

Documentation of Data Linkage between MIDUS Milwaukee 1 (MKE1) Survey and O*NET 2004 (7.0) Database

This document provides details on the data linkage performed between the MIDUS Milwaukee 1 (MKE1) Survey sample and the 2004 O*NET 7.0 database that produced the following standalone dataset:

MKE1_ONET2004_N350_20220222.sav

Specifically, this dataset was derived by linking MKE1 cases' survey responses (collected between 2005-06) and their resulting 2000 Standard Occupation Classification (SOC) codes with the 2004 Occupational Information Network (O*NET 7.0) 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 1 sample used SOC codes from 350 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 12 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 350 MIDUS Milwaukee 1 cases matched with the 2004 O*NET database.

Table 1 listed number of SOC codes available in each O*NET dataset and number of cases directly matched with MIDUS Milwaukee 1 sample (out of the 350 valid cases).

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

2004 O*NET datasets	# of SOC codes Available in O*NET data	# of Cases Matched with MKE1 Sample	% of Cases Matched with MKE1 Sample
Abilities (IM & LV)	916	286	82%
Interests	900	280	80%
Values	900	280	80%
Styles (IM)	280	n.a.	n.a.
Skills (IM & LV)	916	286	82%
Knowledge (IM & LV)	916	286	82%
Activities (IM & LV)	916	286	82%
Context (CX & CT)	916	286	82%

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 ‘11-3031’ did not have a direct match ‘11-3031.00’ in the O*NET data, but O*NET did have ‘11-3021.01’ and ‘11-3031.02’. MIDUS used the mean values of the SOC codes ‘11-3031.01’ and ‘11-3031.02’ to substitute the scores for ‘11-3031.00’ and match them with MIDUS ‘11-3031’.
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 ‘11-9030’ referred to ‘education administrators,’ ‘11-9031’ referred to ‘education administrators, preschool and childcare center,’ ‘11-9032’ referred to ‘education administrators, elementary and secondary school,’ ‘11-9033’ referred to ‘education administrators, postsecondary,’ ‘11-9039’ referred to ‘education administrators, all other’. ‘11-9030’ is a broad occupation code, while ‘11-9031’, ