the items had to be dichotomised as the questions had a different number of response categories and the PSSRU research team decided in draw the line between the top two levels of response. This was on the basis that this line seemed to offer a genuine distinction between users' experiences. The movement between these levels in the 2006 dataset supports the validity of this assumption. However, it may be that a scoring system that reflected all four levels would offer greater discrimination, allowing us to pick up the movement between the top two levels and the relative lack of change at the bottom. Indeed the literature suggests that where possible levels of response should be maintained in the analysis rather than collapsed in order to improve discrimination (Nunnally, 1967). This cannot be considered as an option unless all the questions have the same number of response categories and so is in the present UES not a possibility.

5.3 Validity of the quality measures

Approaches that investigated the validity of the service quality measure generally had positive findings. The measure was found to have good reliability and good construct validity as measured by correlation with the satisfaction variable and with the 2003 service quality variable. Assessment for content validity revealed that the measure covered the majority of the process quality domains lacking only an item covering staff skills and knowledge and communication, although the latter is a result of the communication item having changed between years and there is some room to debate the extent that service users will be able to appreciate the technical skills of their carers. Historically, it has been suggested that consumers do not have the capacity to judge the technical skills of staff (Davies and Ware, 1988). It follows from this that quality may be better assessed through measures that capture the qualification levels of staff in relevant courses, such as NVQs. However, recent work by PSSRU indicates that an indicator related to qualifications such as the NVQ would be a poor indicator of quality as it was found to be associated with lower quality (Netten, et al. forthcoming). Care recipients also tend to emphasise characteristics such as being older and well motivated as important rather than qualified (Francis and Netten, 2003; Harding and Beresford, 1996). There is also room to debate the appropriateness of questions that assess the attitudes of staff as certain attitudes may be related to assumptions about how carers ought to behave that may vary across service users (e.g. Bowers, et al., 2001).

In summary, the major omission from the service quality measure is outcome quality. However, as we have noted in 2003 several measures were derived and these areas captured those that are not covered (or not covered in detail) by the service quality measure, picking up on outcomes (the outcomes quality measures) and also staff knowledge as well as staff attitudes (the carer quality measure). In combination service quality, carer quality and

outcome quality capture the majority of the conceptual space of quality of home care and ideally we would want to validate these measures, as well as the measure of overall quality.

The service quality measure was dominated by items on flexibility and reliability (three items each) with only one item for each of the domains of continuity and staff attitudes. The balance of items covering each domain is important for considering the meaning of the measure, but it is also important to remember that there is some overlap of items between domains. For example, continuity can be thought of in terms of continuity of care so that a service user always receives care from the same care workers, but it can also be thought of in terms of continuity of information. In this latter situation continuity might be maintained in the absence of regular care workers where the new care worker knows what they need to do so the service user does not need to explain to the new care worker. This may be picked up in questions asking about staff knowledge. Unfortunately, there are no questions picking up on these aspects in the current questionnaire. For future surveys it will be worth considering the adequacy of coverage of each domain.

In addition to the derived measures we have also conducted some analysis to validate the weighted satisfaction measure. As we have explained, we have used the service quality measure to weight satisfaction rather than the overall quality measure because of problems with some of the items that compose the overall quality measure resulting from non-response. Analysis of the structure of the relationship between satisfaction and service quality indicated four significantly different levels of quality, represented by extremely satisfied, very satisfied, quite satisfied and neither satisfied/dissatisfied. This structure was identical to that reported in 2003, although the weighting derived from the mean value of service quality at each significantly different level of satisfaction was different between the years, being lower at each level for the 2006 dataset compared with the 2003. The structural similarity is a positive finding, indicating the validity of the structure of the measure, but the change in the value of the weightings indicates a change in the relationship between satisfaction and service quality that needs to be explored in more detail.

We examined whether any change between 2003 and 2006 could be found using the weighted satisfaction measure, weighted using the 2003 weights for both datasets, the 2006 weights for both datasets and the corresponding weights for each dataset (i.e. 2003 weights for 2003 data and 2006 weights for 2006 data). Different results were obtained for each option. Application of the 2003 weights led to a much higher estimated level of quality in 2006, but the change in quality – a slight increase – was only just significant when the 2003 weights were applied. The slight increase was not significant for the 2006 weights, and quality for both 2003 and 2006 was estimated to be much lower. However, when the 2003 weights were used for the

2003 dataset and the 2006 for the 2006 dataset, a highly significant change in quality was found between 2003 and 2006. Quality had by this method declined significantly between 2003 and 2006. These findings are important as it makes the interpretation of the weighted satisfaction measure difficult: Has quality declined or not? Has the relationship between service quality and satisfaction changed between 2003 and 2006? And if so why? Is service quality a less important predictor of people's satisfaction in 2006 compared with 2003? All of these questions seem to follow logically from these findings.

5.4 Has quality changed?

Looking at quality change using the unweighted measures of service quality and satisfaction can help to explain why these results are observed. Analysis of responses to the satisfaction question between 2003 and 2006 found no change in satisfaction between the two years when either the PI or the mean value was used. However, when all valid cases were used (i.e. excluding those where a person had not responded to either the satisfaction question or any of the service quality items), mean satisfaction is found to improve whilst mean service quality is found to decline; the PI shows no change between years. There is a clear difference then in the change in the pattern of responses to satisfaction and the service quality items, which seem to move in different directions. It is likely that this explains the different and incompatible weights that are produced from each of the datasets.

The question is what explains this phenomenon? Why does satisfaction with services improve when these individuals report a decline in perceived service quality? There are several possible explanations to this phenomenon:

- It is possible that the service quality measure does not capture aspects of quality that have improved over the period.
- It is possible that the satisfaction measure captures things other than quality and that it is too 'noisy' an indicator to use to measure quality.
- Another possible reason is that there is gaming associated with the satisfaction performance indicator.
-

We consider each of these in turn:

5.4.1 The measure of quality does not capture the entire universe of meaning of quality of home care

If we assume that satisfaction is a measure of quality *and nothing else*, then the incomplete content validity of the service quality measures might explain the observed finding. We can test this hypothesis by looking at the change in responses to the areas not included in the service quality measure. If the areas not covered show improvement, then we could assume

that it is improvement these areas that is contributing to the improvement (or lack of change in satisfaction). However, that being the case we would still have to observe either considerable improvement in the areas or assume that these areas are more important to people than those measured by the service quality measure. There is very little evidence to support the former: some outcomes have improved, but others have worsened; the latter position also seems unlikely. Evidence from other studies suggests that all the areas covered in the service quality indicator are important to people (Francis and Netten, 2004). However, before we could reject this hypothesis it would help to investigate the relationship between satisfaction and carer quality and outcome quality and look at the interrelationships to assess whether certain domains are more important determinants of satisfaction. It does seem however that it is rather the conception of satisfaction as measuring only quality that is perhaps more relevant.

5.4.2 Satisfaction is a 'noisy' indicator

The assumption relating to the first hypothesis is that satisfaction measures only quality and we have suggested that this is probably not the case. There are many theoretical reasons to suggest that satisfaction is a noisy measure. CSCI have suggested it might reflect the gratitude of service users (Commission for Social Care Inspection, 2006b) or as members of the Social Services User Survey Group have suggested their fear of losing services. It is possible to theoretically conceive of satisfaction in many different ways and the literature reflects this theoretical multiplicity (Hom, 2000; Linder-Pelz, 1982; Sitzia and Wood, 1997; Williams, 1994). What is clear from the literature is that we should consider satisfaction as a type of attitude and therefore it is a measure that reflects not only the service users' experience but their *evaluation* of that experience. As Coulter (2005) argues, satisfaction can be influenced by many other variables including expectations, personal preferences as well as the realities of the care received and these are in turn influenced by other variables that may be well beyond the control of any individual organisation. In summary, Coulter describes satisfaction as "an ill-defined concept" and concludes that it cannot provide reliable information from which to measure trends in performance (Coulter, 2005).

Indeed it would seem from our data that it is exactly in this capacity (measuring trends) that the satisfaction measure is weak. The important consideration though is whether this affects its appropriateness and usefulness as a PI of quality as Coulter suggests it does. Evidence from the 2003 survey suggested that it correlated well with quality and the current research corroborates this finding, but this research, having the benefit of two years worth of data, also finds that the relationship between service quality and satisfaction has changed. If the satisfaction PI is not picking up changes in people's experiences of services then it cannot be considered to be an adequate measure. As we observe from the relationship between satisfaction and service quality it does appear that the satisfaction measure is not picking up

changes in people's experiences of their care, but further analysis with the outcomes and carer quality measures would help to confirm this finding.

