

## Understanding Society User Support - Support #1905

### Inapplicable codes on the W12 hresp helpbuy questions

05/19/2023 12:26 PM - Rory Coulter

<b>Status:</b>	Resolved	<b>Start date:</b>	05/19/2023
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>	Understanding Society User Support Team		
<b>Category:</b>	Data inconsistency		
<b>Description</b>			
Hi,			
This question might be of general interest unless I've missed something really obvious...			
I'm trying to use the W12 hresp 'l_helpbuy' suite of variables but can't figure out why there are so many inapplicable codes. According to the questionnaire, this question is answered by those households reporting on hshownd they are some form of owner-occupier (codes 1-3 on hshownd route respondents through to helpbuy which asks about source of purchase funds for new owners and for those who've recently paid off their mortgage). However, when I restrict the W12 hresp file to only those with codes 1-3 on hshownd, I get around 45% inapplicable values. Some are indeed to be expected but this seems a high percentage of owner-occupying households.			
What seems more confusing is that when I break the W12 hresp sample down by interview month and year I find that we have around 20% inapplicables per month throughout 2020 but from January 2021 this jumps up to 70% inapplicable and stays in this region for the remaining months of the year. I'm struggling to understand why this should be the case and can't find an obvious explanation in the questionnaire or in the W12 technical report.			
FYI attached is an R script that replicates the issue using the W12 hresp file from UKDS SN6614.			
Thanks for all help, Rory			

#### History

##### #1 - 05/19/2023 02:38 PM - Understanding Society User Support Team

- Status changed from New to In Progress

- % Done changed from 0 to 50

Hi Rory,

I have looked into this and found the same type of problem cases as you did. But after reading the universe again, I found that the Universe syntax doesn't match text description of the Universe: "If (HsOwnd = 1|2|3) // Owned outright, mortgaged or shared ownership at new address or has become owned outright, mortgaged or shared ownership". So, I created the Universe syntax based on the text description: ((inlist(l\_hshownd,1,2,3) & l\_origadd==2)|(inlist(l\_hshownd,1,2,3) & l\_origadd==1 & !inlist(l\_ff\_hshownd,1,2,3))). Using this syntax I found only 0.7% problem cases (as opposed to around 35% with the earlier syntax).

I have asked the data & questionnaire team to look into this and confirm whether that is what happened. I will get back to you as soon as I hear back from them.

Best wishes,  
Alita

##### #2 - 05/22/2023 09:39 AM - Rory Coulter

Hi Alita,

Many thanks for looking into this so swiftly. I've rerun the code with your more restricted universe and it does indeed get the percentage of inapplicable codes on helpbuy down to ~1%. That makes sense: essentially these questions are being answered by (1) owners interviewed at a new address OR (2) owners who aren't at a new address but for whom their fed forward tenure isn't homeownership (mostly the fed forward value is missing/inapplicable so we are picking up those converting renting to buying and new owner interviewees).

I'm still a bit puzzled by the time patterning on this variable though. When we restrict the sample to this new universe around 89% of cases were interviewed in 2020 and just 11% in 2021-22. This compares with a 59% 2020 to 41% 2021-22 split in the full sample. I've done a bit of digging around and it appears to be due not to the l\_origadd variable (% interviewed at new address is fairly similar in 2020 v 2021/2) but rather to l\_ff\_hshownd. l\_ff\_hshownd is missing/inapplicable in a much larger share of 2020 interviews than for 2021-22 interviews and this means that a lot more 2020 owning cases have valid 'helpbuy' responses.

Do you have any insight into why this should be the case?

Best  
Rory

Understanding Society User Support Team wrote in [#note-1](#):

Hi Rory,

I have looked into this and found the same type of problem cases as you did. But after reading the universe again, I found that the Universe syntax doesn't match text description of the Universe: "If (HsOwnd = 1|2|3) // Owned outright, mortgaged or shared ownership at new address or has become owned outright, mortgaged or shared ownership". So, I created the Universe syntax based on the text description: `((inlist(l_hshownd,1,2,3) & l_origadd==2)|(inlist(l_hshownd,1,2,3) & l_origadd==1 & !inlist(l_ff_hshownd,1,2,3)))`. Using this syntax I found only 0.7% problem cases (as opposed to around 35% with the earlier syntax).

I have asked the data & questionnaire team to look into this and confirm whether that is what happened. I will get back to you as soon as I hear back from them.

Best wishes,  
Alita

### #3 - 05/22/2023 10:21 AM - Rory Coulter

Just a quick update - I'm wondering if this temporal trend is related to the broader patterning of inapplicable cases on ff\_hshownd. On the web documentation ([https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/ff\\_hshownd](https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/variable/ff_hshownd)) for waves 2-10 there are few inapplicable codes on this variable. Then in W11 the % inapplicable jumps markedly, and when I've gone back to the raw data this appears to be down to those W11 interviews conducted in 2019. I'm not sure why this is the case.

