Understanding Society
Working Paper Series
No. 2020 - 01
January 2020

Understanding and reducing income reporting error in household surveys

Jon Burton (University of Essex)

Thomas F. Crossley (European University Institute)

Paul Fisher (University of Essex)

Alessandra Gaia (University of Milan-Biocca)

Annette Jäckle (University of Essex)

Understanding Society

UKRI Economic and Social Research Council

Non-technical summary

Income data collected in household surveys are critical for the study of living standards. Survey respondents are known to misreport their income, although the types of error are not well documented. Against this backdrop, we experiment with the use of an Editable Summary Screen (ESS) during data collection to learn about the types of error in income data and to improve the data quality. An ESS reflects back to respondents information they previously supplied by category and a total, and then allows them an opportunity to change their answers.

We test two version of the ESS in a large scale panel survey - *the Understanding Society* Innovation Panel. In both versions, respondents report on a detailed set of income questions in an individual interview. In version A, respondents get an individual ESS (Ind-ESS) at the end of the interview and the corrected individual amounts are then summed to give a total for the family. In version B, respondents are invited to take part with their partner in an additional family interview where they review their answers from their individual interviews together (Fam-ESS).

We find that many families were willing to participate (85.0 percent of all eligible families). Moreover, they made substantial revisions to their income in both versions and across a wide range of income sources. The extent and magnitude of the revisions was non-trivial. For example, 17.3 percent of participating families corrected their income and the average absolute correction was £1748 per month. The revisions were large enough to induce changes in some measures of income inequality.

We also shed light on the types of reporting errors made by survey participants by classifying the changes they made. We find that some errors that researchers worry about - slipped decimal places, misallocation by category - are not very common. But simple mis-reporting of amounts and period codes is prevalent.

We do not find that one version of the ESS is clearly better than the other. The main advantage of the Fam-ESS is that it can identify joint receipt problems, i.e. where the same income source is reported by more than one individual in a family and thus counted twice, or where each member expects the other to report a particular income source, and so neither does. Our results indicate that this is relatively common (5.6 per 100 participating families). While the Fam-ESS has a higher correction rate conditional on participants seeing the summary screen (18% vs 16%), it achieves a slightly lower unconditional correction rate (13% vs. 15%). This is because some respondents do not consent to take part in a joint interview with their partner – a requirement of the Fam-ESS. Thus the Ind-ESS improves data quality for a large share of the sample, whereas the quality improvements of the Fam-ESS are arguably larger, but for a select group of consenters.

Understanding and reducing income reporting error in household surveys

Jon Burton (University of Essex)
Thomas F. Crossley (European University Institute)
Paul Fisher (University of Essex)
Alessandra Gaia (University of Milan-Biocca)
Annette Jäckle (University of Essex)

Abstract: Household surveys remain a key data source for measuring living standards and inequality. Survey data on incomes are known to suffer from reporting and non-response errors. In this paper, we implement and test a method that attempts to reduce errors directly in the data collection process, rather than mitigating them econometrically at the analysis stage. We experimentally test two versions of an Editable Summary Screen. We find that both versions elicit improvements in data quality that are large enough to change measured inequality. The experiment also provides new evidence on the nature of errors in survey income data.

Keywords: income, measurement error, misreporting.

Acknowledgements: Author order is alphabetical. We gratefully acknowledge funding from the ESRC that support the Understanding Society Innovation Panel, grant numbers RES-586-47-0001 and RES-586-47-0002; and a grant from the ESRC Transformative Research scheme and the National Centre for Research Methods (ES/N006534/1). Crossley also acknowledges support from the ESRC-funded Research Centre on Micro-Social Change (grant number ES/L009153/1). We thank Mick Couper and participants at the 2018 Methodology of Longitudinal Surveys conference, the 2018 Research Methods festival, the 2019 Family Finance User Survey conference, and the 2019 Royal Statistical Society conference for helpful comments.

Corresponding author: Paul Fisher, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, pfishe@essex.ac.uk

1. Introduction

Income data collected as part of household surveys are critical for the study of material living standards, poverty and inequality. Survey respondents are known to misreport, or decline to report, income information. This may be because the information is not encoded in memory, or is forgotten, or is of a sensitive nature. Such data problems can affect inferences from survey data on income (Meyer and Sullivan, 2003; Meyer and Mittag 2019; Lynn *et al.*, 2012; Brewer, Etheridge, and O'Dea, 2017). Administrative income data (from tax or social security records) are an important alternative. But such data can suffer from problems of coverage (e.g. where tax-filing is not universal) and availability. Moreover, recent research has documented important measurement error in administrative income data as well (Bollinger *et al.*, 2018; Wilhelm, 2018). Perhaps for these reasons, income statistics continue to be commonly estimated on household survey data. Recent examples include Hoynes and Stabile (2019), Fetzer (2019), Liu (2019), Milligan and Schirle (2019), Yang (2018), Chetty *et al.* (2017), Arellano and Bonhomme (2017), Hoynes, Schanzenbach and Almond (2016), Blundell *et al.* (2016), Haushofer and Shapiro (2016), and Armour, Burkhauser, and Larrimore (2013)). Moreover, many countries' official estimates of the income distribution are based on survey data, e.g. the Current Population Survey in the US and the Family Resources Survey in the UK.

A particular challenge for the estimation of inequality is the presence of outliers and at the heart of the problem is the difficulty in distinguishing misreports from leverage points (Cowell and Flachaire, 2007). Bollinger and Chandra (2005) show that trimming or "winsorizing", which are common methods of dealing with extreme values in applied work, are incompatible with a wide range of plausible error-generating mechanisms and may exacerbate, rather than reduce, bias. More generally, measurement error corrections tend to rely on strong and often untestable assumptions about the nature of errors (Bound, Brown, and Mathiowetz, 2001). Similarly, statistical imputation to complete missing data can introduce problematic biases (Bollinger *et al.*, 2019; Bollinger and Hirsch, 2013; Heckman and LaFontaine, 2006). Thus, there is considerable advantage in eliminating erroneous extreme values and other errors or missingness at the data collection stage.

In this paper we report on a survey experiment which tested the use of Editable Summary Screens (ESS) in conjunction with the recommended approach to household income collection. The "gold-standard" approach to collecting income in a household survey, for most types of income, is to ask about each income source received by each individual; and then to aggregate, first over sources and then over members of the household (see the recommendations of the Canberra group (United Nations, 2011)). This can be done in individual interviews with each household member, as in *Understanding Society*: the UK Household Longitudinal Study (UKHLS), the German Socio-economic Panel (SOEP), the US Survey of Income and Program Participation (SIPP) and the Household Income and Labour Dynamics in Australia (HILDA) survey,

or with one household informant who reports on each individual household member, as in the US Panel Study of Income Dynamics (PSID). An ESS reflects back to respondents previously supplied information by category, and an aggregate, and then provides an opportunity for responses to be revised.

One mechanism by which an ESS may reduce errors is that an aggregate or total derived from responses to category-specific questions may make an error in one or more categories salient. An ESS may also facilitate the reporting of sensitive items that may otherwise go unreported, because of the possibility to add unspecified income sources at the review and revision stage. Finally, if data are aggregated over multiple respondents, an ESS may help reduce problems in the reporting of joint income sources.

Our experiment has two purposes. First, we use the corrections that survey respondents make to draw inferences about the types of errors that were present in their initial reports. Thus we provide new evidence on the nature of reporting errors in survey income data collected in a recommended and widely used format. Second, we assess the ability of different ESS versions to improve data quality.

Two previous papers report implementing an ESS in the collection of household spending data. Like household income, such data are also often collected by asking about specific categories (e.g. food, or holidays) and then aggregating over categories (and sometimes over individuals.) Hurd and Rohwedder (2013) report that a spending ESS was effective in reducing outliers and in refining reported spending in the American Life Panel (a longitudinal US web survey). Blake *et al*. (2014) report similar results for a "reconciliation question" (asking respondents to confirm or correct total spending) implemented in the sixth wave of *Understanding Society*'s Innovation Panel. To our knowledge this is the first paper to apply an ESS to income data, to use the corrections made in an ESS to draw inferences about the nature and incidence of reporting error, and to experimentally compare alternative ESS formats.[1]

We tested two version of the ESS in a large-scale, mixed-mode longitudinal household survey – Wave 9 of *the Understanding Society* Innovation Panel. Our focus is on total monthly benefit unit income, where a benefit unit is a single individual or a couple, plus any dependent children. This is the unit recognized by the UK tax and benefit system and similar constructs exist in other jurisdictions; a complex household may contain more than one benefit unit.[2] For expositional purposes, we refer to a benefit unit as a "family" henceforth.

---

[1] Past research has shown that item-level checks of responses to source-specific income questions can improve responses. Within-wave checks can reduce under-reporting and in longitudinal surveys, between-wave checks can reduce inconsistences over time (Lugtig and Jäckle, 2014.)

[2] For example, the benefit unit or "family" that we study is similar to the US Census Bureau "family" in that it excludes unrelated individuals, however, the benefit unit additionally excludes related individuals who are adults, such as, grown-up children. See: https://www.census.gov/programs-surveys/cps/technical-documentation/subject-definitions.html#family

In both arms of the experiment respondents report on a detailed set of income questions in an individual interview (the standard set of income questions in *Understanding Society*). In Version A, respondents are then asked to complete an individual ESS (Ind-ESS) at the end of the income section. In Version B, respondents are invited with their partner to take part in an additional family interview where they review their reports from their individual interviews together (Fam-ESS). For couple families, Version B of the summary screen may facilitate the correction of double-counted or under-counted income sources – i.e. where the same income source is reported by more than one individual in a family and thus counted twice when deriving aggregate family income, or where each member expects the other to report a particular income source, and so neither does. Against this, for couple families, Version B required both members of a couple to consent to participate in the family interview, as elaborated further below, and so it could be that fewer families review their income with this procedure.

We begin by documenting whether respondents are willing to review and revise their income reports using ESSs, and if so, what sources of income are revised, by how much and in which direction? Second, we carefully categorize the corrections that respondents make to study the nature of reporting error in the uncorrected income data.

