MIDUS
Midlife in the United States
A National Study of Health & Well-Being

Documentation of Data Linkage between MIDUS Milwaukee Refresher 1 (MKER1) Survey and O*NET 2012 (17.0) Database

This document provides details on the data linkage performed between the baseline MIDUS Milwaukee Refesher (MKER1) Survey sample and the 2012 O*NET 17.0 database that produced the following standalone dataset:

MKER1_ONET2012_N329_20220217.sav

Specifically, this dataset was derived by linking MKER1 cases' survey responses (collected 2012) and their resulting Standard Occupation Classification (SOC) codes with the 2012 Occupational Information Network (O*NET 17.0) database. If a respondent was not currently working or did not provide adequate occupational information, the case was excluded from the MIDUS-O*NET linkage. Military occupation codes were also excluded from the linkage since they are not included in the O*NET database. The MIDUS Milwaukee Refresher sample used SOC codes from 329 valid cases to link with the O*NET database.

MIDUS SOC codes were classified using six digits in the '12-3456' format, while the O*NET SOC codes classified using eight digits in the format of '12-3456.78', including a 2-digit extension of the decimal. To reconcile this difference, each O*NET SOC code was split into two parts, one containing the six digits before and another containing the two digits after the decimal.

The 2-digit extension was only used in performing the linkage and not included in the SOC codes of the final dataset.

The actual linkage was performed in two steps:

- Step 1 – *for matched cases*: If a MIDUS SOC code perfectly matched the first six digits of an O*NET SOC code, and the O*NET SOC extension code was coded '.00'. The assumption was that the MIDUS code '12-3456' is equivalent to the O*NET code '12-3456.00'. For all the matched cases, MIDUS-O*NET data were linked directly.
- Step 2 – *for unmatched cases*: There was no perfect match between MIDUS and O*NET SOC codes for some cases. Since the number of SOC codes represented in each of the 13 O*NET datasets varied, the SOC codes available for linking could also vary. Table 1 below shows the number of available SOC codes in each O*NET dataset and how many of the 329 MIDUS Milwaukee Refresher cases matched with the 2012 O*NET database.

Table 1 listed number of SOC codes available in each O*NET dataset and number of cases directyly matched with MIDUS sample (out of the 329 valid cases).

*Table 1: SOC codes in O*NET datasets and matching status with MIDUS sample*

2012 O*NET datasets	# of SOC codes Available in O*NET data	# of Cases Matched with MKER1 Sample	% of Cases Matched with MKER1 Sample
Abilities (IM & LV)	903	258	78%
Interests	897	255	77%
Values	974	283	86%
Styles (IM)	902	258	78%
Skills (IM & LV)	903	258	78%
Knowledge (IM & LV)	903	258	78%
Activities (IM & LV)	903	258	78%
Context (CX & CT)	898/902	257	78%

For the cases where a direct link between SOC codes was not available, a series of different mean value substitution adjustments were used to replace the values of the O*NET summary score variables for those of the unmatched SOC codes. In applying the mean substitution values, four scenarios were identified which required special treatment:

1. Mean substitution scenario 1: If a parent code '.00' was missing from the O*NET dataset, but one or more 2-digit extension codes were available (i.e. '.01', '.02'), the mean of the scores with the 2-digit extension O*NET values were calculated and substituted into the variables for the MIDUS parent '.00' value.
   - For example, MIDUS SOC code '11-3071' did not have a direct match '11-3071.00' in the O*NET data, but O*NET did have '11-3071.01', '11-3037.02' and '11-3037.03'. MIDUS used the mean values of the SOC codes '11-3071.01', '11-3071.02' and '11-3037.03'to substitute the scores for '11-3071.00' and match them with MIDUS '11-3071'.
2. Mean substitution scenario 2: For the six-digit SOC codes, if the last digit was '0', it was called a broad occupation code; if the last digit was '1', '2', '3', etc. called a detailed occupation code. For example, SOC code '13-2070' referred to 'credit counselors and loan officers,' '13-2071' referred to 'Credit Counselors, '13-2072' referred to 'Loan Officers.' '13-2070' is a broad occupation code, while '13-2071' and '13-2072' are detailed occupation codes. When a broad occupation code was not available from an O*NET dataset, but several detailed occupation codes within a broader occupation code were available, then the missing broad occupation codes were substituted with the mean of the multiple detailed occupation codes.
   - For example, MIDUS SOC code '13-2070' did not match any O*NET codes. 13-2070' is a broad occupation that includes two detailed occupations ('13-2071', '13-9072'). Therefore, the mean values of scores for SOC codes '13-2071', '13-9072' were used to substitute the scores for SOC code '13-2070'
3. Mean substitution scenario 3: If a detailed occupation code was missing from the O*NET dataset, but several other detailed occupation codes within the same broad occupation code were available, then the scores of the missing detailed occupation codes were

substituted with the mean values of the other detailed occupation codes within the same broad occupation code.

