

Adjusting the annual household income in the ITC 4-country UK dataset

1. Background

The derived income strata of *low*, *moderate*, and *high* (DE212v - based on total annual household income) in the International Tobacco Control (ITC) Policy Evaluation Project 4-country (Canada, United States of America, United Kingdom, and Australia) dataset were not updated since the first wave in 2002 up to wave 9 in 2013, and none of the waves are adjusted for household composition. This is problematic when using income as an indicator of poverty or socioeconomic status. For example, the lifestyle afforded by a family of four with an annual household income of £30 000 is likely to be more austere than that of a single adult on the same income. Similarly, due to inflation, £30 000 in 2002 had more purchasing power than in 2013. This document describes the processes used for adjusting the income variables in the ITC 4-country dataset to control for these two factors (household composition and inflation). The focus will be on the UK component of the data, however much of the coding, issues, and compromises will be generalisable to all four countries.

Note that this document is based on the ITC 4-country dataset (with preliminary release of wave 10 including UK and Australian participants only) downloaded on 1/7/2015. This release contains the wave 9 replenishment sample for the USA and Canada but no USA or Canadian data for wave 10. The surveys for the USA and Canadian replenishment sample were quite different to the previous waves and to the UK and Australia, which will cause a number of issues for coding the equalization variables for these participants (see the data anomaly notes - for the UK this will not be relevant). Some of these issues may be resolved when the wave 10 recontact data becomes available, so the coding decisions for the wave 9 replenishment sample from Canada and the USA will need to be reviewed when this occurs.

2. Adjusting for inflation

2.1. Relevant variables

At every wave, DE212v is derived from a wider range of income responses that are specific to each of the 4 countries. In the UK, the variable is DE211y "*Which of the following categories best describes your ANNUAL household income, that is the total income before taxes, or gross income, of all persons in your household combined, for one year?*"

1	under £6 500
2	£6 500 to £15 000
3	£15 001 to £30 000
4	£30 001 to £40 000
5	£40 001 to £50 000
6	£50 001 to £65 000
7	£65 001 to £95 000
8	£95 001 and over

DE212v is derived as follows: *Low* = (1, 2), *Moderate* = (3), *High* = (4, 5, 6, 7, 8) and this does not vary over the first 9 waves (spanning 2002 to 2013) of the data. In wave 10, although the original categories remain the same, the stratification changes somewhat: *Low* = (1, 2, 3), *Moderate* = (4, 5), *High* = (6, 7, 8).

If participants were present in a prior wave of the data ("recontacts") they are only asked the income question if they report any changes to their demographics since last asked (DE121), and specifically

changes to their income since last asked (DE123). If no changes are reported, then income is set to the value reported at the previous survey wave. However, DE121 and DE123 were only introduced at wave 3. At wave 2, only new recruits from the replenishment sample were asked about their income, and recontacts were not asked. Thus there is no way of knowing what the income of recontacts in wave 2 was, nor whether it changed since wave 1.

CODING DECISION 1: Set income for all recontacts with no income data in wave 2 to the values provided in wave 1. (note that for any coding decisions, sub-groups with different coding should always be flagged to enable diagnostic and comparative analyses).

2.2. Choosing inflation adjustment method and strata cut-points

Strata cut-points were based on the *Households Below Average Income (HBAI): An analysis of the income distribution 1994/5 - 2013/14* report published by the Department for Work and Pensions (UK). The report provides annual estimates of income-related data, based on nationally representative surveys of around 20 000 UK households. "Low income" is defined in the HBAI as 60% of the median UK household income. The UK household income figures, both before (BHC) and after (AHC) housing costs are taken into consideration, are provided. Income in the ITC 4-country dataset is measured *before* housing costs are considered, so only the BHC values from the HBAI document are of interest. There is no definition of "high income" provided in the HBAI document, so it was decided that a reciprocal definition would be used, that is, 140% of the median UK household income. Like the definition of low income, this definition was appealing because it uses a consistent criteria at every wave, rather than being based on more variable measures such as taxation-related high income thresholds. The latest HBAI document that was used reports median income for all years in 2013/14 prices, so the ITC data was adjusted using the Consumer Price Index (CPI) obtained from the UK Office for National Statistics at <http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/june-2015/index.html>. Income figures were converted to 2014 prices using the formula.

$$\text{INCOME}_{2014} = \text{INCOME}_{\text{YearX}} \times \text{CPI}_{2014} / \text{CPI}_{\text{YearX}}$$

The HBAI median income (BHC) values, CPI values, and corresponding ITC 4-country survey years are provided in Table 1.