Rory

Rory Coulter wrote in [#note-2](#):

Hi Alita,

Many thanks for looking into this so swiftly. I've rerun the code with your more restricted universe and it does indeed get the percentage of inapplicable codes on helpbuy down to ~1%. That makes sense: essentially these questions are being answered by (1) owners interviewed at a new address OR (2) owners who aren't at a new address but for whom their fed forward tenure isn't homeownership (mostly the fed forward value is missing/inapplicable so we are picking up those converting renting to buying and new owner interviewees).

I'm still a bit puzzled by the time patterning on this variable though. When we restrict the sample to this new universe around 89% of cases were interviewed in 2020 and just 11% in 2021-22. This compares with a 59% 2020 to 41% 2021-22 split in the full sample. I've done a bit of digging around and it appears to be due not to the l\_origadd variable (% interviewed at new address is fairly similar in 2020 v 2021/2) but rather to l\_ff\_hshownd. l\_ff\_hshownd is missing/inapplicable in a much larger share of 2020 interviews than for 2021-22 interviews and this means that a lot more 2020 owning cases have valid 'helpbuy' responses.

Do you have any insight into why this should be the case?

Best  
Rory

Understanding Society User Support Team wrote in [#note-1](#):

Hi Rory,

I have looked into this and found the same type of problem cases as you did. But after reading the universe again, I found that the Universe syntax doesn't match text description of the Universe: "If (HsOwnd = 1|2|3) // Owned outright, mortgaged or shared ownership at new address or has become owned outright, mortgaged or shared ownership". So, I created the Universe syntax based on the text description: `((inlist(l_hshownd,1,2,3) & l_origadd==2)|(inlist(l_hshownd,1,2,3) & l_origadd==1 & !inlist(l_ff_hshownd,1,2,3)))`. Using this syntax I found only 0.7% problem cases (as opposed to around 35% with the earlier syntax).

I have asked the data & questionnaire team to look into this and confirm whether that is what happened. I will get back to you as soon as I hear back from them.

Best wishes,  
Alita

### #4 - 05/23/2023 03:47 PM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

Thanks for highlighting this pattern.

First, I need to amend my answer above. The syntax of the Universe for helpbuy in the questionnaire is correct. But because we combine the responses from ff\_hshownd and hshownd into hshownd, rather than creating a new variable such as hshownd\_dv, if you check the data it appears that there is a problem with the Universe syntax in the questionnaire. So, we needed to create a new syntax for the routing rule, as I have shown above.

The second issue is that ff\_hshownd that was fed into the interview scripts, had many inapplicables although there were valid values from last wave. So, many hh were asked hshownd even though they were NOT eligible as they had already answered this question earlier and were still living at the original address. Because of the direction of the error, instead of loss of information there was more (not needed) information. After this problem was identified it was corrected. Hence a lot more hhs qualified for this question (around 30%) in 2020, and after that this % dropped off to around 10% in Jan-Feb 2021 and then steadily down to around 2%.

```
generat intdate=intdate*100+intdatem
generat askqs=0
replace askqs=1 if (inlist(hshownd,1) & origadd==2) | (inlist(hshownd,1) & origadd==1 & !inlist(ff_hshownd,1,2,3))
ta intdate askqs, row nofreq
```

Hope this answers your quesitons. Please let us know if you have further questions.

Best wishes,  
Alita

#### #5 - 05/26/2023 05:08 PM - Rory Coulter

Hi Alita,

Thanks for this - all makes sense now!

Best  
Rory

Understanding Society User Support Team wrote in [#note-4](#):

Thanks for highlighting this pattern.

First, I need to amend my answer above. The syntax of the Universe for helpbuy in the questionnaire is correct. But because we combine the responses from ff\_hshownd and hshownd into hshownd, rather than creating a new variable such as hshownd\_dv, if you check the data it appears that there is a problem with the Universe syntax in the questionnaire. So, we needed to create a new syntax for the routing rule, as I have shown above.

The second issue is that ff\_hshownd that was fed into the interview scripts, had many inapplicables although there were valid values from last wave. So, many hh were asked hshownd even though they were NOT eligible as they had already answered this question earlier and were still living at the original address. Because of the direction of the error, instead of loss of information there was more (not needed) information. After this problem was identified it was corrected. Hence a lot more hhs qualified for this question (around 30%) in 2020, and after that this % dropped off to around 10% in Jan-Feb 2021 and then steadily down to around 2%.

```
generat intdate=intdate*100+intdatem
generat askqs=0
replace askqs=1 if (inlist(hshownd,1) & origadd==2) | (inlist(hshownd,1) & origadd==1 & !inlist(ff_hshownd,1,2,3))
ta intdate askqs, row nofreq
```

Hope this answers your quesitons. Please let us know if you have further questions.

Best wishes,  
Alita

#### #6 - 05/29/2023 04:49 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved
- % Done changed from 50 to 100
- Private changed from Yes to No

#### Files

---

hhresp_helpbuy.R	3.06 KB	05/19/2023	Rory Coulter
------------------	---------	------------	--------------