Office for National Statistics

Living Costs and Food Survey

National Statistics Quality Review: Series (2)

April 2016

Report No.: **3**

Editors: **Jeff Ralph and Silvia Manclossi**

Office for National Statistics

NSQR Series: 2

Report: 3

About us

The Office for National Statistics

The Office for National Statistics (ONS) is the executive office of the UK Statistics Authority, a non-ministerial department which reports directly to Parliament. ONS is the UK government's single largest statistical producer. It compiles information about the UK's society and economy and provides the evidence-base for policy and decision-making, the allocation of resources, and public accountability. The Directors General of ONS report directly to the National Statistician, who is the Authority's Chief Executive and the Head of the Government Statistical Service.

The Government Statistical Service

The Government Statistical Service (GSS) is a network of professional statisticians and their staff operating both within ONS and across more than 30 other government departments and agencies.

Contacts

This publication

For information about the content of this publication, please contact:

Dr Jeff Ralph
Email: jeff.ralph@ons.gsi.gov.uk

National Statistics Quality Review Programme

For information about the National Statistics Quality Review Programme, please contact ONS Quality Centre.

Email: ons.quality.centre@ons.gov.uk

Other customer enquiries

ONS Customer Contact Centre
Tel: 0845 601 3034
International: +44 (0)845 601 3034
Minicom: 01633 815044
Email: info@statistics.gsi.gov.uk
Fax: 01633 652747
Post: Room D.265, Government Buildings, Cardiff Road, Newport, South Wales NP10 8XG www.ons.gov.uk

Media enquiries

Tel: 0845 604 1858

Email: press.office@ons.gsi.gov.uk

Copyright and reproduction

Any enquiries regarding this publication should be sent to: info@statistics.gsi.gov.uk

This publication is available for download at: www.ons.gov.uk

National Statistics Quality Reviews

The Office for National Statistics (ONS) is committed to assuring the quality of its statistics. ONS has a number of established mechanisms for assuring quality, one of which is the programme of National Statistics Quality Reviews (NSQR).

The previous NSQR programme was put on hold in 2008 following the creation of the UK Statistics Authority and the launch of the new assessment process for Official Statistics. In December 2012, after the first round of Statistics Authority assessments was completed, the National Statistician launched a new series of rolling, in-depth methodological reviews.

This review of the Living Costs and Food (LCF) Survey is the third review in this second series of NSQRs.

Acknowledgments

This report has been produced by ONS in collaboration with Thomas Crossley, Professor of Economics at the University of Essex, and Paul Smith, Associate Professor of Official Statistics at the University of Southampton. Their contribution is gratefully acknowledged.

Thanks are also extended to ONS colleagues who contributed to this review via the NSQR Review Board and Review Team. Further details of the people involved can be found in Appendix B.

Table of contents

Executive summary

(i) Introduction

The third report of the Office for National Statistics (ONS) re-established series of National Statistics Quality Reviews (NSQRs) has examined the Living Costs and Food (LCF) Survey, which is an annual survey designed primarily to measure household expenditure on goods and services and gather information about the income of household members.

The main reason, historically, for carrying out surveys of the expenditure of households was to provide weighting information for consumer price indices. This use continues today; the LCF provides weights for the Consumer Prices Index (CPI), the Retail Prices Index (RPI) and for Purchasing Power Parities (PPPs) for international price comparison. LCF information on income and expenditure is used in the analysis of how taxes and benefits affect household income and LCF spending data is used to compile national estimates of household final consumption expenditure which feed into the National Accounts. In addition to the uses within ONS, LCF data is used by the Department for Environment, Food and Rural Affairs (DEFRA), the Department for Transport (DfT), the Department of Energy and Climate Change (DECC), Her Majesty's Revenue and Customs (HMRC) and Eurostat.

The objectives of the LCF NSQR have been to:

- assess the current methods against 4 European Statistical System (ESS) quality dimensions (relevance, accuracy, comparability and timeliness) in order to define fitness for purpose
- highlight areas that:
    - have not kept up to date with international best practice
    - require some improvement
    - could impact on ONS's reputation
- carry out an international comparison in relevant areas
- look at other ways that the data could be collected
- make recommendations for work that could be carried out to the survey to make immediate improvements
- make recommendations for work that could be carried out to the survey in the future if funding were to be available

LCF is an influential survey and it is important that it meets the needs of users. For this reason, this quality review has assessed the degree to which user requirements are met and whether the data produced is of sufficient quality for the purposes to which it is put. It has reviewed various methodological aspects of the survey, including the data collection instruments, the collection process, the sample design, weighting, imputation and coding.

UK practice has been compared to the characteristics and methods of similar surveys carried out in other countries. The way the data is collected has also been compared to the approach taken by other countries and a market research company; in particular, the review has looked at increasing the use of technology. Alternative sources of expenditure data have been explored and their

potential for future use assessed. It has also taken into consideration the concerns of other, recent reviews.

The review acknowledges that, in the future, methods for collecting this data may change but concludes that, at least in the medium term, there will be the need for a survey of this kind. The recommendations, therefore, cover both how the current survey could be improved and possible alternative sources to be explored.

This NSQR notes the recommendations from 2 other reviews in this area – the Johnson Review and the Bean Review. These had a different purpose and emphasis to this review but the findings are broadly similar.

Because the scope of the NSQR has to be limited it has, in some cases, uncovered questions which it has not had the time to answer; therefore, recommendations have been made where further methodological analysis would be beneficial. It should also be noted that the following areas were out of scope:

- The position of LCF relative to other surveys (for example, it was not considered whether LCF could be merged with financial surveys).
- Detailed data on food expenditure, including weights and measures, collected and passed to the DEFRA who use it to produce analysis on food consumption and nutritional uptake of the population.

This summary starts with an overall assessment of how fit-for-purpose the survey is, it then highlights the main points identified as part of the review and then lists the 30 recommendations for improvement.

(ii) Overall judgement

The assessment of this review is that the LCF needs improvements to make it entirely fit for purpose for expenditure estimates; however, it is still the best source of expenditure information for most expenditure categories. It is also an important source of income data, but sample size limitations mean it needs calibration controls to be applied in order to manage volatility. Work to investigate integration of LCF income data with EU-SILC should also continue to ensure it is fit for this purpose in the longer term.

There is scope for improvement in the collection instruments and to a lesser degree the collection process. The overall quality of statistical processing is high.

The declining achieved sample has implications for the precision of data supplied for key uses in Price Statistics and National Accounts. A significant sub-set of this data has high coefficients of variation and has to be supplemented by administrative and market research data. Under-reporting of certain expenditures is also accommodated by using administrative sources.

A major concern that the growing sampling error of expenditure weights might be resulting in unacceptably low precision of consumer price index numbers does not appear to be the case. Preliminary analysis has shown the effects are small.

The quality of expenditure information for specific categories of goods could be improved by increasing the LCF sample size as could the precision of data for National Accounts. However, an enormous increase in sample size would be needed to bring the precision of all outputs to acceptable levels; this is not practical. Additional detail on expenditures for Price Statistics should be sought from alternative data sources - initially market research data and, further on, supermarket scanner data.

It is highly likely that response will continue to decline and this will have an impact on the quality of LCF outputs. Up to now, ONS business areas have coped with the data provided by the LCF, the level of precision should not be allowed to decline further. Efforts to improve response should be made and, if these are not successful, the drawn sample should be increased to maintain the size of the achieved sample.

Developments in electronic data collection and the use of administrative and “big data” may have a significant impact in the longer term.

(iii) Main points identified

Meeting user needs

While in some areas user needs are met to a high degree, there are important areas of user need which are not being met fully by the survey. There is under-reporting of some components of expenditure and income; this has been identified through comparison with other sources and is not unique to the UK. International research has suggested diary fatigue as one cause. In addition, analysis for this review shows that categories which are under-reported include those where respondents find it hard to supply data. This review recommends work to improve the questionnaire and investigate whether a shorter diary period would be effective.

The LCF provides weighting information for consumer price indices. While data is provided for most of the required expenditure categories, additional sources are needed where expenditure categories are requested at fine levels of detail. Prices Division, who produce consumer price indices, already use market research data (and other sources) to supplement the LCF expenditure data. While the LCF is recognised as the best source of expenditure data, this review recommends a further look at the potential for making more use of market research data in conjunction with LCF expenditure estimates.

The LCF is also an important source of income data, but estimates of some income components from the LCF are not sufficiently precise on their own. Sample size limitations mean it needs calibration controls to be applied in order to manage volatility. An option for the future could be the integration of aspects of the LCF with the EU Statistics on Income and Living Conditions (EU-SILC) in order to obtain a larger, combined sample with a core set of questions. This work is in progress.

Regular meetings between the LCF team and users ensure that good relationships are maintained. The importance of the LCF and the uses to which its outputs are put means that issues are identified and, where possible, resolved quickly.

Statistical processes

The data collection process was reviewed by conducting focus groups comprising interviewers and field managers. Making contact with a sampled household, encouraging participation, managing the interview and diary completion and ensuring data is of good quality are highly challenging. While the process is effective and well managed, there are aspects where improvements can be made, including work allocation and the way quality assurance is carried out.

The main collection instrument has been developed to appropriate standards and undergoes continuous improvement, though this activity is constrained by limited resources. The survey team, questionnaire/diary design experts from the Data Collection Methodology (DCM) team and interviewers have identified aspects where it has fallen behind trends in consumption and an improved process for revising and updating the questionnaire has been proposed.

The review examined the diary and concluded it is outdated; this view was supported by the interviewer focus groups. A full review of the paper diary is required. While the diary could adopt a mixed mode approach, the length and complexity of the questionnaire would make an effective web version very difficult to achieve under the current design.

The checking of collected data is comprehensive and great effort is put into ensuring it is complete and accurate. The coding of expenditure from the entries in the diary to the classification of individual consumption by purpose (COICOP) categories is mostly manual, but is also highly skilled and accurate. Discussions with a market research company have identified the potential for some automated coding, though this would be most effective if ONS had access to supermarket classification data. Current research at ONS on the use of big data for price statistics is investigating the use of machine learning techniques for auto-classification of transactions. The combination of supermarket look-up data and machine learning tools would not remove the need for skilled manual input as corrections would always be needed, but could lead to a higher rate of processing.

A brief review of the overall approach to imputation concluded that it is of high quality though a detailed review has not been carried out for some time. Non-response bias has been studied for LCF and adjustments are made through the weighing system. No new analysis was carried out for this review. A follow-up study of both responding and non-responding households would be a valuable source of further information, though it would require significant additional resources.

Sample size, response rates and accuracy

An important question is whether the current sample size is sufficient to meet the needs of users; that is, does it supply the data needed at an acceptable level of accuracy? For the majority of expenditure categories the answer is yes; however, the expenditure estimates for detailed categories show a significant proportion with high Coefficients of Variation (CVs). About 15% of the expenditure estimates published in Table A1 associated with the Family Spending Release[1] have a CV that exceeds 20%. The expenditure categories required by National Accounts for Household

[1] www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/compendium/familyspending/2015/aboutthiseditionoffamilyspending2015

Final Monetary Consumption Expenditure have a larger proportion with high CVs, at about 40%. Both Prices Division and National Accounts are able to manage this position although it is not ideal. The more aggregated expenditure categories have small CVs and other sources, such as good administrative data and market research data, are used to supplement the LCF expenditures at the detailed level.

Estimates of some income components from the LCF are not sufficiently precise on their own. There have been instances of volatility in estimates that affect the usability of the data for some purposes and additional calibration weighting has been needed. Further work on calibration controls is required. There is also work in progress to investigate integrating the LCF and the EU-SILC; this would lead to a larger, combined sample with a core set of questions.

A particular concern is the effect of the sampling error of expenditure estimates on consumer price indices. With the decline of the achieved sample, there is a concern that the falling precision of expenditure weights might be leading to inaccurate price index numbers. This topic has been studied for this review at the “all-items, division and group” levels and preliminary results indicate that the effect of expenditure sampling error is small.

There is a further question relevant to the accuracy of expenditure weights when calculating consumer price index numbers, that is the emerging requirement for exploring the inflation experiences of household types. Is the current sample size sufficient to produce expenditure weights of sufficient accuracy for this requirement? It has not been possible to investigate this question for the review; however, it is a recommendation for follow-up work.

International comparisons

The UK LCF is comparable to other EU countries in terms of sample size and response rates. However, Canada, Australia and the US use larger samples and achieve better response rates. In the case of Australia, the survey is compulsory.

UK practice differs from that of other countries in a number of ways. Some countries use a web questionnaire only, though this achieves very low response rates; others use mixed web/Computer Assisted Personal Interviewing (CAPI). The diary is also implemented differently in other countries – some use a web collection instrument. ONS is currently undertaking work to move all surveys on-line. In its current form, this would be challenging for the LCF and it is likely that there will have to be a complete re-design if this is to be achieved.

The survey team is aware of international practice and development programmes in other National Statistics Institutes (NSIs) and needs to continue to be engaged with developments elsewhere.

Alternative sources

The review has considered the potential for using alternative sources including market research data and supermarket scanner data[2]. Market research data on household purchases is already used in producing consumer price indices in ONS. Larger market research datasets have been

[2] Electronic records of purchases and prices created at the tills in supermarkets.

used for index number research at ONS for a number of years, and greater production use of this type of data for estimating expenditure weights for very specific categories of commodities should be investigated. This source has not been specifically designed for statistical purposes like the LCF; however, it could be combined effectively with the LCF expenditure data by constraining detailed market research expenditures to LCF totals.

Supermarket scanner data has been used by some countries to produce components of their price indices for some years. Relative progress of NSIs is strongly dependent on how easily obtainable the data is and only a small amount has been made available in the UK so far. ONS is continuing efforts to acquire it. Its availability would allow a much richer, detailed set of expenditure weights to be obtained, though as for the market research data, it is likely to supplement the LCF expenditure data rather than replace it.

ONS has been scraping price data from retail websites and creating experimental price indices from it. This data does not contain quantities sold, so LCF expenditure weights are combined with scraped prices. The use of scraped data and other forms of “big data” have the potential to impact on the data collection as these methods develop.

(iv) Recommendations

A number of potential changes were considered by the review along with their order of priority. The Review Team identified 30 recommendations, which are listed here in brief and have been grouped into the functional categories below. A more thorough discussion of each recommendation is included in the appropriate chapters.

Sample design and estimation

Recommendation 1 (Medium priority): Explore whether a change of design could deliver improved estimates for the same cost.

Recommendation 2 (Medium priority): Investigate the effect on the precision of levels and change of introducing oversampling areas with high proportions of high income households.

Recommendation 3 (High priority): Carry out further work on the use of LFS and other controls. Include research on the benefits of using the composite calibration framework to improve the quality of LCF estimates on income estimates.

Recommendation 4 (Medium priority): Continue the preliminary analysis of the effect of sampling error in LCF expenditure estimates on price index numbers and publish the results. Extend the analysis to sub-populations.

Data collection

Recommendation 5 (Medium priority): Carry out a follow-up survey of households included in the LCF sample, both responding and non-responding, to provide further information on non-response bias in the LCF.

Recommendation 6 (High priority): Consider whether a change in the incentive structure can be used to address low survey participation rates in a cost effective manner.

Recommendation 7 (Medium priority): Review the methodology for allocating field work to geographic areas to ensure a more even distribution within interviewer areas across the year without violating the survey's design.

Recommendation 8 (Medium priority): Review the effectiveness of the new interviewer training materials by carrying out interviewer and interviewer manager focus groups towards the start of the 2016/17 financial year.

Recommendation 9 (Medium priority): Review the quality assurance processes carried out by interviewers by considering whether any of their tasks could be carried out more efficiently by office based staff.

Recommendation 10 (High priority): Carry out further work to investigate the trade-offs between maintaining a 2 week diary period and adopting a shorter diary period.

Recommendation 11 (Medium priority): Carry out further work to review the level of detail needed for combined payment breakdowns (where bills may not provide the level of detail requested).

Recommendation 12 (High priority): Carry out further analysis on a larger scale to explore the extent of interview length on interview fatigue (including more questionnaire sections asked at different points in the interview and potentially a further survey year to increase robustness of results).

Recommendation 13 (High priority): Carry out further analysis to understand other potential causes of under-reporting in the LCF survey.

Recommendation 14 (High priority): Allocate additional resources to the LCF Research Team (including securing funding for DCM resource) to develop and implement a more robust questionnaire and testing process and ensure that the questionnaire design keeps pace with ongoing changes in consumer spending/behaviour.

Recommendation 15 (High priority): Review the adult, child and pocket diaries and update them once every 5 years (to ensure the content remains up to date). The first review (in paper form) should be implemented by April 2017.

Recommendation 16 (High priority): Carry out in-depth interviews as a follow-up survey with LCF respondents to understand how respondents go about answering questions of concern, and to assess the accuracy and completeness of diary recording.

Recommendation 17 (High priority): Review the collection of income data within ONS Social Surveys to reduce discrepancies with other sources of income data.

Recommendation 18 (Low priority): Carry out more work to understand reporting of holiday expenditure.

Recommendation 19 (High priority): Consider web diary collection for the LCF alongside other ONS web collection initiatives.

Processing

Recommendation 20 (Low priority): Submit the LCF edit strategy and the general principles underlying the imputation strategy into the development of the ONS Business Process Model for Social Surveys.

Recommendation 21 (Low priority): Review and document how editing and imputation are carried out, paying particular attention to the efficacy of the diary imputation.

Recommendation 22 (Medium priority): Discuss potential benefits from and access to supermarket product information with market research companies and supermarkets.

Recommendation 23 (Medium priority): Explore the possibility of semi-automated coding of purchase information from scanned supermarket receipts.

International comparisons

Recommendation 24 (Medium priority): Liaise with other countries and organisations undertaking web data collection, receipt scanning, and automatic coding to understand more about the data quality obtained. If the quality is satisfactory, undertake experiments to evaluate these approaches in the LCF.

Recommendation 25 (Medium priority): Engage with ONS programmes to enhance the use of administrative data on income, though it should be acknowledged that record-level linkage is likely to be a long-term aspiration.

Recommendation 26 (Medium priority): Maintain links and promote information sharing with other NSIs and organisations (particularly the Understanding Society team at Essex[3]) with respect to the development of expenditure surveys. In particular, progress by the US and Norwegian statistical offices should be monitored.

Future developments

Recommendation 27 (Medium priority): Explore the differences between market research data and LCF expenditure data, comparing expenditures at multiple levels of disaggregation. Evaluate the accuracy of panel data constrained to higher level LCF expenditure values.

Recommendation 28 (Medium priority): Investigate the practicality of a market research company running a specific panel to provide ONS with additional expenditure information to supplement LCF data.

Recommendation 29 (High priority): Continue working to secure access to store scanner data.

Recommendation 30 (High priority): Review international use of store scanner data and plan a programme of research to determine the best future use of this data.

ONS will respond to the recommendations listed above with a plan for taking them forward. Stakeholders and users will be informed about the implementation plan and progress against recommendations will be monitored by the ONS's Quality Centre.

[3] More information about Understanding Society can be found at: www.iser.essex.ac.uk/understanding-society

1. About this National Statistics Quality Review

1.1 Background

In December 2012, the Office for National Statistics (ONS) re-established its programme of National Statistics Quality Reviews (NSQRs). This NSQR is the third in the re-established series and assesses the Living Costs and Food (LCF) Survey, focusing on certain fundamental methodological processes used to design and deliver the LCF.

The remit and scope of the review are defined by the principles set for the NSQR programme, which aims to adopt a risk based, proportionate, tailored and efficient approach. A proportionate view has been taken, balancing potential benefits of investment in a methodological review and the associated cost. A tailored approach has ensured that outputs have undergone reviews relevant to their complexity and profile.

NSQRs are regulated by the ONS's Quality Centre, which has oversight of quality in official statistics across the Government Statistical Service (GSS), and are overseen by a Review Board which includes senior managers from the relevant statistical output areas, ONS methodologists and independent external expertise. The role of the Review Board is to ensure that the quality and methodology underpinning ONS outputs have kept pace with changing methods and users' needs, and that the review has been sufficiently inquiring and challenging in its approach and assessment.

The Review Board is supported by a Review Team which is responsible for collating and assessing much of the information gathered on the methodologies used. They also write the review report and develop any recommendations.

The external expert for this NSQR is Professor Thomas Crossley; he is Professor of Economics at the University of Essex, a Research Fellow at the Institute of Fiscal Studies and an associate editor of the Journal of Applied Econometrics. The review has also been assisted by Associate Professor Paul Smith from the University of Southampton.

A full list of Review Board and Review Team members is given in Appendix B.

1.2 Review objectives

The objectives of the LCF NSQR are to:

- assess the current methods against four European Statistical System (ESS) quality dimensions (relevance, accuracy, comparability and timeliness) in order to define fitness for purpose
- highlight areas that:
    - have not kept up to date with international best practice
    - require some improvement
    - could impact on ONS's reputation
- carry out an international comparison in relevant areas
- make recommendations for work that could be carried out to the survey to make immediate improvements
- make recommendations for work that could be carried out to the survey in the future if funding is available

1.3 Review scope

The scope of the review includes assessments of the following topics:

- the degree to which the LCF meets user requirements
- the design of the survey and possible alternative designs
- the issue of under-reporting
- the mode of data collection and the potential for alternatives
- the availability and potential use of alternative data sources
- statistical processing – in particular imputation and coding
- a comparison with international practice

The review is based on the evaluation of the objectives principally carried out by internal ONS experts, with quality assurance provided by independent experts.

1.4 Out of scope

The project did not provide a review of all aspects of the methodology used; just those areas which relate to sample design and estimation, data collection and processing. Moreover, the following areas were out of scope:

- the position of LCF relative to other surveys (for example, it was not considered whether LCF could be merged with financial surveys)
- detailed data on food expenditure, including weights and measures, collected and passed to the Department for Environment, Food and Rural Affairs (DEFRA) who use it to produce analysis on food consumption and nutritional uptake of the population

1.5 Methodology

The review work was carried out throughout 2015 by a variety of ONS staff with appropriate survey and methodological expertise.

This review examined a number of aspects of the LCF and its uses. It assessed the degree to which user requirements are met and whether the data produced is of sufficient quality for the purposes to which it is put. It reviewed various methodological aspects of the survey, including the data collection instruments, the collection process, the sample design, weighting, imputation and coding. UK practice was compared with the characteristics and methods of similar surveys carried out in other countries. The way the data is collected was also compared to the approach taken by other countries and a market research company; in particular, the review looked at the use of technology. Alternative sources of expenditure data were explored and their potential for the future assessed.

In some cases, it was possible to carry out new work. For example, experts from the Data Collection Methodology (DCM) team provided observations about the LCF data collection design and undertook an expert review of a prioritised list of expenditure topics; the data collection process was reviewed by conducting focus groups comprising interviewers and field managers; analysis was carried out to further investigate issues in relation to under-reporting expenditure and income, sampling and the propagation of sampling error into consumer price index numbers. The report contains the results of this work.

1.6 Review Team

The Review Team for this NSQR comprised:

- **Dr Jeff Ralph** (lead reviewer) – Jeff is the current head of Analysis and the Methodology Advisory Service within the Methodology Group at ONS. Jeff joined ONS in 2004 and has spent time managing a number of statistical projects and leading the Index Numbers team in Methodology. He sits on the one of the Advisory Panels on Consumer Prices.

- **Dr Giles Horsfield** – Giles was head of the LCF research team from 2010 until the end of 2015. He joined ONS in 1999 and worked on population statistics, including population projections and migration statistics. He moved to ONS Social Surveys in 2007. He has recently taken up a new post, heading the International Passenger Survey (IPS) research team.

- **Joanna Bulman** – Joanna manages day to day survey processes for the LCF, including development of the questionnaire and delivery of outputs. Joanna joined ONS in 1999 and has worked on a number of projects within Business Surveys as well as a range of Social Surveys.

- **Professor Thomas Crossley** (external reviewer) – Thomas is Professor of Economics at the University of Essex, a Research Fellow at the Institute for Fiscal Studies, and an elected member of the Conference on Research in Income and Wealth (CRIW). His research focuses on both consumer behaviour and survey methods, particularly for measuring household spending and finances. He recently co-edited the CRIW volume "Improving the Measurement of Consumer Expenditures" for the University of Chicago Press.