Given that ITC income was quoted for the year, the weekly median incomes from the HBAI document were transformed to yearly figures using the formula:

$$\text{YEARLY INCOME} = \text{WEEKLY INCOME} / 7 \times 365$$

It should be emphasised that the median weekly incomes from the HBAI document are already equalized for household composition, so they should only be compared with versions of the ITC UK income variable that have also been equalized for household composition.

Table 1. ITC 4-country survey collection periods, and corresponding weekly median incomes and CPI figures for the United Kingdom (2002 - 2014).

ITC 4-country		HBAI		CPI	
wave	Collection period	Median weekly income	Survey date	CPI	Year
wave 1	2002	461	2002/3	95.4	2002
wave 2	2003	462	2003/4	96.7	2003
wave 3	2004	466	2004/5	98	2004
wave 4	2005/6	470	2005/6	100	2005
wave 5	2006/7	474	2006/7	102.3	2006
wave 6	2007/8	474	2007/8	104.7	2007
wave 7	2008/9	478	2008/9	108.5	2008
wave 8	2010/11	467	2010/11	114.5	2010
wave 9	2013	453	2012/13	126.1	2013
wave 10	2014	453	2013/14	128	2014

Note: ITC = International Tobacco Control; HBAI = Houses Below Average Income (Median income is weekly household income in 2013/14 prices, equalized for household composition and before adjusting for housing costs); CPI = Consumer Price Index.

Income in the ITC 4-country survey was obtained as a range rather than a set value. This was problematic because a single figure for income was needed for performing the calculations outlined above.

CODING DECISION 2: Take the mid point of the income range for calculation, and for the highest bracket, take the lower bound. The values used for calculation for each income range were as follows:

1	under £6 500	single figure = £3 250
2	£6 500 to £15 000	single figure = £10 750
3	£15 001 to £30 000	single figure = £22 500
4	£30 001 to £40 000	single figure = £35 000
5	£40 001 to £50 000	single figure = £45 000
6	£50 001 to £65 000	single figure = £57 500
7	£65 001 to £95 000	single figure = £80 000
8	£95 001 and over	single figure = £95 001

3. Adjusting for household composition

Household composition can be broken up into three components, the number of adults in the household, the relationship status of those adults (e.g. married versus friends flat-sharing), and the number of children in the household

3.1. Number of adults

For waves 1-6 of the ITC 4-country survey, the number of adults in the household was ascertained using question BK411: "How many people aged 18 or older live in your household, including yourself?" However, only new recruits at each wave were asked this question, and recontacts were not. Furthermore, there was no question to determine whether there had been any changes to the number of adults living in the household since last being surveyed.

CODING DECISION 3: for waves 2 to 6, set the number of adults living in the household to the value provided at recruitment for recontacted participants with no new data. For those participants who are also present at wave 7, use values provided at wave 7 if the distance between response wave and survey wave is shorter than taking the values provided at recruitment (see matrix in Table 2) and they report changes to the number of adults in their household in the past 2 years (DE128).

Coding decision 3 was particularly problematic because situations could arise where participants' responses provided in 2002 (wave 1) were being used to determine the number of adults in their household in 2007/8 (wave 6). Some alternative solutions to this problem could be to just leave the number of adults in the household as missing for these participants and have them be a separate group for analyses, or to instead use marital status to estimate if it is an "at least" 1 or 2 adult household.

Table 2. Distance (in waves) between response given and response used for number of adults in the household, when responses used are from wave of recruitment versus wave 7.