We then investigate the efficacy of the Ind-ESS and Fam-ESS summary screens. This requires assessing whether revisions improve the accuracy of the reported data. Some previous studies of survey procedures (e.g. Hurd and Rohwedder (2013)) take an informal approach, assuming that reductions in variance and, in particular, the elimination of extreme values represent improvements in accuracy. Alternatively, linked administrative income data might be employed as a standard against which to assess alternative survey procedures as in Lynn *et al.* (2012) or Jäckle and Eckman (2019). The accumulating evidence of measurement error in administrative income data make such comparisons more complicated (Bollinger *et al.*, 2018; Wilhelm, 2018), and as a practical matter, we do not have linked income data for this sample and period. We take a different approach.

It is an accounting identity that net family income, suitably defined, should equal the sum of family spending and family saving (which can be negative) over the same period. If the income is measured on a receipt basis (excluding unrealized capital gains) then the saving concept should be active saving (also excluding unrealized gains). However, over a short period of a month, the active saving will be well approximated by changes in net wealth. Wave 9 of *the Understanding Society* Innovation Panel also collected family spending and changes in net wealth. This allows us to calculate the balance between the two sides of a family's accounting identity. We then use this balance to assess whether and to what extent respondents' corrections to their initial reports represent quality improvements.

Finally, as an illustrative application, we report how the revisions induced by the two ESSs affect measured income inequality in these data.

To preview our findings, most families were willing to participate in the ESS experiment (85.0%) and a substantial share of participating families revised their reported income (17.3%). The revisions induced by the ESSs provide new evidence about the nature and extent of measurement error in survey income data. First, while the literature has emphasized under-reporting of state benefits and transfers (Meyer and Mittag, 2019; Brewer, Etheridge, and O'Dea, 2017; Meyer, Mok, and Sullivan, 2015; Lynn *et al.*, 2012), we observe revisions across a range of income sources. Second, a majority (61.1%) of correcting families actually revised *down* their total income. Together, these two facts challenge the view that under-reporting by lower income households is the major reporting error of concern. Third, whilst the prevalence of errors at both the bottom (Meyer and Mittag, 2019; Brewer, Etheridge, and O'Dea, 2017) and top of the distribution (Cowell and Flachaire, 2007) is known, the corrections extended throughout the distribution. Fourth, joint reporting errors in couples are fairly prevalent (5.6 errors per 100 participating families) and this is likely a direct result of us following best practice interviewing guidelines that recommended individual level interviewing in couples (United Nations, 2011). Our evidence suggests a qualification to these guidelines is warranted. Fifth, some types of errors that researchers and analysts have worried about - slipped decimal places, misallocation by category - are not very common. However, simple mis-reporting of amounts, and period codes is prevalent. Sixth, item non-response can be reduced and is not a fixed feature of data collection. Finally, the corrections are large and large enough that they lead to statistically significant changes in measured inequality.

The flip-side of the reporting errors identified by the ESS is that they were effective at remedying them. The revisions made at the ESSs led to improvements in data quality as measured by reductions in the imbalance between the two sides of a family's accounting identity (imbalance=income-spending-saving). Correcting families had a higher absolute imbalance than all families according to their pre-corrected reports but had a similar imbalance in their corrected reports. Therefore, the ESSs were particularly effective at improving the data quality of those families who initially reported the lowest quality data. Both ESSs led to a reduction in outliers which further points to them usefully improving data quality.

In terms of which version of the ESS performed better, the Fam-ESS was slightly better in identifying errors but it suffered from lower participation due to the requirement that both members of a couple consent to the joint review of their income reports. The overall correction rate, conditional on participants seeing an ESS, was higher in the Fam-ESS (18% vs. 16%), however the unconditional correction rate was actually lower in the Fam-ESS (13% vs. 15%). Therefore, we do not find large differences in the efficacy of the different versions.

The paper proceeds as follows. Section 2 describes our experiment and data in greater detail. Section 3 presents the results and Section 4 concludes.

2. Data and Experimental Design

2.1. The Innovation Panel Survey

To study ESSs for income we implemented an experiment in Wave 9 of the *Understanding Society* Innovation Panel (University of Essex, 2019), henceforth IP9. *Understanding Society*: the UK Household Longitudinal Study (UKHLS) is a large longitudinal study of the UK population.

The Innovation Panel is a sample used to test methodological innovation in longitudinal surveys. Respondents receive the core content but experimental content replaces the rotating content of the main survey. The Innovation Panel was initially composed of 1,500 households. In addition to the original sample, at IP4 and IP7 refreshment samples were added; these refreshment samples were composed of 465 and 443 households respectively. The three samples (original sample, IP4 and IP7 refreshment samples) are stratified, clustered, probability samples of adults living in Great Britain – with the exclusion of the northernmost part of Scotland – at the time of the first wave of fieldwork for each sample. For more information on the survey design, see Lynn (2009) and Jäckle *et al.* (2019).

Individuals aged 16 and above are eligible for an adult interview, which lasts on average 40 minutes. The study aim is to interview every adult member of the household. We excluded dependent children, which are young adults aged 16-19 and in full time education (4.7% of the entire sample).[3] Sample members are invited to complete the survey annually. IP9 interviews took place in spring and summer of 2016.

IP9 had a mixed-mode design, implemented experimentally. The allocation to mode is at the household level. Approximately 2/3 of households were firstly invited to complete the interview by web ("web first" group), and non-respondents were followed-up with a face-to-face interview; the remaining 1/3 of households were invited to complete the survey by face-to-face interviews ("face-to-face first" group). At the very end of the fieldwork period, non-respondents from both groups were invited in a mop-up phase where they were given the possibility to complete the survey by web or telephone. Randomisation of allocated mode occurred at IP5 for the original sample and the IP4 refreshment sample, and at IP9 for the IP7 refreshment sample; more information can be found in Jäckle *et al*. (2019).

Understanding Society includes a detailed set of questions on individual income (henceforth 'standard income data collection' (SIDC)). These are also collected in the Innovation Panel, and are the source of data for our ESSs. While both before and after tax incomes are collected, it is always the net

[3] A very small number of respondents (31 cases, 1.4% of the sample) completed the survey by telephone (after refusing alternative modes), and another small set of interviews were administered to a proxy respondent (63 cases, 2.8%). Both of these groups were excluded from the experiment (and not allocated to one of the summary screens modules.) We also exclude 75 families (4.7% of the sample) which the fieldwork agency classified as having technical problems in the field (the share in each treatment arm is not statistically different).

amounts that are the source data for our ESSs. *Understanding Society* collects a 'current income' concept where respondents can choose the period to which their income amount relates.[4] In our ESSs we convert all income amounts to monthly equivalents.

The SIDC is considered to be of high quality, for example, it forms the basis for official UK estimates of income dynamics. The questions cover a comprehensive set of income sources that are: employee earnings, self-employee earnings, second jobs, and up to 48 unearned income sources (henceforth "other income") including social security benefits, state and private pensions, private transfers and investment income. To maximise the quality of the ESS source data, we further supplemented the SIDC with additional in-interview checks. For the interested reader, the exact details of the SIDC are provided in the appendix.

2.2 Editable Summary Screens for Income

Panel A of Appendix Figure A1 displays an example of an Ind-ESS as first presented to respondents. It displays the amounts the respondent reported for each income source in their individual interview, converted to a monthly equivalent. First presented are amounts for earnings, self-employment, and second jobs, followed by each reported income source from social security benefits, state and private pensions, private transfers, investment income etc. The bottom row of the summary screen displays the sum of the individual sources to show the total net individual income the respondent reported.

Panel A of Figure A2 shows an example Fam-ESS. The main differences with the individual summary screen is that income from employment is listed for each adult in the family and that for the other reported income sources, only the family level total is displayed (calculated as the sum of reports by all adults in their individual interviews). The final row of the summary screen displays the total family net income calculated as the sum of the income amounts reported by all adults in the family.

When respondents are presented with an ESS they are first asked whether it is an accurate summary of their income last month. If they answer "no", they can make a number of possible edits to the summary. They can: add or edit amounts from employment income; edit the amounts on any other sources they reported in the individual interviews; or add unspecified income as either "any other income" or "any other benefits" (See Panel B in Appendix Figures A1 and A2).

We hypothesise that an ESS will improve reporting through a number of mechanisms: First, it gives respondents a chance to spot and rectify straightforward reporting errors/outliers that would have otherwise gone unnoticed. Second, it will aid recall by confronting respondents with the total income implied by their individual income reports. Third, the possibility of adding amounts to unspecified income

[4] The one exception is for second job income where respondents reporting second job(s) are specifically asked about the "usual month". One reason not to collect income over a longer and fixed reference period like a year is that family composition is less likely to be stable.

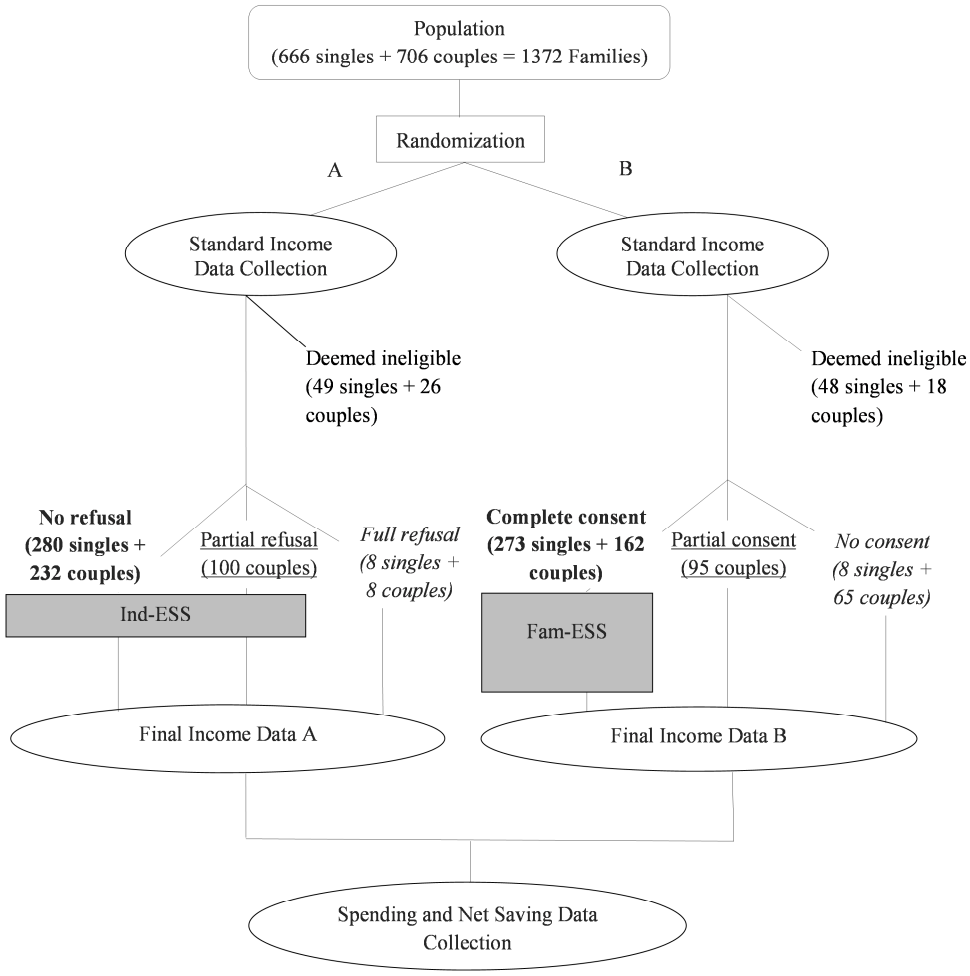
sources will facilitate the reporting of sensitive items that may otherwise go unreported. Fourth, and uniquely for the Fam-ESS, we expect the presentation of total reported income for the full family to reveal joint receipt errors to respondents i.e. where the same income source is reported by more than one individual in the family and thus counted twice when deriving family income, or where each member expects the other to report a particular income source, and so neither does.

