Quality of expenditure data collected with a mobile receipt scanning app in a probability household panel

Alexander Wenz (University of Mannheim)

Annette Jäckle (University of Essex)

Jonathan Burton (University of Essex)

Mick P. Couper (University of Michigan)

Brendan Read (University of Essex)

Non-technical summary

The accurate measurement of household consumption expenditure is important for official statistics as well as economic research across a wide range of areas. This paper reports on a novel approach using smartphone technology to collect expenditure data in a probability household panel of the general population in Great Britain. Members of the *Understanding Society* Innovation Panel were asked to download an app on their smartphone and report their purchases of goods and services over the period of one month. The app directed respondents to use the built-in camera to photograph all paper receipts that they received at a point of sale. In a separate diary section of the app, they were able to manually enter other expenditures, such as non-receipted payments. In this paper, we compare the quality of the reported expenditure with benchmark data from the Living Costs and Food Survey, the national budget survey in the United Kingdom.

The results suggest that total expenditure reported with scanned receipts plus direct entry aligns closely with the national budget survey whereas app data from scanned receipts only clearly underestimate expenditure. Examining category-level expenditure similarly shows that for most categories, the reported expenditure from scanned receipts plus direct entry aligns more closely with the benchmark than scanned receipts only. In addition, the app data align more closely with the national budget survey for respondents who are older, male, have an above-median income, and live in rural areas. The implications of measurement differences vary: comparisons of estimated budget shares are closer to the benchmark for some categories than others.

Quality of expenditure data collected with a mobile receipt scanning app in a probability household panel

Alexander Wenz (University of Mannheim)

Annette Jäckle (University of Essex)

Jonathan Burton (University of Essex)

Mick P. Couper (University of Michigan)

Brendan Read (University of Essex)

Abstract

In this paper, we compare the quality of expenditure data collected from a mobile receipt scanning app in the *Understanding Society* Innovation Panel with benchmark data from the Living Costs and Food Survey, the national budget survey in the United Kingdom. The results suggest that expenditure reported with scanned receipts plus direct entry aligns closely with the benchmark whereas app data from scanned receipts only clearly underestimate expenditure. In addition, the app data align more closely with the benchmark for respondents who are older, male, have an above-median income, and live in rural areas. The implications of measurement differences vary: comparisons of estimated budget shares are closer to the benchmark for some categories than others.

Keywords: data quality, measurement error, spending data, smartphone app

JEL classification: C81, C83

Corresponding author: Alexander Wenz, University of Mannheim, Mannheim Centre for European Social Research, 68131 Mannheim, Germany, a.wenz@uni-mannheim.de.

1. Introduction

The accurate measurement of household consumption expenditure is important for official statistics as well as economic research across a wide range of areas (Browning, Crossley, & Winter, 2014). National budget surveys, such as the Consumer Expenditure Survey in the United States or the Living Costs and Food Survey in the United Kingdom, are the traditional method of collecting data on household expenditure, typically employing a combination of recall surveys and expenditure diaries (Browning, Crossley, & Winter, 2014; Silberstein & Scott, 1991). In recall surveys, respondents are asked to report how much they spent on different expenditure categories over a specified period, which is typically used to capture relatively infrequent expenses. Diary-based methods complement the recall approach and are designed to collect data about more frequent expenses: respondents are asked to keep a diary over a specified period in which they can record detailed information about every single expense close in time to the actual purchase event.

Both types of expenditure surveys are burdensome, often asking about a large number of expenditures over an extended period of time. In recent years, there are growing concerns that the quality of expenditure survey data is declining (Browning, Crossley, & Winter, 2014). In line with the general trend of decreasing response rates in surveys (Luiten, Hox, & de Leeuw, 2020; Williams & Brick, 2017), the response rates in national budget surveys were found to decline across several countries (Barrett, Levell, & Milligan, 2015). Another concern is measurement error that has been reported for both recall surveys and expenditure diaries (Browning, Crossley, & Winter, 2014; National Research Council, 2013). In recall surveys, measurement error arises if respondents forget to report expenses, report incorrect amounts, or report expenses that occurred prior to the specified recall period, also known as *telescoping* (Fricker et al., 2015; Geisen et al., 2011; Maki & Garner, 2010; Neter & Waksberg, 1964). Diary-based methods aim to facilitate respondents' recall of expenditures by shortening the recall period. In practice, however, a large proportion of respondents complete their diary at the

end of the study period, resulting in similar recall errors as in retrospective survey questions, although the recall period in surveys is typically longer than that for diaries (Browning, Crossley, & Winter, 2014; Silberstein & Scott, 1991). An additional problem with diary-based methods is that the level of reporting tends to decline over the course of the study period, also known as *diary fatigue*. Previous research on two-week expenditure diaries, for example, has shown that the level of reported expenditure is significantly lower in the second week compared to the first week, and lower on later days compared to earlier days of either week (Ahmed, Brzozowski, & Crossley, 2006; Brzozowski, Crossley, & Winter, 2017; McWhinney & Champion, 1974; Silberstein & Scott, 1991; Stephens, 2003; Turner, 1961).

This paper reports on a novel approach using smartphone technology to reduce the burden of reporting on expenditures and improve the quality of reporting in a probability household panel of the general population in Great Britain. Rather than reporting the details of each expenditure, respondents were asked to download an app on their smartphone to scan receipts, thereby limiting direct data entry to non-receipted payments. Respondents were asked to use the app to report their purchases of goods and services over the period of one month. The app directed respondents to use the built-in camera of their mobile device to photograph all paper receipts that they received at a point of sale. In a separate diary section of the app, they were also able to manually enter other expenditures, including online payments, regular payments made by standing order or direct debit, non-receipted payments outside structured shopping environments, for example at market stalls, or payments for which they did not retain or lost the paper receipt before scanning. The data collection approach aims to combine the advantages of existing forms of expenditure measurement. The receipt-scanning component allows capturing expenditure information directly from sales receipts and, thus, collecting detailed data about each purchase as well as reducing burden and measurement error due to recall. The diary component aims to ensure that non-receipted purchases are also covered. Since most study participants are likely to carry around their smartphone throughout the day (Keusch,

Wenz, & Conrad, 2022), implementing this data collection procedure within a mobile app, rather than relying on external scanners or using paper diaries, has the advantage that they have access to the receipt scanning app at all times. Obtaining paper receipts at a point of sale might serve as a physical cue to record the spending activity, and respondents are able to scan the receipts or manually report their expenses shortly after the purchase event. Scanning whole receipts enables collecting information about multiple expenditures in one action, rather than requiring respondents to enter each item purchased separately. In addition, regular notifications can be implemented within the app that remind and encourage respondents to report their expenditures every day rather than at the end of the study period. This receipt scanning study is, to the best of our knowledge, the first to be implemented on a sample that is representative of the general population.

We have previously reported on participation in the app study and potential non-participation bias (Jäckle et al., 2019). In this paper, we focus on the quality of data provided by those who participated in the app study. We compare the quality of expenditure data with benchmark data from the Living Costs and Food Survey, the national budget survey in the United Kingdom. We address the following research questions:

RQ1. How do total expenditure and category-level expenditure recorded in the mobile receipt scanning app align with expenditure data from the national budget survey?

RQ2. Are there differences between population subgroups in how the expenditure data from the mobile app align with data from the national budget survey?

RQ3. Do the data from the mobile app lead to different estimates of budget shares than the national budget survey?

2. Background

Major research programmes have been initiated to improve the measurement of consumer expenditure in budget surveys (Browning, Crossley, & Winter, 2014). Examples include the Gemini Project initiated by the U.S. Bureau of Labor Statistics in 2009 to redesign the Consumer Expenditure Survey[1] and the Conference on Research in Income and Wealth sponsored by the National Bureau of Economic Research in 2011 (Carroll, Crossley, & Sabelhaus, 2015). At the same time, the collection and use of expenditure data from other sources has rapidly expanded (Browning, Crossley, & Winter, 2014; Jäckle et al., 2021). Measures of consumer expenditure are, for example, created based on process-generated data from online financial aggregators that link individuals' financial accounts and provide summaries of their income and expenditures (e.g., Angrisani, Kapteyn, & Samek, 2018; Baker, 2018; Gelman et al., 2014; Kuchler & Pagel, 2021). Similarly, transactional data from store loyalty cards scanned at a point of sale (e.g., Andreyeva et al., 2012; Felgate et al., 2012; Newing et al., 2014; Panzone et al., 2016; Tin et al., 2007) or credit and debit cards (e.g., Agarwal et al., 2007; Gross & Souleles, 2002) provide new forms of expenditure measurement. Home scanner data collected by market research organisations are also being used for research on consumer expenditure: consumers are asked to scan the barcodes of all items they purchased with a barcode scanner installed in their home and the barcode data are then linked to other data such as prices and nutritional information (e.g., Leicester, 2015; Leicester & Oldfield, 2009; Aguiar & Hurst, 2007; Broda et al., 2009; Griffith et al., 2009; Griffith & O'Connell, 2009; Lusk & Brooks, 2011; Zhen et al., 2009). The advantage of these data sources over existing forms of expenditure measurement is that they do not rely on the study participants' ability to recall and report information and are, thus, not susceptible to the associated recall errors (Browning, Crossley, & Winter, 2014; Jäckle et al., 2021). In addition, they can collect

---

[1] https://www.bls.gov/cex/geminiproject.htm

expenditure data at a much more detailed level and a higher frequency, which reduces burden on respondents and allows measuring changes in expenditure patterns over time. However, there are also a number of limitations to these data, which mean that they are not necessarily suitable as substitutes for consumer expenditure surveys. Users of consumer expenditure data vary in their requirements: some require full COICOP (Classification of Individual Consumption by Purpose; United Nations, 2000) expenditure classifications, for others, less detailed classifications are sufficient; some require data that identify the spending of individual households, others only need expenditure for the consumer sector as a whole; some require the geolocation of the consumer or of the point of sale where the purchase was made; and some require additional information about the characteristics of the household, such as its composition or income. While bank transactions, credit and debit card data could cover all electronic payments made by consumers in a country if data from all banks/card issuers can be obtained, the data only include the total value of a purchase, and no information on individual items. The Merchant Category Codes identify the type of retailer where a card payment was made, and can be coded to high-level COICOP classifications, which may be sufficient for data users (e.g., Alatrista-Salas et al., 2021; Hoseini & Valizadeh, 2021). Payments by direct debits and standing orders are included in bank transactions data, but the string that identifies the transaction is not necessarily codable. In addition, the data do not identify consumers, or include information about their characteristics, and do not include cash purchases. Studies based on financial aggregator data have the advantage that the accounts of a consumer are linked across banks; however, they mostly rely on convenience samples that are not representative of the general population and have limited sociodemographic information (see Angrisani, Kapteyn, & Samek, 2018 for an exception). Store scanner data include the product barcode or Universal Product Code (UPC), quantity and price of goods sold, but do not identify individual consumers. Till receipt data include all information on the shopping receipts of a retailer, but again do not identify consumers. Loyalty card data are linked to an individual consumer and

can include information about the account holder from when they signed up for the card but provide incomplete data on expenditure. Home scanner data can be expensive to purchase, rely on convenience samples, and have been shown to record lower levels of spending in comparison with consumer expenditure surveys in both the United Kingdom and the United States (Leicester, 2015; Zhen et al., 2009).