Survey wave	Recruitment Wave				
	wave 1	wave 2	wave 3	wave 4	wave 5
wave 2	1 versus 5	---	---	---	---
wave 3	2 versus 4	1 versus 4	---	---	---
wave 4	3 versus 3	2 versus 3	1 versus 3	---	---
wave 5	4 versus 2	3 versus 2	2 versus 2	1 versus 2	---
wave 6	5 versus 1	4 versus 1	3 versus 1	2 versus 1	1 versus 1

Note: The table shows (survey wave - recruitment wave) versus (wave 7 - survey wave), smaller values are better, with recruitment wave used when the two are equal.

From wave 7 to 9, two questions were used to determine the number of adults in the household. The wording for both was the same as for BK411 used in prior waves, but one variable was for new recruits (BI320) and the other for recontacts (BI322). Recontacts were only asked BI322 if they reported changes to their demographics in the last 2-3 years (DE121 - the time-frame of the question varies somewhat from wave to wave), and specifically to the number of people 18 or older in the household (DE128 in wave 7, and DE131 thereafter). In wave 10, the question BI320 (the question asking new recruits the number of adults in the household) was omitted, so there is no way to determine the number of adults for these participants. The best available estimate is probably marital status.

Data anomaly: At wave 7 and wave 8, BI322 is equal to 0 for 6 and 36 participants, respectively (i.e. indicating there are no adults aged 18 or over in the household). This technically should not be possible because all participants were 18 or older when recruited. At wave 7, there are also a considerable number of recontacted participants (mostly from Canada and the USA) who have no data on BI322 (most report having had changes to their demographics since last asked (DE121) but have missing or "not applicable" data on DE128 indicating whether or not they had any changes to the number of people aged 18 or over in their household. A similar issue again occurs at wave 9 where data is missing or "not applicable" on DE131. Most of these are from Australia, however a very large number of participants (4779) who are all new recruits from either the USA or Canada have no data on either DE121 or DE131, and no information about the number of adults in the household. These participants were not asked any questions relating to their marital status either, so

as far as I can tell there is no way to ascertain the number of adults in the household, aside from imputing "1" to represent the respondent.

CODING DECISION 4: Set the number of adults in the household for wave 10 new recruits to = 1 if they are single or = 2 if they are married or in a common-law relationship (see section 3.3 on marital status). Also do this for the participants in waves 7 and 8 for whom BI322 == 0 or in waves 7 and 9 where BI322 is missing and DE128 is also missing/ "not applicable" but there is data available for marital status. For the new recruits from the USA and Canada with all data missing, input the number of adults = 1.

The number of adults in the household was built up using these variables, setting the value for recontacts to the value reported at the previous wave if they reported no changes to the number of adults in their household. It should be noted that much of the missing data for this variable is a result of participants reporting that there were changes to their demographics and/ or number of people aged 18 or over in their household, but having no updated data (in BI322) for the current wave. The numbers missing are considerable in some waves, but they are mostly from the USA or Canada so are not problematic for the UK data.

Data anomaly: It should be noted that for waves 7.5 and 8.5, there appears to be an error in BI322, in that the majority respond as having 7 adults in the household. Given that this seems pretty unusual and 7 is a frequently used missing value code for "question not asked" in the ITC data, it is likely that these responses are actually those who did not report any changes to the number of adults in their household at waves 7.5 and 8.5. Cross tabulation with DE121 and DE131 support this. This is not relevant to the UK data, however, as only participants from the USA took part in wave 7.5, and only Australians in wave 8.5. Nevertheless it is recommended that any responses of "7" for BI322 at wave 7.5 and 8.5 be treated as "missing- not asked".

3.2. Number of Children

3.2.1 Relevant variables

For the first wave of ITC 4-country survey there was no question regarding the number of children living in the household.