- **Professor Paul Smith** (external reviewer) - Paul is Associate Professor in Official Statistics at the University of Southampton, where he has worked since 2014. Before that, he gained extensive experience of survey methodology through working in Methodology in ONS for many years, leading a range of teams including sample design and estimation, time series, index numbers and the Methodology Advisory Service. He maintains a strong interest in UK official statistics.

2. Introduction

2.1 Introduction to the LCF

The LCF is a voluntary sample survey of private households. The basic unit of the survey is the household. A household comprises one person living alone or a group of people (not necessarily related) living at the same address who share cooking facilities and share a living room, sitting room or dining area.

Each individual aged 16 and over in the selected household is asked to keep diary records of daily expenditure for two weeks. Information about regular expenditure, such as rent and mortgage payments, is obtained from a household interview along with retrospective information on certain large, infrequent expenditures such as those on vehicles. Children aged 7 to 15 are asked to keep a simplified version of the diary.

Detailed questions are asked about the income of each adult member of the household. In addition, personal information such as age, sex and marital status is recorded for each household member. A copy of the LCF questionnaire is available from the UK Data Service[4].

The survey is continuous and interviews are spread evenly over the year to cover seasonal effects. The questionnaire content is reviewed thoroughly to ensure that it is up-to-date and captures information efficiently. Some changes reflect new forms of expenditure; for example, combined utility packages or new sources of income. Others are the result of new requirements by the survey's users; for example, the addition of questions about mobile phone ownership.

2.2 A brief history of the LCF

A household expenditure survey has been conducted each year in the UK since 1957. From 1957 to March 2001, the Family Expenditure Survey (FES) and National Food Survey (NFS) provided information on household expenditure patterns and food consumption. In April 2001, these surveys were combined to form the Expenditure and Food Survey (EFS).

In 2008, selected Government household surveys, on which ONS leads, were combined into one Integrated Household Survey (IHS). In anticipation of this, the EFS moved to a calendar-year basis in January 2006. The EFS questionnaire became known as the Living Costs and Food (LCF) module of the IHS in 2008, to accommodate the insertion of a core set of IHS questions[5]. In Quarter 2 (April to June) of 2014, the LCF came out of the IHS.

[4] discover.ukdataservice.ac.uk/?q=Living+costs+and+food+survey
[5] www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/sexuality/methodologies/integratedhouseholdsurvey

2.3 The design of the LCF

The LCF sample for Great Britain is a multi-stage, stratified random sample with clustering. It is drawn from the Small Users file of the Postcode Address File (PAF) which is the Post Office's list of addresses. All Scottish offshore islands and the Isles of Scilly are excluded from the sample because of excessive interview travel costs but they are included in the overall estimates (estimated using information from the collected sample).

Postal sectors are the primary sample unit; 638 postal sectors are randomly selected after being arranged in strata defined by regions (sub-divided into metropolitan and non-metropolitan areas) and two 2001 Census variables: socio-economic group of the head of household and ownership of cars. These census variables were new stratifiers originally introduced for the 1996/97 survey, and updated following the results of the 2001 Census and, subsequently, the 2011 Census. The Northern Ireland sample is drawn as a random sample of addresses from the Land and Property Services Agency list[6].

2.4 Uses of the LCF

The main reason, historically, for carrying out surveys of the expenditure of households was to provide weighting information for consumer price indices[7]. This use continues today; the LCF provides weights for the Consumer Prices Index (CPI), the Retail Prices Index (RPI) and for Purchasing Power Parities (PPPs) for international price comparison.

LCF information on income and expenditure is used in the analysis of how taxes and benefits affect household income and LCF spending data is used to compile national estimates of household final consumption expenditure which feed into the National Accounts. In addition to the uses within ONS, LCF data is used by DEFRA, the Department for Transport (DfT), the Department of Energy and Climate Change (DECC), Her Majesty's Revenue and Customs (HMRC) and Eurostat[8].

Clearly, LCF is an influential survey and it is important that it meets the needs of users. As with many statistical outputs, user needs are not static. For example, in 2014, ONS produced a paper on inflation experiences of household types[9] which used LCF data and the Johnson Review of Consumer Price Statistics recommends ONS produces this as an annual analytical output[10]. This

[6] Additional information about the design of the LCF can be found at:www.ons.gov.uk/ons/guide-method/method-quality/specific/social-and-welfare-methodology/living-costs-and-food-survey/index.html www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey

[7] The first official expenditure survey was carried out in 1904 and provided data for what was called a cost of living measure; it was very limited in scope by modern standards.

[8] A more comprehensive list is contained in Living Costs and Food Survey Technical Report 2013, which is available at: www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey#technical-report

[9] www.ons.gov.uk/ons/rel/elmr/variation-in-the-inflation-experience-of-uk-households/2003-2014/index.html

[10] www.statisticsauthority.gov.uk/reports-and-correspondence/reviews/uk-consumer-price-statistics-a-review

raises a question: is the LCF capable of supporting this new use? It has not been possible to investigate this for the review; however, it is a recommendation for follow-up work.

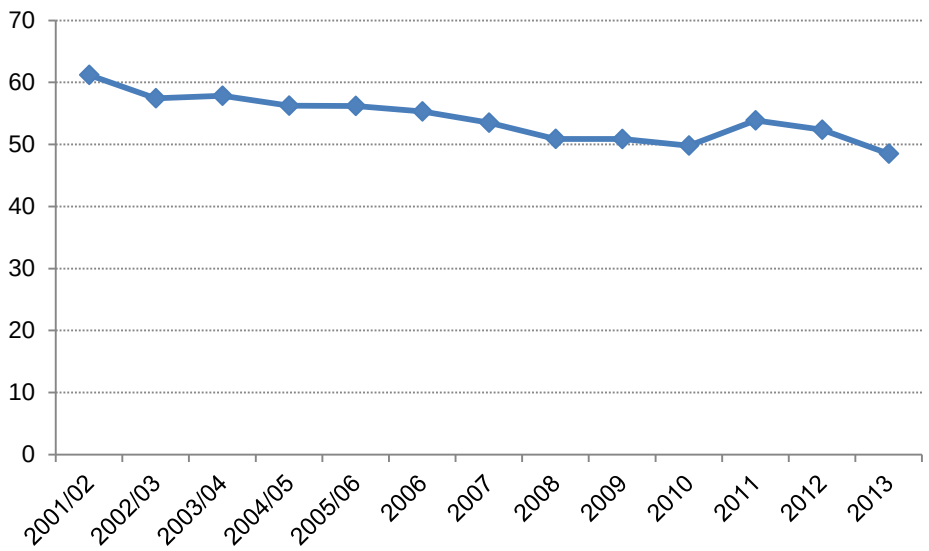
2.5 Response rates in context

A main driver for instigating this review was the concern over the effects of a long-term decline in the response rate and, therefore, the size of the achieved sample for the LCF. The response rate is a commonly used measure of quality in social surveys. Lower response reduces the number of cases achieved and, hence, the precision of estimates. Response may also affect the representativeness of the achieved sample and introduce bias into the estimates. This occurs if non-respondents are more likely than respondents to share certain characteristics.

There are 2 main types of non-response/non-contact: sampled individuals or households may not be available when contacted by an interviewer, and refusal – sampled individuals contacted may refuse to take part. In practice, multiple contact attempts are made rather than just one, and often contact is made only after several attempts. Efforts are also made to contact refusals for another attempt, particularly where their refusal is circumstantial; for example if they report being too busy to respond at that moment.

LCF response rates have been declining over the last 10 years; response rates fell from 62% in 2001 to 48% in 2013.

Figure 2.5.1: UK Response Rate for the LCF from 2001 to 2013



A similar pattern has been seen in other ONS household surveys. Response rates for the Labour Force Survey (excluding imputed households) have fallen faster than for the LCF, from 61% in 2006 to below 50% in 2013 and 2014; it should be noted that this is the response rate for all waves of the survey combined. In general, LCF response rates are lower than for other ONS household surveys. The requirement for respondents to complete a 2 week expenditure diary is a factor in this disparity.

Eurostat's draft technical report of the 2010 Household Budget Survey (HBS)[11] shows that the LCF's response rate (51% in 2010) was close to the median (53%) for participating countries. This excludes the countries using quota sampling (Germany and, currently, Czech Republic). Response rates ranged from 5.6% (Belgium) to 88% (Romania). Comparing response to National Statistical Institutes (NSIs) outside of Europe, the current US design achieves a response rate of around 70%; the Australian expenditure survey achieves a response rate of 72%; and the Canadian survey a response rate of 67%. The Australian survey is compulsory. This brief comparison of response rates achieved in other NSIs confirms that the challenge of addressing low response rates is not unique to the UK.

2.6 Relevance

This section of the review considers whether user requirements are being met and identifies any aspects of the survey outputs where improvements are needed. The Review Board decided that, with limitations on the resources available for this review, analysis of user requirements should be restricted to the core ONS LCF customers. Other reviews have covered a wider range of users, including the UK Statistics Authority Monitoring and Assessment Review[12] and internal reviews.

The 3 business areas examined were:

- Prices: LCF data on spending patterns are used to inform the content and weighting of the basket of goods used to produce the CPI, RPI and PPPs; the latter is used for international price comparisons
- National Accounts: LCF data on spending are an important source used in compiling national estimates of household final consumption expenditure, which feed into the National Accounts and estimates of GDP
- Household Income and Expenditure (HIE): analysis of how taxes and benefits affect household income

There is frequent communication between the LCF team and Prices, National Accounts and HIE. This includes informal day-to-day discussion, consultation over potential questionnaire changes, and formal input via the LCF Steering Group. In addition, HIE have a role as an "intelligent customer" for the LCF, and provide quality assurance, particularly for income estimates. Review meetings are held twice a year with Prices and as required with National Accounts. The LCF team is collaborating closely with both Prices and National Accounts over plans to update the survey's coding frame, as overseen by the UN and Eurostat; the LCF uses the COICOP coding frame. The meetings held as part of this review, therefore, represented an extra opportunity for these customers to present additional areas for improvement.
The 3 business areas were invited to identify any requirements that were not currently being met and any aspects of the service to them that could be improved. The overall position is that the user requirements are met to a high degree. The importance of the outputs dependent on LCF data and

[11] The HBS 2010 quality report is available at: ec.europa.eu/eurostat/documents/54431/1966394/LC142-15EN_HBS_2010_Quality_Report_ver2+July+2015.pdf/fc3c8aca-c456-49ed-85e4-757d4342015f
[12] www.statisticsauthority.gov.uk/publication/statistics-from-the-living-costs-and-food-survey

the regular meetings with users ensures that this is the case. However, there were areas of concern.

(i) Under-reporting

Possible under-recording in the survey is a major area of concern. Discussion helped identify high-priority goods and services for further consideration; that is, those that are thought to be subject to under-recording and that could adversely affect the survey's utility to customers.

This was considered when planning the review work described in Chapter 4 on Data Collection, and informed decisions about which goods and services to select as examples for investigation. The topics selected for further investigation as part of this review were, by necessity, not exhaustive, and those not investigated are listed for potential future work. Examples of expenditure categories investigated are:

- childcare payments
- mobile phone payments
- petrol and diesel

Discussion with HIE revealed the following areas of concern with income measures:

- benefits, focusing on means tested benefit; for example income support, pension credit and tax credits
- reporting of earnings and hours
- response from high earners

Full details of the expenditure categories investigated are provided in Chapter 4, along with an assessment of the extent of under-reporting of expenditure and income.

(ii) Level of detail available

The level of detail available in the expenditure data is an important determinant of its value. Prices reported categories of expenditure where they need to use other sources to break down the data delivered by the LCF team; a lower level of detail would be beneficial. These include:

- men's and women's outer garments, which could be broken down by garment type
- takeaway meals, which could be broken down by meal type

These tend to be instances where there are small numbers of recording households in the survey, and the level of detail offered is limited by the survey's sample size. This is discussed in Chapter 3.

2.7 Precision

As well as considering whether the information supplied by the LCF meets the needs of users, there is also a question of whether the precision of the data is appropriate for the uses to which the data is put.

The LCF team produces standard errors as part of the survey outputs. Table A1, produced as part of the Family Spending release[13], provides a breakdown of household expenditure by category for over 700 categories at 4 levels of aggregation. It also contains standard errors for each category.

Chapter 3 looks at this question. It considers the accuracy of data for household income and expenditure, National Accounts and consumer price indices. A particular concern is the effect of sampling error of expenditure information on price index numbers. The investigation of this requires a complex calculation which has not been done before. Some previous analysis has examined the sensitivity of price index numbers to weights in an approximate way; this review includes a more detailed calculation which is described in Chapter 3.

2.8 Timeliness

The LCF meets the requirements for timely delivery. For consumer price indices, the weights are relevant to a time period between 1 and 2 years before the price reference period, which is the January of each year. The ideal position would be for the weights to be relevant to the price reference period; this would enable a true Laspeyres Index to be used as the index formula rather than the Lowe formula which is currently used. This is not a serious problem; it is recognised that the process of collecting expenditure information takes time and the use of the Lowe index is permitted under international regulation. ONS research has examined the difference between a Lowe index with a weight reference period between 1 and 2 years in the past and an approximate Laspeyres formula. The differences were found to be relatively small[14].

2.9 Frequency/ sample size of survey

While the primary aim of this review is to identify areas of potential improvement to quality, it is also important to consider potential efficiencies and cost savings. An obvious way of reducing costs is to reduce the frequency of the survey's deliveries. Some EU countries conduct expenditure surveys only every 5 years. Reducing the frequency with which the LCF is conducted in the field would be difficult to implement in terms of managing interviewer commitments. It would be more feasible to run the survey continuously with a reduced sample.

A reduced sample might mean delivering estimates for National Accounts use less frequently (every 6 months rather than 3), and the annual sample size would be reduced, with inevitable reductions in quality. Household Expenditure in National Accounts is receptive to exploring the possibility of less frequent deliveries but would require further information from the LCF team in order to design new processes to model the necessary time series data. Only after a quality assessment of the modelled data against the current deliveries could a firm agreement to less frequent deliveries be accepted.

[13] www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/compendium/familyspending/2015/aboutthiseditionoffamilyspending2015

[14] Calculating a Laspeyres Version of the UK Consumer Prices, available at: www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/index.html

The impact of any reduction in sample would certainly have an appreciably negative impact on its use for price indices: CPI, RPI and PPP. The reduction observed in the achieved sample in the LCF in recent years (due to lower response and a 5% cut in sample in 2006) has meant increased volatility in the time series at lower-level, item breakdowns. Another problem is that there has been an increase in the incidence of cases with recorded zeroes for certain categories of expenditure. This highlights that a lower sample size can be seen to reduce the categories where any spending is observed, as well as the precision of estimates obtained. Absence of recorded expenditure at some lower levels has necessitated the increased use of commercial data to break down higher levels of expenditure, as recorded in the LCF. This is methodologically undesirable, and incurs costs of its own since commercial data must be purchased. A reduced sample may necessitate reducing the frequency with which prices weights are updated. It is considered best practice to update weights annually and any move away from this schedule would be a severely retrograde step. It would also run counter to Eurostat guidance.

2.10 Structure of this report

The report consists of 5 substantive chapters. Chapter 3 looks into the LCF survey design and precision of estimates, Chapter 4 investigates data collection instruments, Chapter 5 assesses processing issues, Chapter 6 is concerned with international comparisons and Chapter 7 explores possible future developments. Related reviews are described briefly in Appendix A and Appendix B identifies the members of the Review Team and the Review Board and other stakeholders who have been consulted for this review.

3. Sample design and precision of estimates

This chapter starts by describing the design of the LCF sample; it provides an overview of the current design and considers alternatives. It also looks at how the achieved sample size has changed over time and the effect on the precision of estimates.

The effect of sampling error is, of course, propagated into statistics derived from LCF outputs. One topic of particular concern is the impact on consumer price indices, where LCF provides the expenditure weights. Calculations have been made to investigate the size of the effect and the impact of further reductions in the achieved sample size.

The chapter concludes with a look at weighting and calibration.

3.1 Overview of the existing design

The LCF survey has a stratified 2-stage sample design, with the stratification by region representing the major stratum. Postcode sectors are the primary sampling units; within each major stratum, the postcode sectors are sorted by census factors (percentage of households without a car, percentage of households in the National Statistics Socio-economic Classification, NSSEC, 1-3, and percentage of pensioners).

638 postcode sectors are selected across Great Britain using systematic sampling with probability proportional to the size of their population. Within each postcode sector, 18 addresses are selected using systematic random sampling. All adults in a household are interviewed.

The selected postcode sectors are assigned at random to the months on the year, with roughly the same number of clusters assigned to each month. An independent sample is selected each year. This design leads to an equal probability sample for all households in Great Britain. The Northern Ireland (NI) sample is not clustered; a systematic random sample of households is selected. Further details can be found in the LCF Technical Reports[15].

Currently, a sample of 11,484 addresses is issued into the field; about 12% are usually found to be ineligible and about 50% of those eligible co-operate with interviewers, which resulted in an achieved sample of about 5,100 households in 2013/14.

Before 2007, 672 Primary Sampling Units (PSUs) were drawn and 18 addresses were selected at random from each PSU to form the monthly interviewer quota. A decision was made late in 2006 to reduce the LCF sample by 5%; it was decided to remove 34 PSUs (that is, 5% of the originally selected PSUs) from the sample rather than reduce the size of each quota.

[15] See, for example, Living Costs and Food Survey Technical Report 2013, which is available at: www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey#technical-report

3.2 Response rates

In addition to the reduction in sample size since 2007, response rates have decreased by about 5 percentage points between 2007 and 2013 as highlighted in Table 3.2.1 below.

Table 3.2.1: Response rates in LCF/EFS from 2000/01 to 2013

Year	GB sample	GB eligible households	GB response Rate (%)	NI eligible households	NI response Rate (%)	UK response Rate (%)
2000/01	11,424	10,406	58.8	990	52.7	58.2
2001/02	12,096	10,945	62.2	1,030	50.6	61.2
2002/03	12,096	11,019	57.6	1,039	56.3	57.4
2003/04	12,096	11,120	57.8	1,064	57.9	57.8
2004/05	12,095	11,053	56.7	1,032	51.6	56.3
2005/06	12,097	11,014	56.8	1,057	49.9	56.2
2006	12,096	10,929	55.4	1,081	54.2	55.3
2007	11,484	10,397	53.3	1,080	55.2	53.5
2008	11,484	10,419	50.6	1,065	53.9	50.9
2009	11,482	10,366	50.4	1,084	55.5	50.9
2010	11,484	10,314	49.6	250	58.8	49.8
2011	11,484	10,303	53.7	261	61.7	53.9
2012	11,484	10,386	52.2	300	57.0	52.4
2013	11,484	10,354	48.2	251	60.6	48.5

3.3 Precision of estimates

The combined effect of sample size reduction and decreasing response rates on the precision of expenditure estimates can be seen in Table 3.3.1; it shows that the impact is rather minor. It is lower than what would be expected from the fall in the achieved sample; it is not clear why this is the case - one explanation could be that there is less variation between returns over time.

Table 3.3.1: Coefficients of variation of total expenditure and some components from 2003 to 2013

Year	Achieved sample	Food & non-alcoholic drinks	Clothing and footwear	Total expenditure
2003	7,048	*0.8*	*2.1*	*1.0*
2004	6,798	*0.8*	*2.0*	*0.9*
2005	6,785	*0.9*	*1.9*	*1.2*
2006	6,650	*0.8*	*2.4*	*1.1*
2007	6,140	*0.8*	*2.1*	*1.0*
2008	5,850	*0.8*	*2.2*	*1.2*
2009	5,830	*0.8*	*2.4*	*1.1*
2010	5,260	*0.9*	*2.7*	*1.2*
2011	5,690	*0.9*	*2.4*	*1.0*
2012	5,600	*0.9*	*2.5*	*1.1*
2013	5,140	*1.1*	*2.6*	*1.2*

The precision of level estimates for the Effects of Taxes and Benefits (ETB) for all households is given in Table 3.3.2; the relative precision of these estimates, shown as Coefficients of Variation (CVs), are relatively small. A CV exceeding 20% indicates that an estimate should be treated with caution.

Table 3.3.2: Precision of ETB estimates for gross household income and disposable income – based on 2013/14 LCF data; the bounds are for a 95% confidence interval

	Gross income				Disposable income			
	Lower bound	Published estimate	Upper bound	CV	Lower bound	Published estimate	Upper bound	CV
Mean (£ per year)	38,156	39,200	40,244	1.4	31,029	31,786	32,543	1.2
Bottom quintile (mean, £ per year)	11,181	12,918	14,655	6.9	10,323	11,665	13,007	5.9
Top quintile (mean, £ per year)	79,140	83,370	87,600	2.6	60,768	63,763	66,758	2.4

The estimates of relative change, or growth, are not as accurate as the estimates for levels, as it can be seen from the wide confidence intervals shown in Table 3.3.3.

Table 3.3.3: Precision of ETB estimates of growth of gross household income and disposable income – based on 2012/13 and 2013/14 LCF data; the bounds are for a 95% confidence interval

	Gross income			Disposable income		
	Lower bound	Published estimate	Upper bound	Lower bound	Published estimate	Upper bound
All households	-1.1	2.9	6.8	-0.3	3.2	6.8
Bottom quintile	-17.6	0.9	19.4	-14.6	1.2	16.9
Top quintile	-8.1	-0.5	7.1	-7.4	-0.1	7.2

The width of the confidence intervals indicate that estimates of single year-on-year growth of gross household and disposable income quintile means cannot be estimated from the LCF with a high degree of precision.

3.4 Overlapping samples design

Given that household income between consecutive years is fairly well correlated, the precision of the estimates of change could be improved by adopting a longitudinal design; for example, overlapping samples with multiple waves. The Review Team attempted to estimate the improvements in precision that could be obtained under a multiple-wave design where the achieved sample is similar to the current size.

A 50% response rate in wave 1 and 20% attrition rate between successive waves have been assumed. Using these rates, the overlap would be 44% in a 2-wave design and 70% in a 5-wave design. Table 3.4.1 shows the overlap rates for rotating other designs and indicates that the sampling error of the estimate of change in mean household income in a 2-wave design would be 18.4% lower than that under the current design. The reduction in sampling error increases with the rate of overlap; it would be 31% in a 5-wave design.

It should be noted that in the calculations:

- the sampling errors of the estimates of mean household income in year 1 and year 2 were assumed to be equal to those computed in LCF data from 2012 and 2013

- the Pearson correlation coefficient of household income between 2 consecutive years was computed using 2011 and 2012 Statistics on Income and Living Conditions (SILC) data and was found to be equal to 0.76

Currently, a sample of 11,484 addresses is issued into the field and an achieved sample of about 5,100 households is obtained. To achieve the same sample under the 2-wave design, a sample of about 9,262 in a year would need to be issued (some for the first time and some for the wave 2 interviews). The sample to be issued decreases with the rate of overlap; it drops to 7,818 in a 5-wave design.

Table 3.4.1: Impact of overlapping designs in LCF on precision of estimate of change in mean household gross income

Design	Overlap rate	Total issued sample	Wave 1 issued sample	SE (estimate of change in mean HH[16] income) (in £)	Percentage reduction in SE of change
Current	0%	11,484	11,484	14.8	
2-wave	44%	9,262	6,432	12.1	18.4
3-wave	59%	8,498	4,742	11.0	25.7
4-wave	66%	8,129	3,920	10.5	29.4
5-wave	70%	7,918	3,444	10.2	31.6

Applying a 4-wave design will lead to narrower confidence intervals for the estimates of growth, as it can be seen by comparing Tables 3.3.3 and 3.4.2, but they are still rather wide.

Table 3.4.2: Expected confidence intervals of growth estimates of mean household income in a four-wave design

	Gross income			Disposable income		
	Lower bound	Published estimate	Upper bound	Lower bound	Published estimate	Upper bound
All households	0.1	2.9	5.6	0.8	3.2	5.7
Bottom quintile	-12.0	0.9	13.8	-9.9	1.2	12.2
Top quintile	-5.8	-0.5	4.8	-5.2	-0.1	5.0

The results of overlapping designs have some desirable effects and are worth exploring further. The implications for respondent burden would need to be included.

[16] HH stands for “Household”

Recommendation 1: Explore whether a change of design could deliver improved estimates for the same cost.