Receipt scanning apps have become popular methods of data collection in the market research industry, with the primary aim of studying shopping behaviour (Jäckle et al., 2021). Examples of such apps include ReceiptPal[2], Receipt Hog[3], Ibotta[4], and Worldpanel Plus[5]. In academic research, however, there has only been a small body of literature using receipts for data collection (e.g., Cullen et al., 2007; DeWalt et al., 1990; French et al., 2009; Martin et al., 2006; Rankin et al., 1998; Ransley et al., 2001; Smith et al., 2013a; Smith et al., 2013b; Tang et al., 2016; Weerts & Amoran, 2011). In these studies, rather than using receipt scanning apps, respondents have been asked to keep and annotate paper receipts and return them via postal mail. A major limitation of existing studies is that they are typically based on small-scale volunteer samples and conducted in small geographic areas, such as food shoppers recruited at local supermarkets.

Prior research on the quality of data generated from receipt scanning studies is even further limited and has mainly focused on representation error (Jäckle et al., 2019; Ransley et al., 2001; Smith et al., 2013a). For example, Jäckle et al. (2019), relying on the same receipt scanning app as the present study, examined potential biases in which types of sample members have access to mobile devices (coverage bias) and in which types of sample members participate in the app study (participation bias). They found evidence for extensive coverage bias: mobile device users differed from non-users in terms of sociodemographic characteristics

---

[2] https://www.receiptpalapp.com/
[3] https://www.receipthog.com/
[4] https://home.ibotta.com/
[5] https://www.kantarworldpanel.com/worldpanelplus/

and financial behaviours. Conditional on coverage, comparatively less participation bias was found, but the differences between study participants and non-participants mirrored the coverage bias patterns. Our study builds on this previous work by focusing on the quality of the data provided by respondents who used the receipt scanning app, by comparing the reported expenditure with benchmark data from the Living Costs and Food Survey.

3. The *Understanding Society* Spending Study

3.1 Sample

The Spending Study was implemented on the *Understanding Society* Innovation Panel, a nationally representative household panel in Great Britain (University of Essex, Institute for Social and Economic Research, 2022). The Innovation Panel is based on a stratified, clustered sample of households in England, Scotland, and Wales (Lynn, 2009). The original sample, first interviewed in 2008 (wave 1), consists of approximately 1,500 households, supplemented with refreshment samples of approximately 500 households each in waves 4, 7, 10, 11, and 14. The interviews are conducted annually with all household members aged 16+ and focus on individuals' socio-economic, health, housing, and family situation. The Spending Study was implemented on the Innovation Panel sample between waves 9 and 10. Fieldwork for wave 9 ran from May to September 2016. Sample households were randomly allocated to survey mode: one third of households were allocated to face-to-face interviews and the other two thirds to a sequential mixed-mode (web, face-to-face) design. In the final phase of fieldwork, nonrespondents in both experimental allocations were given the option to complete the survey online or by telephone. The household response rate was 84.7% (AAPOR RR5; The American Association for Public Opinion Research, 2016), with 85.4% of individuals within those households completing the individual interviews (Institute for Social and Economic Research,

2021). More details about the Innovation Panel survey design and fieldwork are available in the online documentation.[6]

3.2 App study design

In October 2016, all sample members aged 16+ in households where at least one adult had completed the wave 9 interview were invited to the Spending Study (Jäckle et al., 2018). They received an invitation letter by post and by email, if their email address was known. The invitation letter informed sample members about the study and contained a URL to an online registration survey, at the end of which they were given login details and instructions on how to download the receipt scanning app on their smartphone or tablet. The app was available for iOS and Android devices and could be downloaded from the Apple App Store or the Google Play Store. Sample members who did not complete the registration survey were followed up with reminder emails, sent twice per week for the duration of three weeks, and a final reminder letter sent by post after four weeks. They were asked to use the app every day for one month to report their purchases of goods and services and received the following incentives for their study participation: £6 or £2 for downloading the app (with households randomly allocated to either of the two incentive conditions), £0.50 for every day on which they used the app, a bonus of £10 if they used the app every day for one month, and £3 for completing a debrief survey.

There were three sections in the receipt scanning app: 1) *Scan receipts*, where respondents were asked to take photos of their receipts; 2) *Direct entry*, where respondents were asked to report purchases for which they did not have any receipts, by indicating the amount and selecting the expenditure categories included in that purchase; and 3) *Report no purchases today*, where respondents were asked to confirm if they did not purchase any goods or services on that day. See Appendix C in Jäckle et al. (2018) for screenshots of the app. The app sent a

---

[6] https://www.understandingsociety.ac.uk/documentation/innovation-panel

push notification every day at 5pm to remind respondents to report their expenses or indicate that they did not have any expenses on that day.

The Spending Study fieldwork was conducted between October and December 2016. Of 2,112 invited sample members, 12.8% used the app at least once (n = 270) and among those, 81.5% used the app for at least 29 days (Jäckle et al., 2019). In this paper, we use data from 262 app users who reported at least one purchase, either by receipt scanning or direct entry. Due to the different data collection periods of the Spending Study (one month) and the benchmark data (two weeks), we restrict the Spending Study data to expenditures reported by respondents in their first two weeks of study participation.

3.3 Coding of expenditure categories

Scanned receipts. The information contained in the receipt images was manually transcribed by Kantar Worldpanel sub-contractors. The data captured include receipt-level information, such as the store name, the purchase date, and the total purchase amount, as well as item-level information, such as the description and price of each item purchased. The scanned receipts also contain information on promotions and price reductions, but these are not structured in a consistent way across receipts and cannot always be attributed to the relevant items: while some receipts record the reductions in the line below the reduced item, others list all reductions at the bottom of the receipt. In the analyses reported in this paper, reductions were, therefore, subtracted from the total purchase amounts but not subtracted from the category-level expenditures.

The items transcribed from the scanned receipts were coded into expenditure categories by matching the item description with the *Volume D: Expenditure codes 2015-16* dataset provided by the Living Costs and Food Survey (Office for National Statistics, 2017). This dataset contains a comprehensive list of consumer items classified according to the Classification of Individual Consumption by Purpose (COICOP; United Nations, 2000). The

coding was done by first collapsing COICOP codes into ten expenditure categories, following suggestions by d'Ardenne and Blake (2012): 1) *Food and groceries*, 2) *Clothes and footwear*, 3) *Transport*, 4) *Child costs*, 5) *Home improvement and household goods*, 6) *Health*, 7) *Socialising and hobbies*, 8) *Other goods and services*, 9) *Holidays*, and 10) *Gifts*. The items on the scanned receipts were then assigned to these categories by using exact string matching. The classifications were checked by manual coders to evaluate the quality of the matching procedure and recode items that were not successfully matched. Of the 13,366 items recorded in the scanned receipts in the respondents' first two weeks of study participation, 74.2% were matched correctly, 21.3% were matched incorrectly or did not match and were coded manually, 2.9% could not be assigned to a category, neither by the matching procedure nor the manual coders, and 1.7% were on a receipt image that was not readable. More details about the matching procedure can be found in Read (2023).

Direct entry. When reporting expenses through direct entry, respondents were asked to indicate the total amount of their purchase and select one or more expenditure categories included in that purchase. The expenditure categories were the same ten categories used to code the scanned receipts. For purchases where respondents selected a single expenditure category (98% of entries), the reported amount was assigned to the respective category. For purchases where respondents selected multiple categories (2% of entries), the reported amount was divided and assigned to the different categories. Rather than simply dividing the reported amount by the number of selected categories, the ratio of category-level expenditures was determined based on data from the scanned receipts and the direct entries with a single expenditure category selected. The reported amount was then distributed according to that ratio.

Descriptive statistics of the reported expenditure are shown in Table 1 for the analysis period (week 1 and 2) and in Appendix Table A1 for the full data collection period. In their first two weeks of study participation, app users submitted 2,092 scanned receipts (63.5% of all submissions) and made 1,200 direct entries (36.5% of all submissions).

Table 1. Expenditure reports in the *Understanding Society* Spending Study in week 1 and 2.

	Scanned receipt		Direct entry	
	n	**%**	**n**	**%**
Total submissions (n = 3,292)	2,092	63.5	1,200	36.5
Total items	13,366	100.0	1,200	100.0
Food and groceries	9,425	70.5	507	42.3
Clothes and footwear	390	2.9	57	4.8
Transport	194	1.5	128	10.7
Child costs	74	0.6	22	1.8
Home improvements and household goods	914	6.8	43	3.6
Health	119	0.9	17	1.4
Socialising and hobbies	631	4.7	217	18.1
Other goods and services	1,000	7.5	141	11.8
Holidays	0	0.0	7	0.6
Gifts	11	0.1	78	6.5
Item cannot be assigned to category	*384*	*2.9*	–	–
Item not readable on receipt	*224*	*1.7*	–	–

Note. The scanned receipts contain the following additional items: promotions or price reductions (n = 920), value-added tax (n = 31), and other items (n = 13). The percentages for direct entry items do not add up to 100% since respondents were able to select multiple expenditure categories per purchase.

The scanned receipts contain a total of 13,366 items and are highly skewed across expenditure categories. The large majority of items (70.5%) are *Food and groceries*, followed by *Other goods and services* (7.5%), and *Home improvements and household goods* (6.8%). The other categories make up a smaller share of less than 5% each. The 1,200 direct entries, in turn, are more evenly distributed across expenditure categories. While *Food and groceries* also make up the largest share (42.3%), other frequently reported expenditure categories include *Socialising and hobbies* (18.1%), *Other goods and services* (11.8%), and *Transport* (10.7%). Since expenditures on *Holidays* and *Gifts* were rarely reported in the sample, we exclude these categories from the analysis.