CODING DECISION 5: Set the number of children in the household for wave 1 participants equal to their responses in wave 2, unless the age of their youngest child is less than the inter-wave-interval between wave 1 and 2, in which case set it to one less than the number of children in wave 2. The age of the youngest child was provided in months in the questions DE826 (those with only one child) and DE836 (those with more than one children), whereas the inter-wave-interval was calculated in days. The inter-wave-interval was divided by 30.42 for comparison, and instances where the inter-wave-interval was equal to the age of the youngest child were treated as no children at wave 1. This strategy might slightly underestimate the number of very young children at wave 1.

The main issue with coding decision 5 was that for those participants who were present in wave 1 but did not continue to wave 2, it is not possible to derive the number of children in the household, so for these participants it would not be possible to obtain the equivalized income (around 536 cases for the UK). So one question is how to treat these participants - should they be treated as a separate group with missing income data, should their non-equivalized income be used, or should equivalization be based solely on the number of adults in the household? For now, I have decided to equalize on the basis of the adults only (equivalent to setting their number of children = 0).

CODING DECISION 5b: Set the number of children for participants in wave 1 who did not continue to wave 2 to zero.

From wave 2 to 7.5, the number and age of the children in the household were ascertained through a number of different questions. DE811 "*Are there any children under the age of 18 currently living in your household?*" was asked first. Those who answered "Yes", were then asked DE816 "*How many children under the age of 18 are currently living in your household?*" From this point on, participants who responded with "1 (*only one child*)" versus any number more than 1 were treated differently. Participants with only one child were asked DE821 "*Is this child...?*" with response options:

1. Under the age of one
2. Between 1 and 5 years old
3. Between 6 and 12 years old, or
4. Between 13 and 17 years old.

Those who responded "1 (under the age of one)" were then asked DE826 "*How many months old is that child?*" (DE826 was coded "0" if the child was less than one month old. So, the information for respondents with only one child was the age in months of the child if it was under 1 year old, or an age range as above if the child was 1 or older.

Participants who responded to DE816 with 2 or more children were asked question DE831 "*How many are under the age of 1?*" Participants who indicated having any number of children under the age of one were then asked DE836: those who indicated only 1 child under the age of one were asked "*How many months old is that child?*", whereas those who indicated more than one child under 1 were asked "*How many months old is the youngest child?*". Again, any responses of under one month were coded "0". If participants with multiple children responded to DE831 with "0 (no children under 1)", they were then asked DE841 "*How many are between 1 and 5 years old?*", DE846 "*How many are between 6 and 12 years old?*", and DE851 "*How many are between 13 and 17 years old?*" If at the end of these questions DE831 + DE841 + DE846 + DE851 \neq DE816, the participant was prompted by the interviewer pointing out that the number of children in each age category do not match the total number of children they reported in DE816, and the questions were asked again. However, it should be noted that there are still a few instances where these numbers do not add up.

CODING DECISION 6: In instances where the total number of children reported did not match the number of children for whom ages were provided for, only those children who had valid ages were included in equivalization calculations. Generally, this was always in the same direction of reported children > children with valid age data, so for these participants the equivalized income may be slightly over-estimated. These cases are generally few in waves 1-7.5 (1 to 5 at each wave) and increase markedly from waves 8 - 9 (up to 54 at each wave), possibly due to the different questions that were asked. It is possible that for those who report more children than they provide ages for that the ones who don't have ages are those who are older than 18 - again they may have already been accounted for in the questions relating to number of adults aged 18 or over living in the household, which would support the decision to not attempt to include them in equivalization calculations.

In summary, the information about children in the household for respondents from waves 2 to 7.5 was the age of their *youngest* child in months if they had any that were under 1, or an age range as above if all their children were 1 or older. It should be noted that, unlike for the number of adults, the number of children in the household was asked of all recontacts regardless of whether they reported any changes to their demographics or to the number of children in the household since last being surveyed (DE129), however if they had missing data on the number of children and reported no changes since the last survey, it was decided to impute the values from the last survey

(very few participants fell into this category, but this also contributes to the issue of the number of reported children being fewer than the number with ages provided).