In the face-to-face mode, our summary screens were displayed on the interviewer's laptop, which was passed to respondents. In web interviews respondents saw the summary screen on their device. In web interviews, respondents living with a spouse or partner were encouraged to review and revise the Fam-ESS with their partner/spouse; they were then asked to report whether they reviewed their answers together, or if only one person revised the summary screen (25.8% reported together). In face-to-face interviews, interviewers coded whether the summary screen was reviewed by one of the partners, or by both of them together (60.8% reported together).

2.3 Experimental Design

Figure 1 shows our experimental design. Our population is made up of 1372 families where at least one member completed an individual interview. All households were randomly allocated to two experimental groups: adults in half of households were allocated to receive an Ind-ESS; the other half were allocated to the Fam-ESS. First, all participants received the standard *Understanding Society* individual interview, including the SIDC as set out in Section 2.1. Families who refused to report any income at this stage were deemed ineligible and so removed from the experiment (141 cases). Participants were then asked to participate in the corresponding ESS - either the individual or the family version. Mostly, we analyse the design as an experiment with non-compliance. We therefore discuss non-compliance below and later also analyse it as an outcome.

Figure 1: Experimental Design



Notes: 'Partial' refers to one of a couple family. In arm B, 8 couples and 6 singles stopped during the family interview and are included in the 'no consent' group.

The Ind-ESS was automatically presented after the income questions in the individual interviews. There were two types of non-participation with the Ind-ESS. First, some survey respondents refused to participate once the summary screen was displayed. Second, in some couples one partner refused to take part in an entire individual interview and so the refusing partner did not report income or see an Ind-ESS. Of the 340 eligible couples, there were 8 where neither partner participated in an Ind-ESS and a further 100 where only one partner participated. Of the latter, 56 occurred because one partner refused to take part in an entire individual interview. Of the 288 singles, only 8 refused to participate in the Ind-ESS. Non-participation is therefore a relatively minor occurrence for families in treatment arm A.

For couples in treatment arm B, we asked each participant to consent to share with their partner/spouse the financial information that they provided during their individual interviews. Only if they

both agreed, was the Fam-ESS presented at the end of the second person's interview. In a handful of cases, consenting families stopped during the family interview and we treat these as non-consenters (8 singles and 6 couples). Of the 322 eligible couples, 160 did not consent to share financial information and so did not participate in the summary screen. Of these, 65 double non-consented and in 95 only one partner consented.

In our analysis we mostly compare net family income after the Ind-ESS for all individuals in arm A to the net family income after the Fam-ESS for all individuals in arm B, excluding the ineligibles, that is we estimate an intent to treat. Our main analysis sample therefore consists of 628 families (288 singles and 340 couples) reporting final income in arm A and 603 families (281 singles and 322 couples) reporting final income in arm B.

A final stage of the experiment collects information on the families' accounting identity (spending and saving) during the family interview. We later use the imbalance in this identity to assess whether and to what extent revisions to the income data represent improvements. As we would like to calculate the imbalance for participants in treatment arm A also, arm A participants were invited to take part in an additional family interview after completion of their individual interview, where we collected the needed information. Details of the questions on spending and net saving are included in the Appendix.

2.4 Comparing pre-treatment Income Across Groups

Although respondents were randomly allocated to the two experimental arms, randomization does not necessarily balance in finite samples and in particular it does not balance outliers. See Deaton (forthcoming) for a recent discussion of the issue. We do not want to remove outliers from the sample as an important function of an ESS is to correct outliers that are errors. We therefore compare reported income *before* the administration of the summary screens across the two randomly allocated groups, A and B and, with one exception, we do not trim the data. Distinctly, one family reported implausibly high income values on a number of state benefits (total monthly income of £144,570 per month) in Arm B. Some of the benefit income was corrected down at the ESS but many of the amounts remained implausible. While winsorizing or trimming are practices an ESS is designed to make unnecessary, in the specific context of evaluating the two ESSs, retaining this family in our analysis sample may skew the findings towards effectiveness of the ESS and of the Fam-ESS over the Ind-ESS. Moreover, this household's responses are so implausible that many editing or checking procedures followed by data providers (including *Understanding Society*) would have dealt with it in some way. They appear to be a non-cooperative household, and would not normally be included in publically provided data. Omitting this observation from the analysis, as we have done, provides a fairer basis on which to assess the ESS.

Table 1 then compares the pre-treatment income distributions of our analysis sample in the two arms of the experiment. A two-sample Kolmogorov-Smirnov test of the equality of the distributions reveals no evidence of a difference in the baseline distributions and has a p-value of 0.288. Means, quintile cut-offs, poverty shares and the Gini are never statistically different across the two.

The difference across the groups in the highest value reflects one extreme observation in arm B. In contrast to the outlier above, this family corrected to a plausible value and showed no signs of being a non-cooperative interview. The presence of this outlier means that across arm comparisons have to be made with care, and we discuss sensitivity of our results to including or excluding this observation below.

Table 1: Baseline balance

	Arm A	Arm B	difference	SE
Quintile cut-points:				
1	979	884	-95	54.46
2	1468	1350	-118	68.64
3	2084	1960	-124	113.80
4	3116	3087	-29	169.00
Highest value	16491	155194		
Gini	0.38	0.46	0.07	0.06
Poverty (share < 60% median)	0.23	0.23	-0.01	0.02
Income components (mean):				
Total	2180	2313	133	270.10
Earnings	1185	1124	-61	85.84
Self-employment	106	118	12	46.40
Second jobs	23	16	-7	6.51
Other	866	1055	189	266.60

Notes: * indicates arm B is statistically different from arm A (p<0.05). Bootstrapped standard errors.

3. Results

3.1. Consent and Participation

In this Section we establish that respondents were willing to participate in both versions of the ESSs, although the participation rates are higher for the Ind-ESS arm, than the Fam-ESS one.

Respondents were generally willing to take part in the experiment. The overall participation rate, the percentage of families where at least one person saw an ESS, was 85.0%. Table 2 provides further details of participation. Participation was always high for singles (97.2% in both arms) but substantial non-

participation occurred for couples in the Fam-ESS arm. While in 68.2% of couples in the Ind-ESS arm both partners were willing to participate, the corresponding figure for the Fam-ESS arm was only 50.2%. Overall couple participation in the Ind-ESS arm was also boosted by a substantial share of couples where only one partner participated (29.4%) while partial participation was not allowed in the Fam-ESS arm. The overall participation percentage for couples was therefore 97.6% in the Ind-ESS arm and 50.2% in the Fam-ESS arm.

The final panel of Table 2 shows that an ESS is an effective tool for identifying misreporting. When presented with a summary screen, 17.3% of families reported that their income was incorrect. While the share was similar for couples in both versions (20.5 and 24.2%), the Fam-ESS identified more errors for singles (10.7 vs 19.8%). This may have to do with the fact that the Fam-ESS formed part of a stand-alone module with its own introduction and warm-up questions. In comparison, the Ind-ESS was inserted as an addition to the existing UKHLS questionnaire and introduced with only a short motivating statement. Respondents may have therefore been better prepped to reflect on their finances in the former, relative to the latter.

Table 2: Participation

	Ind-ESS				Fam-ESS			
	Singles n=288 (45.9%)		**Couples n=340 (54.1%)**		**Singles n=281 (46.7%)**		**Couples n=321 (53.3%)**	
Participation, %	**97.2**	**(280)**	**97.6**	**(332)**	**97.2**	**(273)**	**50.2**	**(161)**
(double participation, %)	-		*68.2*	*(232)*	-		*50.2*	*(161)*
(partial participation, %)	-		*29.4*	*(100)*	-		-	
Non-participation, %	**2.8**	**(8)**	**2.4**	**(8)**	**2.8**	**(8)**	**49.8**	**(160)**
(partial non-participation, %)	-		-		-		*29.6*	*(95)*
(double non-participation, %)	-		*2.4*	*(8)*	-		*20.2*	*(65)*
Conditional on participation:								
Summary Incorrect, %	**10.7**	**(30)**	**20.5**	**(68)**	**19.8**	**(54)**	**24.2**	**(39)**
(partial, %)	-		*18.7*	*(62)*	-		-	
(double, %)	-		*1.8*	*(6)*	-		-	

Notes: Number of families in parentheses.

3.2. Frequency and Magnitude of Corrections

In this Section we establish that participants were willing to correct the reporting errors they identified at the summary screens. As in the previous subsection, we contrast differences between singles and couples and by treatment arm. We demonstrate that the corrections were large in magnitude and occurred across a wide range of income sources.

Conditional on identifying a mistake, most respondents were willing to correct their reports. The figure was 89.0% for the full sample. The figures in Table 3 show that the shares were slightly lower for the Fam-ESS (93.3% vs. 81.5% for singles and 94.1% vs 87.2% for couples). Put differently, the Fam-ESS version was better at getting respondents to identify mistakes (Section 3.1), but it was less effective at translating the identified mistakes into corrections.

Row 4 of Table 3 presents the unconditional correction rates for each arm of the experiment. For singles, the Fam-ESS showed a better performance (9.7% vs. 15.7%). However, for couples we see the reverse and the Ind-ESS performed better (18.8% vs 10.6%). The worse overall performance for couples in the Fam-ESS reflects the low participation rate it achieved for couples. It actually got more corrections than the Ind-ESS conditional on participation (19.3% vs. 21.1%)[5] but this was more than off-set by the large number of couples not willing to participate in the Fam-ESS with their partner.