3.5 Statistical power to detect change

The current design has an 80% statistical power to detect a growth rate of 5 percentage points, whereas the 4-wave design has the same power to detect a growth rate of 3.5 percentage points. To detect lower growth rates with the same power, the sample size would need to be increased substantially. For example, to detect a growth of 2% with 80% power in a 4-wave design, a sample that is about 80% bigger than the current sample would be needed.

3.6 Oversampling of high income households

One way to increase the precision of estimates of both levels and change is to over-sample high income households, as it is currently done in the Wealth and Assets Survey (WAS). This is achieved, in practice, by oversampling areas containing higher proportions of high-income households. This would also improve representativeness. The evaluation of the likely benefits and the optimal oversampling strategy would require quite a significant amount of work. In WAS, HMRC classifies each address in the selected postcode sectors as being in the top 10% of the income distribution or not; the benefits would be greater if the classification was more detailed. The Review Team proposes that the details of the oversampling approach be investigated in further.

Recommendation 2: Investigate the effect on the precision of levels and change of introducing oversampling areas with high proportions of high income households.

3.7 Weighting

The design weights of the responding households are adjusted by applying factors derived from the Census Non-Response Link Study (CNRLS); these factors attempt to correct for differential non-response between sub-population groups. The factors currently used were obtained using 2001 data; factors based on 2011 data were computed and were found to have a modest impact, similarly to the 2001 factors. The application of the non-response factors leads mostly to a decrease in the estimates by a small amount.

The adjusted design weights of responding households are then calibrated so that the weighted sample matches the UK population by age (defined by groups), sex and region; the estimation method is referred to as “calibration”. It is applied using THE specialised software called “Generalised Estimation System” (GES). The calibration factors that are computed can, in principle, be negative, and sometimes quite large; to prevent this, additional constraints on the magnitude of the weights are specified. This increases the variance but the increase is likely to be very small.

An important component of household income is total household wages and salaries. The estimate of the mean household wages and salaries is subject to variation in 2 estimates:

- the number of people earning a wage/salary
- the mean earnings of people with a wage/salary

There are instances where the estimate of the number of people with a wage/salary based on LCF data can vary by a large amount over time, which makes estimates of year-on-year change volatile. Table 3.7.1 shows the quarterly estimates of the number of people with a wage from the LCF, alongside the corresponding LFS estimates of the number of employees. It can be seen that the LCF estimates are volatile; for example, in Quarter 2 of 2012 (April to June), the estimate was 25.5 million, and, for the following 2 quarters, it was 26.8 million and 24.8 million, respectively. By contrast, the Labour Force Survey (LFS) estimates are much less volatile.

Table 3.7.1: Comparing LCF estimates of people with a wage and LFS employee estimates (all figures are in thousands)

Year	Quarter	LCF estimate of number of people with a wage	LFS employee estimates
2012	April to June	25,508	25,124
	July to September	26,766	25,340
	October to December	24,834	25,443
2013	January to March	25,529	25,303
	April to June	25,797	25,397
	July to September	26,747	25,661
	October to December	26,457	25,693
2014	January to March	26,545	25,606

The impact of the volatility of the number of people with a wage in the sample on estimates of growth in mean household wage and income, for example, can be reduced by introducing into the calibration stage additional controls on the number of people in different employment categories from the LFS.

Table 3.7.2 shows the impact of the LFS controls on estimates of growth in mean household wages/salaries between 2012/13 and 2013/14. It can be seen that it had an impact in the last 2 quarters but not in Quarter 3 (July to September); the large growth between Quarter 3 in 2013 and Quarter 3 in 2012 is due to a large growth in the mean earnings of people with a wage – growth rates of such magnitude were seen in the past, as highlighted in Figure 3.7.2. It would be useful to investigate ways to control for this volatility; one option would be to include controls in the calibration for the distribution of the population with respect to NSSEC, which could be obtained from LFS or the Annual Population Survey (APS).

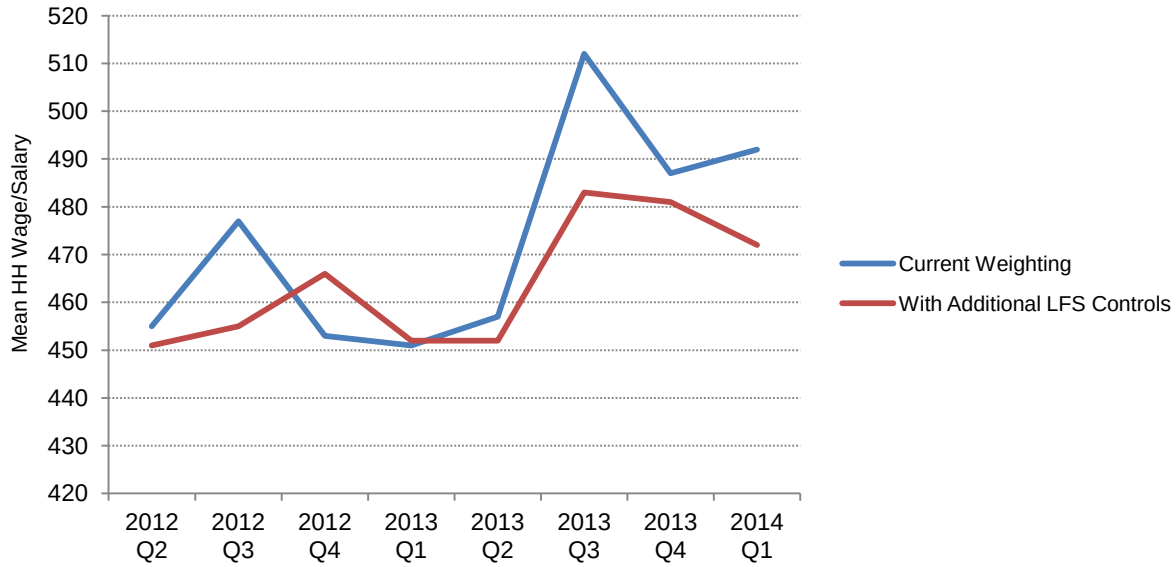
Should income measures for the whole population become available from administrative sources, then the Review Team would strongly encourage investigating the impact of additional controls based on these measures on the precision of LCF estimates, and their quality in general.

Table 3.7.2: Impact of adding LFS controls on estimates of growth in mean household wages/salaries between 2012/13 and 2013/14

		Current weighting		With additional LFS controls	
Year	Quarter	Mean HH[17] wage/salary	Year-on-year % change in mean HH wage/salary	Mean HH wage/salary	Year-on-year % change in mean HH wage/salary
2012	April to June	455.7		451.2	
	July to September	477.9		455.7	
	October to December	453.8		466.5	
2013	January to March	451.5		452.5	
	April to June	457.8	0.5	452.5	0.3
	July to September	512.6	7.3	483.6	6.1
	October to December	487.7	7.5	481.7	3.3
2014	January to March	492.6	9.1	472.1	4.3

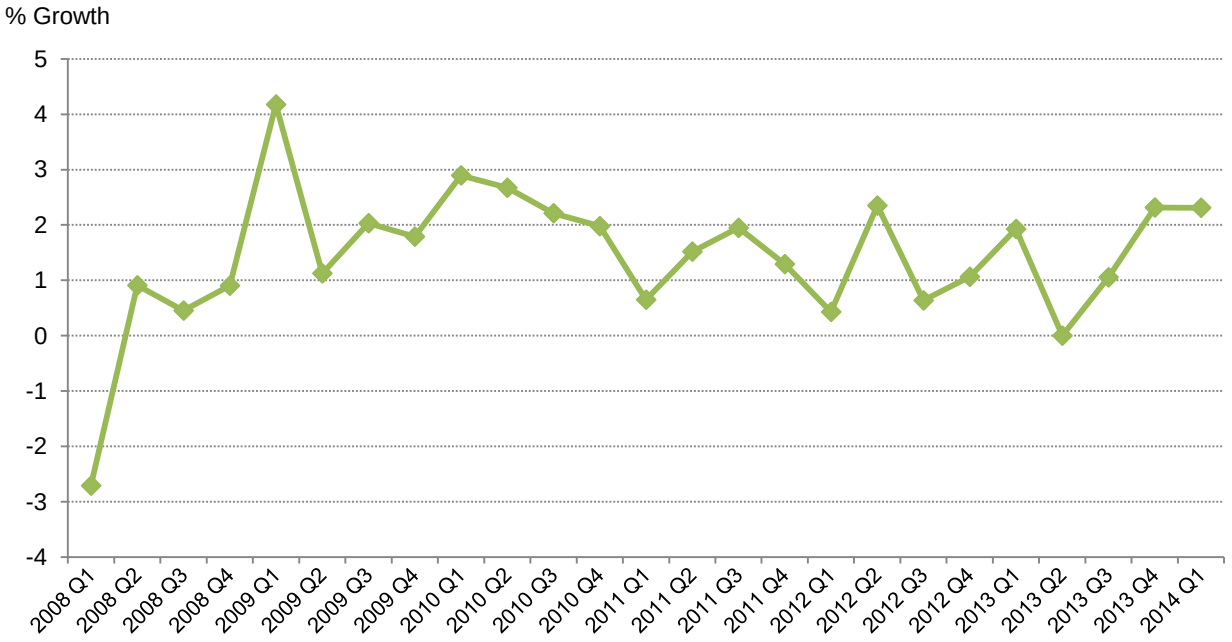
The movements often look contrary; however, if the 2 series are plotted together the patterns do not look so different, as shown in Figure 3.7.1.

Figure 3.7.1: Impact of adding LFS controls on estimates of growth in mean household wages/ salaries between 2012/13 and 2013/14



[17] HH stands for "Household"

Figure 3.7.2: Comparison of year-on-year (2008 to 2014) quarterly growth of average weekly earnings



The original vision for the IHS included using the strength of the large combined sample size to improve the estimates of variables collected only on component surveys. While this vision was not realised, a similar process to utilise the information collected in the LFS/APS to improve the estimation of all variables in the LCF is possible (without affecting the LFS estimates); this would be more embedded in the weighting process than adding additional LFS-based constraints as illustrated in Table 3.7.2. Merkouris (2014)[18] sets out a composite calibration framework for producing consistent estimates for multiple surveys and such an approach could be used to make the LCF consistent with LFS and reduce the variance of LCF estimates. The usefulness of this approach should be investigated as a research project.

Recommendation 3: Carry out further work on the use of LFS and other controls. Include research on the benefits of using the composite calibration framework to improve the quality of LCF estimates on income estimates.

3.8 Expenditure weights for Consumer Prices Indices

Providing weights for consumer prices indices is one of the main uses of the LCF. The decline in the achieved sample size for the LCF has raised concerns about the effects of the increasing sampling error on consumer price index numbers.

[18] Merkouris, T. (2014). Composite calibration estimation integrating data from different surveys. Proceedings 59th ISI World Statistics Congress, August 2013, Hong Kong, 205-210.

3.8.1. Extent and accuracy of expenditure estimates

The expenditure information from the LCF is mostly coded using the COICOP classification system[19] for use in the CPI, though other categories are also provided for use in the RPI which does not use the COICOP scheme. COICOP weights for the CPI are based on National Accounts estimates of household final monetary consumption expenditure which is derived from a variety of sources including the LCF.

The LCF breakdown of expenditure is published as Table A1, which is part of the annual Family Spending publication. The table contains expenditure estimates for categories at up to 4 levels of classification[20]; about 700 categories are listed. The table contains the number of households in the sample reporting each category of expenditure and the percentage standard errors for each category. For categories with few reporting households, the standard errors tend to be large. A small extract from this table is given in Table 3.8.1a below.

Table 3.8.1a: Expenditure data for pets and pet food

Category			Average weekly expenditure all house-holds (£)	Total weekly expenditure (£ million)	Recording households in sample	Percentage standard error
9.3.5	Pets and pet food		4.30	116	1,980	*5.3*
	9.3.5.1	Pet food	2.20	59	1,870	*3.6*
	9.3.5.2	Pet purchase and accessories	0.80	22	630	*13.8*
	9.3.5.3	Veterinary and other services for pets identified separately	1.30	34	190	*11.4*

Of the 700 categories of expenditure reported, for the most recent set of expenditure data, about 100 have CVs (or percentage standard errors, as reported in the table) exceeding 20%.

The LCF does not provide all the expenditure information needed for the weights used in consumer price indices and other sources are required; in particular, for some categories of commodity, a finer level of detail is required. As a rough estimate, the LCF provides 50 to 60% of the expenditure data (by weight) with another 30% using LCF as a partial source. The remainder is found from other sources such as market research data. In some cases, the LCF expenditure estimates are broken down into sub-categories using other sources, using the LCF values as constraining totals. If no other information can be found, expenditures are broken down in equal shares.

As noted in the introduction, there are categories of commodity where more data would be useful, including clothing and some electronic goods. For categories where LCF information is not available, or categories which have large CVs, the further potential for using market research data should be explored; this would supplement LCF expenditure data for consumer price index needs.

[19] Classification of Individual Consumption by Expenditure According to Purpose, available here: unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5&Lg=1

[20] www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey#technical-report

There is potential for sources other than the LCF to provide more of this additional information; this is discussed further in Chapter 7.

3.8.2. The effect of LCF sampling errors on Consumer Price Index Numbers

The calculation of consumer price index numbers uses expenditure estimates from the LCF to weight together price changes for the items in the basket of goods and services to produce an overall measure of price change. As the weights are derived from a sample, they are subject to sampling error which is then propagated into the price index numbers. It is not the only source of sampling error – the prices and the goods and services are also sampled, which introduces further sampling error to the index numbers.

Calculating the effect of the sampling error from both weights and prices is not a simple matter and ONS does not produce measures of uncertainty for either the CPI or the RPI. Some investigation of the effect of varying weights has been carried out before; for example, for the calculation of a superlative index version of the CPI for locally collected data in support of the Johnson Review[21].

To support this review, a preliminary study has been carried out which uses the boot-strapping technique to propagate sampling errors from the expenditure weights through the calculation of price index numbers[22]. The bootstrap involves taking repeated sub-samples from the LCF sample with replacement, and recalculating the estimates of expenditure. These provide new sets of weights, which are used to re-calculate the price index numbers. The variability among the bootstrap estimates enables standard errors of the index numbers (due to sampling variation in the LCF weights only) to be estimated.

A significant difficulty with these calculations arises from the fact that published COICOP weights for the CPI are derived from the National Accounts and involve data sources other than the LCF; the process also involves balancing adjustments. Calculating the effects of uncertainties in other data sources as well as from the LCF (and National Accounts processes) is not practical. Instead, calculations of RPI index numbers were made – the RPI uses LCF weights directly. In order to produce results for COICOP categories (that is, the categories used in the CPI), RPI items were mapped to COICOP classes.

Calculations were made over the period from January 2013 to January 2015 (25 months) at 3 levels of the COICOP structure – the all-items, the 12 divisions and 41 groups. The results show that the effect of LCF sampling error is small; for 24 of the 25 months, the standard error due to sampling variation in the LCF weights was below 0.05 index points. Even at the lower levels of the COICOP hierarchy, the standard errors rarely exceeded 0.1 index points; the exceptions being education, phones and certain categories of household goods.

[21] Sanderson, R. et al (2013), Calculating a Retrospective, Superlative Consumer Prices Index for the UK, available at: webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/index.htm

[22] A paper describing the work will be produced later in 2016.

3.8.3. Consumer Price Index Numbers for Sub-populations

The need for price index numbers for regions and sub-populations was considered by the Johnson Review of Consumer Price Indices[23]. ONS has produced a study that estimates the inflation experience of households within income quintiles based on differences in their weights as derived from the LCF (but not on other differences, which continue to be averaged); it does show differences between inflation estimates for different income quintiles. The Johnson Review (Section 6.9) recommended publishing an annual analytical output which shows the inflation experience of different household types.

It is an additional challenge to provide expenditure weights for regional price indices or for household types – both will use sub-samples from the full LCF sample. The calculation of standard errors of price index numbers resulting from LCF sampling error should be extended to include sub-populations.

Recommendation 4: Continue the preliminary analysis of the effect of sampling error in LCF expenditure estimates on price index numbers and publish the results. Extend the analysis to sub-populations.

3.8.4. LCF data for National Accounts

LCF data on spending is an important source for Household Final Consumption Expenditure (HHFCE) which contributes to the National Accounts. An internal review of this use of LCF data was carried out in 2012 and was considered by the Johnson Review; it noted that LCF is one of 20 sources used in HHFCE, though it represents about two-thirds of the categories of expenditure needed. The precision of the data supplied by LCF was also assessed. Overall, about 40% of expenditure data had CVs exceeding 20%. The current review has re-examined a recent data file; this shows a similar proportion of data with high CVs. As data is needed on a quarterly basis, the annual sample size is correspondingly reduced and results in larger CVs. The 2012 internal review noted that the precision would be improved by increasing the LCF sample size, but the increase would need to be by an order of magnitude to ensure all estimates had CVs below 20%.

[23] www.statisticsauthority.gov.uk/reports-and-correspondence/reviews/

4. Data collection

4.1 Respondent burden and possible bias

This section considers the topics of respondent burden and possible bias in estimates. It provides a qualitative assessment of respondent burden as measured in focus groups carried out as part of this review, discusses evidence of non-response bias in the LCF and, finally, the impact of incentives on improving response.

Focus groups were conducted as part of this review, 3 with LCF interviewers and 1 with interviewer managers. The focus groups with interviewers covered topics such as techniques to achieve response, the survey's incentives, respondent burden through the interview and the diary, as well as quality assurance processes. The focus groups with interviewer managers covered similar topics exploring them from a management perspective as well as looking at issues around work allocation, interviewer training and briefing.

4.1.1. Respondent burden

Respondent burden is the effort, in terms of time and cost, required for respondents to provide satisfactory answers to a survey. When agreeing to take part in the LCF, respondents commit to 2 elements: a face-to-face interview and the completion of a 2-week expenditure diary. It is often assumed that the level of respondent burden impacts on response rates. Respondent burden for the LCF arises from the length and complexity of the questionnaire, combined with the demands of completing the expenditure diary.

The average interview length for the LCF is available from 2005 when the median was 51 minutes. The interview length remained fairly constant until 2008 when it increased to 55 minutes. This coincided with the LCF survey joining the IHS and, therefore, the introduction of additional demographic questions at the start of the questionnaire. The interview length has been reasonably consistent since. Response rates fell over this period, supporting the view that they are impacted by a number of different factors, not just the interview length.

Respondent burden was a main topic for the focus groups with LCF interviewers and their managers. Interviewers reported that the interview length and complexity added to respondent burden – noting that the interview length is increased by respondents having to look for bills/receipts. The complexity of some sections results in a lapse of the respondent's concentration past 1 hour, causing a particular challenge for larger households. The main stumbling block for respondents are sections that ask for combined amounts (for example dual-fuel gas and electricity payments, television bundle payments, benefits and insurance) and, subsequently, proportions for each item, the employment/self-employment questions and questions on cars (for example, car tax, and part exchange). These difficulties are explored in more detail in Section 4.3.4 below.

This evidence points to a need to reduce the interview length. This could be achieved by more aggregation of expenditure categories[24]. A research paper by Browning, Crossley, and Winter[25] suggests that this would reduce non-response but at the expense of increasing under-reporting. Existing user needs have been described in this review introduction, which points to a need to retain (or increase) the fine level of data disaggregation. In addition, under-reporting of expenditure and income is currently a concern as discussed in Section 4.3 "Under-reporting expenditure and income". Interview length is discussed further in Section 4.3.4.

The diary layout/format has remained unchanged since the current survey design was implemented in 2001. The 2-week diary period is designed to optimise the trade-off between obtaining comprehensive expenditure data over the survey sample and maintaining an acceptable level of respondent burden. Respondent burden could be reduced by shortening the diary period, and/or not collecting diary data for every member of the household. Analysis carried out as part of this review (see Section 4.3.4) suggests the existence of diary fatigue, where data collection tails off as time increases or expenditure patterns change due to the act of keeping a diary. This effect will be contributing to the level of under-reporting experienced and points towards a reduction in diary length potentially increasing data quality, but at the expense of some precision. Another option could be to use a single diary per household as is the case in the US (it should be noted that the design of the US survey is under review). This approach appears to reduce non-cooperation, but also increases under-reporting. Some respondents associate "household spending" with shared expenses only, or those necessary to "run the household". This suggests that individual expenditure is more likely to be under-reported with household-level diaries.

4.1.2. Evidence of non-response bias in the LCF

Measuring non-response bias is difficult because of the challenges intrinsic to obtaining information on the characteristics of non-responding households. Furthermore, the relationship between response rates and bias is not clear. It is not certain that improving response would reduce bias, since the "new" respondents may have characteristics that tend to increase the overall bias. There are several strands of analysis that shed some light on bias in the LCF[26].

The main mechanism for adjusting for non-response bias in the LCF is through the weighting system, which has the dual purpose of compensating for differential sampling rates and adjusting responses to match known population totals by age group, sex and broad region. This makes suitable adjustments for variables correlated with age-sex-region distributions; there may, however, be some residual non-response bias where the correlation is weak, or where the probability of response is related to variables measured in the survey and special studies are normally needed to investigate these residual biases.

[24] The US survey has different samples for diary expenditures (predominantly small and regular) and recall expenditures.
[25] Browning, M., Crossley, T., and Winter, J.K. (2014). The Measurement of Household Consumption Expenditures. Annual Review of Economics. Vol. 6, 475-501.
[26] Indicators for the Representativeness of Survey Response, Bethlehem J, et al, 2008, available here: www.statcan.gc.ca/pub/11-522-x/2008000/article/10976-eng.pdf

An important opportunity is the 2011 CNRLS, which matched census records to other survey cases from the period around the census. The results are considered here, both those obtained directly from the LCF and, also, from the LFS study.

(i) CNRLS

The LCF currently uses weighting adjustments for non-response bias based on the 2001 CNRLS. The following characteristics are used:

- region
- household composition
- household reference person of pensionable age
- householder is female and over 85 years of age
- number of cars

The 2011 CNRLS showed that application of these factors has little effect on the LCF expenditure and income estimates. The maintenance of these updated adjustments in the future is under discussion, since their benefits are slight.

This analysis offers some reassurance with respect to bias. However, it should be noted that bias could only be assessed in census variables, choosing variables which had a relationship to variables measured by LCF. Thus, although variables thought to be most closely linked to non-response, such as income, could not be tested directly, variables most strongly related to survey estimates were measured.

(ii) Analysis by household and area characteristics

Analysis of refusal rates across a range of ONS surveys shows that the Output Area Classification system used by the census is a powerful predictor of local non-response rates, with 2 categories within this classification identified as being hard to collect data from – “Multi-cultural” and “City Living”, covering 18% of the UK population. These area types had a greater difference in refusal rate compared with the UK average rate than any single demographic variable. This reflects how their presence is often combined; for instance, younger people in urban areas are more likely to rent and less likely to drive.

There is a known tendency for high-income households to have a low response rate, which can exacerbate under-reporting of expenditure. Statistics Canada oversamples high-income households (based on tax data) for this reason. A similar approach is taken in the US for the Household Assets Survey. Evidence is also present in the LCF that under-reporting exists for high income households (see Section 4.3.7). Data for household income is not available from the census; therefore, an alternative source of data is needed to explore the impact of non-response on this variable more fully.

As part of this review, analysis was carried out using the 3% of diaries for which data was imputed in the 2013 dataset to identify the characteristics of individuals and households whose diaries had been imputed. Non-responding individuals within a responding household, whose diaries have been imputed, were more likely to be male, aged 40 or under, single/never married, and working in the week prior to the interview.

(iii) Further investigation

There is currently some valuable information on non-response bias in the LCF (and other social surveys), but the evidence is not complete. A follow-up survey of households included in the LCF sample, both responding and non-responding, would potentially be a valuable additional source.

Recommendation 5: Carry out a follow-up survey of households included in the LCF sample, both responding and non-responding, to provide further information on non-response bias in the LCF.