From wave 8 to 9, the number of children in the household and their ages was determined using a different set of questions. DE811 was the same as in earlier waves, however those who reported having any children were asked DE800 "*How many children under the age of 18 are currently living in your household?*" and then asked to provide the actual ages of each child, with a maximum of up to 5 children. They were first asked "*How old is the youngest child?*", with answers coded into DE801a if the participant responded in months (0 = less than one month), and DE801b if they responded in years. Depending on how many children the participant had, they were then asked "*How old is the next youngest child?*" (DE802a for responses given in months, and DE802b for years), and similarly DE803a and DE803b for the third youngest child, then DE804 for the fourth youngest child (there was no option to respond in months), and finally DE805 for the fifth child. There was then a final option to provide an open-end response for anyone with more than 5 children, and this was recorded verbatim in DE806o "*What are the ages of any other children under 18, that live in your household?*". However, very few participants with more than 5 children actually responded to this question, so the data for anyone with 6 or more children in these waves is somewhat unreliable. This is not a major issue because there are few participants with 6 or more children.

Data anomaly: In wave 9, responses of "0" for DE801a (age of the youngest child reported in months) seem to actually correspond to "no children" or "not asked/ not relevant". This is because for all all cases where DE801a = 0 in wave 9, when all of the children from the various responses (DE801, DE802, DE803, DE804, DE805 and DE806) are summed, it totals 1 extra to the number of children they report having in DE800. These participants are all from the UK.

Also, in wave 8 DE804 there is one response of a child aged 18, and in wave 8.5 DE803b a response of 19 (both are from Australia). These are not technically errors but do not fit with the definition of child, and it is not possible to deduce if these cases have already been accounted for within the questions relating to number of adults aged 18 or older living in the household.

CODING DECISION 7: In wave 9, set all responses of "0" for DE801a to "missing", and treat responses of children aged 18 or older as "0" or "missing".

A further issue arises with the replenishment sample from the USA and Canada at wave 9. None of these participants were asked the number or ages of the children in the household. Although this is not an issue for the UK data, it is probably better to set the number of children = 0 for these participants (and thus calculate equivalized income based on the number of adults only) rather than have them as missing on equivalized income, because their numbers are substantial (around n = 4779). Of course, this subset of participants have no information about the number of adults in the household either (see section 3.1), so there is no actual data to base equalization calculations on, although adjusting for inflation is still possible. Given that household composition has been imputed as = 1 adult for all these participants, it is likely that for many, the equivalized income will be an over-estimate, so it is especially important to flag these participants to enable comparisons of analyses both including and excluding them.

In wave 10, participants were only asked the number of children under 18 in the household (DE800) and the actual age of only the youngest child living in the household (DE801a and DE801b), but none of the other children. This makes it difficult to weight the children for equalisation based on their ages. If there is only one child, the answer is obvious, but with more than one, the answer will have to be estimated. For those who were present at wave 9 and report no changes to the number of children in their home, we can use their wave 9 responses adjusted by the inter-wave-interval to get the ages.

CODING DECISION 8: For wave 10 participants who were present in wave 9 and report no changes since then, set the ages of each child to the ages provided at wave 9, plus the inter-wave-interval (excluding any children from the calculations for whom this adjustment makes their age be 18 years or older).

3.2.2. Coding strategy.

Because of the way the number of children in the household for waves 1 to 7.5 were coded based on the age brackets they fell into, four variables were needed to capture the ages of each child, these were 1. The number of children under 1 (nUnder1); 2. the number aged 1 to 5 (n1to5); 3. the number aged 6 to 12 (n6to12) and 4. the number aged 13 to 17 (n13to17). These are largely based on DE831, DE841, DE846, and DE851. For wave 8 onwards, the variables coding the actual ages of each child were stratified to fit into these same categories, to enable consistency with the earlier waves. Of course the full information about each child's age was maintained, but just not used for this purpose. Firstly the variables were consolidated to obtain the age of each child. For example, DE801a and DE801b were combined to get the age of the first child, DE802a and DE802b the age of the 2nd etc... The variables nUnder1, n1to5, n6to12, and n13to17 were then built up iteratively using these ages. (Coding note: depending on how the code is written, it is often necessary to have responses of valid participants coded as "0" rather than "missing" if they legitimately were not asked the question, but you are including the question for the purpose of summing across a number of different variables. The ITC dataset does not do this consistently so it is important to double check, especially gDE841, gDE846, gDE851, ggDE846, and ggDE851).