We now turn to the source and size of the corrections (bottom panel, Table 3). We pool couples and singles to increase sample size. We observe a similar pattern in both treatment arms. Conditional on correcting, mean absolute total corrections are £830 and £2831 for the Ind-ESS and Fam-ESS arms, respectively. The corresponding unconditional mean absolute total corrections are £122 vs. £367. These figures compare to mean incomes in the pre-correction data of £2180 in the Ind-ESS and £2313 in the Fam-ESS (Table 1). More total corrections are negative than positive (63% and 59% of corrections are negative in the Ind-ESS and Fam-ESS arms respectively), so that more total family incomes fall then rise.

Specific corrections are seen across all income sources but the highest incidence of correction occurs in "other income": Conditional on making any correction, 71% and 64% of families correct this source. While the literature has emphasized the mis-reporting of state benefits (Meyer and Mittag 2019; Lynn *et al*., 2012), we find that 28% and 26% of correcting families correct their reported earnings. We also see that the magnitude of specific corrections is large in each source and comparable across treatment arms. The mean absolute specific correction ranges from £360 for second job in the Fam-ESS to £3390 for "other income" also in the Fam-ESS (corresponding baseline means in Table 1).

[5] Calculated as the share that report the summary as incorrect conditional on participation (Table 2) multiplied by the share that corrects the summary conditional on the summary being incorrect (Table 3).

We tested the sensitivity of our results to removing the extremely high outlier in arm B. The broad picture remains unchanged although we now see corrections that are smaller in magnitude and that are closer to the changes seen in the Ind-ESS arm (so the mean absolute correction falls from £367 to £112; the mean absolute correction conditional on correcting falls from £2831 to £876; the mean negative correction from £4218 to £904; and the mean correction in "other income" from £3390 to £330).

Table 3: Corrections

	Ind-ESS				Fam-ESS			
	Singles		**Couples**		**Singles**		**Couples**	
Conditional on summary incorrect:								
Corrects Summary, %	93.3	(28)	94.1	(64)	81.5	(44)	87.2	(34)
(partial, %)	-		*85.3*	(58)	-		-	
(double, %)	-		*8.8*	(6)	-		-	
Unconditional correction rate:	9.7	(288)	18.8	(340)	15.7	(281)	10.6	(321)
Absolute correction (all Families), £	121.60	(628)			366.78	(602)		
Conditional on correcting:								
Absolute correction (correctors), £	830.08	(92)			2830.79	(78)		
Share positive	0.35	(32)			0.41	(32)		
Share negative	0.63	(58)			0.59	(46)		
Share offsetting	0.02	(2)			0.00	(0)		
Mean positive, £	1424.02	(32)			836.66	(32)		
Mean negative, £	531.01	(58)			4218.02	(46)		
Share correcting:								
Earnings	0.28	(25)			0.26	(20)		
Self-Employment	0.07	(6)			0.17	(13)		
Second Job	0.08	(7)			0.05	(4)		
Other income	0.71	(65)			0.64	(50)		
Absolute correction, £:								
Earnings	1502.84	(25)			766.70	(20)		
Self-Employment	893.33	(6)			2777.54	(13)		
Second Job	822.43	(7)			360.00	(4)		
Other income	484.68	(65)			3389.72	(50)		

Notes: Number of observations in parentheses.

Figures 2 and 3 illustrate the effects of the summary screens on total family income by means of two scatter plots of corrected vs. original income. A substantial number of the data points are off the 45 degree line, and if corrections are interpreted as quality improvements, they show both under and over reporting of income in both treatment arms. In several cases, and in both versions, the markers are some distance from the 45 degree line and so indicate large revisions. In both versions, corrections are seen in the tails of the distribution (outliers) which may be indicative of reporting improvements. However, corrections occur not only in the tails of the distribution but throughout.

Figure 2: Original *vs*. Corrected Income (Ind-ESS)

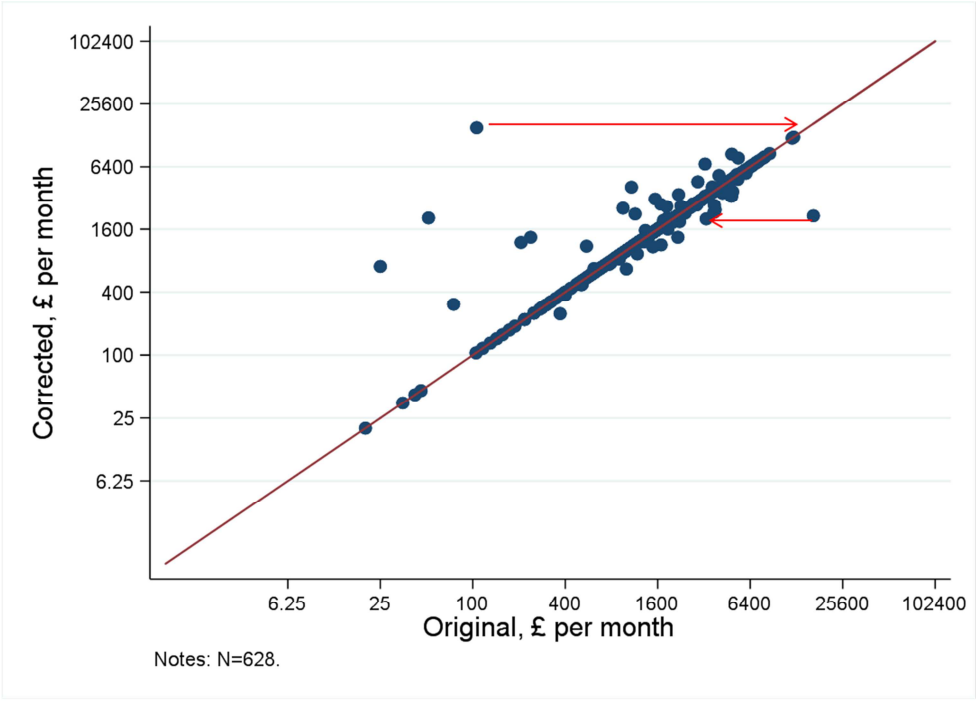
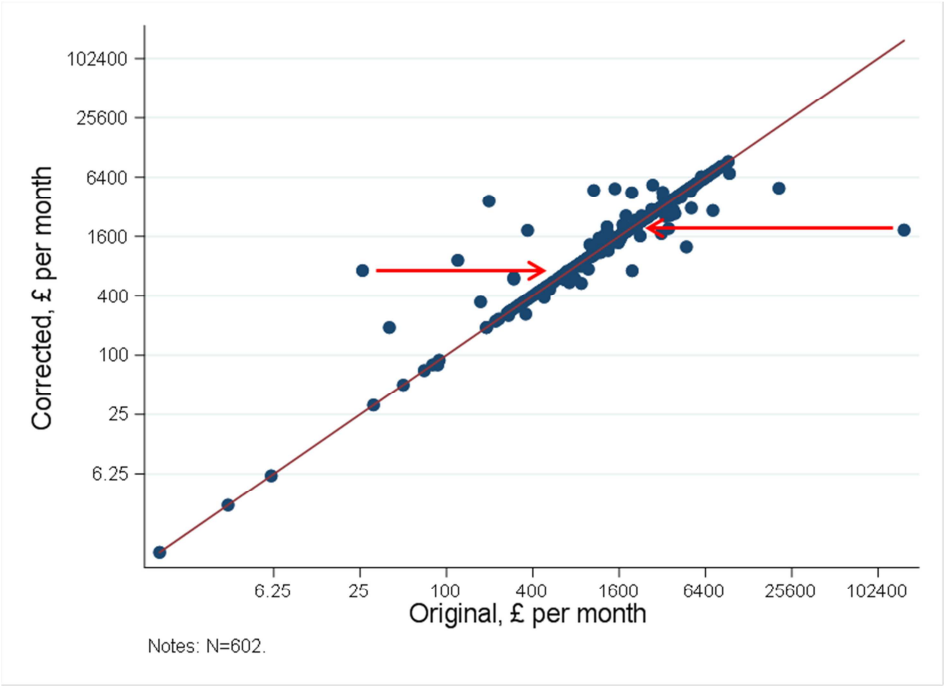


Figure 3: Original *vs*. Corrected Income (Fam-ESS)



3.3 Types of Reporting Errors in Income

We next ask what our experimental results tell us about the types and prevalence of different types of error in income data. We took the universe of corrections made by our participants and classified each correction according to one of eleven types. Table 4 lists the eleven correction types and presents the incidence of each type which we define as the number of corrections of a measurement error of this type per hundred participating families (where again participation is defined as at least one member of the family seeing an ESS). Of course, this is a lower bound on the incidence of each type, as some errors will not have been corrected.

We present the results separately for each treatment arm and by family type (single or couple) as some of our corrections ("joint receipt problem") are only possible for couples and in the Fam-ESS version of the treatment. Note, differences in the correction rates by treatment arm reflect not only the differing effectiveness of each treatment but, as participation is conditional on consent, also compositional differences between the groups participating in each ESS type.

We first focus on the overall rate of corrections in column 5. The two most common corrections were: amount changes and no clear reason (6.8 per hundred participating) and period code confused (6.2 per hundred participating). A detailed inspection of these corrections (not shown) further reveals that amount changes with no clear reason were both positive and negative in magnitude, occurred across a

variety of income sources, and period code confusion is often explained by respondents confusing a month with four weeks (accounting for 86% of the corrections in this category).

It is common for survey respondents to report an income source but then fail to report the amount that is received. Such missing data is often imputed in analysis of income and earnings but can lead to serious bias (Bollinger *et al*., 2019; Bollinger and Hirsch, 2013; Heckman and LaFontaine, 2006). Table 4 shows that our summary screens were highly effective at reducing the missing data problem (2.8 reported and corrected per 100 participating families) and it was the third most common type of correction. Moreover, almost two thirds of these corrections occurred in pay (employee, self-employee, second job).

Adding a source not reported was also a common type of correction (2.4 per 100 participating families). Almost a quarter of this category was made up of additions to an unspecified income source which points to the reduced sensitivity in reporting sensitive items as a key mechanism through which the ESSs operate.