4. Benchmark data: The Living Costs and Food Survey

We use benchmark data from the Living Costs and Food Survey (LCF) (Office for National Statistics. Department for Environment, Food and Rural Affairs, 2018), the national budget survey in the United Kingdom. The survey provides information on spending patterns for the Retail Price Index and is used for National and Regional Accounts to provide estimates of household consumption expenditure (Bulman, Davies, & Carrel, 2017). The LCF is based on a stratified, clustered sample of households in England, Scotland and Wales, and a systematic random sample of private addresses in Northern Ireland. The study has two main components: a questionnaire and a spending diary.

The questionnaire is administered in a face-to-face interview and consists of a household and an individual section. The household section is completed by the household reference person and covers information about the sociodemographic characteristics of household members and regular items of household expenditure. These expenditures include regular payments, such as mortgage or rent payments, utility bills, insurances, education fees, subscriptions of magazines and newspapers, and gym memberships, as well as large but infrequently purchased goods and services, such as vehicle purchases, vehicle service, season tickets, package holidays, and furniture and other home improvements. The individual section is completed by each adult within the household and collects information about income.

All adult household members aged 16 and older are then invited to record their daily expenditure in a paper spending diary for two weeks. Children aged between 7 and 15 receive a simplified version of the diary. The diary is organised into ten sections for different types of expenditures. Six sections cover daily expenditures, including 1) *Food and drink brought home*, 2) *Takeaway meals and snacks eaten at home*, 3) *Meals, snacks and drinks consumed away from the home*, 4) *Clothing and footwear*, 5) *Other payments and purchases*, and 6) *Winnings from lottery, bingo, betting shops, football pools, raffles*. The other four sections cover larger but rare purchases that are not likely to occur on a daily basis, including 7) *Home-grown and*

wild food, 8) *Holidays expenditure abroad*, 9) *Business refunds*, and 10) *Expenditure due to special circumstances*. Respondents are asked to record each item and the amount paid in the appropriate section. In the analysis, we only focus on expenditures reported in the spending diary.

As an incentive for study participation, sampled households receive a booklet of stamps with the advance letter. Adults within responding households also receive an incentive of £20 for completing the questionnaire and spending diary.

In this paper, we use data from the 2016-2017 survey which was conducted between April 2016 and March 2017. Among eligible households, 45% completed the interview and returned at least one spending diary (n = 4,641 households; Williams, 2019). Of these households, 125 provided a partial response, where one or more adults declined to keep the diary, but the diary of the person who does most of the shopping for the household was present. Missing diaries were imputed by the Office for National Statistics with data from a respondent in another household with matching characteristics of age, employment status and relationship to the household reference person (Bulman, Davies, & Carrel, 2017). In total, spending diary data are available from 9,272 respondents aged 16+ in the 2016-2017 survey. To match the time frame and geographic coverage of the Spending Study, we exclude 6,876 respondents who were interviewed outside the period October-December 2016 and 183 respondents who were resident in Northern Ireland, i.e., outside Great Britain. We also exclude 4 respondents who did not report any expenditure, which leaves an analysis sample of n = 2,209.

5. Methods

We use inverse probability weighting (IPW) to match the sample composition of the Spending Study to the LCF (Horvitz & Thompson, 1952). The weights are computed with a two-step approach: First, we fit a logistic regression to estimate the respondents' probability of

being in the LCF sample ($T_i = 1$) as opposed to the Spending Study sample ($T_i = 0$) based on a set of respondent characteristics X collected in both samples:

$$\hat{p}_i = \Pr(T_i = 1|X_i) \tag{1}$$

The following respondent characteristics are included in the model: age (in years), age-squared, gender (male, female), education (degree, no degree), personal monthly gross income (in £), household size, urbanicity (urban, rural), and the interaction of age and education. In the Spending Study, missing values on the respondent characteristics were imputed for 5 respondents with the values from previous waves and for 3 respondents with the median values. Table A2 in the Appendix shows the results of the logistic regression.

Second, we use the inverse of the estimated probability $\hat{p}_i$ to calculate the weights:

$$w_i = \frac{1}{\hat{p}_i} \text{ for LCF respondents} \tag{2}$$

$$w_i = \frac{1}{1-\hat{p}_i} \text{ for Spending Study respondents.} \tag{3}$$

After examining the weight distribution in both samples, we winsorized outliers that are outside the range of mean weight $\pm\ 3 \times$ standard deviation of the weights (Valliant & Dever, 2018) to reduce the effect of excessive weights on the variance.

To assess whether the IPW has successfully matched the Spending Study sample composition to the LCF, we calculate the standardised differences in respondent characteristics between the two samples before and after weighting (Austin, 2009). Figure 1 and Table A3 in the Appendix show that the standardised differences for all respondent characteristics fall between -10% and 10% after weighting; the remaining differences can, thus, be considered a negligible imbalance between the samples (Normand et al., 2001).
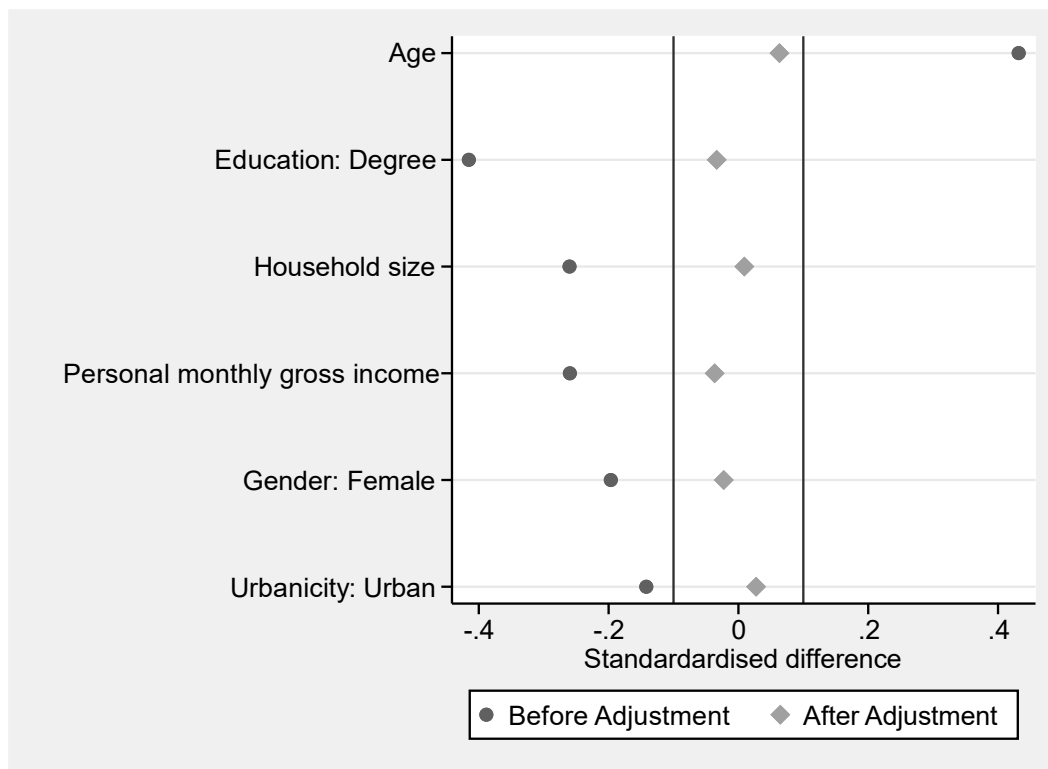
Figure 1. Standardised differences in respondent characteristics before and after inverse probability weighting.

To deal with outliers in total expenditure and category-level expenditure, we recode values greater than the 99th percentile to the 99th percentile separately for the LCF and the Spending Study. The data preparation and analysis were conducted in Stata version 15.1 (StataCorp, 2017).

6. Results

6.1. How do total expenditure and category-level expenditure recorded in the mobile receipt scanning app align with expenditure data from the national budget survey?

To compare the expenditure recorded in the mobile receipt scanning app with that recorded in the national budget survey we proceed as follows. We compare both the data from the scanned receipts only (SR) and the data from the scanned receipts plus the direct entry (SR+DE) with the national budget survey (LCF). Because there are respondents in both samples

who reported zero expenditure for some categories (Table A4 in the Appendix), we calculate weekly expenditure in two ways: (a) including non-zero and zero expenditure, that is, based on all respondents, and (b) including non-zero expenditure only, that is, based on respondents who reported purchases in the relevant category. For each of the four comparisons (SR or SR+DE; non-zero plus zero expenditure or non-zero expenditure only), we test for differences in the overall distribution, mean and median weekly expenditure. We provide the results of all statistical tests in Appendix Tables A5 and A6. The discussion in the text focuses on the distribution and median, unless the test of means leads to different conclusions. Similarly, the discussion focuses on weekly expenditure calculated for all respondents (zero plus non-zero expenditure), unless the findings for non-zero expenditure only show a different pattern.

