Ipsos MORI

Memorandum

To:	Emma McCallum SHS Project Manager SG
Cc:	
From:	Chris Martin SHS Project Director Ipsos MORI
Date:	02/10/2019
Subject:	**Income Imputation in the SHS: Reviews of changes to procedures**

Summary:

- The SHS income section went through a major redesign in 2018. In addition to the income of the Highest Income Household and their spouse, details of the income of up to three Other Adults were asked in order to widen the definition of household income.

- As a result, the imputation routines used in the data processing needed to be redesigned and extended. This note provides details of the routines that have been implemented.

- It also provides analysis to consider the impact of the changes on the existing measure and the new broader measures.

- The analysis suggests that changes made to the questionnaire at the start of 2018 do not appear to have had any unintended impact on this existing measure. Analysis of trends over time should be unaffected.

- With regard to the extension of the household income measure to include the income of Other Adults, the analysis is also very positive. The proportion of household income from Other Adults in the SHS in line with the proportion found in the Family Resource Survey.

- The main area of potential error in the new broader measure of household income is in the imputation of missing earnings of Other Adults. We suggest that consideration is given to amending the questionnaire to help improve this element.

- We conclude that no further adjustments are necessary made to the imputation routines for Other Adults at this stage.

Introduction

The Scottish Household Survey (SHS) provides information on a wide range of topics and allows links to be made between different policy areas at both a national and local level. While the survey wasn't designed with the primary aim of collecting robust estimates of income, income is both a topic of, and a resource for, policy analysis.

The income section in the SHS questionnaire went through a major redesign at the start of 2018. The benefits section was overhauled and new questions added to gather information on the income of (up to three) Other Adults to produce a new, broader definition of household income that is closer to the definition used in the Family Resource Survey (FRS).

The new, broader, definition of household income would include the income from the Highest Income Householder (HiH) and their spouse (the current definition of household income) as well as

- Income from Other Adult 1 (summed from 1 earnings component, 40 benefits[1] components, and 10 sources of miscellaneous income)

- Income from Other Adult 2 (again, sum of 51 components)

- Income from Other Adult 3 (again, sum of 51 components).

In order to produce the new measure, the processing routines for the income data needed to be revised and extended. We agreed that this would be undertaken in three phases.

- Phase 1: develop similar routines for Other Adults that are currently used for HiH/Spouse

- Phase 2: review the results and consider any additional adjustments that might be included.

- Phase 3: implement any other necessary adjustments.

Before summarising the routines for Other Adults and considering additional adjustments, this note examines the income components for the HiH and Spouse to confirm that the changes made in 2018 will not impact trend analysis of the old definition of household income.

Summary of imputation routines used for HiH and Spouse

Income from Earnings

The questions relating to earnings of HiH and their spouse remained completely unchanged in the 2018 Questionnaire. Overall, there are four components of income from earnings: for main jobs of the HiH and their spouse: and for other jobs of the HiH and their spouse. Table A1 in Appendix 1 summarises the level of receipt of income from earnings, the level of missing data, and the imputation routines used. Over half of households (54%) reported receiving income from earnings of HiH, and just under a third (30%) received income from earnings of the spouses of the HiH. Information on the amount received was missing (either not known or the information declined to be given) for around a quarter of these cases. Level of receipt was imputed by hierarchical hot deck imputation, the method used previously.

Income from other jobs was much less common: 2% for the HiH and 1% for the spouse of the HiH. For over half of these cases (56% and 51% respectively) the amount was missing and therefore imputed.

Income from benefits

There were two changes to the structure of the income section in the SHS that impacted on the existing measure of household income:

- 1) The list of benefits asked about was substantially revised. Overall, 40 benefits and other sources of income such as welfare loans were included in four sets.

---

[1] The number of benefits asked about increased from 39 to 40 at the start of 2018.

- 2) A 'don't know' response category was added to the initial questions on whether the respondent or their partner received any of the benefits listed.

For all benefits unchanged in 2018, the methods for imputing missing values were reviewed.

Additionally, imputation routines were developed for the new components. Generally, methods of imputation rely on modelling existing data. The more variance that these explain, the better the resulting imputation strategies will be. Where benefits have a sizeable number of recipients, more variance can be explained, meaning more nuanced routines can be employed. In contrast, components received by fewer respondents, or where receipt levels cannot be modelled, are imputed using more simple methods like imputation of the median amount[2]. Appendix 1 gives level of receipt, the number of proportion of cases with missing data, and the imputation methods used for each component. The methods employed closely follow those that have been used previously. (These are detailed in Appendix 1).

The questionnaire asked about benefits in four batches. For each of these, some respondents either did not know or preferred not to say whether any of the benefits listed were received. (Separate don't know and refusal codes had not been included previously). Overall, around 4% of respondents did not give information on receipt for each of the four lists.

The SHS has never imputed receipt of particular components of income, just the amount received. Therefore, we have continued to not impute any income for households who do not know or do not wish to disclose receipt of benefits.

Income from miscellaneous sources

Specific don't know and refusal codes were also added to the initial question that asked about receipt of any sources of miscellaneous income for the HiH or Spouse. Overall, around 4% of respondents gave such a response. Again, no imputation was undertaken for households who said they did not know or did not wish to disclose receipt of miscellaneous sources.

