

Understanding Society User Support - Support #2248

Some queries r.e. issues with SOC 2000 and 2010 variables

04/29/2025 02:57 PM - Thomas Stephens

<b>Status:</b>	Feedback	<b>Start date:</b>	04/29/2025
<b>Priority:</b>	Normal	<b>% Done:</b>	50%
<b>Assignee:</b>	Understanding Society User Support Team		
<b>Category:</b>	Data documentation		
<b>Description</b>			
Good afternoon,			
I'm currently updating some analysis for Waves 13-14 and I noticed that from your variable guide that Understanding Society is investigating issues with previous waves' SOC 2000 and SOC 2010 data, which you've identified following changes to the approach to gathering new SOC 2020 data.			
This makes sense; I just had a few questions:			
1. Does this issue also affect Standard Industrial Classification data, e.g. jbsic07? I currently use the condensed version (jbsic07_cc) to infer workplace health and safety. There is no mention of this issue affecting this variable, but the routing for this question is fairly similar to the old SOC (i.e. it's only asked of people who have changed jobs) so it would make sense if it were also affected. Any reassurance here would be greatly appreciated!			
2. Could you provide any info on the number of SOCs affected by this issue, and which respondents are affected - e.g. would a variable be made available which flags where this is different, or some code be provided to compute it?			
3. An obvious question but could you provide any further info on timeline to resolve this?			
As this doesn't affect the latest (Wave 13-14) SOC 2020s this likely won't affect my planned analysis too much, but I would especially appreciate assurance r.e. the SICs point.			
Thanks in anticipation for any clarification the team can provide.			
Best wishes,			
Tom			

History

<b>#1 - 04/30/2025 11:24 PM - Understanding Society User Support Team</b>
<p>- Category changed from Questionnaire content to Data documentation</p> <p>- Status changed from New to In Progress</p> <p>- % Done changed from 0 to 10</p> <p>- Private changed from Yes to No</p> <p>Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can. We aim to respond to simple queries within 48 hours and more complex issues within 7 working days.</p> <p>Best wishes,</p> <p>Understanding Society User Support Team</p>
<b>#2 - 05/01/2025 10:16 PM - Understanding Society User Support Team</b>
<p>- Status changed from In Progress to Feedback</p> <p>- % Done changed from 10 to 50</p> <p>Hello Tom,</p> <ol style="list-style-type: none"><li>1. Yes, the issue affecting jbsoc is the same for jbsic07. We observed a significantly higher rate of change (approximately 5-6 times higher than in previous waves) in the industry coding between waves. This is due to the absence of dependent interviewing, with data collected and coded from scratch, and much of this change is artificial, mirroring the situation with jbsoc.</li><li>2. Currently, a variable reflecting this correction isn't included. However, we may consider adding it in our next release.</li></ol>

3. Our aim is to have these corrections implemented for the next wave release, scheduled for November 2025.

Regarding the recoding process for jbsoc20, it involves individuals in continuous employment where a genuine job change was identified after comparing the jbsoc text fields, as well as those who were not in continuous employment. Consequently, jbsoc20 is not complete (approximately 9500 observations are coded as -9). Therefore, I would recommend using either jbsoc00 (the best option) or jbsoc10 (which still has a significant number of -9s, but fewer than jbsoc20).

I hope this information is helpful.

Best wishes,

Roberto Cavazos  
Understanding Society User Support Team

**#3 - 05/04/2025 01:31 PM - Thomas Stephens**

Hi Roberto,

Thanks. Super helpful. I've now appraised myself more of this, reading your reply and also the revisions set out by Understanding Society (here: <https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/revisions-to-the-main-current-job-occupation-jbsoc-in-wave-13/>).

I have some more questions, just to be sure I've understood this issue correctly.

For context, I've been using data on workers' SIC 2007, cross-referenced with Labour Force Survey data, to build a matrix of occupational health and safety and introduce this into Understanding Society. It does this by building a profile of workplace accidents and illnesses in comparable years of LFS and then introducing health and safety data based on workers' SIC in the corresponding wave. My data uses Waves 4, 6, 8, 10 and 12 of UKHLS but I wanted to update this to Wave 14. In addition, I also now want to build a matrix of workers' job quality by SOC 2020 - I probably only need at least **some** good data on SOC 2020 to do this.

It seems these issues may make both these things more difficult, but I think I might still be able to find a way through it depending on the severity of the issue - especially whether prior waves' data are affected. Pursuant to this, can I check that:

1. For the most part, this issue isn't deemed to affect pre-Wave 13 data on jbsic07, jbsoc00 and jbsoc10? My initial worry was that the high rate of change has led the UKHLS team to worry about the accuracy of prior waves' data, but in fact it more seems like the absence of dependent interviewing has created a lot of as you say "artificial" changes in occupations amongst people whose jobs actually remained the same? The analysis linked above appears to suggest as much because once you filter out spurious changes you seem to end up with a rate of job change that's similar to prior waves (~15%).

2. For Waves 13-14, can any jbsoc20 data which are **not** coded as missing (-9) be relied on as providing a reasonably accurate reflection of that person's current job? I.e. someone coded -9 presumably reported an occupation change which has since been identified as spurious, meaning they don't have an soc20 (and unfortunately it can't be recovered).

3. For sic 2007 in Waves 13-14, amongst those in paid work, it seems there has been a huge jump in inapplicables (-8) for Waves 13-14 - previously I found no inapplicables for those prior waves amongst this sub-group - but -9 is fairly stable. Which suggests the UKHLS team have assigned spurious SICs to -8, is that right? This prompts the same question as #3: can those who aren't coded as 8 be relied on?

4. For SIC 2007, is it possible that some of these Wave 13-14 workers' data from prior waves is recoverable? I.e. in cases where their change in jobs is spurious, I can presumably write some code which carries forward their prior waves' SIC 2007? I would need to know which of these -9s are genuinely spurious though, which I could presumably find using workers' self-reported job change data.

Any advice on the above would be greatly appreciated!

Best wishes,

Tom