4.1.3. Addressing non-response

There is extensive literature on non-response and how to address it. The book "Non-Response in Household Interview Surveys"[27] is widely cited as a reference for methods of addressing non-response issues. Its main suggestions include:

- more difficult sample units should be tried early in the sample period as, on average, they will need more calls to be contacted
- urban residents are harder to contact even after accounting for other variables, such as crime rate, and could be prioritised for visits early in the sample period
- single-person households refuse more than larger households
- pre-paid incentives are more beneficial than those conditional on survey completion in less complex surveys

The recommendations are not necessarily directly applicable to the LCF, but give a flavour of the factors that can impact survey response.

ONS has carried out investigations into the factors contributing to non-response, and how to apply this understanding to improving response.

(i) Calling patterns

Analysis of optimal calling patterns was carried out using information collected while conducting the 2011 Census Coverage Survey[28]. It was found that, for all households, the best times for making contact were between 6pm and 9pm. However, this did vary depending on attempt, particularly for weekends, and household characteristics. Improvements for increasing the likelihood of contact in surveys were made, including instructing interviewers to vary the times at which they call at the same address. Calls between 8am and 10am should be avoided for the first call attempt and interviewers should be encouraged to record information that may help future contact.

Current work within ONS is looking again at calling patterns and interviewer performance, change in management, change in characteristics of the area, whether there has been an increase in certain types of non-response (non-contacts or certain types of refusals) as well as other factors. The work is focussed largely on the LFS, but many of the findings are applicable to the LCF and will, therefore, be applied to improve surveys processes where possible.

[27] Groves, R.M. and Couper, M.P. (1998). Non-response in Household Interview Surveys. New York; Wiley.
[28] Ross, H. (2014). An Analysis of Optimal Calling Patterns Using the 2011 Census Coverage Survey. Survey Methodology Bulletin. Pp 73-99.

(ii) Impact of incentives on improving response

Currently, LCF sampled households receive a book of stamps with the advance letter as an unconditional incentive. Conditional incentives in the form of a £10 voucher are given to adults within responding households that complete the face-to-face survey and diary. Children who complete the simplified children's diary receive a £5 voucher. This incentives structure has been in place since the current survey design was launched in 2001.

The value of incentives has a bearing on the impact on response and is particularly relevant for the LCF, given the burden placed on respondents to complete the expenditure diary. The US Consumer Expenditure survey experimented with $20 and $40 incentives and found that the $40 incentive (but not the $20) improved response rates and data quality for the interviews; while both had a similar effect on the diaries, primarily improving data quality rather than response rate.
An experiment was conducted on the LCF from July to December 2010 to test whether an increase in the conditional incentive from £10 to £20 for the main diary keeper increased the survey response rate. Results showed that the response rate for the £20 incentive group was 3.5 percentage points higher than the £10 incentive group. This difference was statistically significant. A decision to increase incentives has to be made in light of the higher survey costs that would be incurred. On balance, it was decided not to increase the incentive as there were no funds available to cover the additional costs.

There is evidence to suggest that pre-paid incentives are more beneficial than those conditional on survey completion in less complex surveys[29]. Testing would be required to understand whether this is the case in a complex survey like the LCF.

A trial on the LCF started in November 2015 and will be run for 8 months to test the impact of an unconditional £5 voucher for respondents who are non-contacts on the first interviewer call. This will be a split sample trial with half of quotas not receiving a book of stamps in the advance letter, but instead receiving the £5 if direct contact is not made. The purpose of the trial is to measure whether the use of the £5 unconditional incentive will reduce the average number of calls to first contact. This could free up interviewer resource to target respondents who are harder to contact. The focus groups carried out for this review revealed that interviewers largely agreed that, while the book of stamps has a useful purpose of helping interviewers to build rapport where respondents remember receiving it in the advance letter, stamps are an out-dated incentive and do not provide any reason to participate in the survey. Interviewers thought that an unconditional incentive of more practical value would be more useful or, alternatively, a higher conditional incentive.

As with the book of stamps, interviewers largely agreed that the conditional incentive (£10 voucher for completed interview and diary) does not stand in any relation to the burden put on respondents considering the length of the interview and the diary period. Interviewer managers thought that most respondents participate because of the interviewer's skill in explaining the importance of their contribution. Respondents who participate because of the incentive often do not accurately

[29] Incentive Payments on Social Surveys: a Literature Review, available here: webarchive.nationalarchives.gov.uk/20150904113534/http://ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-53/index.html

complete the diary. Interviewers thought that it would be easier to administer if only the main diary keeper had to keep a diary and this person could receive a higher incentive.

Recommendation 6: Consider whether a change in the incentive structure can be used to address low survey participation rates in a cost effective manner.

In particular, the following question should be addressed: what amount is needed to improve response rates and motivate respondents to accurately complete the expenditure diary? The findings from the LCF unconditional voucher trial (due to report in June 2016) should be considered, although additional trials may be needed. It is possible that the impact of increasing incentives beyond a certain limit can be counterproductive since they may be interpreted more as earned income than courtesy payments.

4.2 Data collection management

The 2-stage element of the survey presents challenges to both respondents and interviewers. The focus groups were used to understand more about the challenges faced by interviewers in successfully administering the survey and achieving a positive outcome.

Interviewers have a range of responsibilities, including managing their workload across multiple surveys. When administering a survey, the stages are making contact, selling the survey and completing a successful interview. The LCF, however, necessitates a range of additional responsibilities, such as getting respondent buy-in to complete the diary, explaining how to fill in the diary, completing diary checking calls to households to ensure procedures are being followed, and checking weights and measures information at supermarkets if the detail is missing from the diary.

In the focus groups, the practical challenges of completing checking calls were raised by interviewers given travelling distance and competing priorities with other survey work. Interviewers put a lot of effort into making contact with a sampled household and following up on respondents regarding their diary completion, but often find it difficult to plan their time efficiently when work is allocated to places that require them to travel far from where they live and/or far from other work to which they have been allocated.

In comparison to other surveys, interviewers find it difficult to “sell” the survey to participants, in particular because of the diary element. Interviewers require a lot of skill to build up a good rapport with the respondent to get consent to conduct the interview and, subsequently, motivate respondents to accurately complete the diary for two weeks. In comparison to single-mode surveys, the LCF offers “too many opportunities to refuse”.

There was some evidence suggesting that some interviewers do not mention the diary up-front in their introduction, leaving respondents feeling “tricked” into participation and others not mentioning that the survey participation is voluntary, leaving respondents to believe they must participate.

The focus group discussions showed there were mixed views amongst interviewers and their managers with regards to providing respondents with more information prior to the interview to

make them aware of what to prepare. Some thought it would help to make the interview quicker, others thought it would lead to more outright refusals.
Focus groups with interviewers and their managers highlighted issues around work allocation. Concerns were raised that, under the survey's design, multiple quotas of work can fall in a single month in a geographic area under the responsibility of an interviewer manager. This makes it difficult to balance the LCF workload alongside other survey work, as well as leading to problems with maintaining levels of contact and response.

Recommendation 7: Review the methodology for allocating field work to geographic areas to ensure a more even distribution within interviewer areas across the year without violating the survey's design.

There is also a need to ensure that a consistent and complete introduction of the LCF survey and its features is delivered to sample members at first contact. Work took place during 2015 to improve the LCF training package for new interviewers as well as introducing refresher training. This includes a focus on selling strategies to increase interviewer confidence to achieve good results on LCF interviews and to reinforce instructions to ensure a consistent approach across the interviewer field force.

Recommendation 8: Review the effectiveness of the new interviewer training materials by carrying out interviewer and interviewer manager focus groups towards the start of the 2016/17 financial year.

Interviewers spend a substantial amount of time on quality assuring the data collected for this survey. This starts at the interview where interviewers are prompted by automatic checks within the questionnaire and seek clarification of figures provided from the respondent where needed. A fuller set of automated checks are applied after the interview, allowing follow up of any discrepancies identified when returning to the respondents' home to carry out a diary checking call. The majority of interviewer time appears to be spent on the diary, including explaining/selling the diary to the respondent, helping the respondent to complete the diary, validating data entries by probing the respondent, and checking data online or by visiting supermarkets. Interviewer managers reported that they had very little involvement in the quality assurance of LCF data and that their focus was more on the overall interviewer performance.

Interviewers highlighted that the delay in follow-ups from coders in Titchfield are often several weeks after the diaries have been submitted and, sometimes, more than once per diary. The length of time between the interview/diary period and the call back means that interviewers as well as respondents struggle to recall the circumstances around particular expenditure and, therefore, to rectify any inconsistencies identified by the coders.

Evidence from this research showed that interviewers reportedly spend a substantial amount of time on quality assuring diary responses with further checking done at the coding stage.

Recommendation 9: Review the quality assurance processes carried out by interviewers by considering whether any of their tasks could be carried out more efficiently by office based staff.

4.3 Under-reporting expenditure and income

4.3.1. Context

There is evidence that diaries suffer under-reporting through "diary fatigue" (less expenditure is recorded later in the diary period). For example, these patterns have been reported for the US Consumer Expenditure (CE) survey[30,31]. Longer collection periods induce greater diary fatigue and, hence, more under-reporting.

Comparison of LCF data to other sources suggests under-reporting of expenditure and income in the survey data. For example, work carried out by Mike Brewer and his collaborators [32] points towards under-reporting within the household interview. Their analysis suggests that the joint distribution of expenditure and income in the survey is implausible. This is a problem shared by the US survey and, to a lesser extent, the Canadian survey. The data is consistent with the hypothesis that poor households under-report income and rich households under-report expenditure. However, this has not been proven or quantified yet. A project to investigate this phenomenon is under discussion in the innovation panel of the Understanding Society survey[33].

Given the detailed disaggregation required by main LCF customers (DEFRA and Prices), the use of recall and diary collection is not under review, as the level of detail needed could not be achieved through recall alone. Evidence from the US is that recall amounts of regular expenditure items are under-estimated compared with diaries[34]. Therefore, recommendations arising from this chapter will cover improvements to both the diary and household questionnaire, assuming that the diary will remain in the paper form. However, some considerations have been included in relation to web data collection, which is viewed by some as a potential way to improve diary collection. Moreover, Chapter 6 reports on developments by other NSIs and Chapter 7 describes potential future developments through the use of technology.

4.3.2. Work carried out as part of this quality review

This quality review includes an assessment of the extent of under-reporting of expenditure and income. Time constraints did not allow a full review of all sections of the data collection instruments or the identification of all possible sources of under-reporting. A focused approach was taken, as described below, in order to make the best use of the resources available.

For expenditure, a qualitative assessment of the extent of under-reporting due to the design of the data collection instruments was made. Experts from the DCM team undertook an expert review of a prioritised list of expenditure topics. The list of topics to be included in the review was identified

---

[30] Silberstein, A.R., and Scott, S. (1991). Expenditure Diary Surveys and their Associated Errors, in Biermer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S., Editors, Measurement Errors in Surveys, Wiley, Hoboken NJ: 1991.
[31] Stephens, M. (2003). 3rd of the Month, Do Social Security Recipients Smooth Consumption Between Checks? American Economic Review 93(1), 406-22.
[32] Brewer & O'Dea paper, published in March 2012, by the Institute for Fiscal Studies (IFS): www.ifs.org.uk/wps/wp1212.pdf
[33] www.iser.essex.ac.uk/understanding-society
[34] Survey instruments and the reports of consumption expenditures: evidence from the consumer expenditure surveys. JRSS-A 179 559–581

with reference to a previous internal National Accounts review of the LCF, as well as consultation with main internal customers (National Accounts and Prices). Feedback from the interviewer focus groups confirmed that the list of topics identified included those which respondents find more difficult to complete.

For income, a more quantitative approach was taken, comparing LCF income estimates with those from the UK EU-SILC survey[35], which is a harmonised survey run under regulation in EU member states and provides a comparable data source for income. Harmonised questions are used in the 2 surveys where the data requirements are the same. However, differences between LCF and EU-SILC should be noted; for example, the sample size for EU-SILC is larger than the LCF, EU-SILC contains a more detailed set of income questions and the surveys' post-data collection systems are different.

In addition to the DCM questionnaire review and analysis of income data, the investigation of under-reporting was supplemented with findings from the focus groups described in Section 4.2. Within the focus groups, questions were included to elicit information about sections of the questionnaire and diary that respondents find difficult to answer/complete, as well as challenges interviewers face when implementing the survey procedures around diary checking and quality assurance.

Moreover, some analysis has been carried out to explore the impact of the interview length on data quality, and also to examine the level of item non-response in questionnaire sections with high cognitive burden.

The sections that follow provide a summary of the observations and findings from each of these strands of investigation. They report evidence underpinning recommendations for changes to the data collection instruments and for further work to understand more about the causes of under-reporting.

4.3.3. Overarching issues relating to measurement and non-response error

The expert review of expenditure topics focused on the detail of specific sections of the questionnaire and diary. In addition, DCM provided general observations about the LCF data collection design, which are applicable to income and expenditure data, as outlined in the bullet points below. These observations will be considered when implementing recommendations for change to specific sections of the questionnaire and diary resulting from this review.
By nature, the data being collected is complex and, therefore, leads to demanding cognitive processing by respondents. Therefore, when considering questionnaire changes, it is important to ensure cognitive tasks are broken down and simplified. Problems encountered during the interview and diary completion process may include (but are not limited to):

- encoding of behaviour in the memory and recall – not all spending habits may be sufficiently encoded in the memory to enable accurate recall, so rounded estimates may be provided

[35] Cross sectional EU-SILC data was used for the analysis as ONS have access to this data. The Family Resources Survey (FRS) is an alternative source of comparable data. However, access to FRS data is more restricted as ONS carries out the survey on behalf of DWP.

- cognitive ability – the level of cognitive processing required to answer questions accurately may be out of reach for some adults in the population
- satisficing[36] – respondents may learn to answer "yes" or "no" to questions in order to avoid further questions
- interview pace – it may be faster than some respondents can cope with
- social desirability – respondents may not answer questions, or may not answer them accurately, in order to conform to perceived norms or conceal behaviour
- interview and diary keeping fatigue – data quality during the interview and over the 2-week diary period may reduce
- change in spending behaviour – in the process of recording their expenditure, respondents may change their spending habits – this is potentially very important because it means even a perfectly representative sample becomes unrepresentative through the act of completing the survey; this is a little understood phenomenon
- level of incentive payment – the payment may not be high enough to motivate respondents to record data accurately
- proxy response – accepting proxy responses introduces an additional measurement error
- questionnaire length – the questionnaire is long and burdensome which may lead to a reduction in data quality in later stages of the interview

The focus groups identified 3 different types of under-reporting with regards to the diary element:

- respondents unintentionally not recording expenditure (in particular meals and drinks consumed when going out)
- respondents intentionally doing smaller shops or postponing their big shop until after the diary period
- respondents intentionally not recording certain expenditure items (for example cigarettes, alcohol, products of an adult nature)

Of course, the problems described are not unique to the LCF survey. More research is needed to confirm the extent of these issues and understand their causes and effects. It should be noted that, in following good practice questionnaire design, the occurrence of such problems is, at least, minimised.

4.3.4. Impact of interview length and complexity on data quality

Descriptive analysis of LCF interview and diary data from 2013 provides evidence of diary keeping fatigue and reduction in data quality within complex sections of the questionnaire.

The average number of items and average net amount recorded in the LCF diary was analysed by week and category (that is, purchase type, online versus off-line expenditure) to look for any patterns of diary keeping fatigue. The average number of items recorded decreased by 12% between week 1 and week 2. Categories most impacted were "takeaways brought home" (23% decrease) and "eating out" (16% decrease). Non-food purchases were 11% lower. Similar patterns

[36] Motivated Under-Reporting in Screening Interviews. Journal of Survey Statistics and Methodology.

can be seen when looking at internet only and non internet purchases. The decrease in items recorded from week 1 to week 2 translates into a reduction in the average total amount spent of 9%. The impact of average total amount spent is particularly marked when looking at internet only purchases, with a decrease of 26%. This analysis supports the theory of diary fatigue described above and suggests further investigation is needed to understand the difference identified in type and method of purchase.

Interviewers were in agreement that the diary length of 2 weeks led in the majority of cases to respondents showing signs of diary fatigue or, in some cases, respondents even refusing to complete the diary altogether. It is possible that a shorter diary period could lead to increased response rates and higher data quality. However, a shorter diary period could reduce purchase frequency too much and result in increased zero recording for some items and, therefore, higher variability.

Recommendation 10: Carry out further work to investigate the trade-offs between maintaining a 2 week diary period and adopting a shorter diary period.

In particular, the following questions should be addressed: could a shorter diary period produce the same level of accuracy whilst increasing response rates and data quality[37]? Or would a shorter diary period reduce purchase frequency too much resulting in increased zero recording for some items and higher variability?

Evidence from the interviewer focus groups indicated that respondents struggle with questionnaire sections that ask for combined and then individual amounts of expenditure. Interviewer experience suggested respondents are more likely to remember (or look-up bills for) combined payments, but then struggle to split this into proportions. Analysis focused on sections of particular concern such as gas/electricity and television/phone/internet bundles.

The following tables confirm interviewers' feedback, as the proportion of item non-response – based on cases that were imputed and/or abated[38] – was low for combined payments or single payments, but considerably higher for questions that ask for the proportion of an item that was part of a combined payment.

[37] A shorter total period would result in an increase of sampling variance but an increased response and decreased measurement error may more than compensate.
[38] For example, expenditure has been reduced to account for business expenses received.

Table 4.3.4a: Non-response in the combined payments versus separate sections of the LCF household interview, 2013

Gas and electricity

Combined payment section variable	Non-response[1]	Non-response[2]	Separate payment section variable	Non-response[1]	Non-response[2]
Gas and electricity combined amount	2.0%	0.8%	N/A		
Gas portion	13.3%	12.4%	Gas amount	2.5%	1.4%
Electricity portion	13.5%	12.4%	Electricity amount	2.5%	1.1%

Telephone, television and internet

Combined section variable	Non-response[1]	Non-response[2]	Separate payment section variable	Non-response[1]	Non-response[2]
Television/ phone/ internet combined amount	2.8%	0.3%	N/A		
Television portion	20.3%	20.2%	Television amount	0.2%	0.2%
Phone portion	19.8%	18.1%	Phone amount	2.4%	0.5%
Internet portion	20.0%	18.4%	Internet amount	2.8%	0.1%

Notes:

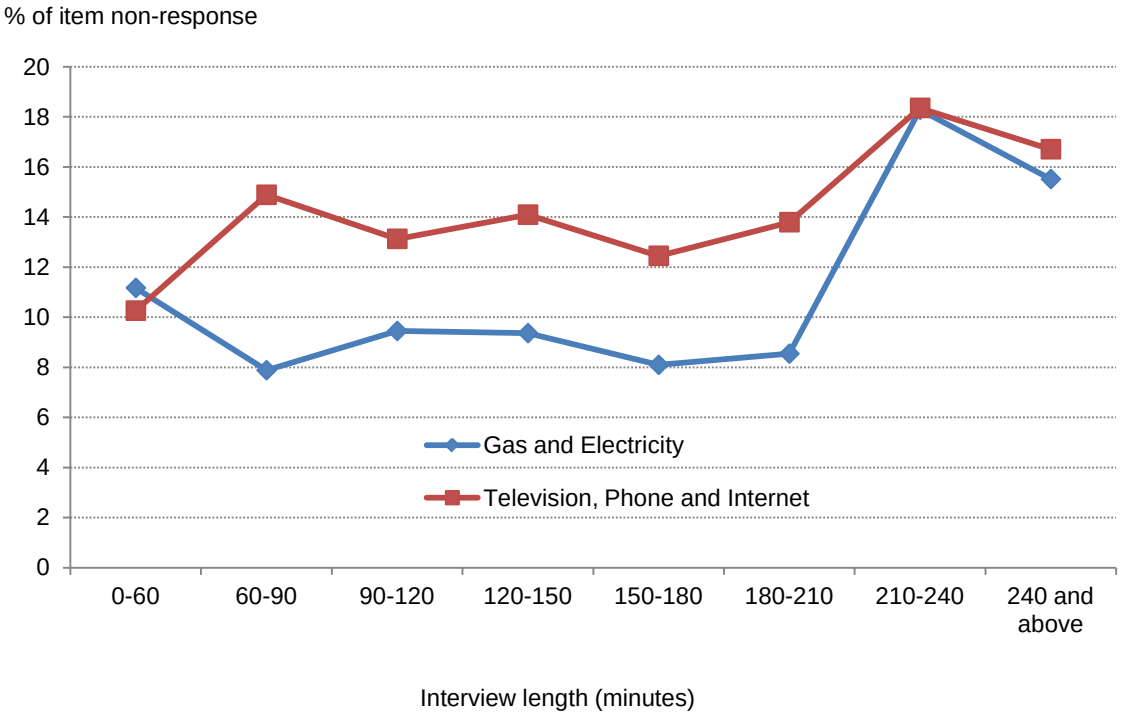
1. Non-response defined as any abatement or imputation to original values
2. Non-response defined as imputation to original values only
3. Proportions of non-response considers all iteration of each variable
4. Other potential comparisons initially suggested in scoping are not included due to low number (less than 100): gas or electricity standing order, rebate or meter amounts

Recommendation 11: Carry out further work to review the level of detail needed for combined payment breakdowns (where bills may not provide the level of detail requested).

The information gathered should be used to inform future questionnaire development. For example, would it be possible to simplify the questionnaire, particularly for combined payment breakdowns or is the need for the detailed breakdown sufficient to justify the impact on respondents?

Interview fatigue was also assessed by looking at the level of item non-response in 2 selected sections of the household interview in relation to interviewer estimates of interview length. The average proportion of item non-response (based on cases that were imputed and/or abated) is similar for both sections for interviews of up to 60 minute length. The proportion of item non-response for both sections levels off between 60 and 210 minutes before it sharply increases for interviews of 210 minutes or longer. Overall, the proportion of item non-response is lower for television/phone/internet bundles than for gas and electricity payments. The analysis suggests that data quality reduces as questionnaire length increases; however, more detailed analysis would be needed to confirm this, accounting for households with longer interviews.

Figure 4.3.4a: Average proportion of item non-response within section of interview length



Recommendation 12: Carry out further analysis on a larger scale to explore the extent of interview length on interview fatigue (including more questionnaire sections asked at different points in the interview and potentially a further survey year to increase robustness of results).

The additional analysis should use questionnaire audit trail data instead of interviewer estimates on the interview length, which was used as a proxy for this preliminary analysis. The findings of this analysis should be used to inform a decision on whether the questionnaire should be shortened in order to increase data quality as well as reduce respondent burden. It is unlikely that reducing the level of detail collected would be a practical option because it has the potential to introduce under-reporting. However, modularising the existing questionnaire is an alternative option, whereby not every household would be asked all areas of expenditure. Estimation of totals would still be possible, but depending on the modularised design, covariances may be lost which are very important to some users. A design with overlapping modules would allow covariances to be estimated but with reduced quality. Main users would need to be consulted on a re-design of this nature.

4.3.5. Under-reporting of expenditure within the data collection instruments

The expenditure topics included in the expert review were as follows:

- fuels (petrol and diesel)
- mobile phone account payments
- insurance
- clothing and footwear
- food and non-alcoholic drinks
- alcoholic drinks and tobacco

- satellite/cable television
- package holidays
- childcare payments
- purchase of high value items

A detailed internal report has been produced by DCM describing their findings, including recommendations for changes to the questionnaire and diary, along with potential further work. A summary is provided below. Due to time constraints, no testing has been carried out to measure the impact of the changes suggested, either on data quality or the additional burden the changes may place on the respondent. Further work will be necessary to test the impact of proposals, as well as to fully understand the cause of under-reporting for each topic.

(i) Changes proposed with further testing/analysis

Changes were proposed, with the need for further testing and/or analysis, for 3 of the 10 topic areas:

- fuels (petrol and diesel) – changes to the diary layout and/or content, for example add a separate section for fuels or include petrol on the “usual purchase” page
- mobile phone account payments – suggested changes include recording expenditure in the individual section of the questionnaire rather than the household section, and retaining Pay As You Go top-up expenditure within the diary
- insurance – potential measurement problems were identified within some areas of insurance and suggestions for change proposed

(ii) Changes proposed but further information is needed

Changes were proposed for the following topics:

- childcare payments – data collection should move from the diary to the questionnaire
- purchase of high value items – possible causes of under-reporting were identified, but further review of both the questionnaire and diary is necessary to ensure purchase of high value items is recorded accurately

Further input is needed from users to understand the exact data collection requirements before formal suggestions for change can be made.