Apart from this, there were no changes to the section collecting income from miscellaneous sources for the HiH and their spouse where relevant. The imputation routines for all 10 components were reviewed but no changes were made. These are again detailed in Appendix 1.

Comparison of household income distribution (old definition) between 2016, 2017 and 2018

In order to check that the changes to the questionnaire and income processing for the 2018 survey did not have any unintended consequences, two measures were compared to previous years: the distribution of household income and the proportion of households where imputation was undertaken.

Table 1 shows the mean, median and selected percentiles of household income in 2015, 2016, 2017 and 2018 of the unweighted data. Overall, the year-on-year change in the income distribution between 2017 and 2018 appears to be in line with the changes in the previous two years. For example, the increase in mean household income between 2018 and 2019 was 2.1% compared to 4% between 2016 and 2017. The median increase was 2.3% between 2017 and 2018 compared to 3.2% between 2016 and 2017 and 2.7% between 2015 and 2016.

Table 1: Mean household income (old definition) and selected percentiles by year.

	2015	2016	2017	2018	Change 15-16	Change 16-17	Change 17-18
Mean	£26,790	£27,210	£28,290	£28,910	101.6%	104.0%	102.2%
P'ciles							
16.7%	£11,490	£11,700	£11,990	£12,390	101.8%	102.5%	103.3%
33.3%	£15,960	£16,400	£17,060	£17,520	102.8%	104.0%	102.7%
Median	£21,930	£22,520	£23,240	£23,780	102.7%	103.2%	102.3%
66.7%	£30,000	£30,370	£31,720	£32,250	101.2%	104.4%	101.7%
83.3%	£42,000	£42,850	£44,000	£45,500	102.0%	102.7%	103.4%

---

[2] Further detail on approaches to imputation and the fundamental principles underlying the routines used in the SHS can be found in https://www2.gov.scot/Resource/Doc/933/0079811.pdf

A similar picture emerges when looking at the extent of imputation undertaken. Overall, at least one component of income was imputed in 43.6% of households in 2018. This is very similar to 2017. There is also a similar pattern for earnings, where some imputation was undertaken for 17.3% in 2018 compared to 17.9% in 2017. There has been a slight rise in the proportion of cases where some benefits income has been imputed (from 24.5% to 27%). This might be due to the revisions to the list of benefits asked about.

Table 2: Summary of receipt of income and level of imputation undertaken by year.

	2015	2016	2017	2018
Household income (old definition)				
None imputed	52.4%	54.7%	52.5%	52.2%
Some imputed	44.3%	41.6%	43.8%	43.6%
Missing (<£25 per week)	3.4%	3.7%	3.6%	4.2%
All	100.0%	100%	100%	100%
Income from Earnings				
No earnings received	41.7%	42.5%	42.9%	43.1%
Income from earnings - no imputation	41.7%	41.0%	39.2%	39.6%
Income from earnings - some imputation	16.6%	16.4%	17.9%	17.4%
All	100.0%	100.0%	100.0%	100.0%
Income from Benefits				
No benefits received	36.6%	38.3%	37.4%	37.3%
Income from benefits - no imputation	37.5%	37.2%	38.1%	35.7%
Income from benefits - some imputation	25.9%	24.5%	24.5%	27.0%
All	100.0%	100.0%	100.0%	100.0%

Income of Other Adults

Structure of the income section for Other Adults.

The income section for Other Adults broadly follows the structure of the questions asked of the HiH and spouse. The sources of income asked are the same – income from earnings, income for 40 different benefits and income from 10 miscellaneous sources. There are, however, a couple of notable differences to the structure of the questionnaire in relation to earnings of Other Adults. Firstly, no distinction was made between earnings from main jobs and earnings from other jobs. Secondly, if the respondent was unable to provide an estimate of the exact amount of earnings, they were asked if they could give estimate earnings using bands. Figure 1 shows the follow-up questions on banded income.

Figure 1: Follow-up question asking about banded income when exact amount cannot be given.

IF YOU DON'T KNOW EXACTLY, COULD YOU LOOK AT THIS TELL ME WHAT THEIR APPROXIMATE PAY IS? JUST READ OUT THE NUMBER THAT APPLIES.

Note: This can be either before of after deductions like tax and national insurance or their gross pay or salary before deductions like tax.

	PER YEAR	PER MONTH	PER WEEK
1	Less than £5,200	Less than £433	Less than £100
2	£5,200 to £10,399	£433 to £899	£100 to £199
3	£10,400 to £15,599	£900 to £1,299	£200 to £299
4	£15,600 to £20,799	£1,300 to £1,699	£300 to £399
5	£20,800 to £25,999	£1,700 to £2,199	£400 to £499
6	£26,000 to £36,399	£2,200 to £2,999	£500 to £699

7	£36,400 to £51,999	£3,000 to £4,349	£700 to £999
8	£52,000 to £77,999	£4,350 to £6,499	£1,000 to £1,499
9	£78,000 or more	£6,500 or more	£1,500 or more

If amount given….
IS THAT AMOUNT GROSS PAY BEFORE DEDUCTIONS OR TAKE HOME PAY AFTER DEDUCTIONS LIKE TAX AND NATIONAL INSURANCE?