5.4.3 Gaming

The potential for gaming where data collected by organisations is used to control their behaviour is now well documented in the academic literature (e.g. Bevan and Hood, 2006b; Bird, et al., 2005; Hood, 2006; McLean et al., 2007). Bevan and Hood (2006b) describe three different effects of gaming that were identified in the former Soviet Union, known as ratchet effects, threshold effects and output distortions. They describe the effects as follows:

- Ratchet effects occur when the next years' targets are based on last year's performance. The incentive is to improve slightly but not too much to make next years targets too demanding.
- Threshold effects are effects on the distribution of performance among a range of and within production units (in this instance local authorities) putting pressure on those performing below the target level to do better, but also providing a perverse incentive for those doing better than the target to allow their performance to deteriorate to standard, and more generally to shift reported performance towards targets.
- Output distortions are the result of attempts to achieve targets at the cost of significant but unmeasured aspects of performance.

Of course the satisfaction PI is not itself a target, but poor performance across PIs can have a negative effect on the authority. It can lead to a poor star rating and consequently increased monitoring, as well as bad publicity. In effect PIs are subject to the same problems as targets and we would expect to observe the same problems. The construction of the PI is based on the proportion responding 'extremely'/ 'very satisfied' but the evaluation of the performance of the council is based on which banding the proportion falls within. The five bandings are determined by the responses to the 2003 UES – each band included 20 per cent of the councils in 2003 – for the specific purpose of making it possible for councils to improve. However, without a single goal, it can be more difficult to detect some of these effects. As McLean and colleagues (2007) note threshold effects are harder to uncover and there is less scope for exceptional performers to not be rewarded.

However, since much emphasis is placed on the goal of year-on-year improvement and tables of PIs are published, it is probably not a gross simplification of the situation to assume that the proportion responding extremely/very satisfied acts like a target. Authorities should try to improve on this proportion, if not improving on their banding, or if they do slip backwards they should not fall out of their band. The aggregate of responses to 'very satisfied' and above is in effect a threshold. Where worsening of the score existed, especially where it

threatened a councils' banding, there would be an incentive to increase the proportion in the very/extremely-satisfied category. Since there is no incentive to report 'extremely' as opposed to 'very satisfied' as they have the same value, if there is gaming we would expect data to be misrepresented for the 'very satisfied' response. In these instances we would expect to see an increasingly peaked distribution at the very satisfied response level. This is indeed what we observe in the distribution of the responses. There is a much larger peak at the very satisfied response with fewer respondents at either edge of the distribution.

We have spent some time discussing gaming because it seems to be a possible explanation for all the observed effects. Similarly to Bevan and Hood (2006a), we do not advocate dropping performance measures altogether as we believe that they are central to any strategy of quality improvement: without information about quality, how can an organisation think about where to direct its resources towards improvement? However, we do feel that the current data suggest that some degree of gaming was present in the data and for this reason we feel that changes to the data collection should be considered to ensure the validity of statistics.

Further analysis of the dataset would certainly help us to conclude with more certainty which of the three explanations is most likely (and this is something we intend to do following manipulation to adjust for non-response and the issues with some variables that arose because of problems in this area). In particular, analysis of the relationship between satisfaction and the other derived measures from 2003 that capture other aspects of quality would be helpful; further analysis to investigate other gaming effects would also be beneficial.

5.5 Measuring performance

We have put forward three alternative explanations and have argued that gaming and the noisiness of the satisfaction indicator are the most likely candidates. Therefore, it seems important to consider the consequences of the problems we have raised.

The most important consequence is what these findings mean for the validity of the PI and the use of the PI as a measure of quality for adjustment of national accounts. Where gaming is present the PI will not provide completely accurate information, as it does not truthfully represent the current situation. Of course the accuracy will be a function of the extent of gaming and this may vary by organisation. The general trend at a national level may be more reliable than trends within organisations and this may lead to false views of the success (or otherwise) of certain organisations.

It would be difficult to justify using a measure that was subject to any degree of gaming to adjust for quality change in National Accounts. The National Accounts are produced by the

ONS and are subject to strict rules governing the collection and use of statistics. Indeed it was the feeling among statisticians that PIs were not abiding by these guidelines that led to the convention of a working party by the Royal Statistical Society on the subject of performance measurement (Bird et al., 2005). A measure of quality used for this purpose would need to meet the criteria set out by Bird and colleagues.

As we have suggested these problems are not reasons to reject PIs or the collection of this type of data outright. Bevan and Hood (2004; 2006b) have argued that it may be possible to retain the social benefits of PIs whilst minimising the social costs. They have argued cogently that greater uncertainty in the specification of PIs would make it more difficult for LAs to take part in gaming. They identify the principle of transparency as working in opposition to this, but note that it is only an extended version of the principle that is incompatible with PIs. Introduction of greater uncertainty would only need to mean that the PIs would not be transparent in real time. They could still be transparent in process and retrospect. They also suggest introducing an independent body that could monitor the accuracy of performance data. (Hood, 2006) also suggests that third parties, without vested interests in the results, could be contracted to collect the data. This would likely increase costs and issues around data protection would need be carefully thought through, but, as well as improving validity, it would remove the burden of collection from LAs. All of these alternatives work as anti-gaming strategies; they accept it is happening and try to introduce mechanisms to make it more difficult for organisations to partake in such activities.

However, it is possible to think of other alternatives, although these alternatives would work outside the PI system. Since incentives to gaming operate where the data collected by organisations are being used to control their behaviour. Another solution would be to divest the data collection of this usage and so remove the incentives to gaming. This would mean that data would not be used as PIs but would rather be collected by LAs for their own benefit to inform their own improvement, a form of continuous improvement. Whether such an approach to quality improvement is effective is an empirical question to be tested, but as Bevan and Hood note, attempts in Wales to improve performance of NHS services without a star system were criticised by the auditor general for Wales as providing "neither strong incentives nor sanction to improve…performance" (quoted in Bevan and Hood, 2006a) and there is evidence to suggest that the PI system has had its successes so removing it entirely may not be in the interest of the wider society.

The explanation that the satisfaction measure is noisy also has consequences for its status as a PI as it affects its interpretation. We have suggested that it is not capturing changes in users' experiences of services that are related to domains of quality such as reliability, continuity

and so on, which was indicated through a comparison of change measured using the derived service quality measure and the satisfaction measure. These conclusions are thus dependent on the adequacy of the service quality measure and as we have suggested more work could be carried out to explore the validity of this conclusion. However there are several reasons to suggest that this measure captures quality more adequately than a measure based on user satisfaction:

- The service quality measure is a composite measure, calculated through combining the responses to several items. This means that it is more reliable than a single item measure as the error in response is partly compensated for by asking multiple questions (Nunnally, 1967). A composite measure also helps with discrimination as it lengthens the response scale, drawing out individuals or LAs along its length rather than, as for a single item, bunched towards one end. The effect of bunching is demonstrated by the finding that the current PI is less sensitive than the mean of the satisfaction measure to changes in quality, which is probably a result of collapsing the seven-point scale further to form a two-point scale.
- There are also theoretical reasons to suggest that the composite measure is more accurate. Since the items for this measure ask about experiences, they are less susceptible to influence by more general attitudes towards or expectations about the service. The items are designed to elicit reports of what actually happened, such as 'did the care workers arrive on time', and not the service users' evaluation of what happened. Although they might be affected by other problems such as recall and so on, the satisfaction measure also suffers from these general problems associated with self-reported data.

Being able to measure and monitor quality using any indicator represents a big step forward, but even if gaming is not a problem, the issues identified above suggest that there is room to improve the current measure. As Bird and colleagues (2005) discuss it is vitally important that PIs meet agreed standards for publicly reported statistics and these include the potential to identify change and being straightforward to interpret. We have identified potential problems in both of these areas with the current satisfaction indicator. The PI is generally quite insensitive to changes and the measure on which it is based does not seem to be identifying change in quality, as measured by the service quality measure. We do not know what change it is capturing and further analysis would help our interpretation of the measure in this respect. If we want to consider using these data on quality for National Accounts then we need to be sure that the adjustment used is in fact an adjustment for quality change.

As a remedy to both the gaming and noisy indicator problems, we suggest that an alternative measure might be developed that is based on the service quality measure. As we have

suggested a composite measure is more reliable and more likely to capture quality. The service quality measure also seems sensitive to change and it is less transparent so could potentially obviate incentives to game. The present measure would need some development in order that it captures the entire universe of meaning of quality of home care. The questionnaire would also need development, replacing the grid questions with new questions that capture the domains they cover but have better response rates. Of course composite measures introduce their own problems regarding how different domains should be combined (Jacobs and Goddard, 2007). However, it seems theoretically justified to assume that quality is a uni-dimensional construct, so items could be simply added together, although this hypothesis would need to be tested. Overall the service quality measure, or a successor in the same vein, seems to be a solution to the problems we have identified.