(iii) No obvious questionnaire/diary design-related change

The data collection review did not identify any obvious questionnaire/diary design-related explanation for under-recording in the following topic areas:

- clothing and footwear
- food and non-alcoholic drinks
- alcoholic drinks and tobacco

However, interviewers provided anecdotal evidence to suggest respondents are under-reporting certain items in these categories, including cigarettes and alcohol.

Further analysis will be needed to understand more about the source of the under-reporting and, therefore, inform a possible solution. For example, the following topics should be considered: are internet purchases of clothing and footwear as well recalled as purchases from other shops or physical locations? And, is “alcohol brought home” more comparable to other sources than “alcohol bought and consumed outside of the home”?

Recommendation 13: Carry out further analysis to understand other potential causes of under-reporting in the LCF survey.

Additional quantitative research should be carried out, including further comparisons with National Accounts data and comparison of diary and recall totals**.** This analysis should include (but not be limited to) the topics of clothing and footwear, food and non-alcoholic drinks and alcoholic drinks and tobacco.

(iv) Recent changes: checking they are fit for purpose

The prioritised topic list included 2 questionnaire sections that have undergone recent review to check recent changes made to complex sections of the questionnaire:

- satellite/cable TV
- package holidays

The review concluded that both sections have been improved by the changes made, but with some qualifications and uncertainties.

4.3.6. Future development of the LCF data collection instruments

(i) Questionnaire development

The expert review acknowledged the ongoing work to develop the LCF questionnaire and addressed problems identified within data processing and/or analysis including the updates to the satellite/cable television and package holidays questions. Feedback is regularly provided from the processing and analytical functions of the survey team and used to inform the questionnaire design process. However, the review also highlighted the challenge of ensuring that the questionnaire keeps pace with changes in the real world, for example changes in consumer spending/behaviour. This view was supported by interviewer feedback which stated that they found the questionnaire to be “a bit dated”, although it generally flowed well and the household questionnaire in particular sections around mortgages and housing work well.

Recommendation 14: Allocate additional resources to the LCF Research Team (including securing funding for DCM resource) to develop and implement a more robust questionnaire and testing process and ensure that questionnaire design keeps pace with ongoing changes in consumer spending/behaviour.

The following best practice development and testing method for the questionnaire/diary has been suggested:

- initial question development involving qualitative research (for example, focus groups or in-depth interviews)
- expert review of draft questions by DCM/subject matter experts
- cognitive testing of draft questions

- quantitative experiments (and follow-up interviews if results are not conclusive) should be used where it is unclear which design is more effective
- quantitative analysis after a design change has been implemented
- gathering of interviewer and coder/editor feedback (as currently takes place)

The questionnaire development process should be informed by keeping pace with developments and trends (social, economic and technological) that impact upon behaviours (expenditure and consumption). This could be done through additional user consultation and/or seeking input from consumer groups.

The process could be set up as a rolling programme which aims to cover all sections of the questionnaire over a specified number of years or the identification of priority questions and topics to review first.

(ii) Diary

The expert review concluded that the diary (both adult and child) is outdated, including the price information and images used; this is unsurprising as the current diary was introduced in 2001. The detailed topic review also pointed to improvements needed to the diary content and layout. Moreover, the expert review noted that the pocket diary requires a review to make it usable and effective.

This view was supported by interviewers and their managers who unanimously agreed on the design of the diary coming across as “cumbersome”, “overwhelming”, “daunting” or “old-fashioned”, as the instructions at the beginning are too detailed and do not reflect current shopping behaviours. The diary is seen by respondents as “a big chore”, giving interviewers the feeling they ought to apologise for it, and requiring a lot of effort from the interviewer to not only sell the diary and explain how to complete it but, in a lot of cases, also to help respondents complete it or to do it for them. Consideration should be given to simplifying the instructions and making it more user-friendly and suitable for modern-day shopping.

Various diary expenditure items were found to be difficult to record or difficult to verify from an interviewer’s perspective; for example items without weights and measures, items bought at corner shops or farmers’ markets. Interviewers thought that the diary would need to be adapted and modified to accommodate frequent expenditure (for example fuel) and to explain how to record unusual circumstances (for example free meals).

Recording items in the diary that are already recorded in the questionnaire adds to the burden, but simplifies the instructions because respondents do not have to decide what to omit. Consideration should be given to dropping the collection of these expenditure items in the diary or, alternatively, improving instructions to interviewers and respondents to make it clearer why these type of expenditure items need to be collected through both modes.

Recommendation 15: Review the adult, child and pocket diaries and update them once every 5 years (to ensure the content remains up to date). The first review (in paper form) should be implemented by April 2017.

(iii) Use of qualitative research to increase understanding of measurement error

The expert review identified a need to understand more about the cause of under-reporting in some of the topic areas as described above. Comparison of existing data sources will be essential in this process.

Recommendation 16: Carry out in-depth interviews as a follow-up survey with LCF respondents to understand how respondents go about answering questions of concern, and to assess the accuracy and completeness of diary recording.

As well as proving valuable information to inform questionnaire design, the follow-up survey would provide a measure of bias and, therefore, be beneficial to improve the survey's estimation processes. The scope of the survey would have to be well-managed, it could have multiple uses so would need to be kept focused to ensure participation.

4.3.7. Under-reporting of income: analysis carried out

The analysis conducted for this review focused on income topics that were identified as being at risk of under-reporting. These were chosen through consultation with main internal customers (the HIE branch, that is, the primary users of the income data) and by reviewing previous analysis, focused on:

- benefits – looking at means tested benefits, for example income support, pension credit and tax credits; reporting of earnings and hours
- response from high earners

Some of the following analysis is based on the Institute for Fiscal Studies (IFS) working paper "Measuring living standards with income and consumption: evidence from the UK"[39]. It is referred to as the "Brewer paper" throughout this document. The aim was to extend the data range of the paper's analysis and to broaden it by consideration of other sources. The analysis presented below should be regarded as exploratory.

(i) Comparison of income distribution by equivalised disposable income percentiles

Table 1 of the Brewer paper conducted an analysis of income, as reported in the Family Expenditure Survey/LCF and the FRS, for all financial years between 1994/95 and 2008/09. It found that:

> In 1994/95, the 2 datasets gave very similar estimates of the income distribution but, in recent years, the estimate from the LCF has been higher than that from the FRS across the distribution (page 12).

The aim was to establish whether that trend had continued in subsequent years. The EU-SILC datasets between 2005 and 2013 were used to allow comparisons to be made. The EU-SILC datasets report by calendar year, so the LCF Quarters 1 to 4 datasets for each year were employed to ensure as direct a comparison as possible for the full time series.

[39] IFS Working Paper W12/12, Mike Brewer/Cormac O'Dea, available here: www.ifs.org.uk/wps/wp1212.pdf

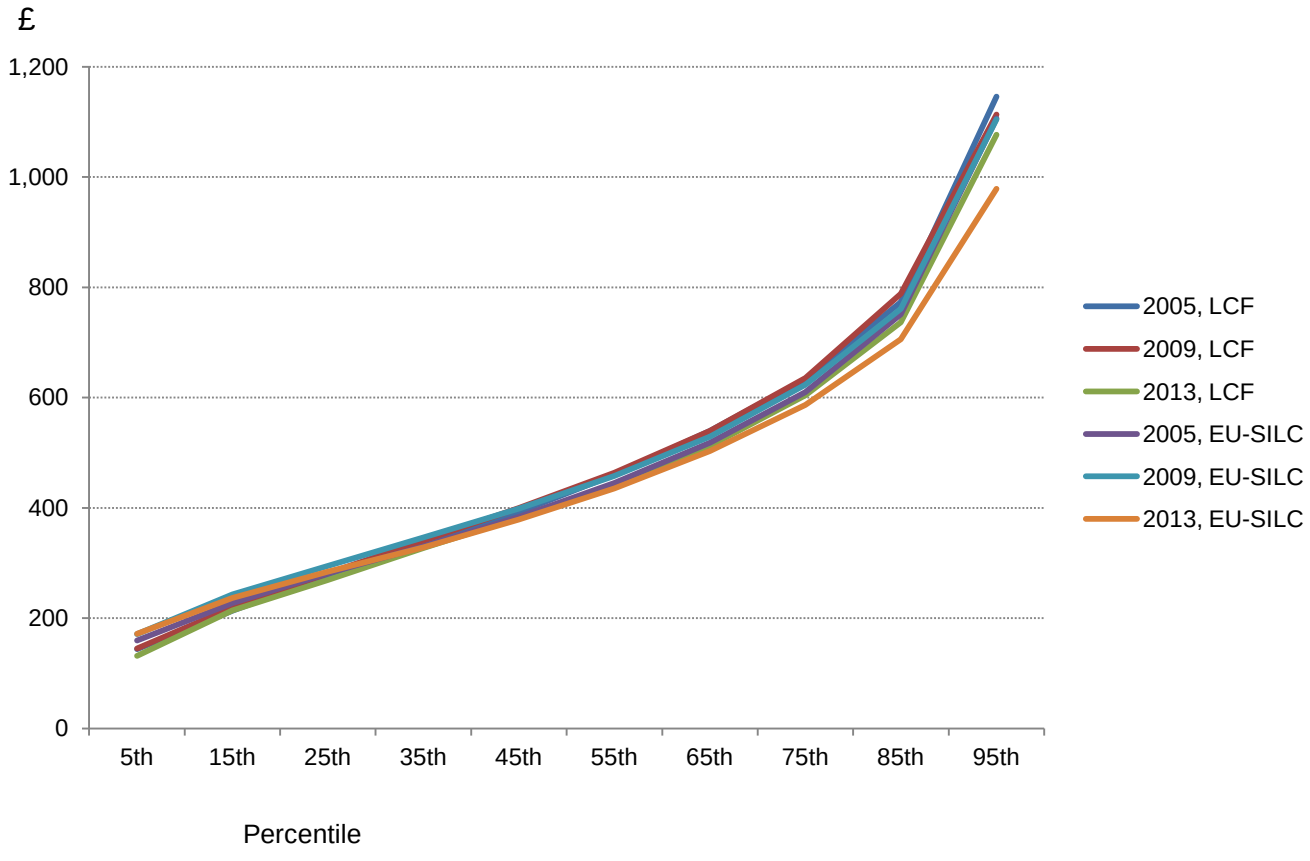
Although it was not possible to exactly recreate the figures in the Brewer paper[40], the table did provide a basis on which income distributions could be compared:

- use of the same percentiles
- the same equivalisation was applied (this adjusts for the different income requirements of households of different sizes and compositions)
- comparable deflation methods were used

The values at the percentiles were calculated for each calendar year from 2005 to 2013 using weighted data. The values were plotted as line graphs for each year and the graphs for LCF and EU-SILC datasets laid over one another for comparison. The overlaid charts show a slight but consistent pattern: LCF income values are lower than EU-SILC in the lower percentiles, broadly similar in the middle (35th to 55th percentiles) and higher in the upper percentiles (not tested for significance). In the 2013 LCF, there is a noticeable increase in average disposable income in most income decile groups when compared with 2012, whereas the majority of EU-SILC income decile groups are similar to or slightly lower than 2012 once inflation is taken into account.

The figure below shows LCF and EU-SILC data for selected years (2005, 2009 and 2013) to illustrate the differences.

Figure 4.3.7a: Overlaid charts of mean equivalised disposable income, for calendar years 2005, 2009 and 2013, LCF and EU-SILC



[40] Due to minor data and definitional differences.

(ii) Comparison of observed social security benefits claimant rates

Another comparator for LCF benefits figures is the DWP administrative data on claims. Initial analysis identified the benefits where there was the least consistency between the sources for further investigation:

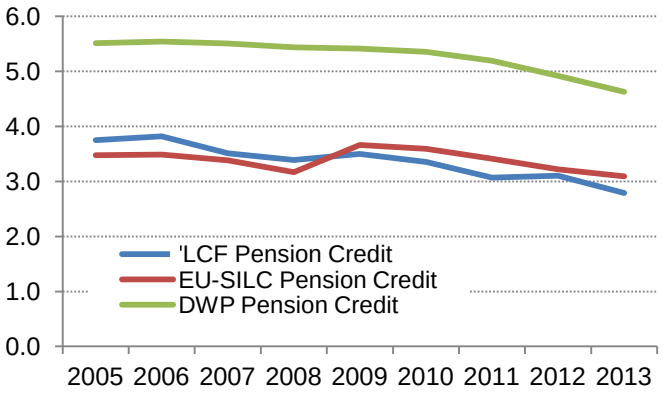
- Pension Credit
- Disability Living Allowance[41] (mobility element)
- Disability Living Allowance (care element)
- Attendance Allowance
- Employment Support Allowance

The Review Team compared the claimant rates as a proportion of the total population in both LCF and EU-SILC between 2005 and 2013. The EU-SILC datasets provided for analysis contained benefits data from the Family Resources Survey. DWP figures obtained using the tabulation tool[42] and ONS mid-year population estimates were also used to calculate the proportion of the population aged 16 years or over who are claimants. For most of the benefit types investigated, EU-SILC and LCF benefit rates have been comparable in most years, although the decline in the proportion of claimants for both types of Disability Living Allowance is at odds with EU-SILC and DWP figures[43]. In general, estimated claimant rates in both EU-SILC and LCF fall short of actual claimant rates for all benefits.

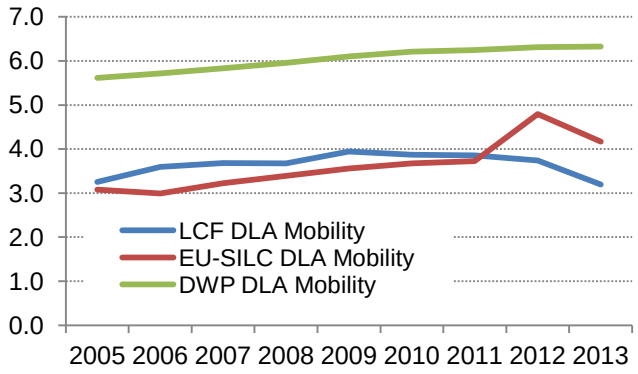
There was a notable change in the claimant rate for Attendance Allowance in EU-SILC between 2007 and 2008; unfortunately, it was not possible to establish why this is the case. The suite of charts below, showing weighted EU-SILC and LCF claimant rates together with DWP claimant rates (based on DWP data and ONS population estimates), illustrates the main findings.

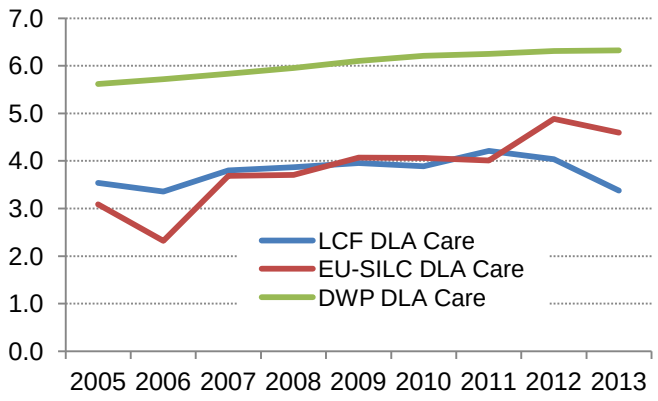
Figures 4.3.7b: Benefits claimed in LCF and EU-SILC datasets, and DWP count of benefits claimed, as % of UK population Aged 16 or over, 2005 to 2013

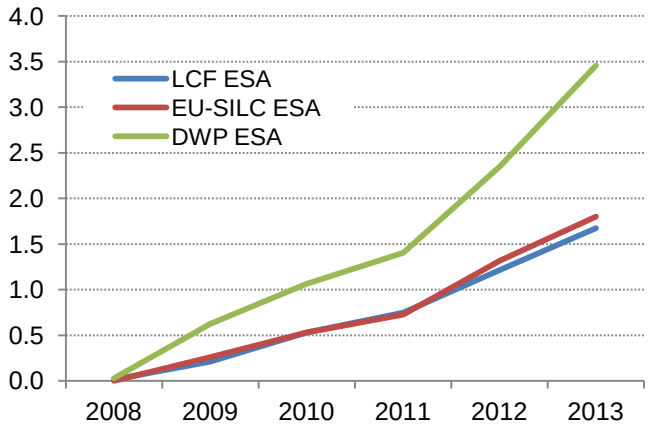
NOTE: LCF and EU-SILC data is weighted to reflect UK population (aged 16 or over). DWP claimant counts are given as a proportion According to ONS mid-year population Estimates (aged 16 or over)

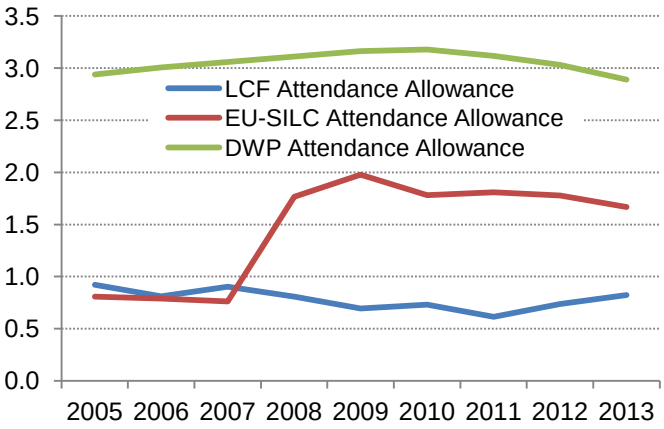


[41] Disability Living Allowance has been replaced by Personal Independence Payments for new adult claimants since June 2013, so there should be a decline in claimants for these benefits in future years.
[42] tabulation-tool.dwp.gov.uk/
[43] Disability Living Allowance has been replaced by Personal Independence Payments for new adult claimants, so there should be a decline in claimants for these benefits in future years.









It is not immediately clear why claimant rates in these surveys should fall so far behind actual numbers; further analysis could be carried out to see whether non-response, or inaccurate response, by a certain household type is having an effect. It should be noted again that these are the benefits that exhibit the greatest difference between the sources[44]. There is much closer alignment between the sources for other benefits.

(iii) Comparison of income and expenditure within LCF data

An expenditure "tick" has been observed by analysts (as it has been highlighted in the Brewer paper). This "tick" occurs when expenditure among the lowest-earning households was higher than among those with slightly higher incomes, which is a counter-intuitive outcome.

To examine recent LCF datasets, analysis focused first on expenditure by income decile. Expenditure by disposable (rather than gross) income decile group first appeared in Family Spending in 2012, so fresh calculations were made to ensure the same analysis was carried out for all datasets from 2005/06 onwards.

The pattern of mean expenditure in each disposable equivalised income decile group is broadly the same in each year, showing similar expenditure by the first and second decile groups. Then, a steady increase in expenditure up to the ninth decile group and a marked increase between the

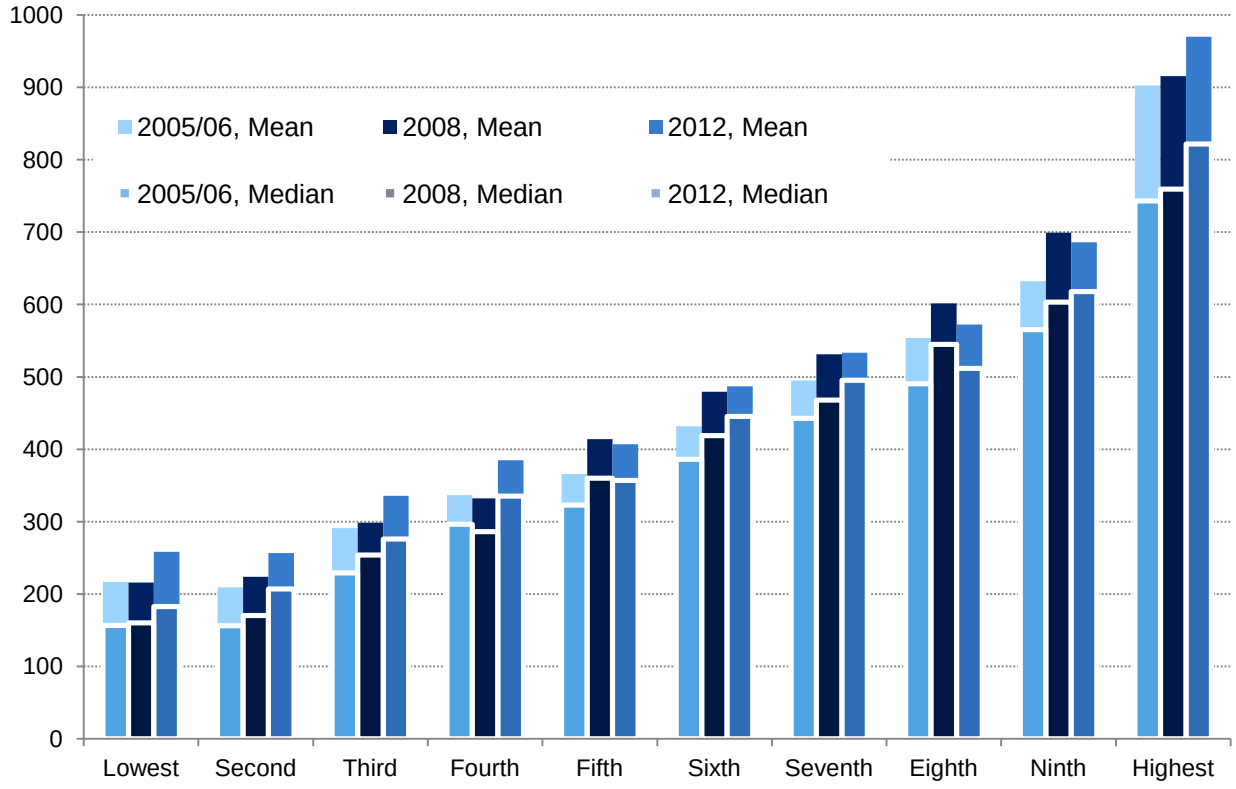
[44] ONS field and office staff working on the LCF are aware of benefit rules and they check with respondents if they think a benefit could be received.

ninth and highest group is observed. In 2013, spending was higher in the ninth income decile group compared with previous years.

Taking into consideration deciles, the expenditure "tick" is present in 2005/06, 2008 and 2012 data. In 2009, spending was identical between the first and second income decile groups.

For all years except 2005/06, there was no "tick" when using median expenditure within each income decile group. Median expenditure was lower than the mean value for all decile groups and all years analysed. The chart below presents mean and median expenditure values for selected years (2005/06, 2008 and 2012):

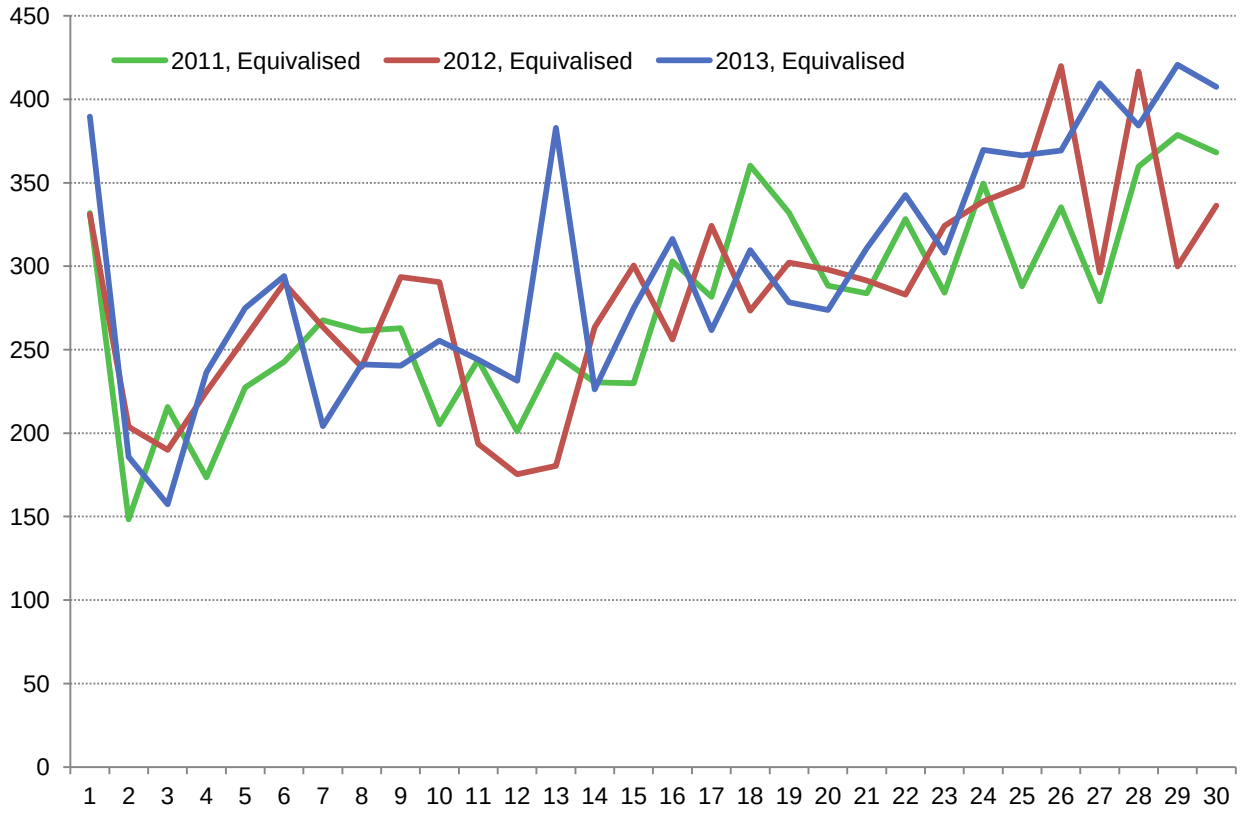
Figure 4.3.7c: Mean, median expenditure by disposable equivalised income decile group, 2005/2006, 2008 and 2012



The analysis so far has been conducted on expenditure by income deciles. Previous research suggested that the unexpected joint distributions of income and expenditure are most clearly manifested at the very lowest end of the income distribution, that is, within the lowest decile.

