

Understanding Society User Support - Support #1963

Missing LSOA for hidp across waves

09/01/2023 02:46 PM - Richard Belcher

Status:	Resolved	Start date:	09/01/2023
Priority:	High	% Done:	100%
Assignee:			
Category:	Data linkage and consents		
Description			
Hello,			
I have been granted a special license to access the hidp LSOA codes. I am conducting a nationally representative analysis on waves 1-9, and using data calculated from a respondents LSOA is essential to my analysis. However, I have found some hidps that have non-zero yearly cross sectional weights (134) do not have a corresponding LSOA code. I am wondering if this is a known issue or something that could/has been fixed, and the best way to deal with it whilst still conducting a nationally representative analysis. Some extra information is that this issue seem to be evenly distributed across waves 1-9.			
All the best,			
Richard			

History

#1 - 09/01/2023 02:50 PM - Richard Belcher

My path forward is probably to impute the LSOA derived values to maintain it being nationally representative but I was just wondering if I had missed anything and to highlight the issue.

#2 - 09/01/2023 04:25 PM - Understanding Society User Support Team

- Category set to Special license

- Status changed from New to In Progress

- Private changed from Yes to No

Hi Richard,

Would you be able to provide some more details, please? To investigate this would be helpful to know:

- 1) the scale of the problem - for how many households this information is missing and in which waves?
- 2) what kind of missingness you're observing, are these empty blank fields or negative values?
- 3) a few hidps for which LSOA is missing for at least some waves would be very helpful.

Many thanks,
Piotr Marzec
UKHLS User Support

#3 - 09/01/2023 04:41 PM - Richard Belcher

Hi Piotr,

Thanks for your quick reply. Information below:

1) the scale of the problem - for how many households this information is missing and in which waves?

- 134 households don't have an LSOA. This permeates to 306 cases (across waves).

- Missing cases from each wave:

- 1 2 3 4 5 6 7 8 9

15 73 25 35 41 27 26 25 39

- Also FYI I tried to merge the non-wave specific hidp and I it is still 306 cases (so this is an issue with the same households through multiple waves)

2) what kind of missingness you're observing, are these empty blank fields or negative values?

- This comes after merging the special license LSOA files. The hidp from the main US dataset isnt found in the special license LSOA file (so it returns NA)

3) a few hidps for which LSOA is missing for at least some waves would be very helpful.

- Here are thehidps that are in the main US dataset, but not in the special license LSOA file and are across multiple waves. Not sure if this impacts identifiability if their is some systematic missingness by location here so may be worth deleting before this gets published online:

#4 - 09/04/2023 08:54 AM - Understanding Society User Support Team

- Private changed from No to Yes

#5 - 09/04/2023 08:58 AM - Understanding Society User Support Team

Hi Richard,

Thank you for providing the information. Regarding 3), as hidp is a wave specific variable, I'd need to know in which waves these hidps are missing.

Thanks,
Piotr
UKHLS User Support

#6 - 09/04/2023 09:11 AM - Richard Belcher

- File hidpwithnolsoa.csv added

Hi Piotr,

Attached is the hidp*wave combinations with missing LSOA.

All the best,

Richard

#7 - 09/04/2023 04:52 PM - Understanding Society User Support Team

Thank you. I'll investigate this further.

Best wishes,
Piotr

#8 - 09/04/2023 04:52 PM - Understanding Society User Support Team

- % Done changed from 0 to 20

#9 - 09/05/2023 02:59 PM - Understanding Society User Support Team

- Category changed from Special license to Data linkage and consents

- % Done changed from 20 to 80

- Private changed from Yes to No

Hi Richard,

I checked this with the team generating the LSOA files.

The process for each wave works on the basis of matching a file of postcodes/hidp combinations used in a wave to w_hhsamp. If an hidp is missing it is because they have not matched at some point in that process. Sometimes we don't have a postcode for an hidp, or it could be that the postcode we have is incorrectly formatted or that it doesn't match a postcode in the ONSPD that we use. The released geography files do not contain any non-matches, hence the discrepancy between the files. So, you should treat that as missing data.

Best wishes,
Piotr
UKHLS User Support

#10 - 09/05/2023 02:59 PM - Understanding Society User Support Team

- Private changed from No to Yes

#11 - 09/05/2023 03:04 PM - Richard Belcher

Hi Piotr,

Thanks very much for clearing that up and solving it so quickly, it makes sense.

This could be something to consider when US creates the weighting files or perhaps an additional weight to be generated for those using a special/secure license. Also is it possible to delete the attached file and references to specific HHID's?

All the best,

Richard

#12 - 10/19/2023 02:29 PM - Understanding Society User Support Team

- Status changed from *In Progress* to *Feedback*

Hello Richard,

We have removed the pidps from your earlier post, and also we had not made this post public.

We will pass on your suggestion to the weighting team.

Best wishes,
Alita

#13 - 11/30/2023 01:26 PM - Understanding Society User Support Team

- Status changed from *Feedback* to *Resolved*

- % Done changed from 80 to 100

- Private changed from *Yes* to *No*

Files

hidpwithnolsoa.csv	2.44 KB	09/04/2023	Richard Belcher
--------------------	---------	------------	-----------------