Gross pay	[1]
Take home pay after deduction	[2]
Don't know	[3]

Number of Other Adults in the household and receipt of different components of income.

Table 3 gives details of the number of Other Adults in the household and a summary of the information on their earnings. Overall, 17% of households have one Other Adult, 4% also have a second Other Adult, and 0.7% have three Other Adults in the household.

Table 3: Number of Other Adults in the household.

	N (households)	% of all households	N (Other Adults)
No Other Adult	8,753	83.1%	0
One or more Other Adults	1,316	16.9%	1,316
Two or more Other Adults	390	4.4%	780
Three Other Adults	73	0.7%	219
Total	10,532	100%	2,315

More than half of Other Adults, (53.7%) were receiving earnings: 55% of first Other Adults, 51% of second Other Adults and 36% of third Other Adults. (Table 4). For a further 3% of Other Adults, the respondent did not know or did not want to disclose if they were in receipt of earnings. This means that around 10% of household overall had at least one Other Adult who was earning, and 0.3% had another adult where it was not clear if they were earning.

Table 4: Households with an ' Other Adult' who was earning

	N earning	% of households	Total OAs	% of OAs
First Other Adult	1,008	9.6%	1,779	55.1%
Second Other Adult	246	2.3%	463	50.9%
Third Other Adult	28	0.3%	73	35.6%
All Other Adults	1,282	9.8%	2,315	53.7%

Other sources of income were received by far fewer Other Adults. Of the 50 components of benefit and miscellaneous income asked about, only 16 were received by more than 1% of Other Adults. These are shown in Table 5. The most commonly received components of income by Other Adults were Student loans (received by 6.9% of Other Adults), State Retirement Pensions (3.4%), Personal Independence Payments (3.3%) and Winter Fuel Payments (3.2%).

Table 5: Benefits and miscellaneous components of income received by more than 30 Other Adults in 2018.

	OAs in receipt	% of OAs	% of all households
Student loan (M8)	159	6.9%	1.5%
State Retirement Pension (C2)	78	3.4%	0.7%
Personal Independence Payments (B1)	77	3.3%	0.7%
Winter Fuel Payments (D7)	73	3.2%	0.7%
Disability Living Allowance (B2)	58	2.5%	0.6%
Employment and Support Allowance (A8)	56	2.4%	0.5%
Universal Credit (A1)	42	1.8%	0.4%
Grant (M9)	37	1.6%	0.4%
Child Benefit (A10)	36	1.6%	0.3%
Occupational/employer (non-State) pension(s) (M1)	35	1.5%	0.3%
Council Tax Reduction (A3)	32	1.4%	0.3%
Cold Weather Payments (D8)	32	1.4%	0.3%
Carers Allowance (A9)	30	1.3%	0.3%
Jobseekers Allowance (A7)	25	1.1%	0.2%
Regular non-work income, from any other source (please specify) (M10)	23	1.0%	0.2%

There is some evidence that levels of receipt of different benefits might be understated for Other Adults. This is difficult to quantify for almost all benefits as it is impossible to tell if someone is eligible to receive a benefit. However, Table 6 compares receipt of State Retirement Pension among HiH 65 and over and Other Adults 65 and over. While 90% of HiH aged 65 plus are in receipt of this benefit, the figure drops to 78% of Other Adults aged 65 plus.

Table 6: Level of receipt of State Retirement Pension among those aged 65 and over: comparison of Highest Income Householders and Other Adults.

	HiH aged 65 plus	OAs aged 65 plus
Information on receipt not known or refused	4.4%	6.9%
Not received	5.5%	14.9%
Received	90.1%	78.2%
N	2,935	87

Ipsos MORI

Summary of imputation routines for Other Adults

Imputation routines were developed for all elements of income of Other Adults. The overall approach used follows that of the routines used for the HiH and Spouse, namely:

- All variables that hold the income information for each component were created.

- Linked flag variables for each component were created. These indicate if each is…

    o Not received;

    o Received, and amount is given by the respondent;

    o Or received, but amount is not given and needs to be imputed.

- All amounts changed to annual equivalents.

- All amounts for each component inspected for any clearly wrong values.

- Remaining outliers from use in the imputation excluded.

- Missing values imputed separately for each component using a range of methods…

    o …from simple methods such as imputation of the median value…

    o …to more complex methods such as hierarchical hot-deck imputation. (For example, for HiH earnings from main job. In the first round the imputation groups used are based on SIC, NSSEC, whether in receipt of means-tested benefits, sex, full-time or part-time work, and tenure. If no match is found, they go into the pot for a second round of imputation where all fields except one (tenure) are used. The process is repeated several times with fewer variables included until all cases with a missing income have a match and a value is imputed).

- Different components summed together to create the final income variable and associated banded variables produced.

Full details of all the imputation routines developed and used for Other Adults are provided in Appendix 2.

For earnings of Other Adults, the banded income question was also used to impute a value where possible. For cases where the respondent said the Other Adult was earning but could not give any indication of the level of earnings, imputation was undertaken by hot deck imputation with imputation groups based on age, sex, economic status, relationship to the head of the household, household type, and whether in receipt of any means-tested benefit. (This is discussed further in the conclusions).