The next stage examined expenditure by a finer income breakdown (centile) within the lowest-income households. The aim was to tease apart the relationship between income and expenditure within these income groups. This is presented in the following figure for 2011, 2012 and 2013.

Figure 4.3.7d: Average total expenditure by equivalised income centile, UK from 2011 to 2013



This shows clearly that higher spending among lower earners is focused on the very lowest earning-households, that is those with little or no income recorded. Further analysis of the circumstances of these households should provide good indications of whether they are likely to have accrued substantial assets, which would lend credibility to the figures, or whether a more likely explanation is under-recording of income, which is likely to relate to receipt of benefits.

(iv) Analysis of people on the national minimum wage

The LCF data at person level[45] were compared with analysis presented in the Brewer paper using LCF data on the proportion of those aged 23 or over reporting implied wages less than the national minimum wage.

A new variable, pay rate, was calculated in the person-level LCF dataset by dividing the variable for usual gross pay by usual hours worked. The weighted) percentage of cases below the national minimum wage figure for that year [46] is reported in the table below. The figures did not match those in the Brewer paper, but followed a similar trend.

[45] The LCF publishes person level datasets as well as household. However, the survey's primary outputs are at household level.
[46] www.gov.uk/national-minimum-wage-rates

Table 4.3.7a: Weighted percentage of cases below the national minimum wage figures, IFS versus LCF from 2001 to 2014

Year	IFS %	LCF %	National Minimum Wage (NMW) for employees aged 23+in year* (£)
2001	5.9	3.8	3.70
2002	6.8	5	4.10
2003	6.7	4.4	4.20
2004	8	5.5	4.50
2005	7	5.5	4.85
2006	8.5	6.5	5.05
2007	8.6	6.7	5.35
2008	8.8	6.6	5.52
2009	10.5	9.1	5.73
2010	-	8.6	5.80
2011	-	7.8	5.93
2012	-	9.9	6.08
2013	-	9.3	6.19
2014	-	-	6.31

Source Brewer and Cormac O'Dea's calculation from LCF
* NMW is usually applied each October (for example October 2014), so figure given is that which applies for the majority of the year

It could also be interesting to compare the LFC data with the data collected through the ONS's Annual Survey of Hours and Earnings (ASHE).

Recommendation 17: Review the collection of income data within ONS Social Surveys to reduce discrepancies with other sources of income data.

4.3.8. Areas for further investigation

Within the time frame of the review it has not been possible to investigate all potential sources of under-reporting. This section describes topics identified during the review where further investigation would be valuable.

(i) Expenditure

Anecdotal evidence from the interviewer focus groups suggests that respondents going on holiday during the diary period are less likely to complete the diary in the time frame defined, so holiday expenditure could be under-estimated. Recording of holiday expenditure was also raised as a concern by one of the expert reviewers of this report.

Recommendation 18: Carry out more work to understand reporting of holiday expenditure.

Some of the points that could be investigated further are reported below:

- analysis of weekly interview distribution: does this reveal under-reporting during holiday periods?
- overseas spend by UK households: do comparisons of overseas expenditure by UK households from the LCF and IPS reveal a mismatch?
- what proportion of households contain absent spenders over the diary period to give an estimate of missed expenditure? – this would give evidence of the possible effect on under-reporting if spending while absent is higher than when present
- would it be possible to include a question to ascertain whether the householder plans a holiday in what would be the diary period? – this could be a good predictor of missing expenditure
- is it possible that potential respondents are less likely to take part if they will be on holiday for part of the diary period?

4.4 Web data collection

This section reports on a short study on the possibility of using a mixed mode approach for both the LCF questionnaire and the diary.

The following have been investigated:

- the questionnaire and the challenges that would be faced in using a mixed mode
- the respondent diary and the feasibility of using a web diary

Adopting a mixed mode approach can bring cost savings over time, though there will be development and set-up costs. Savings could, of course, be used to improve the quality of a survey.

All main survey design issues must be taken into account when considering moving a survey to a mixed mode design, including sampling and coverage, participation and engagement, and measurement issues. Successful transition requires good survey design, taking into account the total survey error, other quality dimensions (such as relevance, accuracy, timeliness) and cost. It is feasible that the LCF could move to a mixed mode design, but it would need significant survey re-design, potentially including the use of administrative data to reduce respondent burden. The LCF collects a lot of detailed information. Research would be needed to understand whether the quality of the information collected would be lower with a mixed mode approach (particularly if the use of administrative data was not possible). It is likely that a large scale pilot would be necessary.

The main re-design issues are discussed below.

4.4.1. Full household response

Income is collected from all household members in the LCF. Proxies are permissible, but this would be more challenging via a web model. Expenditure is collected at household level and income and expenditure is required for the same units. This limits the scope for splitting the questionnaire into shorter “person” versions. Piloting would be required to investigate what proxy

rate would result from a mixed mode design and whether respondents could work through the complex expenditure questions via a web format.
A method would need to be developed that enables all adults in the household to complete the survey online. There may be some instances where only one questionnaire is submitted from the household. It would be necessary, in these cases, to ensure that the respondent is well placed to answer questions about household circumstances (both income and expenditure).

4.4.2. Interviewer support

Respondents find many of the questions in the LCF difficult to answer. Because of this, the survey has comprehensive interviewer instructions and question-by-question help; also, show cards are used on a number of questions. Feedback would need to be sought from the interviewers about how often they have to help a respondent, and on what questions, to inform the web design. Piloting would be required to investigate whether respondents could work through the complex expenditure questions via a web format.

4.4.3. Survey length

The LFS online development work has a criterion that interview length should not exceed 30 minutes; this is based on learning from previous pilots. Whilst there are no fixed rules about this, there is a relationship between length of questionnaire and data quality due to fatigue effects. The point at which this happens varies depending on factors such as the salience of the survey topic, the study population and user-friendliness of the questionnaire. Similar conclusions can be made for the relationship between questionnaire length and response rates and break-off rates.

The LCF is a complex survey involving an interview with a median interview length of just under an hour. ONS interviewers are skilled at delivering the survey efficiently; it may take longer for respondents to complete online. As the interview length is beyond the maximum of 30 minutes recommended for web, modularisation would be required. The appropriateness of this will depend on the analysis requirements of the users of LCF data.

Interview length could be reduced if the survey were supplemented with HMRC and DWP income and benefits administrative data. Respondents would be even less likely to give financial information in a web design without an interviewer present, but a consent rate to data linkage might be higher than item non-response if financial questions were kept within the survey. This is something that would need to be tested in a pilot.

4.4.4. Open questions

There are open questions in the employment section; these questions are used to code Standard Occupational Classification (SOC)/Standard Industrial Classification (SIC). Census and the Social Surveys Transformation Programme are currently investigating the potential to use “look-up” functions for respondents to complete these questions online.

4.4.5. Respondent diary

It may be possible to develop the diary element of the LCF online even if the main survey remained as Computer Assisted Personal Interviewing (CAPI). Any mixed mode design is likely to have an

adverse effect on diary response rates and thorough testing would be needed to understand the likely magnitude of this. Under the current survey structure, the interviewer plays an important role in collecting high quality accurate data, particularly from the diary. Amongst other things, interviewers undertake the following for the diary element:

- accurately identify which members of the household should complete a diary
- secure the help of someone from outside of the household for respondents living on their own who would otherwise be incapable of completing the diary
- explain the diary to respondents prior to completion and complete a number of administrative procedures, such as ringing the correct day on each page
- leave a pair of scales with the household for accurate weight information (when required)
- make a checking call 3 days after placement
- check missing prices/ volumes/ weight information, either at the household or at the local supermarket
- reference 9 pages of clarification notes to ensure accurate information is collected
- return to collect the diary after 2 weeks and post it back to the head office
- make a judgement about whether a “diary buddy” qualifies for an incentive

Although most of the above could be covered within a web diary by written instructions, it is likely that the quality of data would be lower and less complete compared with an interviewer performing these functions. Cost savings would be limited if it was concluded that an interviewer was needed for the process of administering the diary.

If the main interview was online and the diary remained on paper then there could be some cost savings made by limiting the role of interviewers in some of these areas. Examples include:

- the diary with full written instructions could be posted to respondents immediately after the online interview rather than being placed by interviewers
- checking calls could be conducted by Telephone Operations staff rather than interviewers
- diaries could be posted back by respondents at the end of the two weeks rather than posted by interviewers

Each of these is still likely to have some adverse effect on data quality, but perhaps less of an effect compared with a fully mixed mode diary design.

There are a number of electronic data collection features that could reduce reliance on interviewers in the future, but all of these would require significant development work:

- applications could be used to code-in expenditure items – this would produce very accurate information and reduce the need for any in-house coding or checking; however, there would be costs associated with providing respondents with the necessary hardware and software and it could be burdensome to respondents
- receipts could be read in by QR code and transferred to the head office. This would need the cooperation of supermarkets but would bring, at least, some of the diary collection online
- paper receipts could potentially be scanned, calling up contents automatically for coders at ONS head office to code

4.4.6. Subject matter

For sensitive data collection, this can sometimes lead to better/more complete survey data. Web can be less suitable for the collection of data that requires considerable motivation (for example, uninteresting information) or cognitive effort (for example, difficult information). However, some development work is being done to improve motivation in web surveys (for example, using gamification, which does have other problems. Some researchers argue that questions should be re-designed so that they are not cognitively burdensome, regardless of mode. More research is required in this area.

4.4.7. Population of interest

Internet access and internet use tend to be less common among older age, lower income and lower education groups. Web survey response tends, therefore, to be biased towards younger, richer, more educated people. Currently, the use of web will under-represent older age groups (either coverage or ability) but this is likely to change over time (less than 10% of under 55 age group, but over 30% of 55 and over age group, currently do not have internet access). It will exclude the illiterate if you rely on visual stimuli only but it is possible to use audio and video (for example, avatar or even a remote interviewer who can assist if needed). The web could be useful for people with some types of disability (for example, control over print size for those with visual impairment). Those who are not computer literate could also be excluded.

4.4.8. Lessons from other surveys

Understanding Society[47] is moving to mixed mode. Methodological research has been carried out and has included the challenge of dealing with a full household. Future work on moving the LCF to mixed mode would benefit from this experience.

4.4.9. Conclusions

Although a mixed mode questionnaire/interview is technically feasible, it would be very challenging; moving to mixed mode diary would be simpler. Work by other NSIs would inform future ONS work.

Recommendation 19: Consider web diary collection for the LCF alongside other ONS web collection initiatives.

[47] Information about Understanding Society can be found here: www.iser.essex.ac.uk/understanding-society

5. Processing

This chapter looks at 2 important aspects of statistical processing: “editing and imputation” and “coding”. Although the review is mainly concerned with quality, consideration is given to efficiency of processing; this is the case, in particular, in the section on coding.

5.1 Edit and imputation strategy

Only a brief, preliminary study has been possible, so this section concentrates on an overview of the principles of processing and methods that make up the strategy and any consequent risks to survey estimates.

5.1.1. Edit strategy (error localisation)

The principles behind the design of the LCF edit strategy, which are inherently linked with main data collection principles, are sound. For example, the strategy comprises:

- a strong emphasis on avoiding the uncertainty associated with imputation through a number of strategies that push towards collecting real data in the first place
- a well-defined edit strategy covering all of the main areas expected, including hard edit rules, soft edit rules, range checks and internal consistency checks
- automated implementation at collection and beyond in software[48]
- a comprehensive documentation strategy, including issues logs, a review process, and an archiving strategy
- a dedicated editing team

There are no significant omissions; the strength of this approach means that the LCF edit strategy should form the basis of an office-wide set of standards and procedures. There are other important aspects of the approach used in the LCF which could have wider application. The use of a dedicated, centralised editing would fit naturally within a wider, service-oriented architecture and the current user-interface would serve as the basis for a more generic interface.

5.1.2. Imputation strategy

There are some good overarching principles that guide the LCF imputation strategy:

- well-defined rules describing the method to be applied to both routing and target variables, demographic and income/expenditure variables
- well-defined rules for managing changes that occur during the survey cycle in the questionnaire or in the expected range of data related to income such as benefits
- strict adherence to specified imputation methods and a policy that dictates that changes have to be considered, approved and documented
- strict policy against the removal/over-imputation of observed data unless considered and approved

[48] Blaise version 4.8 is used.

- documentation of the context associated with all discrete applications of a Clerical Fallback method (when automated edit/imputation is not satisfactory) so the same method can be applied to similar situations
- a well-defined strategy for managing period codes which can be problematic for imputation

As for the edit strategy, in a broader context, these principles should feed into the design of a comprehensive set of “best practice” criteria and the current re-design of the ONS Business Process Model for Social Surveys.

Recommendation 20: Submit the LCF edit strategy and the general principles underlying the imputation strategy into the development of the ONS Business Process Model for Social Surveys.

5.1.3. Imputation methods

Some of the variables in the LCF have consistent interdependencies and mathematical relationships between them. In general, and where appropriate, the LCF imputation strategy takes advantage of these relationships by replacing missing data deterministically based on well-defined rules and functions. In this respect, deductive imputation is a sound method for replacing missing survey data.

Where missing data cannot be based on calculations that draw on the internal consistency of the data, the LCF imputation strategy relies on 3 primary imputation methods:

- imputation using administrative sources
- group aggregate imputation
- nearest neighbour hot deck imputation

All 3 of these methods are valid. Imputation tends to be implemented manually, on a record-by-record, variable-by-variable basis. In principle, any of these methods could be automated to some extent, if not entirely and this should be explored further.

5.1.4. Imputation and survey estimates

The general aim of any imputation method is to adjust the statistical properties of a survey data base. The adjustment should serve to account for any non-response bias and improve the use of the survey data. These factors are directed at improving accuracy of estimates based on the survey data. However, selected or applied incorrectly, any of the LCF imputation methods could introduce error or bias into survey estimates through the adjustment process.
There are 2 particular areas where imputation may lead to risk to the accuracy of survey estimates:

- where there is a mismatch between the imputation method selected and the survey’s analytical aims and outputs
- where the parameters of the imputation model are specified inaccurately

(i) Risk based on a mismatch between the imputation method selected and the survey's analytical aims

"When one wants to apply imputation to improve the quality of the data, one has to be clear about what quality aspect of the data one wants to improve[49]".

The analytical aims of a survey are perhaps the most important consideration when choosing one particular imputation method over another. Imputation is fundamentally a process that adjusts the statistical properties of a survey database. However, some imputation methods have a more limited scope than others. For example, a relatively simple imputation method may be sufficient for analyses limited to a simple point estimate based only on survey means. In contrast, a more complex method is likely to be needed where an estimate of variance is also required.

(ii) Risk based on inaccurate specification of imputation model parameters

All imputation methods are based fundamentally on other survey variables that serve to predict the values or distribution of plausible values of the target variable or variables being imputed, which are the imputation model parameters. Typically, the imputation model parameters will consist of other variables from the survey that have 2 fundamental properties:

- they should account for any non-response bias identified in the data
- they should be good predictors of the target variables(s)

Poorly specified model parameters will lead to error or bias in survey estimates.

It is important to note that these 2 factors become more significant if the generic aim of the imputation strategy is to establish a unique utility dataset that can be used for a wide range of estimates based on post imputation stratification. In other words, the imputation model parameters need to be most accurate when the analytical aim of the survey is to establish a "general purpose" clean and consistent database.

5.1.5. Potential risks to the accuracy of LCF estimates through the imputation process

Taking into account the range of published LCF outputs, the 2 risks outlined above are evident. Without a more detailed analysis, it is impossible to know whether the risks are realised, and how much impact that will have on survey estimates. This can only be determined by exploring the outcome of the various imputation methods. The section below provides an outline of the main risks associated with each of the 3 imputation methods.

(i) Imputation based on external administrative algorithms, administrative look-up tables (such as social benefit calculators and council tax band indicators)

Imputation based on this type of administrative source is a particularly interesting method for replacing missing survey data as it sits outside of the "traditional" statistical framework. The general problem with this strategy is that it may yield imputed data with discrete values and/or

[49] Handbook of Statistical Data Editing and Imputation (2011), de Waal, Pannekoek, Scholtus, pp225, Wiley

statistical properties that differ from those in the observed survey data which, in turn, will have an impact on survey estimates. What makes the approach interesting is that those differences could arise for 2 different reasons.

If an administrative source is inaccurate, it is likely to lead to the introduction of error in survey estimates through the imputation process in a similar way to poorly specified imputation model parameters. Alternatively, for “fixed income variables”, such as benefits and tax bands, it may actually be more accurate than the survey data. In other social surveys, variance in survey data associated with variables of this type is likely to be error arising through the respondent’s uncertainty about the benefit amounts effectively received.

The risk to be avoided is imputing error into survey estimates through the use of inaccurate administrative sources; it is extremely important to have in place a set of well-defined criteria for evaluating the quality and accuracy of a particular source before this method for replacing missing data is applied.

One of the advantages in establishing that an administrative source yields high quality and accurate data is that it opens up the opportunity to reduce respondent burden by dropping the survey question altogether; it also may be amenable to automation. It is also important to note that if the administrative source is simply equivalent to the survey data in terms of quality and accuracy, it may be more efficient to impute the data in a traditional way rather than refer to an external administrative source on a record-by-record basis.

(ii) Imputation using group aggregates

Group aggregate imputation methods are applied to particular variables in the LCF survey set through look-up tables. Typically, the tables represent either the group mean or the group median of observed data, stratified by relatively simple imputation model parameters consisting of 1 or 2 imputation classes. It is important to note that, in principle, this is a valid imputation method, particularly when working with only a limited amount of observed data. However, it is a limited method and there are a number of risks associated with the accuracy of estimates based on the LCF data.

First, there are a number of risks associated with the analytical aims of the survey. Group aggregate imputation is designed to estimate only the central tendency of the missing data. Consequently, it will suppress the natural variance in the survey data and introduce spikes into observed distributions. If one of the analytical aims of the survey is to publish estimates of variance or the distributional properties of the data as well as point estimates, this may not be a good method.

The LCF output tables include the distributional properties of the data such as deciles. If any of the tables are linked to data imputed with these methods, the variance in the published distributions is likely to be suppressed, converging on either the median or mean. If the impact is considerable this method may need to be replaced with one that also serves to estimate the variance of the missing data.

As a method that focuses on point estimates, group aggregation can also be problematic when applied to data that is skewed. Applied to income-based data that is characterised typically by a

long tail to the right of the distribution, a group mean imputation method will pull the median towards the mean, increasing its value. Conversely, a group median imputation will pull the mean towards the median, decreasing its value. The size of this impact depends largely on how skewed the data is but any skew will have an effect. Consequently, if another analytical aim of the survey is to publish both the means and medians of a given variable, neither may be a good overall method.

LCF outputs include both distributional properties of the data and averages. Consequently, in addition to understanding how the natural variance in the data may be suppressed, it is also important to understand the impact that using a group median imputation has on the mean, and vice versa. If there is a significant shift in either of the estimates of central tendency, an alternative imputation method should be considered. Again, it is likely to be one that also serves to estimate the variance of the missing data.

A second area of risk associated with the accuracy of published estimates relates to the various imputation model parameters that underpin the imputation classes serving to tabulate the range of values used for imputation in the look-up tables. In general, the LCF look-up tables describe a very simple model of the observed data based typically on no more than 1 or 2 auxiliary variables. While simplicity is not an inherent risk, the 2 fundamental questions that should be considered are:

- do the imputation model parameters lead to an appropriate adjustment for any non-response bias that may be present in the data?
- are they the best set of predictors for the target variable?

In general, there is no doubt that poorly specified imputation model parameters will introduce bias or error into survey estimates through the imputation process. Consequently, it is extremely important that the variables contributing to the imputation model are based on an understanding of any non-response bias in the data and the identification of an accurate set of predictors. Appropriate specification of the imputation model parameters is even more important if the overarching analytical aim of the survey is to produce a utility dataset that can be used to establish a wide range of estimates using post-imputation stratification, particularly if those estimates include those derived from joint distributions between variables in the data.

The reasoning behind the imputation model parameters associated with the LCF look-up tables should be revisited. It may well be the case that the imputation model parameters have been established in an appropriate and considered way. A revisit would look to identify any non-response bias in the data and to identify a sufficiently accurate set of predictors, especially for high impact variables in the data.

As a final point in this section, the only other salient risk evident in the LCF imputation strategy is that at least one of the look-up tables is based on data drawn from another survey, ASHE. Typically, this approach is referred to as a “cold deck” imputation strategy. In general, this is a risky approach that is not usually recommended without careful consideration. Firstly, there are often differences in the distributional properties of the data from different surveys, even when the survey is attempting to measure the same thing. Consequently, this will lead to the introduction of bias into survey estimates through the imputation process. This can be seen in merged longitudinal data from the FRS and EU-SILC and have to control for it using a survey indicator in the imputation

model parameters. Conversely, if the statistical properties of the 2 surveys are the same, there does not appear to be a reason to impute based on observed data from another survey.

(iii) Nearest Neighbour Hot Deck (NNHD) imputation

As an imputation method a NNHD can be considered to be a natural extension to the group aggregate methods discussed in the previous section in that, if set up appropriately, it will serve to estimate the variance of the missing data in addition to a point estimate. In the LCF, a relatively simple NNHD is applied to impute missing diaries with a relative small set of predictive variables. This is the most interesting section of the LCF imputation strategy and probably the area that presents the highest risk of introducing bias or error into survey estimates.

What sets this aside from other applications of imputation is that the method is not explicitly targeting the statistical properties of particular variables or simple joint distributions between a few variables. Instead, the imputation is being applied to impute an entire section of the data associated with a particular individual. This is called “partial record” rather than “item level” imputation. The challenge of designing an imputation strategy here is similar to imputing for whole blocks of data missing in other social surveys, or the post-national adjustment imputation applied to the census.

Despite these similarities, imputing the diary is a unique challenge in that it represents information relative to the individual in the household that can be considered to be a cohesive entity in itself. The information in the diary is likely to consist of a raft of nested data structures, each with distinctive statistical properties. The statistical properties of total expenditure are likely to be at the top, but there are likely to be ratios and relationships between particular types of expenditure, particular retail outlets. At the individual item level, there may also be other properties in the data related to things such as regional differences in the cost of a given item and even the things that people are likely to buy. Overall, these structures combine to form what is likely to be a relatively complex whole. In principle, it becomes even more complex when it is considered that all of these internal relationships may depend on other demographic variables outside of the diary itself.