For total expenditure (Figure 2), the Spending Study distribution from SR+DE (dashed line) aligns closely with the LCF distribution (solid line). In contrast, the Spending Study distribution based on SR only (dotted line) clearly underestimates expenditure. Focusing on median expenditure similarly shows that SR+DE comes closer to the LCF than SR (Appendix Table A5): the estimated total expenditure for the LCF benchmark is £122.80, compared with £101.30 for SR+DE and £70.10 for SR. Both differences are statistically significant (p=0.048 and p<0.001, respectively). Looking at mean total expenditure, the estimate from SR+DE does not differ significantly from the LCF estimate (£149.50 vs. £156.50, p=0.258). A Kolmogorov-Smirnov (KS) test for equality of distributions, however, shows that both Spending Study distributions differ significantly from the benchmark (p=0.007 for SR+DE; p<0.001 for SR only).
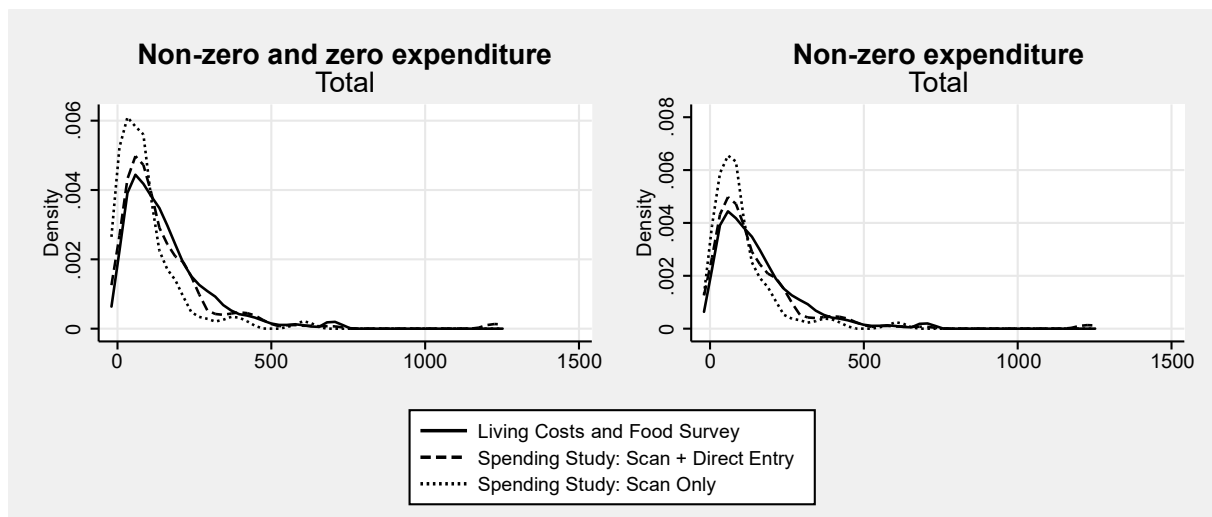
Figure 2. Distribution of average weekly total expenditure.

We next examine the extent to which category-level expenditure aligns between the app and benchmark data. Across all expenditure categories, the percentage with zero expenditures in the two-week period is higher when only looking at scanned receipts. When direct entry is included, this proportion is generally lower, and closer to the LCF benchmark (see Table A4). For example, 32.4% reported zero expenditures for *Transport* in the LCF, compared with 40.8% for SR+DE and 58.8% for SR. In the case of two categories (*Food and groceries* and *Clothes and footwear*), the percentage with zero expenditure is actually lower for SR+DE than LCF.

Non-zero and zero expenditure
Food and groceries

Clothes and footwear

Transport

Child costs

Non-zero expenditure
Food and groceries

Clothes and footwear

Transport

Child costs

Living Costs and Food Survey
Spending Study: Scan + Direct Entry
Spending Study: Scan Only

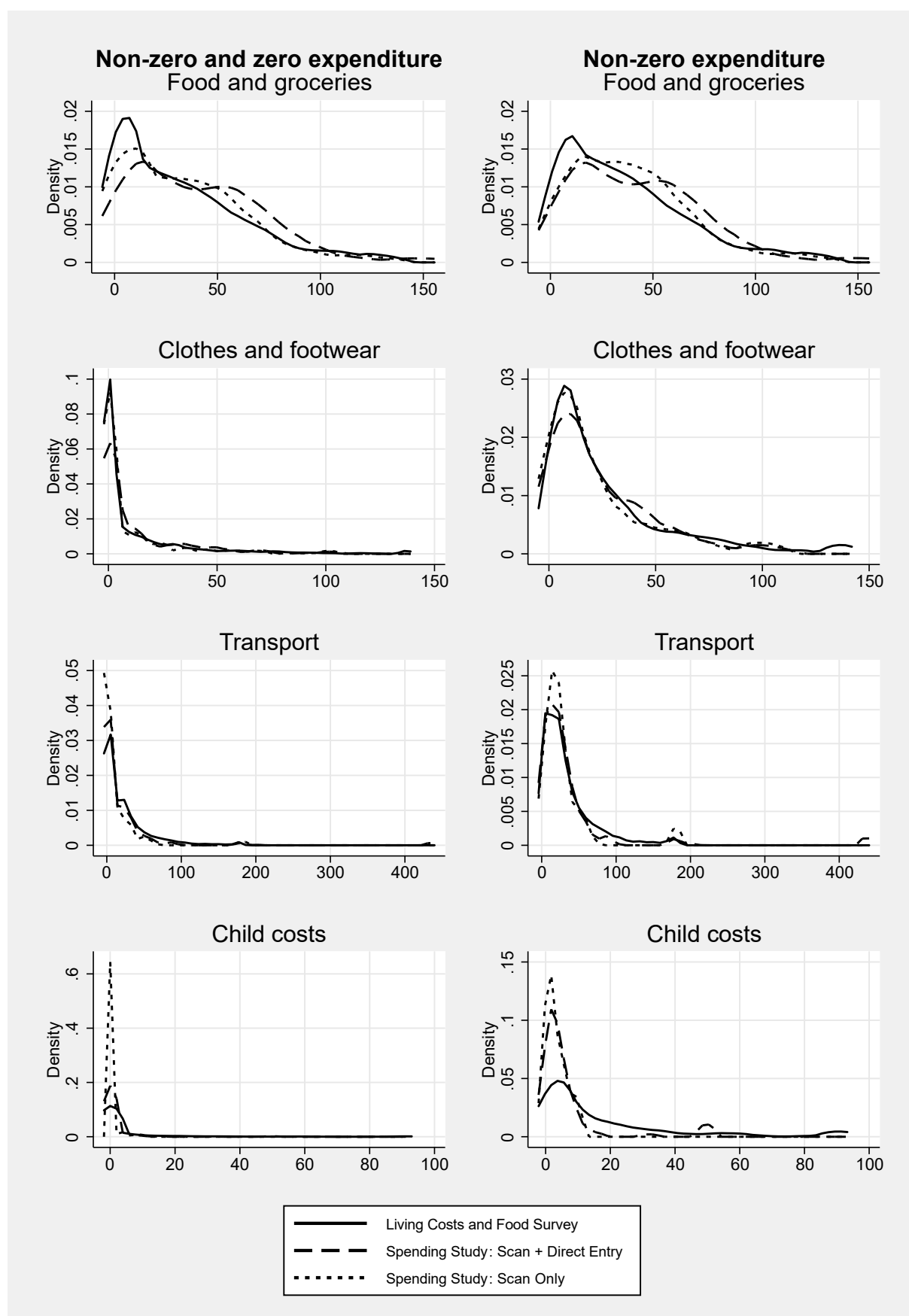Figure 3. Distribution of average weekly expenditure on food and groceries, clothes and footwear, transport, and child costs.

18

For most categories, the reported expenditure amount from SR+DE aligns more closely with the benchmark than SR only. For example, from Appendix Table A6, the LCF median for *Socialising and hobbies* is £27.60, compared with £12.40 for SR+DE and £7.50 for SR. However, the differences in the expenditure distribution and median expenditure between the Spending Study and LCF are still significant for most of the categories (Figure 3 and Figure 4; Table A5 and Table A6 in the Appendix). Two exceptions are *Clothes and footwear* and *Transport*; in both categories, the median expenditure does not differ significantly from the benchmark, either for SR or for SR+DE (see Appendix Table A6).

The pattern is different for the category *Food and groceries*. Here the median expenditure for SR (£27.50, $p=0.402$) is not significantly different from the benchmark (£24.80), whereas SR+DE significantly overestimates the benchmark expenditure (£35.30, $p=0.014$). Similarly, the expenditure distribution from SR is not significantly different from the LCF distribution ($p=0.540$) whereas that from SR+DE is ($p<0.001$).
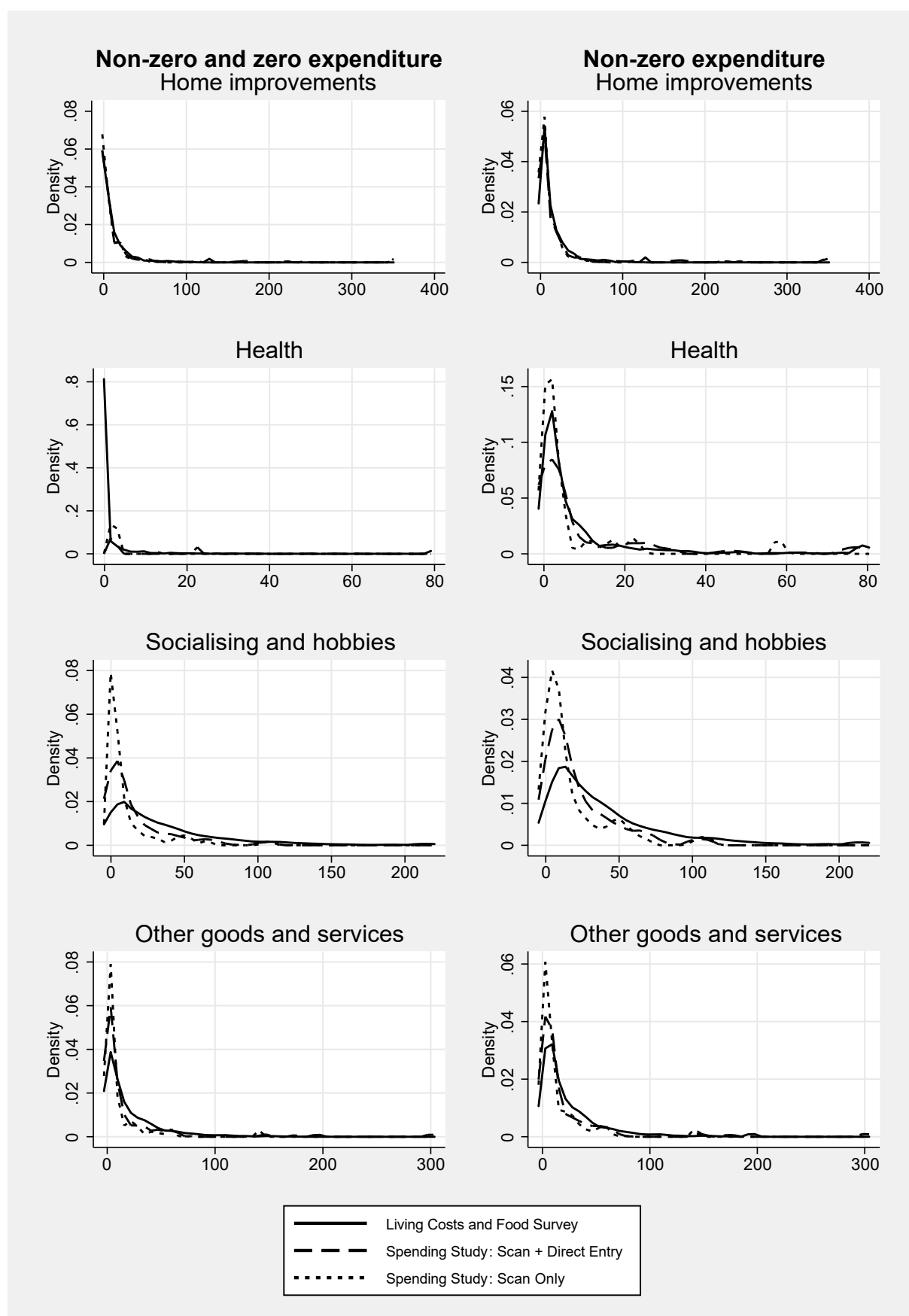
Figure 4. Distribution of average weekly expenditure on home improvements and household goods, health, socialising and hobbies, and other goods and services.