- For example, MIDUS SOC code '21-1029' did not match any O*NET codes. The mean values of scores for '21-1021' thru '21-1023' were used to substitute SOC code '21-1029'.

4. Mean substitution scenario 4: When a 6-digit SOC code ended with '99', it was the last code in a broad category and meant to encompass all cases not listed separately in the broad category. For example, '21-2099' referred to 'Religious Workers, all other,'' '25-3099' referred to 'Teachers and Instructors, all other.' When this type of SOC code was missing, its scores were populated with the mean scores of all the specific detailed occupation codes.
   - For example, MIDUS SOC code '21-2099' did not match any O*NET codes. We treated this code as '21-2000' and used the mean values of the scores for SOC codes '21-2011' and '21-2021' to substitute SOC code '21-2099'.

Table 2 below lists all the SOC codes for which mean score substitution was used when linking MIDUS Milwaukee Refresher data with the O*NET 2012 dataset. The SOC codes that were used to compute the mean scores are also listed.

Table 2. Comprehensive list of unmatched SOC codes and the mean substitution adjustments.

Unmatched SOC codes	SOC Codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
11-3071	11-3071.01; 11-3071.02; 11-3071.03	
11-9199	11-9199.01; 11-9199.02; 11-9199.04; 11-9199.08	
13-1031	13-1031.01; 13-1031.02	
13-2011	13-2011.01; 13-2011.02	
13-2070	13-2071; 13-2072	
13-2071	13-2071.01	No substitution when link with Value data
17-3029	17-3029.01; 17-3029.02; 17-3029.03; 17-3029.04; 17-3029.05; 17-3029.06; 17-3029.07; 17-3029.09	
21-1029	21-1021; 21-1022; 21-1023	
25-3099	25-3011; 25-3021	When link with Value data, scores for 25-3099.02 are used for substitution.
29-9099	29-9099.01	
31-1014	31-1011; 31-1013	No substitution when link with Value data
31-9097	31-9091; 31-9092; 31-9093; 31-9094; 31-9095; 31-9096; 31-9099	Mean substitutions are used only when link with Interests data
33-1021	33-1021.01; 33-1021.02	
33-1099	33-1011; 33-1012; 33-1021	
33-3051	33-3051.01; 33-3051.03	
43-4031	43-4031.01; 43-4031.02; 43-4031.03	
43-4041	43-4041.01; 43-4041.02	

Unmatched SOC codes	SOC Codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
43-4199	43-4111, 43-4121, 43-4131, 43-4141, 43-4151, 43-4161, 43-4171, 43-4181	
43-5081	43-5081.01; 43-5081.02; 43-5081.03; 43-5081.04	
43-9041	43-9041.01; 43-9041.02	
43-9199	43-9011; 43-9021; 43-9022; 43-9031; 43-9041; 43-9051; 43-9061; 43-9071; 43-9081; 43-9111	
47-2031	47-2031.01; 47-2031.02	
47-2152	47-2052.01; 47-2052.02	
49.9069	49-9061; 49-9062; 49-9063; 49-9064	
49-3023	49-3023.01; 49-3023.02	
49-9021	49-9021.01; 49-9021.02	
49-9069	49-9061, 49-9062, 49-9063, 49-9064	
51-2099	51-2091; 51-2092; 51-2093	
51-3099	51-3091; 51-3092; 51-3093	
51-4121	51-4121.06; 51-4121.07	
51-9199	51-9199.01	When link with Interests data, mean scores of codes from 51-9191 to 51-9198 are used for substitution
53-7081	53-7011; 53-7021; 53-7031; 53-7032; 53-7033; 53-7041; 53-7051; 53-7061; 53-7062; 53-7063; 53-7064; 53-7071	Mean substitutions are used only when link with Context CX data

Once linked, variables from the O*NET 2012 datasets were renamed to conform with MIDUS variable naming conventions. The example in Table 3 shows how original O*NET variable names were retained and incorporated into new variable labels. In this example, variable 'RAAABIM1A1A1' represents the Abilities-Important scores for each SOC code to which it was linked.

Table 3: Variable renaming example

Original O*NET 2012 Variable Name	O*NET-MIDUS Merged Variable Name	O*NET-MIDUS Merged Variable Label
@1.A.1.a.1	RAAABIM1A1A1	O*NET 2012 Element ID @1.A.1.a.1: Abilities-Important: Oral Comprehension