For any imputation method to account for this complexity is a quite comprehensive challenge. How is the data imputed in such a way that the statistical properties of all of these data structures, and their relationships, are estimated appropriately? The problem is further complicated in that there is likely to be variability and outliers associated with each and every way the diary data is broken up. Even with small amounts of missing data there is a risk of introducing error or bias into survey estimates.

With respect to the current imputation method, the question is whether or not it is likely to be adequate to impute in a way that leads to a utility dataset where there is minimal risk of bias in estimates that may cover a wide range of analyses? A revisit would examine this.

5.1.6. Concluding remarks on imputation methods

Within this brief review of imputation methods, it is difficult to make recommendations about baseline methods that may be more appropriate or improved parameterisations for the current approach. Consistent with the theme that has developed throughout this review of imputation, the first task is to specify clearly the range of estimates that are likely to be drawn from the diary data. Once that has been defined, some research should be conducted to understand the statistical

properties of the data associated with those analytical aims. This can be used to evaluate the impact of the current strategy or begin the process of designing an imputation strategy that may be more accurate and comprehensive. This is an interesting, but non-trivial, task.

Recommendation 21: Review and document how editing and imputation are carried out, paying particular attention to the efficacy of the diary imputation.

5.2 Coding and quality checking

This section looks at 2 important aspects of data processing. Firstly, the checking of the wider expenditure answers collected via the questionnaire by field staff; secondly, the checking and coding of expenditure data from the diary. Most attention was directed to the latter.

5.2.1. Checking of interview responses

The data collected by field staff during the interviewing stage of the survey is subject to quality checks at the point of capture and further checking is carried out by specialist staff at ONS head office. Questions are referred back to the field worker and the household if required.

The checking of the interview answers is extensive and comprehensive. This process can take many hours as the number of questions is substantial. Staff are highly skilled and spot subtle inconsistencies in responses. They have a variety of reference materials which they consult; for example, in checking the value of benefits. In some cases, the household will be re-contacted to clarify information. There is a strong commitment to ensure that the data is as accurate as possible.

5.2.2. Capture of diary information

For the diary, field staff provide guidance to the households on what is required at the start of the diary period. They return to the household mid-way through the 2-week diary period to review the use of the diary and to provide further guidance. The diaries are posted to ONS head office when the survey period finishes. The diaries are reviewed by specialist staff; entries in the diaries are checked for completeness and are manually coded to expenditure categories.

The diaries are paper booklets with supermarket receipts stapled onto the pages; other expenditure items are handwritten. In most cases, item sizes are written onto the receipts where they are not already present. For items such as restaurant bills, respondents are asked to specify details; if these are not present, the coding staff will try to identify typical items from the restaurant menu if it is available online.

The staff code items to COICOP categories manually by using the descriptions on the receipts and written in the booklets; the coding is captured in their IT systems. Supermarket receipts display very short summaries of the item purchased and some include detailed product codes. In some cases, it is hard to identify the items from these descriptions though coding staff recognise many of them. Where sizes and weights are missing, the staff will try to identify the values from the description and price. Staff are highly committed to recording the information as accurately as possible. The processing is manual and diaries can take 8 hours to code; however, this results in

high quality outputs. The volume of diaries to be processed does occasionally lead to staff having to work additional hours.

5.2.3. The diary process

For the purposes of this review, the diary processing can be summarised in the following steps:

- general checks are made to ensure that the diary has been filled in correctly; these are in addition to the checking carried out by the interviewer in the household
- many receipt entries do not show the weights of items bought; the interviewer will ensure that, in most cases, this information is added and the coder will check that this information is present and correct; if not, they will add it
- data is then entered into an IT system by the coder by:
    - selecting a receipt entry and interpreting the description
    - searching for the relevant term on a look-up list selecting expenditure classification
    - entering price and weight

There is a skill in interpreting the receipts and knowing what to search for to assign an expenditure code. In many cases, the item description for a product is very brief and it is the experience of coding staff that enables them to process the entries quickly and accurately.

Some supermarkets (for example, Asda and Iceland) print unique identifiers on receipts which allow coders to search for these products online.

The coding team are highly effective; their knowledge is extensive with a clear commitment to high quality work.

5.2.4. Access to additional data for process improvement

Expenditure classification is not a simple task; it is time-consuming and requires a detailed and evolving knowledge of supermarket receipts and the COICOP classification. Although the process is highly manual, it does ensure a high quality outcome. A crucial question is whether the process can be partially automated without compromising quality; for example:

- can information from the diary be entered automatically? As supermarket receipts are attached to the food diary, can these receipts be scanned and the items extracted?
- can the classification of receipt entries be automated?

As part of the review, members of the Review Team discussed expenditure data capture and processing with Kantar Worldpanel, a worldwide market research company that captures expenditure information from households for commercial purposes[50]. They capture individual purchases by scanning the product barcodes; households are provided with barcode scanning devices. As most supermarkets are customers of Kantar, they provide the reference information that links product barcodes with product descriptions.

[50] See: www.kantarworldpanel.com

ONS could provide households with barcode reading devices; however, professional and robust barcode reading devices would require a significant investment together with the technology to process the information. Smartphones can also scan barcodes though this would only be suitable for very low volume product processing.

For the processing of supermarket receipts, scanning and extraction of individual product purchases would be a relatively straightforward matter. However, in most cases, the weight of the product would not be part of the description and would need to be added manually. The allocation of an expenditure code could be partly automated from the product description with manual intervention where needed. The automation of some of this process could be helped if ONS gained access to supermarket datasets listing products; this might be obtained either from market research companies or directly from supermarkets.

Recommendation 22: Discuss potential benefits from and access to supermarket product information with market research companies and supermarkets.

The question of whether a paper diary instrument is still the best way to capture expenditure is considered in Chapter 4. Whether households should be asked for their expenditure at all is considered in Chapter 7.

5.2.5. New technology for process improvement

The interest in alternative data sources for producing prices indices has driven a wider interest in the automatic coding of consumer transactions and price data. ONS has been running a project to collect scraped price data from several supermarket websites and to explore methods for producing price indices from this data. Experimental price indices have been published[51].

The Review Team discussed the possibility of using machine learning technology to gradually improve the automated allocation of accurate expenditure codes to capture item purchases with the ONS Big Data team and with the Prices Division. They have an interest in automatic coding to assist with the processing of web-scraped supermarket data; this is discussed in Chapter 7. Machine learning could be used to categorise items from a till receipt. This would not replace coding staff. They would be needed to check that the auto-coding is correct and, if not, to apply the correct classification; this would gradually improve the training of the algorithms.

Success may be limited, at least initially, given the relatively small amount of information available on a receipt. It may speed up the process as the details entered by coders would be reduced. It could offer suggestions for classification which would “self-train” a system.

ONS currently employ machine learning techniques in the classification of web scraped price information from supermarket websites. This approach is applied to a text field utilising a Support Vector Machine (SVM) which is a supervised learning technique[52]. This means that given a dataset

[51] See: webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/research-indices-using-web-scraped-price-data/index.html
[52] en.wikipedia.org/wiki/Support_vector_machine

with classifications, new items may be classified by identifying patterns within the data. The process is iterative with each new influx of data and, if done correctly, will improve the quality of classification over time. Recently, a method to identify mis-classifications using only the price data has been developed within ONS so that the supervision of the process may be partly automated.

Recommendation 23: Explore the possibility of semi-automated coding of purchase information from scanned supermarket receipts.

5.2.6. Cost of change

Moving to a degree of automation would incur set-up and running costs and estimates of these costs would need to be assessed against the current costs to see whether there would be longer-term financial benefits. As noted above, skilled staff will still be needed to operate and intervene to ensure high quality is maintained; however, benefits could derive from more efficient processing.

5.2.7. Expenditure weights from alternative sources

Chapter 7 considers whether data from market research companies or from supermarkets could provide expenditure weights of sufficient quality.

6. International comparisons

6.1 Introduction

The methods used to conduct expenditure surveys have been subject to extensive research both in the UK and internationally. Many countries contend with similar challenges to the UK in terms of trends in response and concerns about under-reporting. Some are exploring the opportunities presented by technological advances, for example in web-based data collection. This chapter reports some of the main findings of this research and considers the potential implications for any future developments to the LCF. Main survey metrics (sample and response) experienced by other countries are considered to give context to the position in the UK.

This chapter first presents topics by theme and then outlines some of the major programmes carried out recently by NSIs. Information has been collated via literature searches and consultation directly with representatives of other countries' NSIs. The Czech Republic statistics office convened a workshop in June 2015 to gather information from representatives of EU countries on their expenditure surveys, both current methods and potential developments; this proved a very useful source of information for this review.

6.2 Sample and response

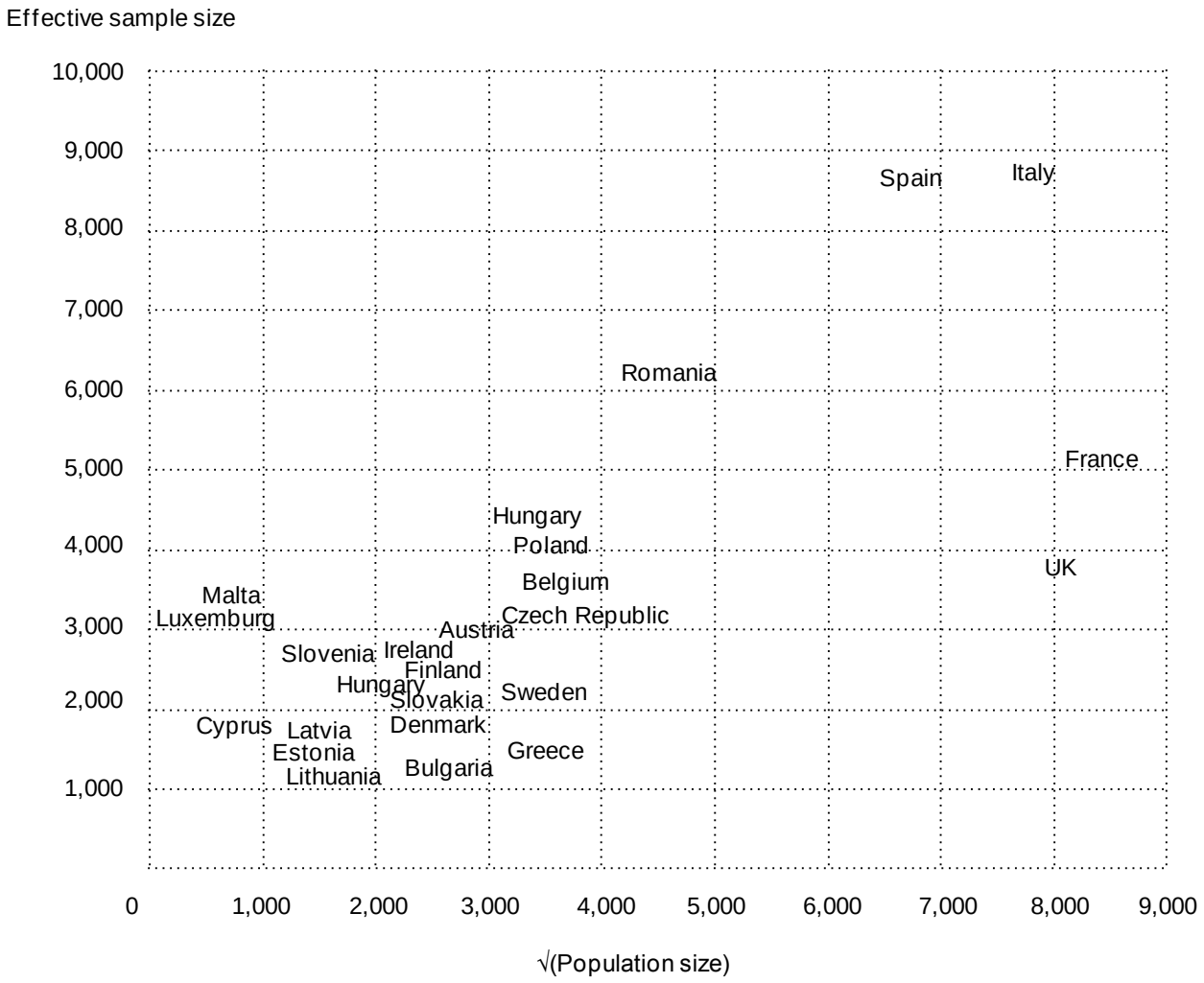
This section illustrates examples of sample sizes and response rates in other countries' expenditure surveys. The methodological implications of the UK's sample and achieved response and how they compare to those of other countries are considered also in another section of the review (see Section 2.5: "Response rates in context").

The LCF's achieved sample size (5,263) was also close to the Eurostat Household Budget Survey median (4,594). The survey's effective sample size, which is the sample equivalence if simple random sampling was used, as calculated by Eurostat, is 3,865, higher than the median of 2,600.

Sampling works in such a way that the sample size for a survey will be approximately proportional to the square root of the population size for the same accuracy. Another way of putting the UK sample size in context is comparing against the square root of the population size. Figure 6.2.1 shows the effective survey sample size plotted against the square root of population taken from the Eurostat Household Budget Survey[53].

[53] Germany and Poland are excluded; Germany has quota sampling. The Netherlands is also excluded as no data is provided in the HBS quality report. europa.eu/about-eu/facts-figures/living/index_en.htm

Figure 6.2.1: Household budget survey effective sample size (2010, calculated by Eurostat) against square root of population size (2015) for EU member states (excludes Germany, Poland and Netherlands)



Source: Eurostat

The surveys' precisions are partly a function of population sizes and it should be considered that the UK's population is greater than the EU average. The achieved confidence intervals for a key estimate are useful as a direct measure of precision. The confidence interval (as computed by Eurostat) for total consumption expenditure in the LCF is ±2.6%, slightly higher than the median value of ±2.2%.

The LCF's sample size is smaller than some other major expenditure surveys outside the EU. The current US design uses separate samples for interview and diary collection. A total of 15,000 addresses are issued annually for the interview, achieving an overall response of around 73%. Responding households participate over 5 consecutive quarters. 12,000 addresses are issued annually for the diary with a response rate of around 71% (the US design is currently under review; see Section 6.5.1 below). The Australian expenditure survey achieves 10,000 cases annually with response reported at 72%, while Canada achieves 18,000 cases with 67% response to the interview.

6.3 Data collection methods

6.3.1. Diary versus recall

Almost all national expenditure surveys use diaries to capture some types of expenditure. Recent work has scrutinised the efficacy of diary-based data collection compared with recall in the questionnaire.

The current US survey has separate diary and recall surveys. Bee, Meyer and Sullivan (2015)[54] reported a series of analyses that seem to suggest that the recall data is of better quality. The debate over diary versus recall data collection was the source of the dissent to the National Academy of Science report in the US. Essentially, the main panel thought the only way forward was to try to improve diary collection with technology, while a dissenting minority questions whether the evidence supported a continued reliance on diaries.

The Canadian survey, which until recently was based on recall, had unusually good coverage of expenditure relative to that recorded in National Accounts[55], compared with the UK, US and Australia (all of which use diaries extensively). The Canadian survey introduced diaries in 2009.

Battistin et al. (2016) suggest that diaries are better for regular purchases though the position is less clear for larger purchases[56]. There is evidence that diaries suffer under-reporting through "diary fatigue" (fewer expenditures are recorded later in the period covered by the diary). This pattern has been reported for the US (Silberstein and Scott, 1991[57]; Stephens, 2003[58]), the Canadian Food Expenditure Survey (Statistics Canada, 1999; Ahmed et al., 2010[59]) and the Family Expenditure Survey (Tanner, 1998[60]). Research suggests that what disappears in the second week are smaller purchases. Analysis of recent LCF diary purchases over the diary period has been presented in Chapter 4 (Data Collection). A possible explanation is that habits may be altered by recording of expenditure.

Recalled expenditure is also subject to error. This can occur when respondents fail to recall purchases or recall the amounts inaccurately. Conversely, respondents can falsely place purchases within the recall period when they actually occurred longer ago (known as "telescoping").

---

[54] The validity of consumption data: Are the consumer expenditure interview and diary surveys informative?, available here: www.nber.org/papers/w18308

[55] Barret, G., Levall, P., and Milligan, K. A Comparison of Micro and Macro Expenditure Measures across Countries Using Differing Survey Methods. Pp 263-286 in Carroll C, Crossley TF, Sabelhaus J, eds. 2015. *Improving the Measurement of Consumer Expenditures*. NBER Ser. Stud. Income Wealth. Cambridge, MA: Natl. Bur. Econ. Res.

[56] Battistin et al (2016). Journal of the Royal Statistical Society A, Vol 179, Issue 2, p559-581

[57] Silberstein, A.R., and Scott, S. (1991). Expenditure Diary Surveys and their Associated Errors, in Biermer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., and Sudman, S. Editors, *Measurement Errors in Surveys*, Wiley, Hoboken NJ: 1991.

[58] Stephens, M. (2003): '3rd of tha Month': Do Social Security Recipients Smooth Consumption Between Checks? *American Economic Review* 93(1), 406-22, available here: www.nber.org/papers/w9135.pdf

[59] Ahmed, N., Brzozowski, M., and Crossley, T. F. (2010). Measurement errors in recall food consumption data. Institute for Fiscal Studies Working Paper 06/21, London.

[60] Tanner, S. (1998). How much do consumers spend? Comparing the FES and National Accounts. In *How Reliable Is the Family Expenditure Survey?*, ed. J Banks, P Johnson, pp. 67-121 London: Inst. Fisc. Stud.

Diaries are currently used by all EU countries. At the HBS conference in Prague, in June 2015, no delegates reported intentions to abandon diary collection.

In summary, research has identified short-comings in the use of diaries to record expenditure; however, the general opinion, on balance, favours their retention. All major national surveys continue to use diaries. In practical terms, it would be very difficult to capture expenditure via questionnaires in the UK at the level of granularity required by Prices and National Accounts customers. There is scope for the design of the diary to be updated and for changes to the model of diary data collection to be improved.

6.3.2. Web data collection

Web data collection has been explored and developed by a number of NSIs. The ONS perspective on the topic has been considered in Chapter 4. Presented here are examples of web-based data collection developed and considered by other NSIs.

Web data collection is used by some countries for the questionnaire component of their expenditure surveys. An example is the Netherlands, where a very low response rate (16%) is accepted. Web-based questionnaires are also used in Germany; however, this is currently in conjunction with quota sampling though there is interest in adopting random sampling in the future. Austria uses a combination of CAPI and web-based interviewing. It would be useful to understand more about the complexity and quality of the data captured via these countries' web-based questionnaires.

Web collection is also used for diaries by the countries (cited above) where web collection is used for questionnaires. Other NSIs have considered using web diary collection to complement face-to-face interviewing. For example, the US is piloting online diary collection as part of its redeveloped methodology (see Section 6.4.1). It should be noted that this approach was adopted against the recommendations of the review panel for the US Consumer Expenditure Survey (CES); the panel recommended the use of a tablet-based instrument instead.

In summary, current examples suggest that web-based questionnaires demand compromises in terms of the quality of the data collected, response rates and departure from random sampling. Web diary collection appears more promising, especially when considered in the context of technological advances.

Recommendation 24: Liaise with other countries and organisations undertaking web data collection, receipt scanning and automatic coding to understand more about the data quality obtained. If the quality is satisfactory, undertake experiments to evaluate these approaches in the LCF.

6.3.3. Scanning and coding of receipts

Paper receipts can be scanned automatically and coded. The Swedish statistical office has developed this technology, but reports few overall gains in efficiency to date. The scanning technology must be formatted to retailer-specific layouts and these frequently need changing. It is hoped that technological advances will enable greater efficiency to be achieved in the future. The current technology is most effective in countries with fewer major retailers than the UK.

Purchase information can also be coded automatically once recorded electronically. In Sweden, scanned receipt information is then sent through an automatic coding process, which is successful in 47% of cases. Automatic scanning and coding is also used in Austria and Finland. Finland also utilise a "fuzzy" matching algorithm to enable automatic coding where precise matching is not achieved.

Scanning technology and automatic coding are used by ONS for its business surveys and it could be useful to draw on this experience to develop the use of these techniques for the LCF.

There was interest among delegates at the June 2015 HBS conference in Prague in the possibility of scanning bar codes or Q codes on receipts to call up all the receipt information automatically. This seems a very promising avenue; once the information is available, it would be more efficient than scanning the content of receipts. This depends on retailers making the information from individual receipts available.

6.3.4. Administrative data

Linkage of record-level administrative data is carried out by Canada and several EU countries (for example Sweden, Finland and Slovenia), obviating the need for collection of income information in the questionnaire. These systems rely on Personal Identification Numbers (PINs) linked to population registers to link survey and administrative records.

ONS Social Surveys Division is involved with programmes aimed at obtaining and utilising administrative data from other government departments. These programmes could well be effective and, in the longer-term, might enable reduced collection of variables by the survey. In the shorter term, administrative data could be useful for improving sampling and for quality assurance.

Recommendation 25: Engage with ONS programmes to enhance the use of administrative data on income, though it should be acknowledged that record-level linkage is likely to be a long-term aspiration.

6.3.5. Access to commercial data

Commercial sales information could potentially have a role in quality assurance. If, later, it could be linked to households, then it might have a more direct use. This would demand the consent both of households and commercial organisations. If this were achieved, it would facilitate, for example, collection of data over a longer period for participating households. This might enable data collection during atypical periods of the year, such as Christmas. Use of commercial transaction data is being developed in Norway (see Section 6.5.2 below) and it will be useful to see how this initiative progresses.

6.4 Investigations into under-reporting

Analyses conducted by ONS in 2012 highlighted the growing gap between expenditure recorded in the LCF and the National Accounts. Similar conclusions were reached by Barrett et al. (2015)[23], who reported that the ratio of expenditure captured by the LCF to household expenditure in the National Accounts has fallen very significantly (from about 90% in the mid-1970s to less than 70%

to date). This performance is arguably the least successful of the Anglosphere countries (the US has had a broadly similar decline; coverage in Canada and Australia seems to have held up better – again see Barrett et al., 2015[61]). While there are differences between the LCF and the National Accounts in both the population and categories of goods covered, it is hard to think that there is a clear enough trend in any of these differences to explain the decline. The decline in coverage has happened even in categories of expenditure where the LCF is the main source of information in the National Accounts; notably food, where coverage seems to have fallen by at least 10 percentage points.

6.5 International expenditure survey development programmes

Numerous countries are reviewing aspects of their expenditure surveys. Some are carrying out more major reviews and fundamental changes in methods. Current designs are quite aged; shopping patterns and technology are changing fast, data collection technology is also changing rapidly and the quality of expenditure surveys is declining in many countries. This section documents 3 large-scale current reviews.

6.5.1. US Gemini project

The US Bureau of Labour Statistics (BLS) launched a major review into their CE survey in 2009; this was named the "Gemini Project"[62]. The Gemini Project aimed to re-design the survey to achieve a verifiable reduction in measurement error while improving or maintaining response rates, without increasing costs.

The review considered at length all aspects of data collection. A main issue was the merits of maintaining an expenditure diary as opposed to using recall to capture all expenditure types. Views varied on this topic among the US panel overseeing the Gemini Project; the majority favoured maintaining diary data collection, but a dissenting minority advocated discontinuing diaries.

The Gemini Project settled on a field model for a pilot. The pilot was in the field from July to October 2015. The diary has been conducted online by default, with paper versions available on request. The results from the field test are currently being analysed and will facilitate a decision on how to move forward with the re-design project. ONS will be keen to see the results of this pilot when they are published. The current US model uses separate samples for diaries and questionnaires, so the piloted method represents a significant change in approach by using a single sample for both elements of the survey.

Visit 1:

- first interview, recalled expenditure (3 month reference period), mainly large items such as home furnishing and appliances
- briefing to collate relevant documentation for the second interview
- briefing on diary-keeping procedures

[61] Barrett G, Levell P, Milligan K (2015). A comparison of micro and macro expenditure measures across countries using differing survey methods. See Carroll et al. 2015.
[62] The project's website can be accessed via the following link: www.bls.gov/cex/geminiproject.htm.

Diary period (1 week)
This utilised an electronic diary.

Visit 2 (one of days 9 to 18):
Review of diary and expenditures; interviewer reviewed diary with respondent and asked follow-up recall questions. Second interview, capturing expenditure on items including housing costs, vehicles and insurances. Income was also recorded.