6.2. Are there differences between population subgroups in how the expenditure data from the mobile app align with data from the national budget survey?

We next examine to what extent the total expenditure recorded in the app align with the benchmark data for different population subgroups. Our analysis focuses on the respondent characteristics that were used in the IPW, including age (recoded into 16-50, 51-82), gender (male, female), education (no degree, degree), personal monthly gross income (recoded into below median, above median), household size (recoded into single, non-single), and urbanicity (rural, urban). Figure 5, Figure 6, and Appendix Table A7 present results for both SR and SR+DE from the Spending Study, along with LCF estimates. Given the finding earlier that the SR+DE estimates are generally closer to the LCF benchmark, our discussion focuses on these estimates.

For respondents aged 16-50, the expenditure distribution from the app (SR+DE) is significantly different from the LCF expenditure distribution (p=0.009), whereas the respective distributions for respondents aged 51-82 are not significantly different (p=0.245). Examining median expenditure similarly shows that for those aged 16-50 the app data (£75.90, p=0.005) significantly underestimate the LCF (£118.0), whereas for those aged 51-82 the app data (£118.30, p=0.632) are not significantly different from the benchmark (£127.30).

We also find gender differences for the alignment between the mobile app and the benchmark data. The expenditure distribution from the app is significantly different from the LCF distribution for women (p<0.001) but not for men (p=0.458). Comparing median expenditure also shows that the app data significantly underestimate the LCF for women (£93.70 vs. £139.90, p<0.001) but not for men (£118.20 vs. £103.70, p=0.426).
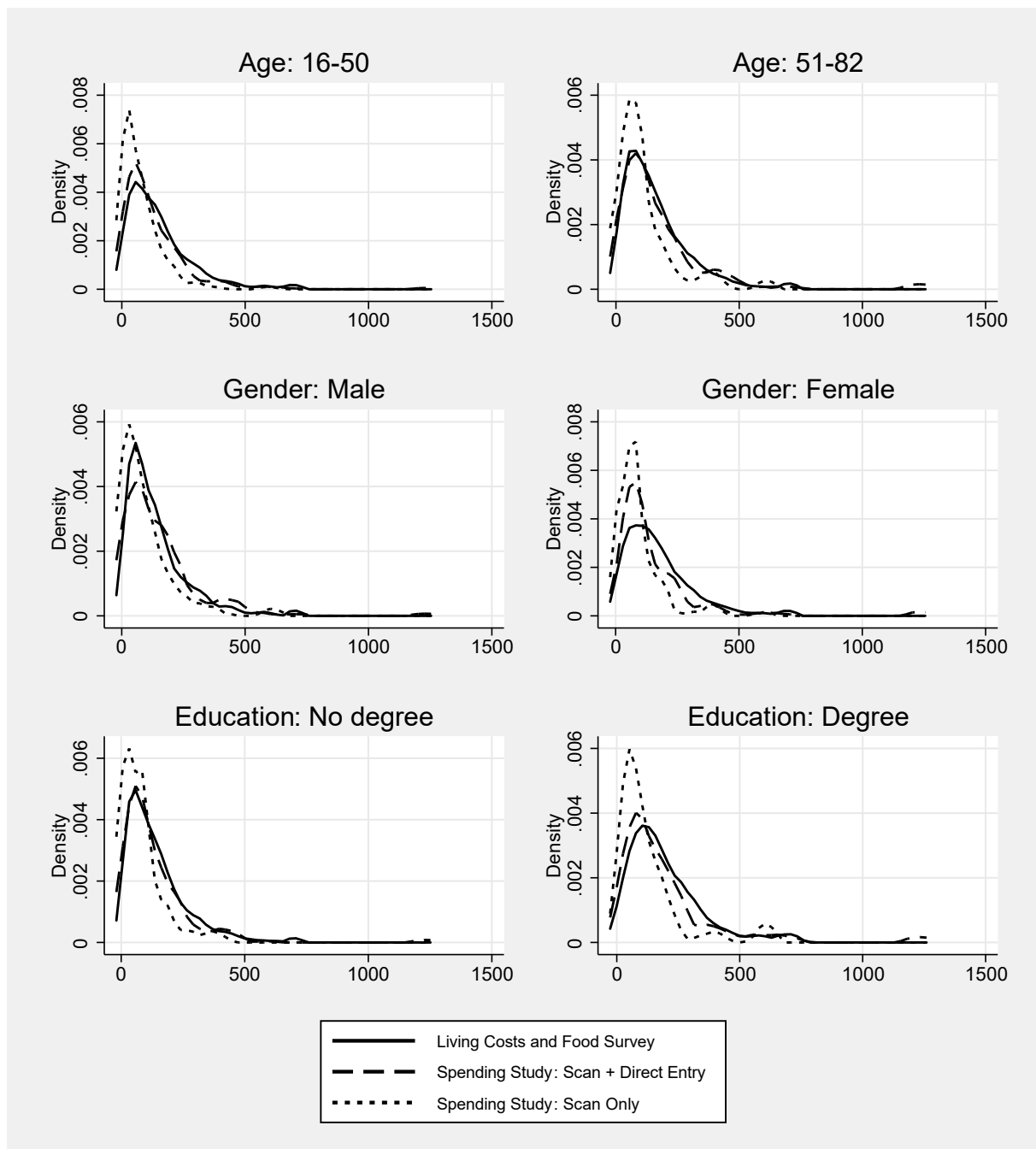
Figure 5. Distribution of average weekly total expenditure by age, gender, and education.

We do not find differences by educational attainment. The expenditure distributions for both respondents with a degree (p=0.039) and those without a degree (p=0.044) differ significantly from the respective LCF distributions. The median expenditure estimates are not significantly different from the LCF for respondents with a degree (£129.60 vs. £157.70, p=0.108), but are for respondents without a degree (£82.40 vs. £109.40, p=0.039).

The alignment between the mobile app and benchmark data differs by personal monthly gross income (Figure 6). For respondents with a below-median income, the expenditure distribution is significantly different from the LCF expenditure distribution (p=0.005), whereas the distribution for respondents with an above-median income is not significantly different (p=0.267). Comparing median expenditure similarly shows that for those with a below-median income, the app data significantly underestimate the LCF (£74.00 vs. £92.40, p=0.004) whereas for those with an above-median income, the estimate is not significantly different from the benchmark (£134.70 vs. £153.20, p=0.216). Comparing the means presents a different picture: the two estimates do not differ for those with lower income (£127.20 vs. £124.00, p=0.735), whereas for those with higher income, the app estimate is significantly lower than the LCF (£168.50 vs. £187.50, p=0.015).

We do not find differences by household size. Although a KS test indicates significant differences in the expenditure distributions for respondents in single households (p=0.010), but non-significant differences for those in multi-person households (p=0.453), a comparison of median and mean expenditure suggests that the app data from both population subgroups similarly align with the respective benchmark data.
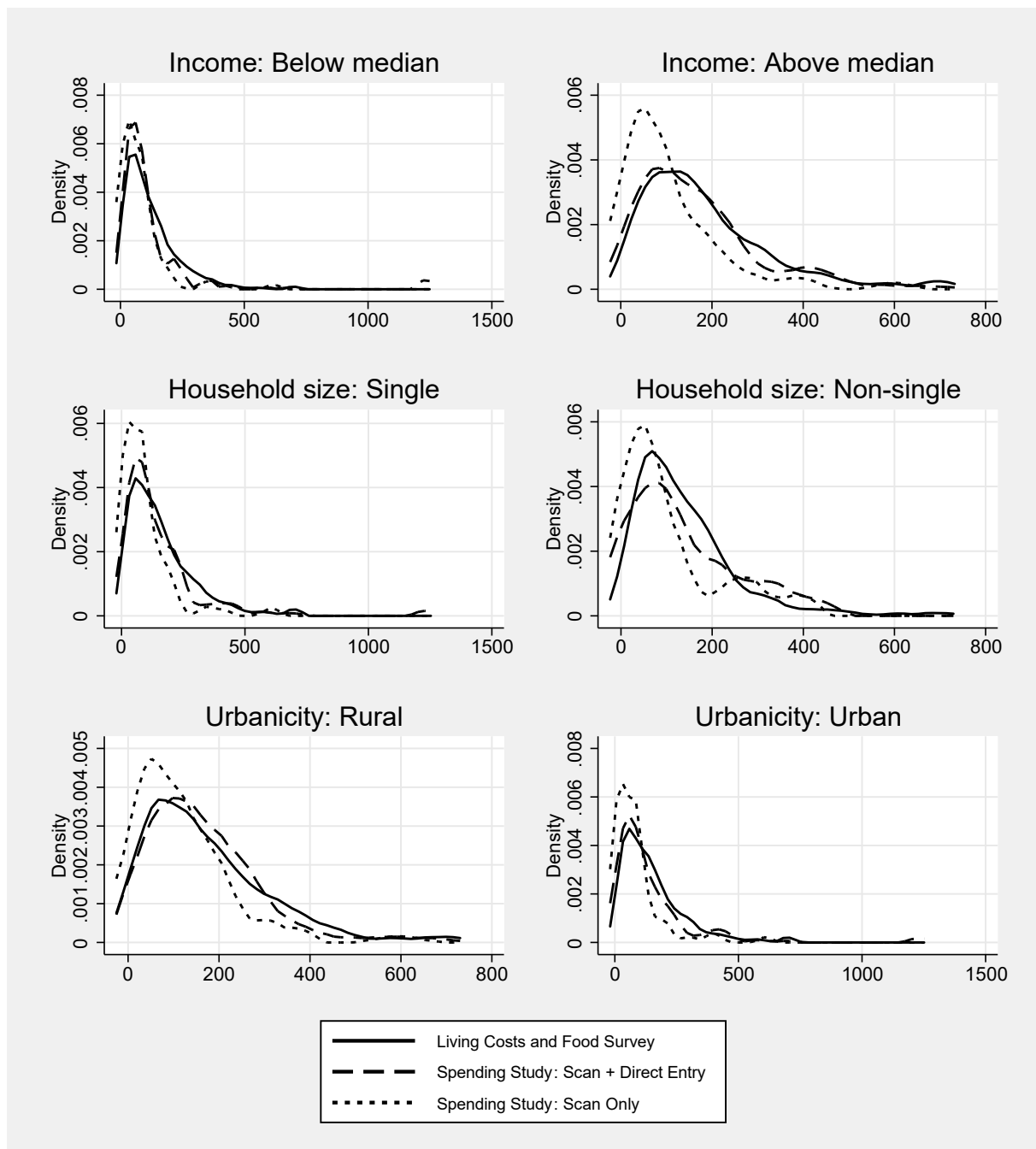
Figure 6. Distribution of average weekly total expenditure by personal monthly gross income, household size, and urbanicity.