Also notable for being uncommon is a type of error that is thought to be problematic for inequality measurement: moving a decimal place. There were no such examples of this correction type. Distinctly, switching sources is also rare (0.5 per hundred). The later included cases switching between pension types (state pension to employer pension) and switching between employment income types (self-employment to second-job income).

Joint receipt problems in couples are relatively prevalent (5.6 per hundred) but they can only be detected in the Fam-ESS arm as, to detect an error, the family needs to observe the reports of both partners at the ESS. The table pools two types of joint receipt problem: the first is where there are conflicting reports of the value of a source and the second is double /half counting where a participant essentially makes the wrong assumption about their partners' individual report. A related source of joint receipt problem is where each member of a couple expects the other to report a particular income, and so neither does. This is captured in the "added a source not reported" category, which both versions of the ESS can detect (2.7 per hundred such corrections were seen for couples in the Ind-ESS and the equivalent number for the Fam-ESS is 3.1).

Table 4: Types of Correction: Number of Corrections per Hundred Participating Families by Allocation and Family Type

	Ind-ESS		Fam-ESS		
Correction Type	**single**	**couple**	**single**	**couple**	**Total**
amount changes and no clear reason	3.6	8.7	7.7	6.8	6.8
period code confused	4.3	9.6	5.1	4.3	6.2
missing amount filled in	1.8	2.1	2.2	6.8	2.8
added a source not reported	1.1	2.7	2.9	3.1	2.4
removed a source (e.g. 150 to 0)	1.8	1.8	2.9	3.1	2.3
joint receipt problem	0.0	0.9*	0.0	5.6	0.9
switching sources (e.g. 150 to 0 and 0 to 150)	0.7	0.6	0.4	0.0	0.5
problem where multiple payments of a source	0.0	0.6	0.7	0.0	0.4
missing period code filled in	0.0	0.0	0.4	0.0	0.1
moving a decimal (e.g. 15000 to be 1500)	0.0	0.0	0.0	0.0	0.0
one-off payment corrected	0.0	0.0	0.4	0.0	0.1

Notes:

* 3 individuals who reported joint amounts divided them by 2 at the individual summary screen. We exclude these from the "total" column as it is not clear if this reflects a quality improvement or a confusion over the summary screen total. A revised version of the Ind-ESS might better address issues of joint receipt.

Some, but not all, of the extreme values in both arms were corrected down at the ESSs. Of the richest five families in each treatment arm, three singles corrected down in the Fam-ESS treatment and one couple corrected down in the Ind-ESS treatment. Two of the correcting families in the Fam-ESS arm saw large corrections in pension income that were classified as “changing an amount with no clear reason”. The third correcting family in the Fam-ESS arm corrected their self-employment pay and the correction was classified as “period code confused”. The correcting family in the Ind-ESS arm did a large revision to their child benefit income which was classified as “changing an amount with no clear reason”. Whereas the latter reporting error could be identified in data checking procedures, the former errors would have gone undetected as, in the absence of the ESS, it is not clear whether they are errors or true high values. The three corrected amounts in the Fam-ESS arm fell from £155194 to £1854 (corresponding to the outlier discussed in Section 2.4), £20581 to £4750 and £6914 to £4500. The corrected amount in the Ind-ESS arm fell from £16491 to £2150.

3.4 Effect on Data Quality

We present evidence on whether the income revisions induced by the ESSs led to data quality improvements by examining whether they brought the reported family budgets closer to balance. We construct the family budget imbalance for the last month as:

Imbalance = family income – family spending – family saving

which as an identity must equal zero. The additional family budget variables (spending and saving) were collected in the family interview for the Fam-ESS arm and families in the Ind-ESS arm were invited to take part in an additional family interview where the required variables were collected (Section 2.3). We can therefore construct the imbalance variable for participating families in both groups but only in arm A if participants agreed to take part in the additional family interview. As some participating families in arm A were not willing to complete the additional family interview, they do not appear in our comparison sample. For this reason, we may overstate the effectiveness of the Fam-ESS relative to the Ind-ESS compared to a wider comparison that included all eligible families. We also drop from our analysis families who reported zero monthly spending.[6] This leaves us with a sample of 791 families (406 in arm A and 385 in arm B).

Participant data quality improved under both treatments according to the imbalance data and a similar improvement was seen in both arms. The top panel of Table 5 shows that the mean absolute imbalance fell for participants in both the Ind-ESS and Fam-ESS arms (£1428 to £1380 and £1863 to £1380); the share of families in balance[7] increased (23.4% to 24.9% and 22.6% to 24.2%) and a positive fraction of families saw a reduction in their imbalance (8.4% and 10.6%). We checked the sensitivity of arm B to removing the outlier of Section 2.4 and the results were little changed apart from the fall in the absolute imbalance which becomes of slightly smaller magnitude (the absolute imbalance is £1466 in the pre-revision data and revised to £1381). According to these results, the revisions induced by the ESSs are consistent with data quality improvements in both arms.

The initial reports of correcting families are on average of lower quality than those of all participants in each arm, but following the revisions, the pattern reverses or there is little difference in data quality on average. The lower panel of Table 5 shows that the correctors initially had a higher mean absolute imbalance than all participating families in their respective arms (£1539 vs. £1428 in arm A and £4833 vs. £1863 in arm B) and a lower share were in balance (16.4% vs. 23.4% in arm A and 14.5% vs. 22.6% in arm B). The revisions were such that the majority of correcting families reduced imbalance (61.8%

[6] Of the 612 participating families in arm A, 159 did not provide the needed data in a family interview and 4 participants in arm B did not answer the budget balance questions during their family interview. We drop a further 92 families (47 in arm A and 45 in arm B) who reported zero monthly spending.

[7] Defined as absolute (Imbalance/income) less than 0.1.

reduced imbalance in arm A and 59.4% in arm B) and their final reports were on average of similar or higher quality than all participants in their arm. In arm A, the corrected income of the correcting families had a lower mean absolute imbalance compared to all participants in that arm (£1184 vs. £1380) and a higher share in balance compared to all participants in that arm (27.3% vs. 24.9%). In arm B, the corrected income of the correcting families had a similar share in balance compared to all participants in that arm (23.2% vs. 24.2%) but a higher mean absolute imbalance compared to all participants in that arm (£2137 vs. £1380). Put together these results indicate that the revisions at the ESSs improved data quality and particularly for those who initially reported the lowest quality data.

Table 5: Effects on Income-Spending-Saving Imbalance

	Ind-ESS		Fam-ESS	
	original	corrected	original	corrected
Participating families				
Absolute imbalance (mean)*	1428	1380	1863	1380
Share balanced†	23.4	24.9	22.6	24.2
Share reduced imbalance	-	8.4	-	10.6
N	406		385	
Correcting families				
Absolute imbalance (mean)*	1539	1184	4833	2137
Share balanced†	16.4	27.3	14.5	23.2
Share reduced imbalance	-	61.8	-	59.4
N	55		69	

Notes:
Sample of families participating in a family interview.
*Imbalance = monthly income – monthly spending – monthly saving
†Absolute (Imbalance/income) less than 0.1.

An issue to acknowledge is that improvements to the income data will improve balance unless the income and spending or saving errors are very strongly positively correlated (so that they offset in the uncorrected data but not in the corrected data – in that case improvements to one but not the other could actually reduce balance.) Correlation in measurement errors across income, spending and saving is beyond the scope of this paper but an important question for future research.

3.5 Consequence for Measured Income Inequality

Income distribution measures are commonly estimated on household survey data. This Section examines the effects of the ESSs on measured income inequality.

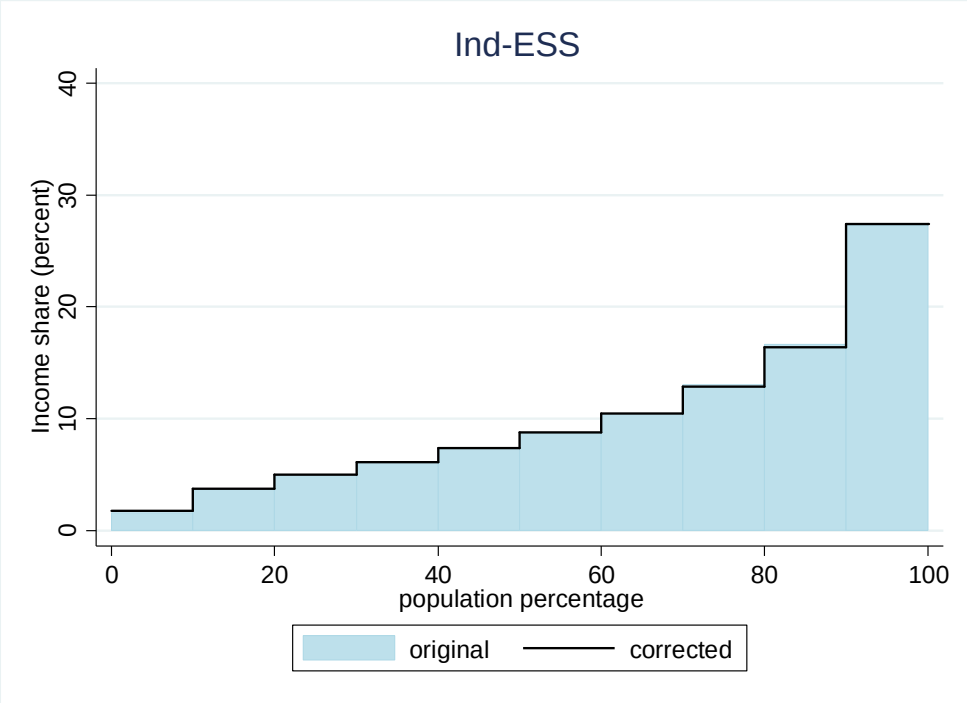
Figure 8 presents decile shares of total income before and after respondents had the chance to correct their reports at the ESSs. Such a decile plot would be useful, say, for making judgments about the distributional effects of tax reforms (a recent example of such a plot is seen in Emmerson, Farquharson, and Johnson (2019)). The upper half of the figure corresponds to the Ind-ESS treatment where we see fairly minor differences in the decile shares before and after corrections. In contrast, we see that the Fam-ESS led to substantial differences in the decile shares in the lower half of the figure. In particular, the pre-correction data indicates that the richest decile held 36% of all income, but after correction, the share falls to 26% (the beneficiaries of this change were largely covered by deciles 2-9). The corrections induced by the Fam-ESS have therefore given us a different picture of inequality.