One issue to consider is the willingness of LAs to adopt such a composite measure. LAs are currently unwilling to send out long questionnaires and a composite measure would need to have a number of items. Most similar types of measures seem to have at least 15 items and to cover all the domains of quality identified a measure of quality would have to have a similar number of items to cover all the identified domains of quality of home care. Engagement with eventual users of the measure will clearly be critical to its success. It should be noted though that scale based measures of quality are not new. One such measure, known as SERVQUAL (Parasuraman, et al. 1988), has been used with success in various parts of the private sector and has been tried in public sector organisations (Wisniewski, 2001; Wisniewski and Donnelly, 1996). It has been criticised for theoretical reasons (Buttle, 1996) and its use of items that measure satisfaction with various aspects of quality would seem most damaging in this respect. The findings from the measure have, however, been seen as useful and helpful for organisations in thinking about improving the quality of their services.

6. Conclusion

We have considered the validity of using the weighted satisfaction measure as a measure of quality and suggested that there may be some difficulties in using this as a measure of quality change. These arise, we have argued, from its use as a PI, from its specification as a single point estimate and from it being a measure of satisfaction. These flaws were importantly only apparent when a time series was present, which emphasises the importance of collecting a time series to establish validity of measures designed to capture change.

We have identified multiple and competing interests in a measure of quality change: from local authorities interested in improving their services, managers and central Government interested in monitoring such services and ONS interested in capturing quality change in

order to adjust measures of output in National Accounts. With such a range of potential end users of such a measure, there is the potential for conflicting and incompatible demands. It may be that the unintended consequences of using statistics for performance management, such as gaming, may make using such statistics in other areas where confidence in the validity of the statistics is required difficult. Consideration of such competing interests and how these can be reconciled by alternative modes of collection of data or alternative data collections may improve the measures.

The results from this study have led us to suggest that some of the flaws of the current quality measure could be overcome by introducing a measure based on the same principles as the service quality measure derived from the previous PSSRU extension study. We have discussed that some problems with the measure would need to be overcome through some redevelopment of the questionnaire and remoulding of the measure. The redevelopment required is a reflection of the fact that the survey was not originally developed to provide a scale of home care quality. The end measure would also need to be tested for validity and would clearly benefit from input from end users in its development.

Aside from those issues associated with the measurement of quality change, we have identified that some questions in the survey would benefit from remodelling to improve response rates. In Appendix H we suggest how the questionnaire might be improved. We have also suggested that approaches to sampling may want to be reconsidered to ensure comparability and applicability at a national level, should the latter seem important.

References

Agresti, A. and Finlay, B. (1997) *Statistical Methods for the Social Sciences*, Third Edition, Prentice-Hall, New Jersey.

Arber, S., Gilbert, G. N. and Evandrou, M. (1988) Gender, household composition and receipt of domiciliary services by elderly disabled people, *Journal of Social Policy* 17, 153-175.

Atkinson, A. (2005) *Atkinson Review of Government Output and Productivity for the National Accounts: Final Report*, Palgrave Macmillan, Basingstoke.

Beresford, P., Croft, S., Evans, C. and Harding, T. (1997) Quality in Personal Social Services: The Developing Role of User Involvement in the UK, in A. Evers, R. Haverinen, K. Leichsenring and G. Wistow (eds) *Developing Quality in Personal Social Services: Concepts, Cases and Comments*, Vol. 22, Ashgate, Aldershot.

Bevan, G. and Hood, C. (2004) Targets, inspections, and transparency, *British Medical Journal* 328,**7440**, 598-598.

Bevan, G. and Hood, C. 2006a Have Targets Improved Performance in the English NHS? *British Medical Journal* 332, **7538,** 419-422.

Bevan, G. and Hood, C. 2006b What's measured is what matters: targets and gaming in the English public health care system, *Public Administration* 84, **3**, 517-538.

Bird, S. M., Cox, D., Farewell, V. T., Goldstein, H., Holt, T. and Smith, P. C. (2005) Performance indicators: good, bad and ugly, *Journal of the Royal Statistical Society A* 168, **1**, 1-27.

Bowers, B., Fibich, B. and Jacobson, N. (2001) Care-as-Service, Care-as-Relating, Care-as-Comfort: Understanding Nursing Homes Residents' Definitions of Quality, *The Gerontologist* 41, **4**, 539-545.

Boyne, G. A. (2002) Concepts and Indicators of Local Authority Performance: An Evaluation of the Statutory Frameworks in England and Wales, *Public Money and Management* 22, **2**, 17-24.

Buttle, F. (1996) SERVQUAL: review, critique, research agenda, *European Journal of Marketing* 30, **1**, 8-32.

Calnan, M., Almond, S. and Smith, N. (2003) Ageing and public satisfaction with the health service: an analysis of recent trends, *Social Science and Medicine* 57, **4**, 757-62.

Commission for Social Care Inspection (2006a) *Performance Ratings for Adults' Social Services in England*, Commission for Social Care Inspection, London.

Commission for Social Care Inspection (2006b) *Time to Care*, Commission for Social Care Inspection, London.

Coulter, A. (2005) *Trends in Patients' Experience of the NHS*, The Picker Institute, Oxford.

Darton, R., Forder, J., Bebbington, A., Netten, A., Towers, A.-M. and Williams, J. (2006) *Analysis to support the Development of FSS Formulae for Older People: Final Report.* Discussion Paper No. 2265/2, Personal Social Services Research Unit, University of Kent, Canterbury.

Davies, A. R. and Ware, J. E. (1988) Involving consumers in quality of care assessment, *Health Affairs* 7, **1**, 33-48.

Department for Transport Local Government and the Regions (1998) *Modern Local Government: in Touch with the People*, TSO, London.

Department of Health (1998) *Modernising Social Services: promoting independence, improving protection, raising standards*, CM 4169 Edition: TSO, London.

Ellul, J., Watkins, C. and Barer, D. (1998) Estimating total Barthel scores from just three items: the European Stroke Database 'minimum dataset' for assessing functional status at discharge from hospital, *Age and Ageing* 27, **2**, 115-122.

Evandrou, M. (2005) Health and Social Care, in Office for National Statistics (ed) *Focus on Older People*, The Stationery Office, London.

Fichman, M. and Cummings, J. N. (2003) Multiple Imputation for Missing Data: Making the most of What you Know, *Organizational Research Methods* 6, **3**, 282-308.

Francis, J. and Netten, A. (2003) *Quality in home care: Client and provider views*. PSSRU Discussion Paper No. 2017, Personal Social Services Research Unit, University of Kent, Canterbury.

Francis, J. and Netten, A. (2004) Raising the quality of home care: a study of service users' views, *Social Policy and Administration* 38, 290-305.

Harding, T. and Beresford, P. (1996) *The Standards We Expect: What Services Users and Carers Want From Social Services Workers*, The National Institute for Social Work, London.

Hom, W. (2000) *An Overview of Customer Satisfaction Models* Chancellor's Office, California Community Colleges, California.

Hood, C. (2006) Gaming in Targetworld: the Targets Approach to Managing British Public Services, *Public Administration Review* 66, **4** 515-521.

Information Centre (2005) *Personal Social Services Older People's Home Care User Experience Survey. What Needs Doing for 2005-2006*. Information Centre, London.

Information Centre (2006) *Community Care Statistics 2005. Home care services for adults, England*': Information Centre, London.

Information Centre (2007) *Personal Social Services Survey of Home Care Users in England aged 65 and over, 2005-06*. Information Centre, London.

Jacobs, R. and Goddard, M. (2007) How Do Performance Indicators Add Up? An Examination of Composite Indicators in Public Services, *Public Money and Management* 27, **2**, 103-110.

Jones, K., Netten, A. and Sandhu, S. (2005) *Home Care Provider Survey in Kent: Report for Kent Community Care Association*. PSSRU Discussion Paper No. 2230/2, Personal Social Services Research Unit, University of Kent, Canterbury.

Kalton, G. and Schuman, H. (1982) The Effect of the Question on Survey Responses: A Review, *Journal of the Royal Statistical Society A* 145, **1**, 42-73.

Linder-Pelz, S. (1982) Toward a theory of patient satisfaction, *Social Science and Medicine* 16, 577-582.

Little, R. J. A. (1988) A test of missing completely at random for multivariate data with missing values, *Journal of the American Statistical Association* 83, 1198-1202.

Little, R. J. A. and Schenker, N. (1994) Missing data, in G. Arminger, C. C. Clogg and M. E. Sobel (eds) *Handbook for Statistical Modeling in the Social and Behavioral Sciences*, Plenum, New York.

Malley, J., Sandhu, S. and 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 No. 2360, Personal Social Services Research Unit, Uiversity of Kent, Canterbury.

McLean, I., Haubrich, D. and Gutiérrez-Romero, R. (2007) The Perils and Pitfalls of Performance Measurement: The CPA regime for Local Authorities in England, *Public Money and Management* 27, **2**, 111-118.

Netten, A., Forder, J. and Shapiro, J. (2006) Measuring Personal Social Services Outputs for National Accounts: Services for Older People, PSSRU Discussion Paper No. 2267/3, Personal Social Services Research Unit, University of Kent, Canterbury.