Most components of benefits and miscellaneous income were received by a very small number of cases. As this makes modelling difficult and they have a small effect on the final estimates of income, most were imputed using the median value. (All components that were imputed using this method had 20 cases or less with missing data.) For five components, hot deck imputation was used:

    o Council Tax Reduction (with hot-deck groups based on number of rooms, SIMD, whether HiH/Spouse also in receipt), imputed for 24 cases.

    o Personal Independence Payments (with hot-deck groups based on receipt of means-tested benefit, number of disabled people, rate and type received), imputed for 30 cases.

    o Disability Living Allowance (with hot-deck groups based on receipt of means-tested benefit, number of disabled people, rate received), imputed for 24 cases.

    o State retirement pension (with hot-deck groups based on receipt of means-tested benefit, number of retired in household), imputed for 34 cases.

    o Student loan (with hot-deck groups based on age, sex, economic status), imputed for 90 cases.

Comparison between SHS 2018 and FRS data.

In order to assess the likely level of any under-reporting of income for Other Adults in the SHS, a number of comparisons were made between SHS 2018 data and FRS data 2017/2018[3].

It should be noted that the definition of "Other Adults" is slightly different between the two surveys. In the SHS, these are defined as people who are not the Highest Income Householder (HiH) or their spouse. In the FRS, "Other Adults" are those not in the first Benefit Unit. Overall, the proportion of "Other Adults" in the SHS and FRS is similar. In the SHS, 16.9% of households contained an 'Other Adult', while in the FRS, the correspondent figure was 15.1% in Scotland (and 18% in the UK).

There are several other differences between the surveys:

- The reporting period is different. The SHS is organised by calendar year while the FRS uses financial year.

- The coverage of sources of income is slightly different. Most importantly, the FRS includes income from children. Where possible the comparisons have excluded child income. (Children's income tends to account for a very small proportion of household income and is unlikely to have a major impact on comparisons).

- The FRS attempts to interview all adult members of the household, whereas the SHS only interviews a single household respondent. Additionally, while the FRS is on average 90 minutes long with the primary purpose of measuring household income, the SHS is a multi-purpose survey and the income section in the SHS is around 5-10 minutes long.

However, for the purpose of assessing any under-measurement of income in the SHS data, these differences should have minimal impact.

Table 7 shows the proportion of household income from Other Adults among all households and among households that contain Other Adults. The SHS figures are very much in line with the FRS figures. Overall, 7.7% of income in the SHS comes from Other Adults. The corresponding figure for the FRS in Scotland is 6.5%. In households with an "Other Adult", the overall proportion of household income from Other Adults is 29.9% in the SHS and 31.8% in the FRS in Scotland.

Table 7: The proportion of household income[4] from Other Adults in the FRS and SHS.

	FRS - UK	FRS - Scotland	SHS
Among all households	9%	6.5%	7.7%
Among all households with an Other Adult	36.7	31.8%	29.9%

Table 8 shows a comparison of the breakdown of household income between earnings and other sources of income. Again, the SHS figures are very similar to the FRS figures. Compared to the FRS Scotland figures, the proportion of all household income from earnings of Other Adults is slightly higher in the SHS (6.6% compared to 4.7%) and the proportion from other sources for Other Adults is slightly lower (1.3% compared to 1.7%). This difference is also reflected in the HiH/Spouse/BU1 figures, although the differences are very small.

---

[3] Figures for the FRS were provided by Maike Waldmann of the Scottish Government.

[4] Figures for SHS are unweighted. Firgures for FRS exclude income from children.

Table 8: Breakdown of household income.

	FRS – UK	FRS - Scotland	SHS
Earnings – BU1/HiH&Spouse	62%	62.7%	66.5%
Other income - BU1/HiH&Spouse	29%	30.8%	27.5%
Earnings – Other BUs/Other Adults	7%	4.7%	6.6%
Other income – Other BUs/Other Adults	2%	1.7%	1.3%
Total	100%	100%	100%
N	19,040	2,755	10,120

Finally, analysis was undertaken by household type (Table 9). For most types of household, the proportion of household income from Other Adults was similar between the FRS[5] and the SHS. For example, in households with three or more adults and no children, 37.9% of household income is associated with second and subsequent Benefits Units in the FRS. For the SHS, 33.3% of income came from Other Adults.

For very large households, with three or more adults living with either two children or three or more children, it does appear that the proportion of income from Other Adults is lower in the SHS than in the FRS (9.2% and 11.3% compared to 29.1% and 31.5%). While the SHS figures are based on small sample sizes and some of the difference may be due to definitional issues around 'Other Adults', this might indicate that the income from Other Adults is somewhat underestimated in these types of household in the SHS compared to the FRS.

Table 9: Proportion of household income from BU2+ (FRS, UK, 2015-2018) compared to proportion from Other Adults (SHS, Scotland, 2018) by household type.