3.3 Marital status

Marital status was captured by the variable DE111: "*Are you married, separated, divorced, widowed, living common-law, or single?*". For the purpose of equivalization, responses of "married" and "living common-law" were treated as equivalent. So, this variable was dichotomised into (married or living common-law) versus (separated, divorced, widowed, or single). For some reason, recontacted participants at wave 2 were not asked about their marital status, so have missing data for this variable.

CODING DECISION 9: Set marital status for recontacted participants at wave two to the values provided at wave 1.

3.4. Methods of equivalization.

Various different methods of equivalization based on household composition exist. The UK HBAI 2013/14 document was again used as a guideline to select the method to be used for the ITC UK data. Two methods emerged as the most appropriate for the UK data. The UK government has used the McClements (1977) scale up until 2006, after which it adopted the modified Organisation for Economic Co-operation and Development (mOECD) scale developed by Hagenaars, De Vos, & Zaidi (1994). Use of the McClements scale allows for better comparability with UK government publications, whereas the mOECD is based on more recent data and is the version recommended by Eurostat and the office for official publications of the European Unions, so it is arguably more generalisable (OECD membership also extends beyond Europe). It is easy enough to calculate both and compare and contrast, however, the format of the ITC data does not lend itself precisely to either scale. Equivalization involves weighting members of the household based on certain criteria. In the UK, the reference point (= 1.0) is a household with an adult couple. The McClements differs from the mOECD scale mainly in that it differentiates between adults based on marital status (e.g.

treats a 2 adult household differently depending on whether they are in a romantic relationship v.s. friends sharing a house), and also applies different weights to a much greater range of children's ages (whereas the mOECD only differentiates between under 14 years versus 14 or over). The HBAI document provides a comparable version of the McClements scale to the mOECD by combining some of the McClements children's age groups and calculating weights for these new categories, see Table 3.

Unfortunately, the ages of children in the ITC data (at least up until wave 8) are measured in brackets that do not lend themselves to exact classification either by McClements' or mOECD criteria because some of the ITC brackets "straddle" the age brackets of the mOCED and/ or McClements scales: ITC "under 1 year" (both ok: mOECD = 0.2, McClements = 0.09); ITC 1 to 5 years (mOCED ok = 0.20, McClements problematic = 0.175 and/ or 0.205); ITC 6 to 12 years (mOECD ok = 0.20, McClements problematic = 0.205 and/ or 0.225, and/ or 0.25, and/ or 0.27), ITC 13 to 17 years (both problematic: mOCED 0.20 and/ or 0.33, McClements 0.27 and/ or 0.355).

CODING DECISION 10: Only use two categories for calculating weights for children (either the mOECD or the McClements scale with mOECD weights), but shift the category boundaries slightly to fit with the ITC 4-country data, that is "under 13" instead of "under 14", and "13 or over" instead of "14 or over". Note that it would be possible to use the exact McClements age categories from wave 8 onwards, but consistency of methodology with prior waves is deemed to be more important at this stage.

Table 3. Equivalization weights for the modified OECD (mOECD) and McClements scales, taking an adult couple as the reference point (= 1).

	mOECD	McClements	McClements with mOECD weights
First Adult	0.67	0.61	0.61
Spouse	0.33	0.39	0.39
Other 2nd adult	0.33	0.46	0.46
Third adult	0.33	0.42	0.42
Subsequent adults	0.33	0.36	0.36
Children under 1 year	0.20	0.09	0.2
Children 2 - 4 years	0.20	0.175	0.2
Children 5 - 7 years	0.20	0.205	0.2
Children 8 - 10 years	0.20	0.225	0.2
Children 11 - 12 years	0.20	0.25	0.2
Children 13	0.20	0.27	0.2
Children 14 - 15	0.33	0.27	0.32
Children 16 - 18 years	0.33	0.355	0.32

Note: ITC = The OECD weights for the McClements scale are taken from the HBAI document. Values for the McClements scale are adapted from McClements (1977) by taking the average of the weights obtained for 1971 and 1972 (which never differed by more than 0.02).