6.5.2. Norway

The Norwegian statistics office plans to implement a multi-mode data collection model in 2017. This will involve card purchase data being transferred directly from the purchase point to the statistical office. This requires both the consent of the householder and the co-operation of the retailer. Currently, only one of five major retailers is prepared to co-operate, while research suggests that around 40% of respondents would consent. The survey model also utilises web-based questionnaires and diary functions. Purchases recorded via retailers will automatically populate the web diaries, and respondents can fill in the gaps appropriately. This seems to be a strength of the proposal, because the reduction in burden associated with the automatic population of the diary can be enhanced gradually as more retailers join the scheme. This seems a very promising direction and developments should be monitored closely.

In the longer-term, Statistics Norway is considering using transaction data as a sampling frame for data collection. This would mean a very different approach to sampling: selecting households for which transaction data were available and then approaching these households for interviews. Such an approach intuitively would lead to bias, but could be effective if this were accounted for via weighting or by combining with other sources.

6.5.3. Czech Republic

The Czech Republic statistics office is re-designing its expenditure survey, and convened a workshop to which representatives of EU member states were invited. Representatives of 14 member states attended. The workshop has been a useful source of information about methods and developments in other countries and a summary of current methods and developments is being prepared. This should provide a useful framework for networking and information sharing in the future. The Czech Republic statistics office is seeking to move from quota to random sampling and to adopt a mixed questionnaire and diary approach to data collection.

Recommendation 26: Maintain links and promote information sharing with other NSIs and organisations (particularly the Understanding Society team at Essex[63]) with respect to the development of expenditure surveys. In particular, progress by the US and Norwegian statistical offices should be monitored.

The use of supermarket transactions data by NSIs is discussed in section 7.2.2.

[63] More information about Understanding Society can be found at: www.iser.essex.ac.uk/understanding-society

7. Future developments

The LCF follows the traditional approach to gathering information on expenditure through a household survey. This method of collecting expenditure information is the standard international approach; it is well-established, well-understood and has been specifically designed and run for statistical purposes. However, ONS is not the only organisation in the UK collecting expenditure information from households; market research companies also carry out surveys for commercial purposes. There is also a potential source of expenditure information in data within retail organisations – store point of sale scanner data.

This chapter looks at these 2 alternative sources of expenditure data – home scanner surveys carried out by market research organisations and store scanner data. It also considers whether they could be either an alternative to the LCF or a means to supplement it.

7.1 Market research data

Market research companies run surveys of consumer behaviour in the UK. Kantar run a number of continuous consumer panels covering a variety of consumer markets including food, fashion and telecommunications. Their continuous panel surveys run in a number of countries; in the UK, this consists of a sample of 30,000 households. Participants use home barcode scanners to record their purchases; these are matched with prices, stores and information about the household. This data is used for a variety of purposes, including measuring market information, such as supermarket market shares and brand performance[64].

There are important methodological differences between the LCF and the Kantar continuous panel surveys:

- the LCF uses a stratified random sample while the Kantar panels use quota sampling with weighting adjustments to adjust the panel composition to known demographic controls
- the LCF uses a combination of an interviewer-led questionnaire with a 2-week food diary while the Kantar panels use home barcode scanners and households participate for as long as they want with some telephone capture of household information
- the LCF food diary records purchased items as supermarket receipts or written descriptions, the Kantar data consists of captured product barcodes which are matched to specific items through data provided by supermarkets[65]

7.1.1. Research uses

Beyond the commercial purposes for the market research information, this data has been used by researchers for a wide range of purposes. In particular, the longitudinal nature of the survey makes it well suited to consumer behaviour research and many academic institutions have made good use of the data. ONS has worked with a number of datasets from Kantar and has used them to

[64] See: www.kantarworldpanel.com/en
[65] Leicester, A. (2012).[65] How might in-home scanner technology be used in budget surveys?, IFS Working Paper WP12/01. Also available in Improving the Measurement of Consumer Expenditures (2015) NBER, Studies in Income and Wealth 74, edited by Carroll, C. D., Crossley, T. F. and Sabelhaus, J.

investigate a range of research questions including the degree to which consumers substitute similar products in response to relative price change[66] and the prevalence of discounts for grocery items[67]. This data has proved a useful source of information to ONS for exploring some of the assumptions inherent in the current approach to calculating consumer price indices. The considerable extent to which this data is used in academic research is an indication of its value.

7.1.2. Potential use in official statistics

The potential use of market research data in official statistics has been explored by a number of researchers; this includes supplementing survey-based expenditure information and partial replacement. The work of Leicester[68] reviews this research and reports new work. An important precursor to considering using this data in production use is an understanding of the range of items covered by the market research data and how the differences in methodology affect the derived expenditure.

How do the expenditures captured by market research companies compare to the LCF? The Worldpanel data does not cover the whole of the COICOP classification; Leicester estimates that it covers about 15% of the requirements of the CPI. The datasets acquired by ONS for research show good coverage of food and drink and partial coverage of other COICOP categories; the panel data would cover most of the items recorded in the LCF diary but not all; for example, restaurant meals are not captured.

For categories in common between the LCF and Kantar data, Leicester reports that both UK and US studies show lower average weekly spending on food and alcohol in the Kantar datasets by about 20%. The reasons for the differences are the subject of ongoing research. A number of possibilities have been suggested for the difference, including households forgetting to record all top-up shops in the Kantar homescan data. From a methodological viewpoint there are many differences between the LCF and Kantar data, from the sampling to the mode of capture and multiple factors are likely to contribute to the differences.

ONS has made similar evaluations by comparing panel data from Kantar with LCF data; the figure below shows a consistent, higher level of total expenditure on food items for the LCF data compared with the panel data.

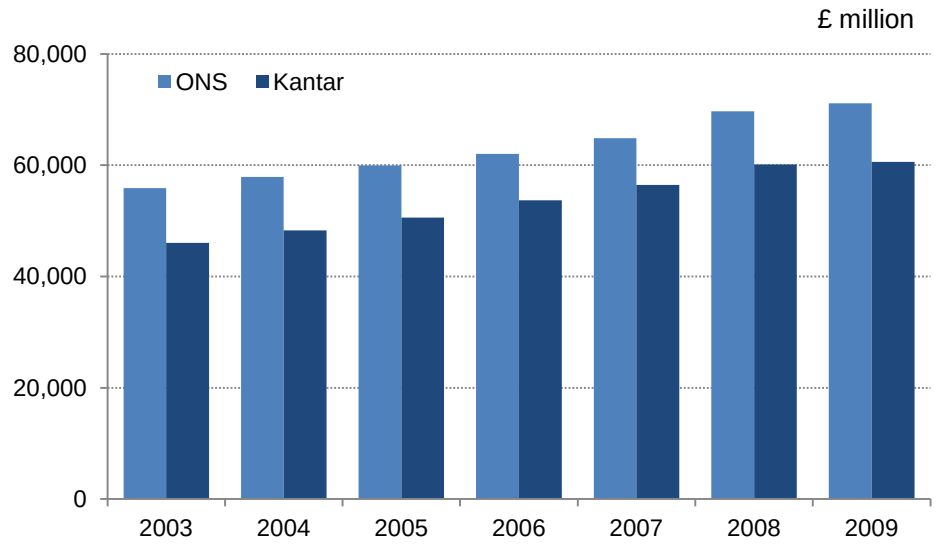
---

[66] Elliott D, O'Neill R (2012). Estimating the elasticity of substitution for alcohol products, ONS Survey Methodology Bulletin No 71, available here: webarchive.nationalarchives.gov.uk/20160204094749/http://ons.gov.uk/ons/guide-method/method-quality/survey-methodology-bulletin/smb-71/index.html

[67] See: Initial investigations into the prevalence of multi-buy discounts and the effect of discount prices on price indices (2014), available here: webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/index.html

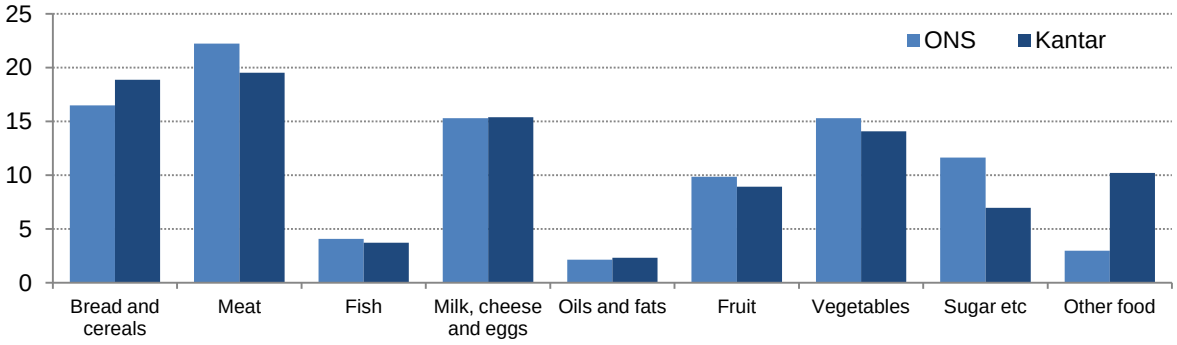
[68] Leicester, A. (2012). How might in-home scanner technology be used in budget surveys?, IFS Working Paper WP12/01. Also available in Improving the Measurement of Consumer Expenditures (2015) NBER, Studies in Income and Wealth 74, edited by Carroll, C. D., Crossley, T. F., and Sabelhaus, J.

Figure 7.1.2a: Total expenditure, Food, UK, 2003 to 2009



A different comparison can be made by looking at the relative expenditure on different categories or commodity groups; this shows similar patterns of expenditure. ONS has also made comparisons of expenditure patterns at a high level within the food and drinks category from Kantar data with LCF, which also show reasonable agreement for most categories.

Figure 7.1.2b: Expenditure share, Food, UK, 2009



The larger differences for the last 2 categories, “Sugar and confectionery products” and “Other food” are likely to be the result of differences in coding which further analysis would resolve. It may also offset some of the other differences shown.

Leicester reports on research studies which compare panel and expenditure survey results and aggregated expenditure estimates against National Accounts expenditures[2]. Both expenditure survey and panel aggregated expenditures are lower than National Accounts values, with expenditure survey values being closer. From the current research findings, partial replacement of the LCF with panel expenditures is not recommended. There is, however, potential for using market research panel data to produce expenditure shares at a lower level of detail than is obtained from the LCF. It would be necessary to ensure that, when summing the detailed

expenditures from the market research data, LCF values are matched. This would be a way of combining traditional survey data with market research data and retaining consistency.

Alternatively, Kantar run specific commodity panels, so it is possible that, if ONS required expenditure information at a very detailed level for certain categories, a market research company could run a specific panel to deliver this. This might provide a cost-effective way to collect additional breakdowns.

Recommendation 27: Explore the differences between Kantar and LCF expenditure data, comparing expenditures at multiple levels of disaggregation. Evaluate the accuracy of panel data constrained to higher level LCF expenditure values.

Recommendation 28: Investigate the practicality of a market research company running a specific panel to provide ONS with additional expenditure information to supplement LCF data.

7.2 Supermarket scanner data

Another source of expenditure information is point of sale scanner datasets from retail organisations. Products are identified at the till by the scanning of the product barcode; this identifies the item and the price; it also records the sale and is used for stock control.

7.2.1. Benefits and challenges

The potential benefits of using this type of data for statistical purposes were identified over 20 years ago; its use was recommended in the influential Boskin Report in the US in 1996 as a way of improving consumer price indices[69]. The potential benefits include a better treatment of goods with volatile prices, better modelling of goods with complex attributes, a better understanding of consumer behaviour, improved quality adjustment and more detailed data on consumer expenditure. For expenditure information, these datasets could deliver information at very low levels of detail. This would apply for particular stores only though, if the major food retailers would be willing to share this data, it would cover a significant proportion of consumer spending in the UK.

There are, of course, challenges that come with this type of data. It is the property of private companies who may not be willing to provide it; the volumes of data can be very large which has implications for IT systems and software to analyse it; information about the customer is not associated with the transactions (though this might be available through associated loyalty card schemes); it is restricted to a narrow range of commodities and retail outlets; it would not cover the full range of retail outlets that households buy from and would be captured in a consumer expenditure survey.

For deriving expenditure information from transaction datasets, the coding of transactions to expenditure categories would, of course, need to be automated. A very significant amount of effort

[69] Boskin M J et al. (1996). Final Report of the Advisory Committee to Study the Consumer Price Index, Washington DC, US Government Printing Office.

would be required to process the transaction records. While the detail of individual transactions would be interesting for research purposes, this level of detail would not be necessary for production or expenditure share needs. It may be possible to obtain the data with some partial pre-processing carried out by the retailer.

7.2.2. Progress with store scanner data

Many NSIs have explored the use of this source of data and a substantial body of research now exists; it continues to be one of the main research topics in price statistics[70]. A number of NSIs have been using this data in their CPIs for some years, including the Netherlands, Norway, Sweden, Switzerland, Australia and New Zealand[71,72,73,74].

NSIs have used the data in different ways. For example, Statistics New Zealand has used scanner data in their CPI to measure price change of 12 consumer electronics products. In this case, the store scanner data contains product characteristics information and is obtained from a market research company, who gets it from retailers[75]. The Australian Bureau of Statistics has a multi-stage development programme that includes the improvement of expenditure information[76].

7.2.3. Improved expenditure data

The quality of the CPI construction could be improved with the use of expenditure information from store scanner data. Timeliness of expenditure information could be improved significantly; expenditures could be produced at a much lower level than is possible currently and recall bias would be removed.

The expenditure weights for the CPI as provided by the LCF pertain to a time period more than a year in the past; they are price updated as part of their incorporation in the Lowe price index formula. This is recognised as an acceptable practice to reflect the time needed to acquire the data from an expenditure survey. With store scanner data, the derived expenditure shares would be almost current and this opens options for using different price index formulas. A true Laspeyres formula could be used; this would match the weight reference period with the price reference period[77]. Weights could be calculated on a month by month basis allowing for a Paasche formula to be used. It would also be possible to go further and use a symmetric price index formula; for example, a Fisher or Törnqvist which uses expenditure information from both the price reference period and the current period. With current practice, it is only possible to approximate a symmetric

---

[70] The Ottawa Group is the main research meeting for index numbers and consumer price statistics; alternative sources were one of the four themes for the 2015 meeting.
[71] Norberg A, Sammar M (2010). Notes on Scanner Data and Price Collection in the Swedish Consumer Price Index, Joint UNECE/ILO Meeting on Consumer Price Indices.
[72] Krsinich F, 2015, Implementation of consumer electronics scanner data in the New Zealand CPI, Presented at the Ottawa Group meeting in Tokyo, www.ottawagroup.org
[73] van der Grient H, de Haan J (2010). The use of supermarket scanner data in the Dutch CPI, Statistics Netherlands paper, www.cbs.nl
[74] Howard A et al (2015). Using transactions data to enhance the Australian CPI, www.ottawagroup.org
[75] van der Grient H, de Haan J (2010). The use of supermarket scanner data in the Dutch CPI, Statistics Netherlands paper, www.cbs.nl
[76] Howard A et al (2015). Using transactions data to enhance the Australian CPI, www.ottawagroup.org
[77] A Practical Introduction to Index Numbers, Wiley

version of a price index retrospectively and with approximations to monthly weights[78]. Work on symmetric price indices would be for research purposes only. It is important to note that the UK CPI adheres to the Eurostat Harmonised Index of Consumer Prices regulations which requires a Lowe (or "Laspeyres-type") index.

There is also the possibility of obtaining expenditure data at a more detailed product level than is currently possible, which could help address a long-standing challenge with price index construction. Expenditure information for elementary aggregates, which are the lowest level at which price relatives are combined in the CPI, is not usually available. This would provide much greater benefits than improving the timeliness of weighting information.

The elementary aggregate level is problematic for consumer price indices. At this level, price quotes are obtained for specific items and, in most cases, no expenditure share information is available. The price quotes are combined without weighting information. Above this level, price quotes are combined with expenditure shares using the Lowe price index formula[79]. ONS has carried out research to see how well unweighted formulas approximate weighted ones[80]. This research found that no unweighted formulas approximate weighted formulas well. Ideally, expenditure information would be provided at all levels in the CPI.

Statistics Norway has been using store scanner data to compute the sub-index for food and non-alcoholic beverages in the CPI. Before using scanner data they captured prices for 260 items; with store scanner data, this has increased to 14,000 items. With weighting information available at both the reference and current periods, the Fisher price index formula is used at the lowest level in the COICOP classification[81].

7.2.4. Availability of data

An important factor that affects how much progress an NSI can make with store scanner data is its availability. While many other NSIs have secured access to this data, this is not the case in the UK. ONS has acquired small store scanner datasets for a narrow range of products which have been used for research; however, wider access is not currently available[82].

As an alternative, ONS is exploring the use of web scraped data; data has been collected for a small number of food items and experimental price indices produced[83].

Recommendation 29: Continue working to secure access to store scanner data.

---

[78] See: Calculating a retrospective superlative CPI for the UK (2014) webarchive.nationalarchives.gov.uk/20160204094749/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/index.html

[79] A Practical Introduction to Index Numbers, Wiley.

[80] Elliott D, Winton J and O'Neill R (2013). Elementary aggregate Indices and lower level substitution, Statistical Journal of the IAOS 29 (1), 11-18

[81] See: www.ssb.no/en///http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5&Lg=1

[82] See: Initial report on experiences with scanner data in ONS (2014) http://webarchive.nationalarchives.gov.uk/20160204094749/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/index.html

[83] Breton R, et al (2015) Research indices using web scraped data, available at: inflationmatters.com/wp-content/uploads/2015/10/ONS-web-scraped-data-article-01092015.pdf

Recommendation 30: Review international use of store scanner data and plan a programme of research to determine the best future use of this data.

7.2.5. Price indices from web scraping

ONS has been scraping price information and additional data from 3 supermarket websites for more than a year. It has been used for research purposes and to produce experimental price indices[84]. Price scraping does not capture information on expenditures shares, so indices calculated using scraped prices use existing expenditure weights, mostly derived from the LCF.

While prices are conventionally collected once a month, web-scraped data is of much higher frequency and this opens avenues for a more frequent reporting of inflation. Research run jointly between ONS and Bristol University is currently looking at combining the established, monthly CPI with high-frequency scraped priced data using inflation models and time series techniques[85].

[84] See: Research indices using web scraped price data (2015) webarchive.nationalarchives.gov.uk/20160204094749/http://www.ons.gov.uk/ons/guide-method/user-guidance/prices/cpi-and-rpi/index.html

[85] Automated High Frequency Index Calculation, Powell, B., and Nason, G. (2016), University of Bristol, in preparation.

Appendix A: Other reviews and programmes

A.1 The Johnson Review

The Johnson Review of UK Consumer Price Statistics was published in January 2015. The brief for the review was to consider what changes are needed for the range of consumer price statistics to best meet current and future needs[86].

The LCF is discussed in Chapter 11 of the Johnson Review; this chapter is concerned with specific design issues in consumer price statistics (Section 11.5 onwards). It emphasises the importance of the LCF by noting that it is the biggest source of weighting information for the CPI and it provides almost all the weighting information for the RPI.

The effect of a continued reduction in achieved sample is discussed in this section; in particular, the issue of sample size and its effect on the variability of weights. The report recognises that the sample would have to increase dramatically to reduce variability to satisfactory levels in all circumstances. It suggests looking at alternative sources instead, where they can provide suitable information. The review also discusses under-reporting.

In Chapter 6, the Johnson Review considers the case for inflation measures for population sub-groups, regions and countries of the UK. To accompany the Johnson Review, ONS has produced inflation measures for household income quintiles based on weights for groups of households taken from the LCF, but not specific price quotes for different households[87].

The Johnson Review recognises that such measures would prove challenging with the LCF as the source of weights and suggests that the user need should be assessed to see whether it could justify additional costs to provide accurate weighting information. However, it did recommend producing inflation measures for a range of household types on an annual basis as an analytical output.

The use of alternative sources is the subject of Chapter 8; it considers the use of 3 types of alternative data – market research data, web-scraped price data and supermarket scanning data. It notes that consumer panel data could provide detailed weighting information to supplement our existing practice; it recommends greater use of these alternative data sources in future.

[86] www.statisticsauthority.gov.uk/reports---correspondence/current-reviews/range-of-prices-statistics.html
[87] webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/elmr/variation-in-the-inflation-experience-of-uk-households/2003-2014/index.html

A.2 The Bean Review

At the time of writing, only the Interim Report of the Independent Review of UK Economic Statistics by Professor Sir Charles Bean has been released. Paragraphs 3.92 and 3.93 refer specifically to the survey and note the continuing decline in response rates and reference is made to the Johnson Review. It notes that a review of the LCF is in progress.

A broader theme of the Bean Review is that ONS should make more use of administrative and alternative data sources.

A.3 Other programmes

There are several work programmes which will involve the LCF over the 2016 calendar year and beyond.

There are proposals in place to "'integrate" LCF and EU-SILC, so a combined sample receives a "core" set of questions, then sub-sets within that sample receive either additional questions required for EU-SILC or the expenditure section of LCF. The "core" questions would mainly cover income and would be sufficient to calculate at-risk-of-poverty or social exclusion (AROPE) estimates required for EU-SILC. This will be explored during 2016.

ONS is embarking on a number of Transformation Programmes from 2016 onwards; one of these will be a Data Collection Transformation Programme. The scope and objectives of this programme have yet to be published.

Appendix B: ONS contributors to this review and groups consulted

B.1 Review Board members

Member	ONS division
Tricia Dodd	Chair, Chief Methodologist, Deputy Director Survey Methodology and Statistical Computing Division
Jeff Ralph	Lead reviewer Survey Methodology and Statistical Computing Division
Emma Wright/ Ria Sanderson	Survey Methodology and Statistical Computing Division
Pete Brodie	Survey Methodology and Statistical Computing Division
Silvia Manclossi	Project Manager and secretariat Survey Methodology and Statistical Computing Division
Dr Giles Horsfield	Review Team, Social Survey Division
Joanna Bulman	Review Team, Social Survey Division
Ed Dunn	Deputy Director, Social Survey Division
Chris Daffin	Social Survey Division
Mike Prestwood	Deputy Director, Prices Division
Philip Gooding	Prices Division
Christopher Jenkins	Prices Division
Ole Black	Deputy Director, Expenditure Household and Economic Indicators Division
Richard Tonkin	Well-being, Inequalities, Sustainability and the Environment Division
Professor Thomas Crossley	External Expert, University of Essex

B.2 Review Team members

Member	ONS division
Dr Jeff Ralph	Survey Methodology and Statistical Computing
Pete Brodie	Survey Methodology and Statistical Computing; Collection and Editing Methods & Statistical Computing
Silvia Manclossi	Survey Methodology and Statistical Computing
Dr Giles Horsfield	Social Survey Division; Living Costs and Food Survey
Jo Bulman	Social Survey Division; Living Costs and Food Survey
Professor Thomas Crossley	External Expert, University of Essex

B.3 Other contributors

Member	ONS division
David Ainslie	Social Survey Division; Strategy and Development Team
Martina Aumeyr	Social Survey Division; Strategy and Development Team
ONS interviewers and interviewer managers	Social Survey Division
Genevieve Groom	Survey Methodology and Statistical Computing Division; Data Collection Methodology
Peter Betts	Survey Methodology and Statistical Computing Division; Data Collection Methodology
Emma Dickinson	Survey Methodology and Statistical Computing; Data Collection Methodology
David Howell	Social Survey Division; Integrated Household Survey
Dr Salah Merad	Survey Methodology and Statistical Computing; Sample and Design and Estimation
Dr Steve Rogers	Survey Methodology and Statistical Computing; Process, Edit and Imputation (Social)
Joe Winton	Survey Methodology and Statistical Computing; Index Numbers

The Review Team is grateful to Paul Smith, associate professor at the University of Southampton, for some very helpful contributions.

B.4 Groups consulted

Within ONS:

- National Accounts;
- Household Income and Expenditure;
- Prices Division;
- Big Data Team;
- Electronic Data Collection Programme;
- Surveys and Life Events Processing.

Outside of ONS:

- Kantar Worldpanel