Netten, A., Francis, J., Jones, K. and Bebbington, A. (2004) *Performance and quality: user experiences of home care services. Final Report*, PSSRU Discussion Paper No. 2104/3 Personal Social Services Research Unit, University of Kent, Canterbury.

Netten, A., Jones, K. and Sandhu, S. (forthcoming) Provider and care workforce influences in quality of home care services in England, *Journal of Aging and Social Policy*.

Netten, A., McDaid, D., Fernandez, J.-L., Forder, J., Knapp, M., Matosevic, T. and Shapiro, J. (2005) *Measuring and understanding social services outputs*, PSSRU Discussion Paper No. 2132/3, Personal Social Services Research Unit, University of Kent, Canterbury.

Netten, A., Ryan, M., Smith, P., Skatun, D., Healey, A., Knapp, M. and Wykes, T. (2002) *The development of a measure of social care outcome for older people*, PSSRU Discussion Paper No. 1690/2, Personal Social Services Research Unit, University of Kent, Canterbury.

Nunnally, J. C. (1967) *Psychometric Theory*, McGraw-Hill Book Company, New York.

Parasuraman, A., Zeithmal, V. and Berry, A. (1988) 'SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality', *Journal of Retailing* 64, 12-40.

Pickard, L., Wittenberg, R., Comas-Herrara, A., Davies, B. and Darton, R. (2000) Relying on informal care in the new century? Informal care for elderly people in England to 2031, *Ageing and Society* 20, 745-772.

Qureshi, H., Patmore, C., Nicholas, E. and Bamford, C. H. (1998) *Overview: Outcomes of Social Care for Older People, Outcomes in Community Care Practice Series*, Social Policy Research Unit, University of York, York.

Qureshi, H. and Rowlands, O. (2004) User satisfaction surveys and cognitive question testing in the public sector: the case of personal social services in England, *International journal of social research methodology* 7, **4**, 273-288.

Robine, J.-M., Jagger, C. and Romieu, I. (2002) *Selection of a Coherent Set of Health Indicators for the European Union. Phase II: Final report*, Euro-REVES, Montpelier.

Särndal, C.-E. and Lundström, S. (2005) *Estimation in Surveys with Nonresponse*, John Wiley & Sons Ltd, Chichester.

Schafer, J. L. and Graham, J. W. (2002) Missing data: our view of the state of the art, *Psychological Methods* 7, 147-177.

Schuman, H. and Presser, S. (1996) *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*, Second Edition, Sage, London.

Sitzia, J. and Wood, N. (1997) Patient satisfaction: A review of issues and concepts, *Social Science & Medicine* 45, **12**, 1829.

Starfish Consulting (2003) *User Survey Analysis Final Report: Report to London ADSS Benchmarking Region*, Starfish Consulting, London.

Tabachnick, B. and Fidell, L. (2001) *Using Multivariate Statistics*, 4th Edition: Allyn and Bacon, Boston.

Watson, D. (2002) A Critical Perspective on Quality within the Personal Social Services: Prospects and Concerns, *British Journal of Social Work* 32, **7,** 877-891.

Williams, B. (1994) Patient satisfaction: A valid concept? *Social Science & Medicine* 38, **4**, 509.

Wisniewski, M. (2001) Using SERVQUAL to assess customer satisfaction with public sector services, *Managing Service Quality* 11, **6**, 380 - 388.

Wisniewski, M. and Donnelly, M. (1996) Measuring service quality in the public sector: the potential for SERVQUAL, *Total Quality Management* 7, **4**, 357.

Wittenberg, R., Comas-Herrara, A., King, D., Malley, J., Pickard, L. and Darton, R. (2006) *Future Demand for Long-Term Care, 2002 to 2041: Projections of Demand for Long-Term Care for Older People in England.* PSSRU Discussion Paper No. 2330, Personal Social Services Research Unit, University of Kent, Canterbury.

Wood, A. M., White, I. R. and Hotopf, M. (2006) Using number of failed contact attempts to adjust for non-ignorable non-response, *Journal of the Royal Statistical Society A* 169, **3**, 525-542.

Appendix A

Your Home Care Service

What we would like you to do
We would like you to help us by taking a few minutes to give us your views about the home care services you receive. If you do not wish to answer the questions, this won't affect the services you receive.

What to do if you need help
If you would like, you can ask a friend or a relative to help you complete the questionnaire. **[Councils can mention here any telephone help line they have through which assistance in completing the survey can be arranged]**

What to do if you have queries or would like to obtain information on the results
If you, or your friend or relative, have questions you would like to ask about the questionnaire, please ring …… on Monday to Friday between 10.00 a.m and 12.00 p.m. or between 2.00 p.m. and 4.00 p.m. **[Councils can vary these hours or expand this sentence eg to say leave a message and someone will get back to you] [Councils can add a paragraph on the availability of Alternative formats of questionnaire here]**

Why you were selected
Your name is just one of many that have been selected at random from [Social Services'] records.

What will be done with the results of the survey
The results of the survey will be used by the Commission for Social Care Inspection, the Department of Health and your local [social services department] to see how happy people are with the home care services, to see whether improvements need to be made to local care services and for research purposes.

Confidentiality
Your answers will be treated as confidential: they will not be passed on to your care workers, your social worker or anyone else responsible for providing you with home care or other help

(except that they may be provided to your home care provider after being anonymised).
If you say on the form that you are being hurt or harmed by anybody, someone (but not your care worker) **[Councils can be more specific if they wish]** will contact you to talk about it.

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.

[Councils may include a comments box after some or all questions]

1. **Overall, how satisfied are you with the help from [Social Services] that you receive in your own home?**

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	☐

2. Do your care workers come at times that suit you?

Please tick [✓] one box

They always come at times that suit me	☐
They usually come at times that suit me	☐
They sometimes come at times that suit me	☐
They never come at times that suit me	☐

3. Are you kept informed, by your home care service, about changes in your care? (e.g. your visit will be late or you'll have a different carer)

Please tick [✓] one box

Someone always lets me know about changes	☐
Someone usually lets me know about changes	☐
They hardly ever let me know about changes	☐
They never let me know about changes	☐

4. Do your care workers do the things that you want done?

Please tick [✓] one box

They always do the things I want done	☐
They nearly always do the things I want done	☐
They sometimes do the things I want done	☐
They never do the things I want done	☐

Please list other things you want your care workers to do in this box

5. Do you know how to make a complaint about the Home Care Service?

Please tick [✓] one box

Yes and I feel I could if I wanted to	☐
Yes but I do not feel I could if I wanted to	☐
No I do not know how to make a complaint	☐

6. Please read the following statements and then put a tick (✓) next to the answer which comes closest to the one you want to give.

	Strongly agree	Agree	Disagree	Strongly disagree
I feel safe in my home				
I have as much contact with other people as I want				
I get up and go to bed at times that suit me				
I am always as clean as I want to be				
I always feel comfortable				
I spend too long with nothing interesting to do				

7. Which of the following statements best describes your present situation?

By 'control over daily life' we mean you have the choice to do what you want when you want to, for example having meals, going to bed and getting up, going out etc.

Please tick [✓] one box

I feel in control of my daily life ☐

Services help me to feel in control of my daily life ☐

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

I have no control over my daily life ☐

8. Do your care workers arrive on time?

Please tick [✓] one box

My care workers are never on time	☐
My care workers are sometimes on time	☐
My care workers are often on time	☐
My care workers are always on time	☐
I never know what time my care workers are going to arrive	☐

9. Do your care workers spend less time with you than they are supposed to?

Please tick [✓] one box

They never spend less time with me than they are supposed to	☐
They sometimes spend less time with me than they are supposed to	☐
They often spend less time with me than they are supposed to	☐
They always spend less time with me than they are supposed to	☐

10. Are your care workers in a rush?

Please tick [✓] one box

They are never in a rush ☐

They are sometimes in a rush ☐

They are often in a rush ☐

They are always in a rush ☐

11. Do you have as many visits from your care workers as you need?

Please tick [✓] one box

Yes, I have as many visits as I need ☐

No, I need a few more visits ☐

No, I need a lot more visits ☐

12. Do you always see the same care workers?

Please tick [✓] one box

Yes, I always see the same care workers ☐

No, but I nearly always see the same care workers ☐

No, I hardly ever see the same care workers ☐

No, I never see the same care workers ☐

13. Overall, how do you feel about the way your care workers treat you? (e.g. whether they are understanding and treat you with respect)

Please tick [✓] one box

I am always happy with the way my care workers treat me	☐
I am usually happy with the way my care workers treat me	☐
I am sometimes happy with the way my care workers treat me	☐
I am never happy with the way my care workers treat me	☐

14. Now, please read the following statements and then put a tick [✓] next to each statement under the answer which comes closest to the one you want to give.