	FRS - % of income from second and subsequent BUs	SHS - % of income from Other Adults	N (FRS 15-18 UK pooled)	N (SHS 18)
One adult, no children, over pension age	-	-	9,028	1621
One adult, no children, under pension age	-	-	8,875	2090
Two adults, no children, both over pension age	0.7%	0.7%	7,439	1194
Two adults, no children, one over pension age	12.3%	12.4%	2,551	444
Two adults, no children, both under pension age	4.3%	3.8%	9,584	2238
Three or more adults, no children	37.9%	33.3%	3,497	774
One adult, one child	-	-	1,835	194
One adult, two children	-	-	1,362	131
One adult, three or more children	-	-	673	30
Two adults, one child	1.7%	2.0%	4,198	500
Two adults, two children	0.4%	0.6%	4,882	419
Two adults, three or more children	0.3%	0.5%	2,061	100
Three or more adults, one child	27.8%	27.9%	1,022	215
Three or more adults, two children	29.1%	9.2%	411	123
Three or more adults, three or more children	31.5%	11.3%	173	48
Total	8.7%	7.7%	57,591	10121

---

[5] To increase the sample sizes, the analysis is based on FRS figures for 3 years across the whole of the UK.

Indeed, when looking at the proportion of household income from these types of household (rather than the within household split), there is very little difference between the FRS and the SHS (Table 10). Overall, 1.5% of household income in the FRS comes from households with 3 or more adults and two children. The corresponding figure for the SHS is 2.2%. For households with three or more adults and three or more children, the FRS figure is 0.6% of all income, and 0.7% in the SHS. This suggests that the differences seen in Table 9 in these two household types may be due less to an underestimation of household income for Other Adults in the SHS but instead to the split between the income from the first Benefit Unit and 'Other Adults'.

Table 10: Breakdown of Household income by household types.

	FRS			SHS		
	BU1	BU2+	All	HiH/Sp	OAs	All
One adult, no children, over pension age	7%		6.4%	8%		7.1%
One adult, no children, under pension age	10%		8.9%	15%		14.2%
Two adults, no children, both over pension age	11%	1%	9.8%	9%	1%	8.4%
Two adults, no children, one over pension age	4%	6%	4.0%	4%	6%	3.7%
Two adults, no children, both under pension age	23%	11%	22.4%	31%	15%	29.7%
Three or more adults, no children	9%	60%	13.8%	9%	57%	13.1%
One adult, one child	2%		1.6%	2%		1.4%
One adult, two children	1%		1.2%	1%		1.2%
One adult, three or more children	1%		0.6%	0%		0.3%
Two adults, one child	11%	2%	9.9%	7%	2%	6.5%
Two adults, two children	13%	1%	11.6%	6%	0%	5.8%
Two adults, three or more children	4%	0%	3.9%	1%	0%	1.2%
Three or more adults, one child	3%	12%	3.9%	3%	15%	3.2%
Three or more adults, two children	1%	5%	1.5%	2%	3%	2.2%
Three or more adults, three or more children	0%	2%	0.6%	1%	1%	0.7%
All	100%	100%	100%	100%	100%	100.0%

Conclusions and recommendations for Stage 3

The results of the analysis examining the *existing* measure of income are reassuring. The changes made to the questionnaire at the start of 2018 do not appear to have had any unintended impact on this measure and analysis of trends over time should be unaffected by the changes.

With regard to the extension of the household income measure to include the income of Other Adults, the analysis is also very positive. The proportion of household income from Other Adults in the SHS in line with the proportion found in the FRS.

The main area of potential error in the new broader measure of household income is in the imputation of missing earnings of Other Adults. We suggest that consideration is given to amending the questionnaire to help improve this element.

Methods of imputation almost always rely on modelling existing data. The more variance that these models explain, the better the resulting imputation strategies will be. There are two reasons why the imputation procedures are less effective for the earnings of Other Adults than those for the earnings of the HiH and the spouse.

Firstly, more respondents are unsure of whether Other Adults receive components of income (or are unwilling to disclose this information) especially when these are unrelated. For Other Adults who were (or might have benen) earning, an estimate of earnings was provided by the respondent for only 34% of cases. This is much lower than for earnings of HiHs

and their spouses where around three-quarters have estimates provided. Overall, around half of the earnings of Other Adults had to be imputed.

Secondly, information on the characteristics of Other Adults is limited and much less extensive than for the HIH. The key variables that exist are from the household grid section: age, sex, economic status, and relationship to other household members. We do not have information on their occupation or any other information about their employment. This means that the ability to model the given data to structure the imputation routines for the missing data is limited.

In the development of the imputation routines, a regression model of income was computed to help choose what factors best account for differences in income [6]. The adjusted R^2 of the model for Other Adult's earnings was only 0.14. In other words, only 14% of the variance can be explained by these factors. Given that earnings account for most of the income of Other Adults, and that the level of earnings for over half of Other Adults needs to be imputed, the more we can predict earnings level, the more effective the imputation techniques will be.

As such, consideration should be given to incorporating additional questions into the questionnaire that help to model (and impute) earnings of Other Adults where missing, namely number of hours worked and the questions used to calculate NS-SEC. This would allow a much greater amount of variance in income levels to be explained and incorporated into the imputation routines. For example, it would allow us to use the income of someone in a higher managerial position to impute a missing income of a person in a similar occupation. Currently it is not possible to distinguish such a person from, for example, someone in routine agricultural work. This would involve additional questions for all workers who are in the household. Three of these are open questions that would also need to be coded[7].