Given the issues with the missing data for the ages of children in wave 10, it was decided to use the "weighted average of weights" for the age categories that were being used. To explain, based on CODING DECISION 10, children under 18 were divided into (0 to 12) versus (13 to 17). Assuming an even spread of ages, this means that 13/18 (= 72%) of children under 18 are in the (0 to 12) category, and 5/18 (= 28%) of children are in the (13 to 17) category. So, the weights for each equivalization method can be multiplied by these proportions and summed to give a "weighted weight" that is applicable to all children under 18:

$$\text{mOECD} = (0.72 \times 0.20) + (0.28 \times 0.33) = 0.236$$

$$\text{McClements} = (0.72 \times 0.20) + (0.28 \times 0.32) = 0.234$$

CODING DECISION 11: For those participants in wave 10 who have no information on the ages of some of their children (due to only the age of the youngest being asked), set the child weights for the unknown children to the "weighted averages" (as above) for each equalization method.

A final small issue is that at each wave, between 3 - 70 participants report that there are more than 5 adults living in the household. Given that this is a pretty unusual situation, it is possible that this is an error, or that the participant would be less likely to accurately state the household income when there are potentially multiple independent income earners (e.g. in a shared-house scenario). One option would be to truncate the number of adults at 5 for these cases and base the equalization calculations on this value. For now, it has been decided to simply calculate the equalized income based on the reported number of adults, but flag these cases so that analyses including and excluding them can be compared.

The general formula for equalizing the income for household composition is as follows:

Equalized income = Income / [(n Adults × Adult weights) + (n Children × Child weights)]

4. Coding Workflow

- Set income brackets to mid point of range
- adjust income for inflation, to 2014 prices using CPI data
- consolidate the number of adults living in the household
- consolidate the number of children living in the household that fall into each of the different age brackets.
- determine marital status of adults
- generate adult modifiers based on number of adults and marital status (one based on mOECD equalization weights, and one based on McClements weights)
- generate child modifiers based on number of children aged under 13 versus 13 to 17 in the household (one each for mOECD and McClements versions)
- Divide the inflation-adjusted incomes by (adult modifier + child modifier). Again there will be two versions, one for mOECD and one for McClements
- stratify the equalized income into low, moderate, high, and not disclosed, based on the historical median income statistics provided by the HBAI in 2014 prices (low = 60% of income, high = 140% of income)

Coding note: it is helpful to set the values for the CPI index modifiers and also the cut-offs for the high and low income brackets to be global variables at the start of the code, and referred to indirectly in the rest of the code. This is so that the data can quickly and easily be updated when a new ITC wave comes out or there is a need to use different cut-offs.

Acknowledgements

Thanks to Ann McNeill, Sara Hitchman, Robert Branston, and Rosemary Hiscock for their valuable feedback.

References

Chanfreau, J., & Burchardt, J. (2008). Equivalence scales: rationales, uses and assumptions. *Scottish Government, Edinburgh*. <http://www.gov.scot/Resource/Doc/933/0079961.pdf>

Hagenaars, A. J. M., De Vos, K., & Zaidi, A. (1994). Poverty statistics in the late 1980s: Research based on micro-data. *Luxembourg: Office for official publications of the European Union*.

McClements, L. D. (1977). Equivalence scales for children. *Journal of Public Economics*, 8, 191-210.

Shale, J., Balchin, K., Rahman, J., Reeve, R., & Rolin, M. (eds) *Households below average income: An analysis of the income distribution 1994/95 - 2013/14* Department for Work & Pensions, National Statistics Report, 25th June 2015. <https://www.gov.uk/government/statistics/households-below-average-income-19941995-to-20132014>