	Strongly agree	Agree	Disagree	Strongly disagree
My care workers are obliging				
My care workers are unfriendly				
As far as I know, my care workers keep any personal details they know about me to themselves				
My care workers gossip to me about other people they care for				
My care workers are excellent at what they do				
My care workers are less thorough than I would like				
My care workers treat me with respect				
My care workers do things in their way rather than mine				
My care workers are careless				
My care workers are honest				

[This is where councils should add optional questions other than about direct payments]

[Councils can choose to include this question if they wish]

15. If you could change one thing about your home care services, what would it be?

Please write your answer in this box

[Councils can choose to include this question if they wish]

16. Please write any other comments you would like to make about the home care you receive in this box

17. Direct payments is the option for you to purchase some or all of the services you are eligible for directly yourself, using money provided by [Social Services]. This should not be confused with welfare benefits that are usually paid directly into an account which are also called direct payments.

Has your social worker or care manager told you about direct payments?

Please tick [✓] one box

Yes ☐

No ☐

Don't know ☐

[This is where councils should add additional satisfaction questions about direct payments for use locally]

The answers to the next group of questions will be used to make sure that we have a balanced sample of home care users.

18. Do you receive any practical help from any friends, neighbours or family members?

Please tick [✓] all those that apply

Yes, from someone living in my household ☐

Yes, from someone living in another household ☐

No ☐

19. During the past month did you use any of the following care services?

Please tick [✓] one box for each service

1.1.1.	Yes	No
Meals on wheels	☐	☐
Day centre	☐	☐
Community/district nursing services	☐	☐
Other care services (eg short breaks/residential care)	☐	☐

Please describe in this box

20. How many hours of home care do you usually receive each week?

Please write your answer in this box

[Councils may choose to include this gender question if they are not confident of their current records]

21. Are you male or female?

Please tick [✓] one box

Male ☐

Female ☐

[Councils may choose to include this age question if they are not confident of their current records]

22. How old are you?

Please tick [✓] one box

Under 65 ☐

65-74 ☐

75-84 ☐

85 or over ☐

[Councils don't have to include this ethnicity question if they are confident of the quality and coverage of the information about the client's assessment of their ethnic origin in their current records; if they do include the question, they may break the categories down further, if they wish to do so]

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

Please tick [✓] one box

a) White (British, Irish, any other white background) ☐

b) Mixed (White and Black Caribbean, White and Black African, White and Asian, any other mixed background) ☐

c) Asian or Asian British (Indian, Pakistani, Bangladeshi, any other Asian background) ☐

d) Black or Black British (Caribbean, African or any other Black background) ☐

e) Chinese ☐

f) Any other ethnic group ☐

24. Did you fill in this questionnaire by yourself or did you have help from someone else?

Please tick [✓] one box

I filled it in myself ☐

I had help from a care worker ☐

I had help from someone else ☐

[Councils may seek further information on who helped here if they wish. They may also insert additional questions eg about additional services the user would like to meet their needs better or seeking permission for information to be shared to follow up a particular matter]

25. For how long have you been receiving help from Social Services in your own home?

Please tick [✓] one box

Less than 6 months ☐

6 months to 1 year ☐

1 to 2 years ☐

2 to 5 years ☐

More than 5 years ☐

26. How is your health in general?

Please tick [✓] one box

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

27. Do you need help from somebody to:

Please tick [✓] one box for each statement

1.1.2.

	Yes	No
Get dressed or undressed	☐	☐
Get in and out of bed or a chair	☐	☐
Wash face and hands	☐	☐
Prepare hot meals	☐	☐

28. If further research about home care services were to take place, would you be happy for us to contact you?

Please tick [✓] one box

Yes ☐

No ☐

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 B

The variables included in each of the scales for 2003 and 2006 are shown in the tables below:

Table 11: Variables in the overall quality scale

	2003	*2006*
Do your care workers come at times that suit you?	Yes	Yes
Do your care workers arrive on time?	Yes	Yes
Do your care workers spend less time with you than they are supposed to?	Yes	Yes
Are your care workers in a rush?	Yes	Yes
Do your care workers do the things that you want done?	Yes	Yes
Overall, how do you feel about the way your care workers treat you?	Yes	Yes
My care workers are understanding	Yes	Dropped
My care workers are miserable	Yes	Dropped
My care workers are obliging	Yes	Yes
My care workers are unfriendly	Yes	Yes
As far as I know, my care workers keep any personal details they know about me to themselves	Yes	Yes
My care workers gossip to me about other people they care for	Yes	Yes
My care workers are excellent at what they do	Yes	Yes
My care workers are less thorough than I would like	Yes	Yes
My care workers treat me with respect	Yes	Yes
My care workers do things in their way rather than mine	Yes	Yes
My care workers are gentle	Yes	Dropped
My care workers are careless	Yes	Yes
My care workers are honest	Yes	Yes
I am always clean	Yes	Changed, but included
I always feel comfortable	Yes	Yes
I don't feel safe in my home	Yes	Changed, but included
I have as much contact with other people as I want	Yes	Yes
I don't spend too long with nothing interesting to do	Yes	Yes
I get up and go to bed at times which suit me	Yes	Yes
The help I get from Social Services has made me more independent than I was	Yes	Dropped

Table 12: Variables in the service quality scale

	2003	*2006*
Do your care workers come at times that suit you?	Yes	Yes
Do your carers arrive on time?	Yes	Yes
Do your care workers spend less time with you than they are supposed to?	Yes	Yes
Are your care workers in a rush?	Yes	Yes
Do you always see the same care workers?	Yes	Yes
Do your care workers do the things that you want done?	Yes	Yes
Are you kept informed, by your home care service, about changes in your care?	Yes	Changed, but excluded
Overall, how do you feel about the way your care workers treat you?	Yes	Yes

Table 13: Variables in the outcomes scale

	2003	*2006*
I am always clean	Yes	Changed, but included
I am always comfortable	Yes	Yes
I feel safe in my home	Yes	Changed, but included
I have as much contact with other people as I want	Yes	Yes
I don't spend too long with nothing interesting to do	Yes	Yes
I get up and go to bed at times which suit me	Yes	Yes

Table 14: Variables in the carer quality scales

	2003	*2006*
My care workers are understanding	Yes	Dropped
My care workers are obliging	Yes	Yes
As far as I know, my care workers keep any personal details they know about me to themselves	Yes	Yes
My care workers are excellent at what they do	Yes	Yes
My care workers treat me with respect	Yes	Yes
My care workers are gentle with me	Yes	Dropped
My care workers are honest	Yes	Yes
My care workers are miserable	Yes	Dropped
My care workers are unfriendly	Yes	Yes
My care workers gossip to me about other people they care for	Yes	Yes
My care workers are less thorough than I would like	Yes	Yes
My care workers do things in their way rather than mine	Yes	Yes
My care workers are careless	Yes	Yes

Appendix C

Two measures of hours of home care provision were recorded: planned as reported by the local authorities and actual hours as reported by service users. These questions were particularly poorly answered, with a 40 per cent mismatch, meaning that when one variable is not missing the other is 40 per cent of the time likely to be missing. With such a high mismatch, the samples are very different and so comparison between these figures is unwise.

For those authorities able to supply the data, a mean of 7.94 planned hours per week of home care (s.d. 8.55 hours) was reported. The median was 5.75 hours per week. When converted to a categorical variable and compared with the national picture (Information Centre, 2006), the picture is quite different (see Table 15, columns two and four). Amongst authorities able to provide data on planned hours of home care per week, they provided fewer individuals with smaller packages of care and more individuals with medium to large packages of care compared with the national picture. This could be due to inaccuracies in reporting or show some real differences between the cases within our sample that were able to provide these data compared with the national picture.

Table 15: Hours of home care per week reported by LAs and service users in the PSSRU sample compared with the national picture

Hours of home care per week	*Percentage of individuals reported by authorities*	*Percentage of individuals by self-report*	*National percentage of households*[18]
2 hours or less	15		23
More than 2 hours and up to 5 hours	31	33	26
More than 5 hours and up to 10 hours	29	30	24
More than 10 hours	25	24	27

Within the questionnaire, individuals were also asked to report how many hours of home care they received per week. The average reported by service users was 7 hours per week (s.d 7.69 hours), with a median of 5 hours per week. When converted to a categorical variable and compared with the national data (see Table 15, columns three and four), the picture is quite

[18] These data are drawn from the HH1 returns for 2005 available at:Information Centre 2006 'Community Care Statistics 2005. Home care services for adults, England': Information Centre.

different again with fewer individuals reporting that they received either small packages of care or large packages of care, and more individuals reporting that they received medium sized packages of care. It is likely that there are some inaccuracies due to interpretation of the question (e.g. reporting of hours when two carers are present for a session, reporting hours on the care plan rather than actual hours received and also the potential for confusion between care provided informally and formally or care organised privately and via the local authority) but it could also demonstrate some real differences.