Table 6 shows results from further inequality comparisons. It compares measures of the income distribution and inequality for the original and corrected data and for each arm of the treatment. For the Ind-ESS, we see the summary screen led to a statistically significant increase in the second decile. For the Fam-ESS, we observe a statistically significant increase in the poverty rate as defined as having a family income less than 60% of the median.[8] The changes in measured inequality remain statistically significant when removing the outlier discussed in Section 2.3.

Extreme values are highly influential in the measurement of inequality (Cowell and Flachaire, 2007) and trimming the data may exacerbate bias (Bollinger and Chandra, 2005). The bottom part of the table looks at the ability of the ESSs to remove extreme errors, and thus decontaminate inequality estimates, by reporting changes in the five highest values in each group. The table shows falls in the extreme values in both arms, with them being larger and more common in the Fam-ESS version. For example, in the Fam-ESS arm the top five values fall from £155194, £20828, £9413, £9231 and £8268 to £9231, £8268, £8250, £7957 and £7497. The ESSs therefore provide a novel way of removing extreme errors from the data while leaving extreme values belonging to the true distribution in place.

[8] This is consistent with the official UK poverty definition, although are unit of analysis is the family (benefit unit) rather than household.

Figure 8: Income Shares Before and After Correction



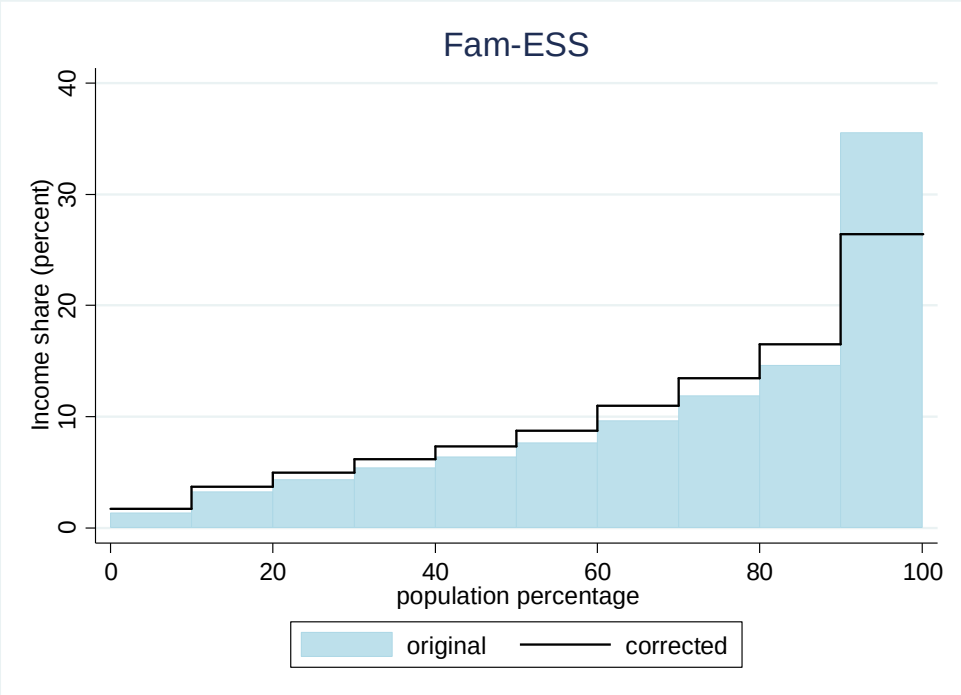


Table 6: Effect of ESS on Measured Inequality

		Ind-ESS		Fam-ESS	
		Original	Corrected	Original	Corrected
Quintiles:					
	1	979	1000	884	888.8
	2	1468	1498*	1350	1366.8
	3	2084	2089	1960	1958
	4	3116	3116	3086.8	3022
Highest value		16491	15106	155194	9231
Mean		2180	2203	2313	2035
Gini		0.38	0.38	0.46	0.38
Poverty (share < 60% median)		0.23	0.23	0.23	0.24*
Five highest values:					
	1	16491	15106	155194	9231
	2	12274	12274	20828	8268
	3	12000	12000	9413	8250
	4	8552	8552	9231	7957
	5	7950	8452	8268	7497

Notes: * indicates corrected is statistically different from the original ($p<0.05$). Bootstrapped standard errors. "Poverty" is defined as having a family income less than 60% of the median.

4. Discussion and Conclusions

We experimentally tested two versions of an editable summary screen (ESS) for family income in a household survey that includes a long-standing and detailed set of questions on respondent income. An ESS reflects back to survey respondents previously supplied information about their income and then allows them an opportunity to revise their initial responses. We tested one version of the ESS at the individual level (Ind-ESS) and the other at the family level (Fam-ESS). In both versions, most respondents were willing to participate and to revise their income.

The revisions reveal errors in the initial reports, and hence in survey income data collected according to expert recommendations (e.g., the United Nations' Canberra Group). We find that reporting errors occur in many income sources (not only, or even primarily, benefits and transfers), that there are over-reports as well as under-reports, and mis-reports occur right through the distribution (not just at the bottom or top). The revisions – and hence implied initial errors – were large in magnitude. They were large enough that their correction induced statistically significant changes in some measures of inequality. For example, the Gini coefficient in the Fam-ESS arm fell from 0.46 pre-ESS to 0.38 in the corrected data. While only marginally statistically significant, the magnitude of this fall is similar to recent US estimates of the fall

in the Gini when moving from tax unit to household measures of inequality. That fall was also considered to be of an economically important magnitude (Larrimore *et al.,* 2019). Of course, an ESS only reveals those errors that a respondent choses to correct, and so the error rates we report are a lower bound.

We proposed and implemented a method for testing whether the revisions of the ESSs represented data quality improvements that does not require linked administrative records. This method, based on the imbalance between the two sides of a families accounting identity (imbalance=income-spending-saving), may be useful in other studies. We find that the ESSs led to modest improvements in data quality as measured by reductions in this imbalance. Reductions in imbalance imply data improvements as long as errors in income and spending are not too strongly correlated. Thus understanding the correlations between measurement errors in different components of the household budget constraint remains an important topic for future research. Both ESSs also led to a reduction in outliers, which further points to them usefully improving data quality.

The ESSs were also successful at reducing non-response to income questions. Respondent confidentiality concerns lead to high non-response rates to income questions and the typical approach of analysts is to replace the missing values with imputed values such as those produced by statistical offices or data producers. Our findings point to the ESSs overcoming respondent confidentiality concerns and so the ESSs operate as an alternative to statistical imputation where respondents fill in items that they previously did not respond to, or complete totals which otherwise would have been unavailable due to item nonresponse.

We do not find that one version of the ESS is clearly better than the other. The main advantage of the Fam-ESS is that it can identify the joint receipt problems. The overall correction rate, conditional on participants seeing an ESS, is slightly higher in the Fam-ESS compared to the Ind-ESS (18% vs 16%), however the unconditional correction rate is lower in the Fam-ESS (13% vs. 15%). This is because many respondents do not consent to take part in a family interview with their partner – a requirement of the Fam-ESS, but not Ind-ESS. The Fam-ESS shows larger quality improvements (a larger mean absolute correction and larger falls in imbalance) but this, at least in part, reflects the particular realization of the randomization in our sample. In particular, there is one exceptionally high value in the Fam-ESS arm of the experiment. As dealing with such values is exactly what ESS is supposed to do, removing it would lead us to underestimate the efficacy of an ESS overall, but it does effect our comparisons across the arms (indeed removing the outlier typically reduces the improvements of the Fam-ESS relative to the Ind-ESS).

We lastly draw some best practice recommendations for data collectors or economists implementing their own surveys on income. First, an ESS can improve data quality in an economically meaningful way and so we recommend them for data collection. In particular, as the Ind-ESS is simpler to

implement than the Fam-ESS, and we do not find strong evidence that the Fam-ESS is more effective, we recommend the Ind-ESS. Second, in the absence of an ESS, it is still useful to ask for aggregates in some way because it overcomes item nonresponse and so reduces the risk of introducing biases in analysis from the use of statistical imputation. Third, joint reporting problems in couples (such as double counting) are a serious concern and the Canberra recommendation (United Nations, 2011) to collect information from each household member and for each source separately should be qualified. Whilst we support the recommendation in terms of maximising the accuracy of each individual's reports, if spouses are not coordinated in their reporting, it can directly lead to joint reporting problems. Therefore, the qualification is that survey protocols should be sensitive to this issue, either through the use of a Fam-ESS or through an alternative mechanism.

References

Arellano, M., & Bonhomme, S. (2017). Quantile Selection Models With an Application to Understanding Changes in Wage Inequality. *Econometrica, 85*, 1-28.

Armour, P., Burkhauser, R. V., & Larrimore, J. (2013). Deconstructing Income and Income Inequality Measures: A Crosswalk from Market Income to Comprehensive Income . *American Economic Review, 103*, 173-77.

Blake, M., Crossley, T., D`Ardenne, J., Oldfield, Z., & Winter, J. (2014). Testing Quick Expenditure Questions. In T. A. Baghal (Ed.), *Understanding Society Innovation Panel Wave 6: Results from Methodological Experiments.* Understanding Society Working Paper Series No. 2014-04.

Blundell, R., Costa Dias, M., Meghir, C., & Shaw, J. (2016). Female Labor Supply, Human Capital, and Welfare Reform. *Econometrica, 84*, 1705-1753.

Bollinger, C. R., & Hirsch, B. T. (2013). Is Earnings Nonresponse Ignorable? *The Review of Economics and Statistics, 95*, 407-416.

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., & Ziliak, J. P. (2018). The Good, The Bad and The Ugly: Measurement Error, Non-response and Administrative Mismatch in the CPS. *Discussion paper*.

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., & Ziliak, J. P. (2019). Trouble in the Tails? What We Know about Earnings Nonresponse 30 Years after Lillard, Smith, and Welch. *Journal of Political Economy, 127*, 2143-2185.

Bollinger, C., & Chandra, A. (2005). Iatrogenic Specification Error: A Cautionary Tale of Cleaning Data. *Journal of Labor Economics, 23*, 235-257.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement Error in Survey Data. In *Handbook of econometrics* (Vol. 5, pp. 3705-3843). Elsevier.

Brewer, M., Etheridge, B., & O'Dea, C. (2017). Why are Households that Report the Lowest Incomes So Well-off? *The Economic Journal, 127*, F24-F49.

Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R., & Narang, J. (2017). The Fading American Dream: Trends in Absolute Income Mobility Since 1940. *Science, 356*, 398-406.

Cowell, F. A., & Flachaire, E. (2007). Income Distribution and Inequality Measurement: The Problem of Extreme Values. *Journal of Econometrics, 141*, 1044-1072.

Deaton, A. (2020). *Randomization in the Tropics Revisited: A Theme and Eleven Variations.* Oxford University Press.

Emmerson, C., Farquharson, C., & Johnson, P. (2019). *The IFS Green Budget.* The Institute for Fiscal Studies.

Fetzer, T. (2019). Did Austerity Cause Brexit? *American Economic Review, 109*, 3849-3886.

Haushofer, J., & Shapiro, J. (2016). The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. *The Quarterly Journal of Economics, 131*, 1973-2042.

Heckman, J. J., & LaFontaine, P. A. (2006). Bias-corrected Estimates of GED returns. *Journal of Labor Economics, 24*, 661-700.

Hoynes, H., & Stabile, M. (2019). How Do the US and Canadian Social Safety Nets Compare for Women and Children? *Journal of Labor Economics, 37*, S253--S288.

Hoynes, H., Schanzenbach, D. W., & Almond, D. (2016). Long-Run Impacts of Childhood Access to the Safety Net. *American Economic Review, 106*, 903-34.

Hurd, M. D., & Rohwedder, S. (2013). Measuring Total Household Spending in a Monthly Internet Survey: Evidence from the American Life Panel. In *Improving the Measurement of Consumer Expenditures* (pp. 365-387). University of Chicago Press.

Jäckle, A., & Eckman, S. (2019). Is That Still the Same? Has That Changed? On the Accuracy of Measuring Change with Dependent Interviewing. *Journal of Survey Statistics and Methodology*. doi:10.1093/jssam/smz021

Jäckle, A., Al Baghal, T., Burton, J., Kaminska, O., & Lynn, P. (2019). Understanding Society The UK Household Longitudinal Study Innovation Panel, Waves 1-11, User Manual. *Colchester: University of Essex*.

Larrimore, J., Mortenson, J., & Splinter, D. (2019). Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions. *Journal of Human Resources*. doi:10.3368/jhr.56.2.0718-9647R1

Liu, K. (2019). Wage Risk and the Value of Job Mobility in Early Employment Careers. *Journal of Labor Economics, 37*, 139-185.

Lugtig, P., & Jäckle, A. (2014). Can I Just Check...? Effects of Edit Check Questions on Measurement Error and Survey Estimates. *Journal of Official Statistics, 30*, 45-62.

Lynn, P. (2009). Sample Design for Understanding Society. *Understing Society Working Paper Series*.

Lynn, P., Jäckle, A., Jenkins, S. P., & Sala, E. (2012). The impact of Questioning Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation Study. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 175*, 289-308.

Meyer, B. D., & Mittag, N. (2019). Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net. *American Economic Journal: Applied Economics, 11*, 176-204.

Meyer, B. D., & Sullivan, J. X. (2003). Measuring the Well-Being of the Poor Using Income and Consumption. *The Journal of Human Resources, 38*, 1180-1220.

Meyer, B. D., Mok, W. K., & Sullivan, J. X. (2015). The Under-Reporting of Transfers in Household Surveys: Its Nature and Consequences.

Milligan, K., & Schirle, T. (2019). Push and Pull: Disability Insurance, Regional Labor Markets, and Benefit Generosity in Canada and the United States. *Journal of Labor Economics, 37*, S289--S323.

United Nations. (2011). *Canberra Group Handbook on Household Income Statistics* (Second ed.). New York and Geneva.

University of Essex, Institute for Social and Economic Research. (2019). Understanding Society: Innovation Panel, Waves 1-11, 2008-2018. *[data collection]. 9th Edition. UK Data Service. SN: 6849*.

Wilhelm, D. (2018). Testing for the Presence of Measurement Error. *Cemmap Working Paper CWP45/18*.

Yang, T.-T. (2018). Family Labor Supply and the Timing of Cash Transfers: Evidence from the Earned Income Tax Credit. *Journal of Human Resources, 53*, 445-473.

Online Appendix /Supplemental Materials

A1. Additional Tables and Figures

Figure A1: Sample Ind-ESS

A.

Thank you for telling us about these types of income. Here is a summary of what you have told us you received last month after tax and deductions. Please take a look and select, whether this summary is accurate. If not, you will have an opportunity to update the amounts in the boxes.

HELP

Employment income: Main job	£	1300	per	month	
Employment income: Second job	£	200	per	month	
Income from self-employment	£		per	month	
Benefit and Pensions					
Carer's Allowance	£	150	per	month	
Working Tax Credit	£		per	month	
Total:	**About £**	1650	per	month	

Does this summary seem correct?

- (●) Yes
- ( ) No

B.

Thinking about your total income after taxes and deductions, please correct the summary below.

Employment income: Main job	£	1300	per	month
Employment income: Second job	£	200	per	month
Income from self-employment	£		per	month
Benefit and Pensions				
Carer's Allowance	£	150	per	month
Working Tax Credit	£		per	month
Any other benefits	£		per	month
Any other income	£		per	month

Figure A2: Sample Fam-ESS

A.

Below is a summary of the income you have reported for the last month. You reported a total of £2750 from earnings, benefits, and other income sources after taxes and deductions. Thinking about the money coming in, that is after any tax or deductions, does this seem correct? If not, you will have an opportunity to update the amounts in the boxes.

HELP

Employment income: Main job			
Matt	About £	1300	per month
Nicky	About £	1100	per month
Employment income: Second job			
Matt	About £	200	per month
Nicky	About £		per month
Income from self-employment			
Matt	About £		per month
Nicky	About £		per month
Benefit and Pensions			
Child Benefit (including Lone-Parent Child Benefit payments)	About £	150	per month
Child Tax Credit	About £		per month
Total net disposable income:	**About £**	2750	per month

Does this seem correct?

- (•) Yes
- ( ) No

B.

Thinking about your total income after taxes and deductions, please correct the summary below.

HELP

Employment income: Main job			
Matt	About £	1300	per month
Nicky	About £	1100	per month
Employment income: Second job			
Matt	About £	200	per month
Nicky	About £		per month
Income from self-employment			
Matt	About £		per month
Nicky	About £		per month
Benefit and Pensions			
Child Benefit (including Lone-Parent Child Benefit payments)	About £	150	per month
Child Tax Credit	About £		per month
Any other benefits	£	0	per month
Any other income	£	0	per month

A2. *Understanding Society* Standard Income Data Collection

Overview

Data collection of the income components occurs in different modules. A 'current employment module' establishes whether a respondent currently is in employment and if so, whether they are an employee or self-employee. An 'employee's' module asks for net (and gross) pay at last payment. Respondents are encouraged to consult a payslip if possible. To reduce reporting errors, we included an in-interview check that asked respondents to confirm their reported amounts where reported gross pay was less than or equal to reported net pay. A 'self-employment' module asks self-employees for their share of the net profit or loss on their most recent accounts or, where not available, an estimate of their net usual monthly or weekly self-employment income. For both employees and self-employee in web-mode, where a respondent failed to provide an amount, we added motivational statements which have been shown to increase respondent willingness to answer the question (Al Baghal & Lynn, 2015).

All respondents receive the 'second jobs' module that asks about net (and gross) income from any second or odd jobs in the last month.

An 'unearned income and state benefits module' collects information on whether a respondent is currently receiving any of 48 income sources; the amount received and whether it is received jointly with a partner. The income sources covered by the module include the main UK state benefits; pensions and investments; family transfers and other income streams. A list of the 48 sources is included below. As our experiment is implemented in a panel, dependent interviewing is used to reduce spurious change in reports of unearned sources between waves (Lynn *et al*., 2012). We further added in-interview checks where the pattern of state benefit reporting was suggestive of errors. Full details are provided below.

Survey questions on income

<u>Current employment module</u>

Q1) Can I just check, did you do any paid work last week - that is in the seven days ending last Sunday - either as an employee or self-employed?

Q2) Even though you weren't working did you have a job that you were away from last week?

Q3) Are you an employee or self-employed?

<u>Employee's module</u>

Q4) If an employee on Q3: The last time you were paid, what was your gross pay - that is including any overtime, bonuses, commission, tips or tax refund but before any deductions for tax, National Insurance

or pension contributions, student loan repayments, union dues and so on? How long a period did that cover?

Q5) And what was your take home pay last time, that is after any deductions were made for tax, National Insurance, pensions, student loan repayments, union dues etc? How long a period did that cover?

Self-employment module

Q6) If a self-employee on Q3: In this job/business are annual business accounts prepared for the Inland Revenue for tax purposes?

Q7) If yes to Q6): What was the amount of (your share of) the profit or loss figure shown on these accounts for this period? (And month/year accounts began and ended)

Q8) Does this figure relate to profit or loss?

Q9) Can i just check, is that figure before deduction of income tax?

Q10) Can i just check, is that figure before deduction of National Insurance?

Q11) If amount in Q7 was before tax and/or national insurance deductions: What was the amount of profit after tax and after National Insurance?

Q12) If no to Q6: After paying for any materials, equipment or goods that you use(d) in your work, what was your weekly or monthly income, on average, from this job/business over the last 12 months?

Q13) Was that weekly or monthly income?

Q14) Can i just check, is that figure before deduction of income tax?

Q15) Can i just check, is that figure before deduction of National Insurance?

Q16) What was the amount of profit after tax and after National Insurance?

Second jobs module

Q17) Do you currently earn any money from a second job, odd jobs, or from work that you might do from time to time, apart from any main job you have?

Q18) If yes to Q17: Before tax and other deductions, how much do you earn from your second and all other occasional jobs in a usual month?

Q19) If gross pay is provided at Q18: What is your take-home pay from second/odd jobs(s) in a usual month? That is, after any deductions were made for tax, National Insurance, pensions, union dues etc?

Benefits and unearned income module

There are two stages to this module. The first stage collects information on whether a respondent is currently receiving each of 48 income sources. This is done by presenting respondents with a series of showcards listing the possible income sources they might be receiving. We added in-interview checks where the pattern of state benefit reporting was suggestive of an error. For the relevant state benefits, we list the in-interview check(s) in footnotes.

At the end of stage one, respondents who fail to report a source at wave t but reported it at wave t-1 are asked:

Q20) Can I just check, according to our records, you have in the past received X. Are you currently receiving X, either just yourself or jointly?