Finally, the alignment between the app and benchmark data differs by urbanicity. For respondents from urban areas, the expenditure distribution is significantly different ($p=0.002$) whereas the distribution for respondents from rural areas is not significantly different ($p=0.816$). For those living in urban areas, median expenditure significantly underestimates the

LCF (£89.50 vs. £117.50, p=0.004), whereas for those in rural areas, the median expenditures do not differ (£144.40 vs. £151.10, p=0.355). Neither of the mean expenditure estimates is significantly different from the benchmark (p=0.355 for those in urban areas, p=0.429 for rural areas).

6.3. Do the data from the mobile app lead to different estimates of budget shares than the national budget survey?

To examine whether the differences between the app and national budget data lead to differences in conclusions about economic relationships, we compare budget shares calculated from the two data sets. Budget shares are indicators of consumer behavior that are frequently used in the economics literature (e.g., Ahmed et al., 2006; Leicester, 2015). They denote the ratio between category-level expenditure on category k and total expenditure:

$$w_{ik} = \frac{x_{ik}}{\sum x_{ik}} \qquad (4)$$

In line with the general pattern observed in RQ1, the mean budget shares of SR+DE are closer to the national budget survey than the mean budget shares of SR only. For example, the budget share of *Food and groceries* is overestimated by 19.9 percentage points with SR and by 14.8 percentage points with SR+DE (Table 2). The categories, however, differ substantially in the extent that expenditure is over- or underestimated in the Spending Study compared to the benchmark.

Table 2. Mean budget shares.

	LCF	Spending Study: Scan + Direct Entry		Spending Study: Scan Only	
	%	%	Δ	%	Δ
Food and groceries	25.1	39.9	+14.8	45.0	+19.9
Clothes and footwear	7.2	9.4	+2.2	9.7	+2.5
Transport	15.8	11.6	-4.2	10.4	-5.4
Child costs	2.1	0.9	-1.2	0.7	-1.4
Home improvements and household goods	7.0	9.7	+2.7	10.8	+3.8
Health	2.1	2.3	+0.2	1.8	-0.3
Socialising and hobbies	26.6	14.4	-12.2	11.2	-15.4
Other goods and services	14.0	11.7	-2.3	10.3	-3.7

Expenditure on *Food and groceries* is substantially overestimated in the Spending Study compared to the LCF, which is reflected by a larger budget share (SR+DE: +14.8pp). Similarly, the budget share is larger in the Spending Study than in LCF for *Clothes and footwear* (SR+DE: +2.2pp) and *Home improvements and household goods* (SR+DE: 2.7pp). In turn, the budget share is smaller in the Spending Study for *Socialising and hobbies* (SR+DE: -12.2pp), *Transport* (SR+DE: -4.2pp), and *Other goods and services* (SR+DE: -2.3pp). Finally, the differences in budget shares are rather small for *Health* (SR+DE: +0.2pp) and *Child costs* (SR+DE: -1.2pp).

7. Conclusions

We report on a novel approach using smartphone technology to collect expenditure data in a probability household panel of the general population in Great Britain. Respondents were asked to report their purchases of goods and services, by using the built-in camera of their device to photograph all paper receipts for one month. In a separate diary section of the app, they could manually enter expenditures such as non-receipted payments. In this paper we

compare the data collected with the app with benchmark data from the Living Costs and Food Survey, the national budget survey in the United Kingdom. The results suggest that the level of total expenditure reported in the app is comparable with the benchmark data. The option to manually enter purchases in the app in addition to the receipt scanning turned out to be crucial: the scanned receipts on their own underestimate expenditure.

The category-level expenditure reported in the app is also comparable with the national budget survey, although the expenditure categories vary in their alignment with the benchmark data. The app-based expenditure data on *Clothes and footwear* and *Transport*, for example, align closely with the Living Costs and Food Survey whereas greater differences are found for categories such as *Socialising and hobbies* and *Child costs*. Similarly, the percentage of zero expenditure reported also varies by category. There are different potential reasons for these differences. First, the expenditure categories have varying likelihoods of generating paper receipts that respondents could scan with the app. Expenditure in the category *Socialising and hobbies*, for example, might be more likely to consist of regular payments made by standing order or direct debit than expenditure on *Clothes and footwear*. In a follow-up study, one option would be to add an explicit question in the app to confirm zero expenditures, especially for frequent items like food and groceries. Second, the overreporting of expenditure on *Food and groceries* and underreporting of expenditure on *Socialising and hobbies* in the app could be due to a lack of guidance on where to report food eaten outside the home, for example in restaurants (see National Household Food Acquisition and Purchase Survey. FoodAPS, 2016). Third, the benchmark data are of course themselves collected with a survey, and therefore also not entirely error free (see Eckman, 2022 for an analysis of under-reporting in the equivalent US Consumer Expenditure Survey).

We find that app-based data collection on consumer expenditure is more effective for specific subgroups of the population. For respondents who are older, male, have an above-median income, or live in rural areas, we find that the app data align more closely with the

national budget survey. These patterns might relate to differences in the familiarity with smartphone technology and willingness to adhere to the app study protocol.

Finally, the implications of measurement differences for economic estimates are likely to vary, depending on the estimates. Our examination of budget shares, for example, suggests that the estimated shares for some categories are closer to the benchmark data than others. Future research could examine the implications for other economic applications.

Overall, the receipt scanning app seems to be a promising method for collecting population-representative consumer expenditure data in a probability sample alongside individual- and household-level characteristics. The scanned receipts can provide information on people's expenditure coded to the full COICOP expenditure classifications, complemented by expenditure reports coded into higher-level expenditure categories. Although the combined app data from scanned receipts and direct entries can only capture higher-level expenditure categories, this level of detail is often sufficient for users of consumer expenditure data. A major limitation of the app study, however, is the low participation rate (Jäckle et al., 2019). Future research is needed on how to increase participation rates as well to reduce any non-participation biases in smartphone-based data collection.

8. References

Agarwal, S./Liu, C./Souleles, N. S. (2007). "The reaction of consumer spending and debt to tax rebates - Evidence from consumer credit data". *Journal of Political Economy, 115*(6), pp. 986–1019.

Aguiar, M./Hurst, E. (2007). "Life-cycle prices and production". *American Economic Review, 97*(5), pp. 1533–1559.

Ahmed, N./Brzozowski, M./Crossley, T. F. (2006). "Measurement errors in recall food consumption data". Institute for Fiscal Studies Working Paper, W06/21. London: Institute for Fiscal Studies.

Alatrista-Salas, H./Gauthier, V./Nunez-del-Prado, M./Becker, M. (2021). "Impact of natural disasters on consumer behavior: Case of the 2017 El Niño phenomenon in Peru". *PLoS ONE, 16(*1): e0244409.

Andreyeva, T./Luedicke, J./Henderson, K. E./Tripp, A. S. (2012). "Grocery store beverage choices by participants in federal food assistance and nutrition programs". *American Journal of Preventive Medici*ne*, 43*(4): pp. 411–418.

Angrisani, M./Kapteyn, A./Samek, S. (2018). "Real time measurement of household electronic financial transactions in a population representative panel". Paper presented at the 35th IARIW General Conference, Copenhagen. http://old.iariw.org/copenhagen/angrisani.pdf.

Austin, P. C. (2009). "Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples". *Statistics in Medicine, 28*: pp. 3083-3107.

Baker, S. R. (2018). "Debt and the response to household income shocks: Validation and application of linked financial account data". *Journal of Political Economy, 126*(4): pp. 1504–1557.

Barrett, G./Levell, P./Milligan, K. (2015). "A comparison of micro and macro expenditure measures across countries using differing survey methods". In Carroll, C. D./Crossley, T. F./Sabelhaus, J. (Eds.), *Improving the measurement of consumer expenditures* (pp. 263–286). Chicago, IL: University of Chicago Press.

Broda, C./Leibtag, E./Weinstein, D. E. (2009). "The role of prices in measuring the poor's living standards". *Journal of Economic Perspectives, 23*(2): pp. 77–97.

Browning, M./Crossley, T. F./Winter, J. (2014). "The measurement of household consumption expenditures". *Annual Review of Economics*, *6*: pp. 475–501.

Brzozowski, M./Crossley, T. F./Winter, J. K. (2017). "A comparison of recall and diary food expenditure data". *Food Policy, 72*: pp. 53–61.

Bulman, J./Davies, R./Carrel, O. (2017). "Living Costs and Food Survey. Technical report for survey year: April 2015 to March 2016. Great Britain and Northern Ireland". Newport: Office for National Statistics.

Carroll, C. D./Crossley, T. F./Sabelhaus, J. (2015). "Improving the measurement of consumer expenditures". Chicago, IL: University of Chicago Press.