As we have suggested it is unwise to compare the figures for service users and LAs directly as, due to the large number of missing values, the samples they come from are very different. However, the difference between planned and actual hours reported by service users is intriguing and we explored this further controlling for differences in sample populations (due to mismatch of cases). A new variable was calculated – the difference between the planned hours of home care per week and the actual reported hours per week (planned minus actual). The distribution of this variable was leptokurtic[19] (133.069, standard error = 0.039), with mean 0.5 (median zero) and standard deviation 7hrs. The range was wide from -167.5 to 104.5. However, the inter quartile range (IQR) was only 1.62 indicating that the values covered at the end points of the range were likely to be a result of inaccuracies in reporting. Extreme outliers can be calculated using the IQR as those values lying beyond the outer fences (Q1 – 3*IQR and Q3+ 3*IQR). The outer fences were calculated as -5.19 and 6.10 respectively. A total of 2,118 values out of 15,477 valid responses (14 per cent) were identified as extreme outliers, indicating a quite high degree of problem with the reporting and recording of these items.

Given all the problems identified we decided it was unwise to pursue any further analysis with these data, as even when we excluded extreme outliers there was still a high degree of variance around the derived variable, with a standard deviation of approximately 2 hours and a very wide range of approximately 11 hours. This seems improbable and largely due to errors in reporting.

[19] A leptokurtic distribution is one that is more peaked, than a normal distribution. A distribution flatter than a normal distribution is described as platykurtic.

Appendix D

Table 16. Percentage of respondents reporting different levels of being satisfied with the service they receive but responding negatively on the survey items

	Extremely satisfied	*Extremely/ very satisfied*	*Extremely/very/ quite satisfied*
Do your care workers come at times that suit you? (Sometimes/Never)	0.6	2.4	8.6
If you ask for changes in the help you are given, are those changes made? (Hardly ever/Never)	1.9	7.7	18.9
Do your care workers do the things that you want done? (Sometimes/Never)	0.2	0.8	3.7
Do you know how to make a complaint about the Home Care Service? (feel couldn't/no)	3.2	10.1	19.6
I feel safe in my home (Strongly Disagree/Disagree)	0.4	1.1	2.4
I have as much contact with other people as I want (Strongly Disagree/Disagree)	1.5	5.4	10.3
I get up and go to bed at times which suit me (Strongly Disagree/Disagree)	0.5	1.8	3.9
I am always clean (Strongly Disagree/Disagree)	0.3	1.5	3.6
I am always comfortable (Strongly Disagree/Disagree)	0.9	3.2	7.2
I spend too long with nothing interesting to do (Strongly Agree/Agree)	8.9	23.8	39.5
Which of the following statements best describes your present situation? (Some control/No control)	1.8	5.7	11.9
Do your carer workers arrive on time? (I never know what time carer is going to arrive/Never/Sometimes)	1.8	6.2	15.2
Do your care workers spend less time with you than they are supposed to? (Always/Often)	0.5	1.9	5.5
Are your care workers in a rush? (Always/Often)	0.5	2.3	7.0
Do you have as many visits from your care workers as you need? (Lot more/Few more)	1.3	3.6	6.9
Do you always see the same care workers? (Hardly Ever/Never)	0.4	1.5	3.8

Overall, how do you feel about the way your care workers treat you? (Sometimes/Never Happy)	0.1	0.4	1.7
My care workers are obliging (Strongly Disagree/Disagree)	0.1	0.2	0.5
My care workers are unfriendly (Strongly Agree/Agree)	0.8	2.1	3.6
As far as I know, my care workers keep any personal details they know about me to themselves (Strongly Disagree/Disagree)	0.5	1.2	2.0
My care workers do gossip to me about other people they care for (Strongly Agree/Agree)	1.0	2.6	5.0
My care workers are excellent at what they do (Strongly Disagree/Disagree)	0.2	0.8	3.2
My care workers are less thorough than I would like (Strongly Agree/Agree)	1.3	4.6	12.7
My care workers treat me with respect (Strongly Disagree/Disagree)	0.2	0.6	1.0
My care workers do things in their own way rather than mine (Strongly Agree/Agree)	3.3	10.6	22.2
My care workers are careless (Strongly Agree/Agree)	0.7	1.8	3.6
My care workers are honest (Strongly Disagree/Disagree)	0.3	1.0	1.7

Table 17: Percentage of respondents reporting different levels of being dissatisfied with the service they receive but responding positively on the survey items

	Extremely dissatisfied	*Extremely/ very dissatisfied*	*Extremely/very/ quite dissatisfied*
Do your care workers come at times that suit you? (Always/usually)	0.1	0.3	1.0
If you ask for changes in the help you are given, are those changes made? (Always/usually)	0.1	0.2	0.8
Do your care workers do the things that you want done? (Always/nearly always)	0.1	0.3	1.3
Do you know how to make a complaint about the Home Care Service? (yes, feel could)	0.2	0.5	1.4
I feel safe in my home (Strongly Agree/Agree)	0.3	0.8	2.6
I have as much contact with other people as I want (Strongly Agree/Agree)	0.3	0.7	2.1
I get up and go to bed at times which suit me (Strongly Agree/Agree)	0.4	0.8	2.5
I am always clean (Strongly Agree/Agree)	0.3	0.7	2.2
I am always comfortable (Strongly Agree/Agree)	0.2	0.6	2.0
I spend too long with nothing interesting to do (Strongly Disagree/Disagree)	0.2	0.5	1.4
Which of the following statements best describes your present situation? (In control/service help)	0.2	0.5	1.5
Do your carer workers arrive on time? (Always/often)	0.1	0.3	1.0
Do your care workers spend less time with you than they are supposed to? (Sometimes/Never)	0.2	0.5	1.5
Are your care workers in a rush? (Sometimes/Never)	0.1	0.4	1.4
Do you have as many visits from your care workers as you need? (As many as needed)	0.2	0.6	2.0
Do you always see the same care workers? (Always/nearly always)	0.3	0.7	2.1
Overall, how do you feel about the way your care workers treat you? (Always/usually)	0.2	0.5	1.8
My care workers are obliging (Strongly agree/Agree)	0.2	0.6	2.1

My care workers are unfriendly (Strongly Disagree/Disagree)	0.3	0.8	2.5
As far as I know, my care workers keep any personal details they know about me to themselves (Strongly agree/Agree)	0.3	0.8	2.5
My care workers do gossip to me about other people they care for (Strongly Disagree/Disagree)	0.3	0.9	2.5
My care workers are excellent at what they do (Strongly agree/Agree)	0.1	0.4	1.3
My care workers are less thorough than I would like (Strongly Disagree/Disagree)	0.1	0.3	1.2
My care workers treat me with respect (Strongly agree/Agree)	0.3	0.8	2.5
My care workers do things in their own way rather than mine (Strongly Disagree/Disagree)	0.1	0.3	1.0
My care workers are careless (Strongly Disagree/Disagree)	0.2	0.6	2.0
My care workers are honest (Strongly agree/Agree)	0.3	0.8	2.5

Appendix E

The table below shows the response rate for all questionnaire items, auxiliary variables and derived variables.

Table 18: Item response rate for all variables and derived variables

		Total responses		*Item response rate*	
		(from total sample)	*(from respondents)*	*(% of total sample)*	*(% of respondents)*
Auxiliary data					
	Gender	46,226	28,511	93	99
	Age	46,218	28,541	93	99
	Ethnicity	43,455	28,115	88	97
	method of collection	49,542	28,840	100	100
	number of providers	37,578	22,341	76	77
	planned number of hours	30,468	18,512	62	64
	type of provider	42,096	25,621	85	89
Item on questionnaire					
Q1	Satisfaction		28,171	57	98
Q2	times that suit		28,083	57	97
Q3	Changes		27,759	56	96
Q4	things want done		27,883	56	97
Q5	Complain		27,454	55	95
Q6a	feel safe		25,858	52	90
Q6b	contact with others		24,475	49	85
Q6c	suitable bedtimes		24,971	50	87
Q6d	feel clean		23,730	48	82
Q6e	feel comfortable		24,121	49	84
Q6f	nothing interesting to do		22,552	46	78
Q7	control over daily life		27,333	55	95
Q8	arrive on time		27,654	56	96
Q9	spend less time		27,049	55	94
Q10	in a rush		27,617	56	96
Q11	visits as needed		27,476	55	95
Q12	see same carers		27,822	56	96
Q13	overall treatment		27,912	56	97
Q14a	Obliging		27,087	55	94
Q14b	Unfriendly		22,996	46	80
Q14c	keep details to themselves		26,012	53	90
Q14d	Gossip		24,035	49	83
Q14e	Excellent		25,980	52	90
Q14f	Thorough		23,297	47	81
Q14g	Respect		26,919	54	93
Q14h	things their way		23,715	48	82
Q14i	Careless		23,683	48	82
Q14j	Honest		26,782	54	93
Q17	direct payments		24,878	50	86

Q18a	practical help in household		28,827	58	100
Q18b	practical help other household		28,828	58	100
Q18c	no practical help		28,824	58	100
Q19a	meals on wheels		23,456	47	81
Q19b	day centre		23,485	47	81
Q19c	community nursing		23,944	48	83
Q19d	Other		21,882	44	76
Q20	hours of home care		23,888	48	83
Q24	help completing		27,558	56	96
Q25	length of time in receipt of services		26,605	54	92
Q26	Health		27,284	55	95
Q27a	dressed/undressed		25,960	52	90
Q27b	transfer from bed/chair		23,717	48	82
Q27c	wash face and hands		24,213	49	84
Q27d	prepare hot meals		25,169	51	87
Derived variables					
	Disability score		22,358	45	78
	Combined practical help		28,816	58	100
	Overall quality		15,106	30	52
	Service quality		25,236	51	88
	Carer quality		19,871	40	69
	Outcome quality		19,473	39	68

Appendix F

Analysis for non-response bias

The aim of this analysis is to determine the extent to which non-response, both unit and item, might introduce bias into the estimates. Having determined any potential sources of bias it is then important to reflect on how this might affect the planned analyses and develop some kind of strategy to deal with the effects of non-response.