The second stage of the module loops over each reported source to ask:

Q21) How much was the last payment of X you received (to nearest £)? What period did that cover?

Q22) Do you receive that solely in your name or jointly with someone else?

Q23) For any reported pensions at stage one: Was this amount received before tax was owed or after tax was paid?

Q24) And are you currently receiving any other payments of X? (If so, repeat Q21)

The 48 income sources collected are:

1 National Insurance retirement/state retirement (old age) pension[9]
2 a pension from a previous employer
3 a pension from a spouse's previous employer

[9] in-interview checks were added: i) if reports receiving when under state pension age, ii) if does not report receiving when of state pension age, iii) check on whether other benefit amounts are included in the reported pension amount (where other relevant state benefits are reported).

4 a private pension/annuity
5 a widow's or war widow's pension
6 widowed mother allowance / widowed parent allowance / bereavement allowance
7 pension credit (includes guarantee credit & saving credit)
8 severe disablement allowance
9 industrial injury disablement allowance
10 disability living allowance
11 attendance allowance[10]
12 carer's allowance (formerly invalid care allowance)[11]
13 war disablement pension
14 incapacity benefit
15 income support
16 job seeker's allowance
18 child benefit (including lone-parent child benefit payments)[12]
19 child tax credit
20 working tax credit (includes disabled person's tax credit)
21 maternity allowance
22 housing benefit
23 council tax benefit
24 educational grant (not student loan or tuition fee loan
25 trade union / friendly society payment
26 child maintenance, alimony or separation allowance
27 regular payments from friends or relatives outside the household
28 rent from boarders or lodgers (not family members) living here with you
29 rent from any other property
30 foster allowance / guardian allowance
31 rent rebate
32 rate rebate
33 employment and support allowance
34 return to work credit
35 sickness and accident insurance
36 in-work credit for lone parents
37 other disability related benefit or payment
38 any other regular payment
39 any other state benefit
40 universal credit
41 personal independence payments

[10] in-interview check where respondent reports it and is under the eligible age (age 65). It is possible the respondent is confusing it with a different state benefit.

[11] in-interview check on whether the respondent cares for someone. If not caring for someone, the respondent is possibly confusing it with a different state benefit such as attendance allowance.

[12] in-interview check if not reported by females who are a responsible-adult for children in the household

42 government training schemes, such as youth training allowance
43 an occupational pension from overseas, paid in foreign currency
44 an occupational pension from an overseas government or company, paid in foreign currency
45 grants from social fund or community care grant from DWP or local authority
46 an annuity (includes home income plan or equity release)
47 income as a sleeping partner in a business
48 one or more disability benefit(s), but not sure which one(s)

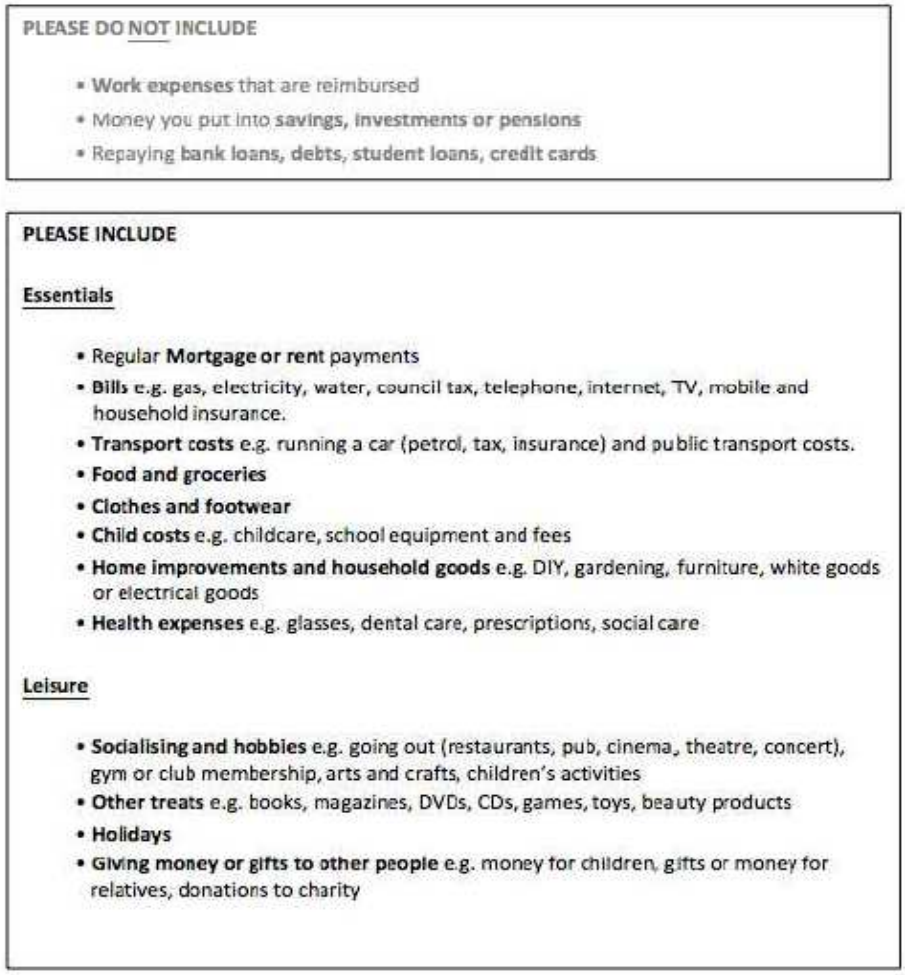
A3. Collecting the families accounting identify (spending and saving)

Overview

The intent of the spending and saving questions was to collect each concept for the same time period ("in the last month"). All families received the same spending question that was based on an experiment in an earlier wave of the Innovation Panel (Blake *et al*. (2014)). The savings (change in net financial assets) questions were experimentally varied. There were two versions with a random half of households receiving version 1 and the other half version 2.

Survey question on spending

Q25) About how much did you spend on everything in the last month? Please exclude work expenses for which you are reimbursed, money put into savings and repayment of loans. Examples of what to include and exclude are shown on this card.



PLEASE DO NOT INCLUDE

- **Work expenses** that are reimbursed
- Money you put into **savings, investments or pensions**
- Repaying **bank loans, debts, student loans, credit cards**

PLEASE INCLUDE

Essentials

- Regular **Mortgage or rent** payments
- **Bills** e.g. gas, electricity, water, council tax, telephone, internet, TV, mobile and household insurance.
- **Transport costs** e.g. running a car (petrol, tax, insurance) and public transport costs.
- **Food and groceries**
- **Clothes and footwear**
- **Child costs** e.g. childcare, school equipment and fees
- **Home improvements and household goods** e.g. DIY, gardening, furniture, white goods or electrical goods
- **Health expenses** e.g. glasses, dental care, prescriptions, social care

Leisure

- **Socialising and hobbies** e.g. going out (restaurants, pub, cinema, theatre, concert), gym or club membership, arts and crafts, children's activities
- **Other treats** e.g. books, magazines, DVDs, CDs, games, toys, beauty products
- **Holidays**
- **Giving money or gifts to other people** e.g. money for children, gifts or money for relatives, donations to charity

Survey questions on savings

Version 1: Net saving is calculated as the difference between "outgoing money" (new savings and debt repayments) and "incoming money" (increases in debts and withdrawals from savings).

Q26) (outgoing money) In the last month, how much, if any, have you deposited into savings accounts, invested in stocks and shares investments or repaid on debt?

PLEASE DO NOT INCLUDE

- Regular Mortgage or rent payments
- Pension payments
- Student loan repayments

PLEASE INCLUDE

Money put into savings

- Savings or deposit accounts (with a bank, post office or building society)
- National Savings Accounts (formerly National Savings Bank or Post Office Accounts)
- Cash ISA (Individual Savings Accounts)
- Stocks and Shares ISA (Individual Savings Accounts; formerly PEPs)
- Premium Bonds
- Children's savings accounts

Money put into investments

- National Savings Certificates / National Savings Bonds (Capital, Income or Deposit)
- Unit Trusts / Investment Trusts
- Company stocks or shares, UK or foreign
- Other investments (e.g., gilts, government or company bonds or securities, stock options)

Money repaid on debts

- Hire purchase agreements
- Credit cards, store cards
- Personal loans (from a bank, building society or other financial institution)
- Loans from a private individual (including family/friends)
- Catalogue or mail order purchase agreements
- DWP/SSA Social Fund loans
- Overdrafts

Q27) (incoming money) In the last month, how much, if any, have you withdrawn from your savings or newly borrowed?

PLEASE DO NOT INCLUDE

- Money withdrawn from debit cards

PLEASE INCLUDE

Money taken from savings

- Savings or deposit accounts (with a bank, post office or building society)
- National Savings Accounts (formerly National Savings Bank or Post Office Accounts)
- Cash ISA (Individual Savings Accounts)
- Stocks and Shares ISA (Individual Savings Accounts; formerly PEPs)
- Premium Bonds
- Children's savings accounts

Money borrowed

- Hire purchase agreements
- Credit cards, store cards
- Personal loans (from a bank, building society or other financial institution)
- Loans from a private individual (including family/friends)
- Catalogue or mail order purchase agreements
- DWP/SSA Social Fund loans
- Overdrafts
- Student loans

Version 2: A series of questions asked about the starting and ending balance in all financial assets (including debts) which were then aggregated to give net saving. The questions are as follows:

Q28) Do you have any current accounts, savings accounts, ISAs, or stocks/bonds/investments listed here?

Current account, savings accounts, ISAs, other stocks/bonds/investments

Q29) For each account type listed in Q28):

Is the overall balance across all [INSERT ACCOUNT TYPE] you have higher, lower or about the same compared with the same time last month?

Q30) For each account type with an increase/decrease in balance in Q29):

By how much did the balance increase/decrease in the last month?

Q31) Do you have any credit card(s), store card(s) or loans listed here?
Credit card, store card (not loyalty cards), personal loans (including from family and friends), pay-day loans.

Q32) For each item listed in Q31):

Is the amount you owe across all of your [INSERT ITEM] higher, lower or about the same compared with the same time last month?

Q33) For each increase or decrease in Q32):

By how much did the balance increase/decrease in the last month?

Additional Appendix References

Al Baghal, T., & Lynn, P. (2015), "Using motivational statements in web-instrument design to reduce item-missing rates in a mixed-mode context", Public Opinion Quarterly, 79(2), 568-579.