Apart from this, we suggest that no further adjustments are made at this stage to the imputation routines for Other Adults.

Finally, we understand that the Scottish Government is exploring the possibility of data linkage between SHS data and HMRC records. If feasible, this would enable analysis of levels of under-reporting of benefits and whether imputation of receipt of benefits would improve the income measures.

---

[6] Income distributions are almost always positively skewed, with a long right tail. In other words, there is a wider range of values above the median than below it. As regression models assume that the distribution of the dependent variable is normal, the model was run on the square root of earnings. This significantly reduced the skew.

[7] Namely, name/title of job, description of job, description of firm/organisation.

Appendix 1: Details of imputation routines used for HiH and spouse

Table A1.1: Imputation of earning for HiH and Spouse

	% households receive	N households receive	Total with amount given	Total with missing data and imputed	% missing	Method of imputation	Key variables
Main earnings - HIH	54.4%	5728	4,312	1416	25%	Heirarchical hot deck	Banded age, Sex, Full-time/Part time, Receipt of means tested benefit, Social Industrial Classification, NS-SEC
Main earnings - Spouse	30.3%	3193	2355	838	26%	Heirarchical hot deck	Banded age, Sex, Full-time/Part time, Receipt of means tested benefit, Employed/self-employed, Tenure, No. of rooms
Other earnings - HiH	1.8%	186	82	104	56%	Heirarchical hot deck	Banded age, Sex, Full-time/Part time, Receipt of means tested benefit, Number of rooms, Access to a car, Tenure
Other earnings - Spouse	0.9%	99	49	50	51%	Heirarchical hot deck	Banded age, Sex, Receipt of means tested benefit, Employed/self-employed, Tenure, No. of rooms, Household type

Table A1.2: Imputation of benefits for HiH and Spouse: First group of benefits

	H'holds in receipt (%)	H'holds in receipt (N)	Total with amount given	Total with missing data	% missing	Method of imputation	Key variables
Universal Credit (A1)	1.9%	201	150	51	25%	Heirarchical hot deck	Income from jobs, Income from other benefits, household type, number of disabled people in household
Housing benefit (A2)	11.4%	1198	674	524	44%	Heirarchical hot deck	SIMD, banded income from jobs, number of rooms
Council Tax Reduction (A3)	21.5%	2261	801	1460	65%	Heirarchical hot deck	SIMD, Number of rooms, whether on means tested benefits
Working Tax Credit (A4)	3.1%	323	213	110	34%	Hot deck	Income from other benefits
Child Tax Credit (A5)	5.7%	604	424	180	30%	Hot deck	Number of children
Income Support (A6)	1.9%	204	139	65	32%	Hot deck	Income from other benefits
Jobseekers Allowance (A7)	1.0%	102	84	18	18%	Median	
Employment and Support Allowance (A8)	4.7%	491	379	112	23%	Hierarchical hot deck	Income from other benefits, type received
Carers Allowance (A9)	2.6%	278	188	90	32%	Median	
Child Benefit (A10)	17.8%	1875	1602	273	15%	Hierarchical hot deck	Number of children, whether single parent household, on means-tested benefits
Guardians Allowance (A11)	0.0%	5	3	2	40%	Median	
Maternity Allowance (A12)	0.2%	19	14	5	26%	Median	
Statutory Maternity/Paternity pay, Statutory Adoption Pay (A13)	0.2%	25	17	8	32%	Median	
Statutory sick pay (A14)	0.2%	17	13	4	24%	Median	

Table A1.3: Imputation of benefits for HiH and Spouse: Second group of benefits

	H'holds in receipt (%)	H'holds in receipt (N)	Total with amount given	Total with missing data	% missing	Method of imputation	Key variables
Personal Independence Payments (B1)	5.2%	548	379	169	31%	Hot deck	Type and rate received
Disability Living Allowance (B2)	5.8%	616	331	285	46%	Hot deck	Type of DLA and rate received
Attendance allowance (B3)	1.9%	202	130	72	36%	Hot deck	Rate received
Severe disablement allowance (B4)	0.5%	50	13	37	74%	Median	
Incapacity benefit (B5)	0.4%	46	15	31	67%	Median	
Industrial Injuries Disablement Benefit (B6)	0.4%	44	37	7	16%	Median	

Table A1.4: Imputation of benefits for HiH and Spouse: Third group of benefits

	H'holds in receipt (%)	H'holds in receipt (N)	Total with amount given	Total with missing data	% missing	Method of imputation	Key variables
Pension Credit (C1)	3.2%	340	175	165	49%	Hot deck	Number of pensioners, Age
State Retirement Pension (C2)	29.9%	3150	2324	826	26%	Hot deck	No. of retired adults, Age
Widows Pension, Bereavement Allowance, or Widowed Parents Allowance (C3)	0.9%	96	34	62	65%	Hot deck	Banded age
Armed Forces Compensation Scheme (C4)	0.1%	10	6	4	40%	Median value	
War Widows/Widower's Pension (C5)	0.0%	3	2	1	33%	Median value	