Analysis of unit non-response

This year several items were requested as auxiliary data. These items were:

- gender
- age
- ethnicity
- type of provider
- number of providers
- planned hours of home care.

However, some authorities found it difficult to provide any information for some of the items. Only 28 authorities were able to provide data for all seven of the items for both respondents and non-respondents. The items recorded least often for non-respondents were planned number of hours (31 authorities provided data), number of providers (40 authorities provided data), ethnicity (41 authorities provided data), type of provider (43 authorities provided data), gender and age (45 authorities provided data). Table 18 in Appendix C presents for each auxiliary variable the proportion of missing data as a percentage of the total sample and as a percentage of those who responded. From this table it can be seen that a significant proportion of auxiliary data is missing for all variables (except method of data collection) and that this proportion is greater for the non-responders compared with responders. This has implications for the usefulness of the data in assessing the characteristics of non-responders.

The extent to which the auxiliary data are useful in examining potential non-response bias depends on the extent to which the auxiliary data items explain variation in the variables of interest, in this case satisfaction and the derived quality variables. The 2003 extension study found that the following variables were associated with perceptions of home care as measured by the derived quality variables: gender, age, ethnicity, receipt of practical help, hours of home care receipt per week, number of providers and type of provider, although it

was noted by the authors that these variables explained a small amount of the variation (Netten et al., 2004)[20]. These findings indicated that the auxiliary data should be useful for the purpose of investigating potential bias but that any such bias was likely be limited. Of course, it may be that there are other more important determinants of quality that we are not measuring.

Given that only 28 authorities were able to provide auxiliary data for all seven variables, the following analysis of the structure of the non-respondent and respondent population focuses on these authorities. It is assumed that if there were problems in this population these would extend to the 22 authorities not included in the analysis, which seems a fair assumption to make. Table 19 shows the population proportions for all of the seven auxiliary variables for non-respondents and respondents.

[20] A recent publication by the Information Centre (2007) of the national results for the 2006 survey largely supports these findings, albeit they were looking at relationships with satisfaction. They found a relationship between age, gender and ethnicity and satisfaction, but no relationship between reported actual hours of care (not planned) and satisfaction. However, the specification of the reported hours of home care variable was different, as in the 2003 extension study a high-low variable was used separating those reporting receiving more than 10 hours from those reporting receiving less then 10 hours.

Table 19: Differences in the characteristics of the respondent and non-respondent population

Auxiliary variable	Response population		Non-response population	
	Number	Percentage	Number	Percentage
Gender				
Male	7,336	25.7%	4,783	27.0%
Female	21,175	74.3%	12,931	73.0%
Age				
65 to 74	4,350	15.2%	3,084	17.4%
75 to 84	11,722	41.1%	7,074	40.0%
85 and over	12,469	43.7%	7,518	42.5%
Ethnicity				
White	27,155	96.6%	14,485	94.4%
Mixed	74	0.3%	29	0.2%
Asian/Asian British	269	1.0%	289	1.9%
Black/Black British	475	1.7%	424	2.8%
Chinese	43	0.2%	41	0.3%
Other	99	0.4%	71	0.5%
Provider type				
In house	7,474	29.2%	4,431	26.9%
Independent	17,552	68.5%	11,689	71.0%
Both	595	2.3%	353	2.1%
Number of providers				
One provider	20,726	92.8%	13,963	91.7%
More then one provider	1,615	7.2%	1,272	8.3%
Planned hours of home care				
0 to fewer than 2 hrs	2,706	14.6%	1,626	13.6%
2 to fewer than 5 hrs	5,547	30.0%	3,304	27.6%
5 to fewer than 10 hrs	5,383	29.1%	3,365	28.1%
More than 10 hrs	4,876	26.3%	3,661	30.6%

Chi-squared analysis was used to test for any association between status as respondent and non-respondent and each auxiliary variable. The respondent and non-respondent populations were found to vary significantly by each of the auxiliary variables at the 1 per cent level of significance (for number of providers, χ^2=15.872, d.f =1; for planned number of hours, χ^2=68.581, d.f.= 3; for type of provider, χ^2=28.426, d.f.=2 ; for gender, χ^2=9.056, d.f.=1; for age χ^2=39.056, d.f.=2 ;and for ethnicity, χ^2=138.790, d.f.=5). Analysis of the standardised adjusted form of the residual[21] was performed to analyse which cells contribute the most to the difference, with a value of 2 standard errors (effectively a 5 per cent significance level) used as the cut-off point. These findings are summarised in Table 20.

[21] The difference between the observed and expected values for each cell is called the residual. An adjusted form of the residual, to reflect the difference in the size of each cell, can be used to tell whether the observed value departs enough from the expected value to be considered different. In effect the adjusted residual reports the number of standard errors that the observed count falls from the expected count. Values above 2 are possible in about 5 per cent of cases and above 3 in very few, roughly 1 per cent of cases (Agresti and Finlay, 1997).

Table 20: Differences between respondents and non-respondents

Auxiliary variable	*More likely to respond*	*Less likely to respond*
Gender	Female	Male
Age	75 to 84 85 and over	65 to 74
Ethnicity	White	Asian/Asian British Black/Black British Chinese
Provider type	In-house	Independent
Number of providers	One provider	More than one provider
Planned hours of home care	2 to fewer than 5hrs 0 to fewer than 2hrs	10 hrs and over

Analysis of the 2003 dataset showed that all of these variables varied significantly with perceived quality and this would suggest that the non-response may affect the randomness of the sample and bias the estimates. However, it is important to recognise that this conclusion is based on the assumption that the people who have not responded from each of these population groupings will answer on average in the same way as those who have responded already. It is, however, possible that those people who have not responded may have different viewpoints to those who have responded or that something unaccounted for may be driving their non-response. This is a theoretical problem and one that therefore needs to be justified in terms of each specific use of the procedure.

In this instance, the population groupings derived from the auxiliary variables are likely to reflect a genuinely different milieu of service delivery and so there is some theoretically justified position from which to assert that those who have not responded are likely to respond in the same way as those who have. For example, ethnicity may affect the way a person experiences the provider's services, in terms of the cultural sensitivity as might gender, if, for example, people have preferences for the gender of the care-giver based on their own gender. However, other variables may reflect different milieus more adequately, for example the provider agency and functional ability. Provider agencies are likely to have their own ways of working that affect the experience of service users and degree of disability may influence the type of service that is provided. Indeed, as we note above, previous work by PSSRU found that the variables we have identified in Table 20 only explained a small amount of the variation in quality (Netten et al., 2004). A search for other variables that explain variation in quality and can be collected as auxiliary data is likely to be profitable. It would also be useful to have some evidence to determine whether non-respondents have different views to respondents. Information about the stage at which a respondent responded,

for example whether they responded to the first or second mail-out of the questionnaire, may help in this regard.

Having established that the differences in the non-respondent and respondent populations on these auxiliary variables could introduce bias into the estimates of perceived quality and satisfaction, the question is how to resolve this issue. A weighting procedure is the only procedure available to adjust for bias due to unit non-response (Särndal and Lundström, 2005) and it would also help to adjust for item non-response. The procedure also compensates for complexity introduced by interrelationships between auxiliary variables using a technique called propensity scoring. The difference between un-weighted and weighted estimates provides an estimate of the extent to which the sample is biased due to unit non-response.

In the analyses planned, we are not attempting to estimate actual levels of quality of home care services in the English 65 plus population, so the weighting procedure is not essential. However, it is the aim of the work that an indicator for quality should be developed from this data to feed into ongoing work in the PSSRU programme that should accurately reflect quality nationally. Any indicator proposed should be adjusted for bias due to unit non-response using the weighting procedure. Before proposing the weights for the quality measure, we will need to examine the difference that weighting the dataset for non-response bias makes to the estimates.