Cullen, K./Baranowski, T./Watson, K./Nicklas, T./Fisher, J./O'Donnell, S./Baranowski, J./Islam, N./Missaghian, M. (2007). "Food category purchases vary by household education and race/ethnicity: Results from grocery receipts". *Journal of the American Dietetic Association, 107*(10): pp. 1747–1752.

d'Ardenne, J./Blake, M. (2012). "Developing expenditure questions: Findings from focus groups". Institute for Fiscal Studies Working Paper, W12/18. London: Institute for Fiscal Studies.

DeWalt, K. M./D'Angelo, S./McFadden, M./Danner, F. W./Noland, M./Kotchen, J. M. (1990). "The use of itemized register tapes for analysis of household food acquisition patterns prompted by children". *Journal of the American Dietetic Association, 90*(4): pp. 559–562.

Eckman, S. (2022). "Underreporting of purchases in the US Consumer Expenditure Survey". *Journal of Survey Statistics and Methodology, 10*(5): pp. 1148-1171.

Felgate, M./Fearne, A./DiFalco, S./Garcia Martinez, M. (2012). "Using supermarket loyalty card data to analyse the impact of promotions". *International Journal of Market Research, 54*(2): pp. 221–240.

French, S. A./Wall, M./Mitchell, N. R./Shimotsu, S. T./Welsh, E. (2009). "Annotated receipts capture household food purchases from a broad range of sources". *International Journal of Behavioral Nutrition and Physical Activity, 6*: 37.

Fricker, S./Kopp, B./Tan, L./Tourangeau, R. (2015). "A review of measurement error assessment in a U.S. household consumer expenditure survey". *Journal of Survey Statistics and Methodology, 3*(1): pp. 67–88.

Geisen, E./Richards, A./Strohm, C./Wang, J. (2011). "U.S. Consumer Expenditure Records Study".

Gelman, M./Kariv, S./Shapiro, M. D./Silverman, D./Tadelis, S. (2014). "Harnessing naturally occurring data to measure the response of spending to income". *Science, 345*(6193): pp. 212–215.

Griffith, R./Leibtag, E./Leicester, A./Nevo, A. (2009). "Consumer shopping behavior: How much do consumers save?" *Journal of Economic Perspectives, 23*(2): pp. 99–120.

Griffith, R./O'Connell, M. (2009). "The use of scanner data for research into nutrition". *Fiscal Studies, 30*(3–4): pp. 339–365.

Gross, D. B./Souleles, N. S. (2002). "Do liquidity constraints and interest rates matter for consumer behavior? Evidence from credit card data". *Quarterly Journal of Economics, 117*(1): pp. 149–185.

Hoseini, M./Valizadeh, A. (2021). "The effect of COVID-19 lockdown and the subsequent reopening on consumption in Iran". *Review of Economics of the Household, 19*(2): pp. 373–397.

Horvitz, D. G./Thompson, D. J. (1952). "A generalization of sampling without replacement from a finite universe". *Journal of the American Statistical Association, 47*: pp. 663–685.

Institute for Social and Economic Research (2021). "Understanding Society: The UK Household Longitudinal Study. Innovation Panel, Waves 1-13, User guide". Colchester: University of Essex.

Jäckle, A./Burton, J./Couper, M. P./Lessof, C. (2019). "Participation in a mobile app survey to collect expenditure data as part of a large-scale probability household panel: Coverage and participation rates and biases". *Survey Research Methods, 13*(1): pp. 23–44.

Jäckle, A./Burton, J./Wenz, A./Read, B. (2018). "Understanding Society: The UK Household Longitudinal Study. Spending Study 1, User guide". Colchester: Institute for Social and

Economic Research, University of Essex. Available from the UK Data Service at https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8749.

Jäckle, A./Couper, M. P./Gaia, A./Lessof, C. (2021). "Improving survey measurement of household finances: A review of new data sources and technologies". In P. Lynn (Ed.), *Advances in longitudinal survey methodology* (pp. 337–367). Hoboken, NJ: Wiley.

Keusch, F./Wenz, A./Conrad, F. (2022). "Do you have your smartphone with you? Behavioral barriers for measuring everyday activities with smartphone sensors". *Computers in Human Behavior, 127*: 107054.

Kuchler, T./Pagel, M. (2021). "Sticking to your plan: The role of present bias for credit card paydown". *Journal of Financial Economics, 139*(2): pp. 359–388.

Leicester, A. (2015). "The potential use of in-home scanner technology for budget surveys". In Carroll, C. D./Crossley, T. F./Sabelhaus, J. (Eds.), *Improving the measurement of consumer expenditures* (pp. 441–491). Chicago, IL: University of Chicago Press.

Leicester, A./Oldfield, Z. (2009). "Using scanner technology to collect expenditure data". *Fiscal Studies, 30*(3–4): pp. 309–337.

Luiten, A./Hox, J./de Leeuw, E. (2020). "Survey nonresponse trends and fieldwork effort in the 21st century: Results of an international study across countries and surveys". *Journal of Official Statistics, 36*(3): pp. 469–487.

Lusk, J. L./Brooks, K. (2011). "Who participates in household scanning panels?" *American Journal of Agricultural Economics, 93*(1): pp. 226–240.

Lynn, P. (2009). "Sample design for Understanding Society". Understanding Society Working Paper, 2009-01. Colchester: Institute for Social and Economic Research, University of Essex.

Maki, A./Garner, T. (2010). "Estimation of misreporting models using micro-data sets derived from the Consumer Expenditure Survey: The gap between macro and micro economic

statistics on consumer durables". *Journal of Mathematical Sciences: Advances and Applications, 4*(1): pp. 123–152.

Martin, S. L./Howell, T./Duan, Y./Walters, M. (2006). "The feasibility and utility of grocery receipt analyses for dietary assessment". *Nutrition Journal, 5*: pp. 6–12.

McWhinney, I./Champion, H. E. (1974). "The Canadian experience with recall and diary methods in consumer expenditure surveys". *Annals of Economic and Social Measurement, 3*(2): pp. 411–437.

National Household Food Acquisition and Purchase Survey. FoodAPS (2016). "User's Guide to Survey Design, Data Collection, and Overview of Datasets". Washington DC: U.S. Department of Agriculture, Economic Research Service.

National Research Council (2013). "Measuring what we spend: Toward a new Consumer Expenditure Survey". Washington DC: National Academy Press.

Neter, J./Waksberg, J. (1964). "A study of response errors in expenditures data from household interviews". *Journal of the American Statistical Association, 59*(305): pp. 18–55.

Newing, A./Clarke, G./Clarke, M. (2014). "Exploring small area demand for grocery retailers in tourist areas". *Tourism Economics, 20*(2): pp. 407–427.

Normand, S.-L. T./Landrum, M. B./Guadagnoli, E./Ayanian, J. Z./Ryan, T. J./Cleary, P. D./McNeil, B. J. (2001). "Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: A matched analysis using propensity scores". *Journal of Clinical Epidemiology, 54*(4): pp. 387–398.

Office for National Statistics (2017). "Living Costs and Food Survey, 2015-2016. User guide. Volume D: Expenditure codes". Newport: Office for National Statistics. Available from the UK Data Service at https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8210.

Office for National Statistics. Department for Environment, Food and Rural Affairs (2018). "Living Costs and Food Survey, 2016-2017". [data collection]. UK Data Service. SN: 8351. doi:10.5255/UKDA-SN-8351-1

Panzone, L./Hilton, D./Sale, L./Cohen, D. (2016). "Socio-demographics, implicit attitudes, explicit attitudes, and sustainable consumption in supermarket shopping". *Journal of Economic Psychology, 55*: pp. 77–95.

Rankin, J. W./Winett, R. A./Anderson, E. S./Bickley, P. G./Moore, J. F./Leahy, M./Harris, C. E./Gerkin, R. E. (1998). "Food purchase patterns at the supermarket and their relationship to family characteristics". *Journal of Nutrition Education and Behavior, 30*(2): pp. 81–88.

Ransley, J. K./Donnelly, J. K./Khara, T. N./Botham, H./Arnot, H./Greenwood, D. C./Cade, J. E. (2001). "The use of supermarket till receipts to determine the fat and energy intake in a UK population". *Public Health Nutrition, 4*(6): pp. 1279–1286.

Read, B. (2023). "Automated coding of data from shopping receipts for survey research". Unpublished Manuscript.

Silberstein, A. R./Scott, S. (1991). "Expenditure diary surveys and their associated errors". In Biemer, P. P./Groves, R. M./Lyberg, L. E./Mathiowetz, N. A./Sudman, S. (Eds.), *Measurement errors in surveys* (pp. 303–326). Wiley.

Smith, C./Parnell, W. R./Brown, R. C./Gray, A. R. (2013a). "Providing additional money to food-insecure households and its effect on food expenditure: A randomized controlled trial". *Public Health Nutrition, 16*(8): pp. 1507–1515.

Smith, C./Parnell, W. R./Brown, R. C./Gray, A. R. (2013b). "Balancing the diet and the budget: Food purchasing practices of food-insecure families in New Zealand". *Nutrition and Dietetics, 70*(4): pp. 278–285.

Stephens, M. (2003). "'3rd of tha month': Do social security recipients smooth consumption between checks?" *American Economic Review, 93*(1): pp. 406–422.

StataCorp (2017). "Stata Statistical Software: Release 15". College Station, TX: StataCorp LLC.

Tang, W./Aggarwal, A./Liu, Z./Acheson, M./Rehm, C. D./Moudon, A. V./Drewnowski, A. (2016). "Validating self-reported food expenditures against food store and eating-out receipts". *European Journal of Clinical Nutrition, 70*(3): pp. 352–357.

The American Association for Public Opinion Research (2016). "Standard definitions: Final dispositions of case codes and outcome rates for surveys". Ninth edition.

Tin, S. T./Ni Mhurchu, C./Bullen, C. (2007). "Supermarket sales data: Feasibility and applicability in population food and nutrition monitoring". *Nutrition Reviews, 65*(1): pp. 20–30.

Turner, R. (1961). "Inter-week variations in expenditure recorded during a two-week survey of family expenditure". *Journal of the Royal Statistical Society. Series C: Applied Statistics, 10*(3): pp. 136–146.

United Nations (2000). "Classification of Expenditure According to Purpose: COFOG, COICOP, COPNI, COPP, Series M: Miscellaneous". Statistical Papers, No. 84. New York.