Table A1.5: Imputation of benefits for HiH and Spouse: First group of benefits

	H'holds in receipt (%)	H'holds in receipt (N)	Total with amount given	Total with missing data	% missing	Method of imputation	Key variables
Funeral Expenses Payment (D1)	0.1%	13	5	8	62%	Median value	
Sure Start Maternity Grant (D2)	0.0%	4	3	1	25%	Median value	
Best Start Grant (D3)	0.0%	0	0	0	-		
Discretionary Housing Payment (D4)	0.4%	46	19	27	59%	Median value	
Loan or grant from DWP (D5)	0.4%	44	24	20	45%	Median value	
Loan or grant from Local Authority (D6	0.1%	9	5	4	44%	Median value	
Winter Fuel Payments (D7)	28.5%	3005	2800	205	7%	Hot deck	Number of pensioners, Age
Cold Weather Payments (D8)	5.5%	574	404	170	30%	Hot deck	None
Extended payment of Housing Benefit (D9)	0.1%	7	1	6	86%	Median value	
Bereavement Payment (D10)	0.2%	18	11	7	39%	Median value	
Return to Work Payment (D11)	0.0%	4	3	1	25%	Median value	
Community Care Grant from the Scottish Welfare Fund (D12)	0.1%	12	3	9	75%	Median value	
Crisis Grant from the Scottish Welfare Fund (D13)	0.2%	19	10	9	47%	Median value	
Budgeting Loan from Social Fund/Budgeting Advances from Universal Credit (D14)	0.4%	40	33	7	18%	Median value	
Healthy Start Vouchers (D15)	0.3%	31	27	4	13%	Median value	

Table A1.6: Imputation of miscellaneous income for HiH and Spouse

	H'holds in receipt (%)	H'holds in receipt (N)	Total with amount given	Total with missing data	% missing	Method of imputation	Key variables
Occupational/employer (non-State) pension(s)	23.4%	2469	1784	685	28%	Hierarchical hot deck	Age, SIMD, Whether on means-tested benefits, how managing financially, SEG
Benefit from annuity, trust or covenant	1.1%	120	92	28	23%	Hot deck	How managing financially, number of cars
Maintenance payments	1.1%	120	105	15	13%	Median	
Rent from property or subletting, including boarders	2.7%	288	228	60	21%	Hot deck	Number of adults, Rooms, SIMD
Dig money from other household members	1.2%	130	108	22	17%	Hot deck	Number of adults in household
Benefit from accident/sickness scheme etc	0.1%	7	5	2	29%	Median	
Investment income (eg Dividends from shares/interest from savings)	4.1%	430	228	202	47%	Hot deck	SIMD
Student loan	1.3%	142	106	36	25%	Hot deck	Age, economic status
Grant	0.6%	60	46	14	23%	Median	
Regular non-work income, from any other source (please specify)	1.2%	124	66	58	47%	Hot deck	Type of source (if given).

Ipsos MORI

Appendix 2: Details of imputation routines used for Other Adults

Table A2.1: Imputation of earning for HiH and Spouse

	% households with OA in receipt of earnings	N	Earnings amount given	Earnings amount missing – Banded level given	Earnings amount missing - imputed	Receipt not known	Method and key variables
Other Adult 1	9.6%	1008	37%	16%	44%	3%	Hierarchical hot-deck imputation – age, sex, economic status, relationship to the head of the household, household type, whether in receipt of any means-tested benefit
Other Adult 2	2.3%	246	26%	14%	56%	4%	
Other Adult 3	0.3%	28	18%	11%	64%	7%	
All Other Adults	9.8%	1,282	35%	16%	46%	3%	

Table A2.2: Imputation of benefits for Other Adults: First group of benefits

	OAs in receipt (N)	%age of OAs	%age of h'holds	Total with amount given	Total with missing data	Method & Key variables
Universal Credit (A1)	42	1.8%	0.4%	30	12	Median
Housing benefit (A2)	18	0.8%	0.2%	6	12	Median
Council Tax Reduction (A3)	32	1.4%	0.3%	4	28	Hierarchical hot deck (Number of rooms, SIMD, whether HiH/Spouse also in receipt)
Working Tax Credit (A4)	12	0.5%	0.1%	9	3	Median
Child Tax Credit (A5)	16	0.7%	0.2%	7	9	Median
Income Support (A6)	20	0.9%	0.2%	7	13	Median
Jobseekers Allowance (A7)	25	1.1%	0.2%	15	10	Median
Employment and Support Allowance (A8)	56	2.4%	0.5%	39	17	Median
Carers Allowance (A9)	30	1.3%	0.3%	17	13	Median
Child Benefit (A10)	36	1.6%	0.3%	19	17	Median
Guardians Allowance (A11)	0	0.0%	0.0%	0	0	
Maternity Allowance (A12)	1	0.0%	0.0%	0	1	Median
Statutory Maternity/Paternity pay, Statutory Adoption Pay (A13)	2	0.1%	0.0%	1	1	Median
Statutory sick pay (A14)	2	0.1%	0.0%	1	1	Median