For comparative work, it is essential that the samples are not biased, as this introduces error into the estimates upon which calculations of differences are made. We would not know whether differences are due to error in the estimates or differences in the populations. Where the comparisons are drawn from sub-samples of the sample of responders the problem is minimal as each sub-sample has the same error bias. However, where samples are compared with other samples, for example between authorities the problem is more acute. Estimates could be adjusted at the level of the authority using the weighting procedure; but, not all authorities have been able to provide complete sets of auxiliary data so this approach would not be possible. To do nothing seems inappropriate as the pattern of non-response varies by authority (see Appendix G), but we are reassured by the small contribution to explaining variation in quality that these auxiliary variables make. Until all authorities are able to provide auxiliary data for all variables doing nothing remains the only viable option.

A similar problem exists for the comparison across time as the 2003 study did not collect auxiliary data for non-respondents. This means it is impossible to examine the extent of bias in the 2003 sample derived from unit non-response. Comparison of estimates for authorities that participated in both years may help in this regard but this avenue still assumes that the

type of non-respondents for each authority were the same for the 2003 and 2006 sample and would produce the same bias in the estimates. This is quite a strong assumption that we have no way of testing, but again we are reassured by the small contribution towards variation in quality that these auxiliary variables explain.

Analysis of item non-response

The SPSS missing values analysis programme was used to investigate item non-response. Little's MCAR test (χ^2 =1299.820, d.f.=458, p=<0.01) indicated that the missing data pattern was not missing completely at random. Further analysis of patterns and separate variance t-tests on the derived quality variables and the satisfaction indicator for all questionnaire items and auxiliary data items confirmed this view. In such a situation the optimal solution is considered to be multiple imputation for missing values (Fichman and Cummings, 2003; Särndal and Lundström, 2005; Schafer and Graham, 2002). This method uses the valid data to estimate, several times over, values for the missing data. Multiple imputations, as opposed to single imputation, of missing values allows for the effect of the imputed values on the estimates to be gauged. A comparison between imputed and non-imputed estimates can reveal the likely extent of bias introduced through item non-response. However, this strategy takes time to implement.

A more simple and quicker approach to dealing with non-response that is preferable to other simple procedures is complete case analysis, where only records with complete response sets are used in the analysis (Little and Schenker, 1994; Särndal and Lundström, 2005). However, where many of the cases are missing some values this can result in a loss of a lot of data, which is inefficient. In this dataset there are only 5,067 cases with valid responses for all variables, about 18 per cent of the original dataset size and a significant loss of information. Another method can be used which uses all the valid cases for each variable or variables pair depending on the analysis. This is known as valid case analysis. It increases the efficiency of the analysis, but can distort relationships between the variables[22]. Since we plan to impute data at a later stage, we identify here which analyses are most likely to be robust to the missing data problem and in particular valid case analysis.

The following are just some of the relationships that were found between missing items and valid items[23]

[22] For example Tabachnick and Fidell (2001) report that correlations can be altered and be outside the valid range.

[23] Analysis was conducted using separate variance t-tests, which do not account for the increases in probability error associated with conducting multiple tests. We therefore have only reported results where there were highly significant results.

- The average value for satisfaction when either provider type or number of providers is missing is significantly lower than in cases where the values for these variables are found. Given that these two variables are missing for whole authorities rather than individuals, this implies that authorities who were unable to provide auxiliary data have less satisfied service users. This suggests very interesting implications for the value of knowledge and data systems in contributing towards high quality services.
- The average value for satisfaction where the derived quality variables were missing, i.e. service quality, carer quality, outcomes quality and overall quality, is lower than in cases where the derived quality variables are present. This implies that less satisfied people are less likely to answer the questions. Therefore estimates of quality derived using these variables are likely to overestimate quality.
- When statements are positively phrased, the average value for satisfaction for all the missing cases is lower. This indicates that less satisfied people are less likely to answer positively phrased statements.
- Conversely, when statements are negatively phrased, the average value for satisfaction for all the missing cases is higher. This indicates that more satisfied people are less likely to answer negatively phrased statements.

It seems that there is relationship between the phrasing of a statement and the value of the missing items, where people who perceive the service negatively are less likely to answer a positively phrased statement and people who perceive the service more positively are less likely to answer a negatively phrased statement. There is also, however, likely to be an interaction between these relationships and the general finding that people who are less satisfied are less likely to answer any question about the quality of their services. These relationships have important consequences for both the functioning of the derived quality variables and the weighted satisfaction indicator of quality. For these reasons we have decided to not continue investigations with the quality variables including negatively and positively phrased statements i.e. the overall quality variable, carer quality and outcomes quality, until we have had an opportunity to investigate the effect of multiple imputation on the estimates.

The service quality indicator, although also exhibiting the same outcome from the missing data problems as the other quality variables, at least does not contain any statements so its missing data pattern is less complex. It also has far fewer missing values compared with the other derived quality variables, meaning that any moderation of the relationship between satisfaction and service quality caused by missing values is likely to be tempered. We therefore decided to replace the overall quality variable with the service quality variable in the weighted satisfaction measure and conduct the validation using this construction of the

weighted satisfaction. Since the overall quality variable contains the majority of the items in the service quality indicator it should provide a good indication of the validity of the approach.

It is still possible that the relationship between satisfaction and service quality is altered by the missing values pattern and therefore possible that this might influence the weighting structure for the weighted satisfaction measure. If the pattern of item non-response to service quality or satisfaction has changed between 2003 and 2006, this could have quite serious consequences for the validation exercise as well as for the analysis of change in perceptions of quality. We therefore examined the patterns of item non-response for service quality in the 2003 dataset and found that the same problem, that people who were less satisfied were less likely to answer the service quality questions, existed. (Interestingly, the relationship between positive and negative statements still remained, but was not consistently significant.) This reassures us that the pattern of non-response has remained the same between 2003 and 2006 and that validation of the weighting structure should not be overly influenced by the item non-response. It also reassures us that the analysis of quality change is likely to be quite robust to non-response.

Another potential problem is that changes in the structure of the population receiving services, for example the move towards increasing intensification of services, could impact on the problem of non-response in the relationship between satisfaction and service quality. Fortunately all variables identified explain a very small amount of variation in quality, so we can be hopeful that they would not interfere in this relationship. However, given the finding from 2003 that independent providers and reported receipt of more than 10hrs of home care were associated with lower perceptions of quality, it is possible that a significant change in the structure of the population along these variables could contribute to a change in the relationship between satisfaction and service quality, and hence weighting structure, as more dissatisfied people lead to fewer people responding to the service quality item. However, the level of item non-response for service quality between 2003 and 2006 was very similar (and in fact non-response was higher in 2003). We can therefore be relatively sure that this is not a problem.

Analysis of mismatched pairs in SPSS missing values analysis shows for each pair of data items the extent to which the items, if missing, were missing together. For some variables, such as age, gender and ethnicity, a very small percentage of the responses are not missing when satisfaction and service quality were missing and vice versa. However, for some other variables, such as the questions asking about planned and actual hours, number of providers, provider type, services and ADLs, quite a large proportion of the responses were missing when satisfaction and service quality were not missing and vice versa. It is possible that using

pairwise correlations for planned analyses, in particular the regression, might adversely affect the results as a result of large differences in the sample size for each correlation. Given the very small proportion of cases with complete data, this option is not viable and it seems most sensible to multiply impute data before going ahead with any multivariate analysis.

Appendix G

Appendix H

We suggest that the questions that follow, drawn from the user experience survey that was developed for adults aged 18-64 with physical and sensory impairments (Malley, et al. 2006), may be considered as a replacement to the grids, questions 6 and 14 (see).

Table 21: Replacements for question 6

Questions	Responses
Which of the following statements best describes how safe you feel?	
Thinking about the way you look and feel, which of the following statements best describes your present situation?	
Which of the following statements best describes your present situation with respect to your social life?	
Which of the following statements best describes how involved you are in activities of your choice?	
Thinking about the meals you eat, which of the following statements best describes your present situation?	
How well do you think your home is designed to meet your needs?	
Thinking about your home, which of the following statements best describes your present situation?	

Table 22: Replacements for question 14

Question	Responses
Do your care workers or personal assistants (PAs) do things in their way rather than yours?	
Are your care workers or personal assistants (PAs) careless, e.g. they put things in the wrong place, are wasteful, etc?	
Do you feel you are treated with dignity and respect by your care workers or personal assistants (PAs)?	
Would you describe your relationship with your care workers or personal assistants (PAs) as…?	
Do you think your care workers or personal assistants (PAs) are professional and do a good job?	
How well do your care workers or personal assistants (PAs) understand how your condition affects you?	

We have chosen these questions as they seem to have face validity. However, they have been developed in the context of a questionnaire for adults aged 18-64 whose use of services is different. They would need to be tested for their validity with this age group and people receiving 'traditional' home care services. Although these questions have been cognitively tested (Malley et al., 2006), they have yet to be tested as a survey on the population for which they were developed. The results from this survey are due to be reported at the end of 2007 and by this time we should have a clearer idea about how well they work.