University of Essex. Institute for Social and Economic Research (2022). "Understanding Society: Spending Study 1, 2016-2017". [data collection]. UK Data Service. SN: 8749. doi:10.5255/UKDA-SN-8749-1

Valliant, R./Dever, J. A. (2018). "Survey weights: A step-by-step guide to calculation". College Station, TX: Stata Press.

Weerts, S. E./Amoran, A. (2011). "Pass the fruits and vegetables! A community-university-industry partnership promotes weight loss in African American women". *Health Promotion Practice, 12*(2): pp. 252–260.

Williams, D./Brick, J. M. (2018). "Trends in U.S. face-to-face household survey nonresponse and level of effort". *Journal of Survey Statistics and Methodology, 6*(2): pp. 186–211.

Williams, T. (2019). "Living Costs and Food Survey Technical report: Financial years ending March 2017 and March 2018". Newport: Office for National Statistics.

Zhen, C./Taylor, J. L./Muth, M. K./Leibtag, E. (2009). "Understanding differences in self-reported expenditures between household scanner data and diary survey data: A comparison of Homescan and Consumer Expenditure Survey". *Review of Agricultural Economics, 31*(3): pp. 470–492.

9. Appendix

Table A1. Expenditure reports in the *Understanding Society* Spending Study in the full data collection period.

	Scanned receipt		Direct entry	
	n	%	n	%
Total submissions (n = 8,954)	5,472	61.1	3,482	38.9
Total items	34,931	100.0	3,482	100.0
Food and groceries	24,345	69.7	1,422	40.8
Clothes and footwear	881	2.5	177	5.1
Transport	472	1.4	397	11.4
Child costs	207	0.6	53	1.5
Home improvements and household goods	2,686	7.7	141	4.1
Health	311	0.9	62	1.8
Socialising and hobbies	1,800	5.2	587	16.9
Other goods and services	2,707	7.8	434	12.5
Holidays	11	0.1	28	0.8
Gifts	30	0.1	243	7.0
Item cannot be assigned to category	*1,030*	*3.0*	–	–
Item not readable on receipt	*451*	*1.3*	–	–

Note. The scanned receipts contain the following additional items: promotions or price reductions (n = 2,331), value-added tax (n = 73), and other items (n = 19). The percentages for direct entry items do not add up to 100% since respondents were able to select multiple expenditure categories per purchase.

Table A2. Logistic regression model of probability of being in the LCF sample as opposed to the *Understanding Society* Spending Study sample.

	Coeff.	(s.e.)	p	Average marginal effects
Intercept	2.84	(0.58)	0.001	–
Age	-0.02	(0.03)	0.512	-0.003
Age-squared	0.01	(0.01)	0.072	0.001
Gender (Ref: Male)				
Female	-0.49	(0.14)	0.001	-0.043
Education (Ref: No degree)				
Degree	1.00	(0.45)	0.026	-0.060
Personal monthly gross income	-0.01	(0.01)	0.001	-0.001
Household size	-0.06	(0.06)	0.267	-0.006
Urbanicity (Ref: Rural)				
Urban	-0.28	(0.17)	0.099	-0.024
Age x Degree	-0.04	(0.01)	0.001	–
McFadden's R^2	0.068			
N	2,471			

Table A3. Sample composition and standardised differences before and after inverse probability weighting.

	Before Adjustment			After Adjustment		
	LCF	Spending Study	Std. diff.	LCF	Spending Study	Std. diff.
Age	50.0	42.8	0.432	49.3	48.2	0.063
Gender	%	%		%	%	
Female	51.7	61.5	-0.197	52.7	53.8	-0.022
Male	48.3	38.5	0.197	47.3	46.2	0.022
Education	%	%		%	%	
Degree	26.3	45.8	-0.416	28.3	29.9	-0.035
No degree	73.7	54.2	0.416	71.7	70.1	0.035
Personal monthly gross income	1,545.9	1,983.0	-0.260	1,582.0	1,643.6	-0.037
Household size	2.6	3.0	-0.260	2.7	2.7	0.009
Urbanicity	%	%		%	%	
Urban	75.5	81.3	-0.142	76.1	74.9	0.026
Rural	24.5	18.7	0.142	23.9	25.1	-0.026

Table A4. Percentage of zero expenditure reported.

	LCF	Spending Study: Scan + Direct Entry	Spending Study: Scan Only
Total expenditure	0.0	0.0	9.5
Category-level expenditure			
Food and groceries	12.4	4.2	14.1
Clothes and footwear	54.8	48.9	56.1
Transport	32.4	40.8	58.8
Child costs	78.1	79.4	84.0
Home improvements and household goods	24.0	24.8	28.2
Health	62.2	68.7	72.5
Socialising and hobbies	10.0	26.0	39.7
Other goods and services	13.6	21.0	30.2

Table A5. Non-zero and zero weekly expenditure.

	LCF		Spending Study: Scan + Direct Entry					Spending Study: Scan Only				
	Mean	Median	Mean	p	Median	p	KS test p	Mean	p	Median	p	KS test p
Total expenditure	156.50	122.80	149.50	0.258	101.30	0.048	0.007	93.70	<0.001	70.10	<0.001	<0.001
Category-level expenditure												
Food and groceries	32.30	24.80	39.90	<0.001	35.30	0.014	<0.001	32.50	0.861	27.50	0.402	0.540
Clothes and footwear	12.40	0.00	12.10	0.779	0.30	0.691	0.550	9.70	0.003	0.00	1.000	0.277
Transport	22.50	10.00	19.10	0.048	1.50	<0.001	<0.001	10.80	<0.001	0.00	<0.001	<0.001
Child costs	3.90	0.00	1.10	<0.001	0.00	–	0.061	0.40	<0.001	0.00	–	0.002
Home improvements and household goods	11.00	3.50	17.40	<0.001	2.50	0.071	0.108	13.00	0.149	2.00	0.007	0.007
Health	3.50	0.00	3.10	0.325	0.00	–	0.006	1.70	<0.001	0.00	–	<0.001
Socialising and hobbies	36.70	23.30	16.00	<0.001	7.50	<0.001	<0.001	10.60	<0.001	2.00	<0.001	<0.001
Other goods and services	22.10	10.00	16.30	<0.001	4.10	<0.001	<0.001	10.00	<0.001	2.30	<0.001	<0.001

Note. Mean: p-values from t-tests. Median: Median regressions were estimated with weekly expenditure as dependent variable and the sample as independent variable. The p-values from t-tests for the "sample" coefficients are reported here. KS test = Kolmogorov-Smirnov test.

Table A6. Non-zero weekly expenditure.

	LCF		Spending Study: Scan + Direct Entry					Spending Study: Scan Only				
	Mean	Median	Mean	*p*	Median	*p*	KS test *p*	Mean	*p*	Median	*p*	KS test *p*
Total expenditure	156.50	122.80	149.50	0.258	101.30	0.048	0.007	105.00	<0.001	75.90	<0.001	<0.001
Category-level expenditure												
Food and groceries	36.80	30.40	42.80	<0.001	40.30	0.013	0.001	38.90	0.095	35.90	0.100	0.067
Clothes and footwear	27.10	16.00	23.80	0.044	14.10	0.537	0.628	22.30	0.004	14.00	0.339	0.196
Transport	33.00	22.10	35.10	0.406	20.00	0.383	0.285	29.20	0.041	19.90	0.260	0.101
Child costs	17.00	7.40	6.60	<0.001	3.50	<0.001	<0.001	3.20	<0.001	2.50	<0.001	<0.001
Home improvements and household goods	14.50	6.80	22.70	<0.001	5.00	0.070	0.033	17.70	0.075	4.70	0.045	0.009
Health	9.30	3.00	10.30	0.372	2.50	0.341	0.133	6.40	0.004	1.80	0.010	0.033
Socialising and hobbies	40.70	27.60	21.70	<0.001	12.40	<0.001	<0.001	17.50	<0.001	7.50	<0.001	<0.001
Other goods and services	25.60	13.50	21.00	0.005	7.70	<0.001	<0.001	14.60	<0.001	4.90	<0.001	<0.001

Note. Mean: *p*-values from t-tests. Median: Median regressions were estimated with weekly expenditure as dependent variable and the sample as independent variable. The *p*-values from t-tests for the "sample" coefficients are reported here. KS test = Kolmogorov-Smirnov test.

Table A7. Non-zero and zero weekly expenditure by population subgroup.

	LCF		Spending Study: Scan + Direct Entry					Spending Study: Scan Only				
	Mean	Median	Mean	*p*	Median	*p*	KS test *p*	Mean	*p*	Median	*p*	KS test *p*
Age												
16-50	152.30	118.00	127.50	0.001	89.90	0.005	0.009	76.40	<0.001	55.90	<0.001	<0.001
51-82	161.00	127.30	173.90	0.182	118.30	0.632	0.245	112.80	<0.001	75.90	<0.001	<0.001
Gender												
Male	138.20	103.70	149.00	0.195	118.20	0.426	0.458	92.20	<0.001	60.60	0.004	<0.001
Female	172.90	139.90	150.00	0.009	93.70	<0.001	<0.001	94.90	<0.001	72.30	<0.001	<0.001
Education												
No degree	141.80	109.40	131.70	0.118	82.40	0.039	0.044	79.90	<0.001	68.10	<0.001	<0.001
Degree	193.60	157.70	191.40	0.870	129.60	0.108	0.039	126.10	<0.001	79.30	<0.001	<0.001
Income												
Below median	124.00	92.40	127.20	0.735	74.00	0.004	0.005	77.70	<0.001	64.30	0.004	<0.001
Above median	187.50	153.20	168.50	0.015	134.70	0.216	0.267	107.20	<0.001	76.40	<0.001	<0.001
Household size												
Single	159.40	124.40	152.20	0.294	104.50	0.071	0.010	91.90	<0.001	69.20	<0.001	<0.001
Non-single	137.70	108.10	134.40	0.791	78.40	0.314	0.453	103.60	0.005	77.30	0.111	0.010
Urbanicity												
Rural	173.40	141.40	165.00	0.429	123.00	0.491	0.816	117.30	<0.001	84.40	0.023	0.013
Urban	151.10	117.50	144.40	0.355	89.50	0.004	0.002	85.80	<0.001	64.30	<0.001	<0.001

Note. Income: personal monthly gross income. Mean: *p*-values from t-tests. Median: Median regressions were estimated with weekly expenditure as dependent variable and the sample as independent variable. The *p*-values from t-tests for the "sample" coefficients are reported here. KS test = Kolmogorov-Smirnov test.