Table A2.3: Imputation of benefits for Other Adults: Second group of benefits

	OAs in receipt (N)	% of OAs	% of h'holds	Total with amount given	Total with missing data	Method
Personal Independence Payments (B1)	77	3.3%	0.7%	47	30	Hierarchical hot deck (Receipt of means-tested benefit, number of disabled people, rate and type received)
Disability Living Allowance (B2)	58	2.5%	0.6%	34	24	Hierarchical hot deck (Receipt of means-tested benefit, number of disabled people, rate received)
Attendance allowance (B3)	22	1.0%	0.2%	10	12	Median
Severe disablement allowance (B4)	5	0.2%	0.0%	0	5	Median
Incapacity benefit (B5)	1	0.0%	0.0%	1	0	
Industrial Injuries Disablement Benefit (B6)	1	0.0%	0.0%	0	1	Median

Table A2.4: Imputation of benefits for Other Adults: Third group of benefits

	OAs in receipt (N)	% of OAs	% of h'holds	Total with amount given	Total with missing data	Method
Pension Credit (C1)	11	0.5%	0.1%	5	6	Median
State Retirement Pension (C2)	78	3.4%	0.7%	44	34	Hierarchal hot deck (Receipt of means-tested benefit, number of retired in household, Age)
Widows Pension, Bereavement Allowance, or Widowed Parents Allowance (C3)	0	0.0%	0.0%	0	0	
Armed Forces Compensation Scheme (C4)	0	0.0%	0.0%	0	0	
War Widows/Widower's Pension (C5)	0	0.0%	0.0%	0	0	

Table A2.5: Imputation of benefits for Other Adults: Fourth group of benefits

	OAs in receipt (N)	% of OAs	% of h'holds	Total with amount given	Total with missing data	Method
Funeral Expenses Payment (D1)	1	0.0%	0.0%	1	0	
Sure Start Maternity Grant (D2)	1	0.0%	0.0%	0	1	Median
Best Start Grant (D3)	0	0.0%	0.0%	0	0	
Discretionary Housing Payment (D4)	0	0.0%	0.0%	0	0	
Loan or grant from DWP (D5)	2	0.1%	0.0%	0	2	Median
Loan or grant from Local Authority (D6	3	0.1%	0.0%	0	3	Median
Winter Fuel Payments (D7)	73	3.2%	0.7%	62	11	Median
Cold Weather Payments (D8)	32	1.4%	0.3%	21	11	Median
Extended payment of Housing Benefit (D9)	0	0.0%	0.0%	0	0	
Bereavement Payment (D10)	0	0.0%	0.0%	0	0	
Return to Work Payment (D11)	1	0.0%	0.0%	0	1	Median
Community Care Grant from the Scottish Welfare Fund (D12)	1	0.0%	0.0%	1	0	
Crisis Grant from the Scottish Welfare Fund (D13)	0	0.0%	0.0%	0	0	
Budgeting Loan from Social Fund/Budgeting Advances from Universal Credit (D14)	0	0.0%	0.0%	0	0	
Healthy Start Vouchers (D15)	0	0.0%	0.0%	0	0	

Table A2.6: Imputation of miscellaneous income for Other Adults

	OAs in receipt (N)	% of OAs	% of h'holds	Total with amount given	Total with missing data	Method
Occupational/employer (non-State) pension(s) (M1)	35	1.5%	0.3%	24	11	Median
Benefit from annuity, trust or covenant (M2)	3	0.1%	0.0%	2	1	Median
Maintenance payments (M3)	8	0.3%	0.1%	6	2	Median
Rent from property or subletting, including boarders (M4)	4	0.2%	0.0%	3	1	Median
Dig money from other household members (M5)	6	0.3%	0.1%	5	1	Median
Benefit from accident/sickness scheme etc (M6)	0	0.0%	0.0%	0	0	
Investment income (eg Dividends from shares/interest from savings) (M7)	5	0.2%	0.0%	3	2	Median
Student loan (M8)	159	6.9%	1.5%	69	90	Hierarchical hot deck (Age, Sex, Economic Status)
Grant (M9)	37	1.6%	0.4%	17	20	Median
Regular non-work income, from any other source (please specify) (M10)	23	1.0%	0.2%	9	14	Median

Appendix 3: Key variables in SHS dataset

Summary variables

	Old (narrow) definition	New Definition
Total Net Annual Income	Annetinc	Annetinc_broad
Summary of Total income	Incsum	Incsum_broad
Net weekly income	Wknetinc	Wknetinc_broad
Banded net annual income - with £0-£6000 split	incband	Incband_broad
Banded net annual income	Bandinc	Bandinc_broad
Summary of benefit income	Bensum_hihsp	Bensum
Summary of income from earnings	Earnsum_hihsp	Earnsum
Summary of income from miscellaneous sources	Miscsum_hihsp	Mscsum

Component variables

Income from earnings components	Hincminc (Hih Main jobs) Hincoinc (HiH other jobs) Sincminc (Spouse Main jobs) Sincoinc (Spouse other jobs) Earninc_oa1 (OA1 earnings) Earninc_oa2 (OA2 earnings) Earninc_oa3 (OS3 earnings)
Income from the 40 benefit components	Benin01 to beninc40 Beninc01_oa1 to beninc40_oa1 Beninc01_oa2 to beninc40_oa2 Beninc01_oa3 to beninc40_oa3
Income from the 10 miscellaneous components	Mscinc01 to mscinc10 Mscinc01_oa1 to mscinc10_oa1 Mscinc01_oa1 to mscinc10_oa1 Mscinc01_oa1 to mscinc10_oa1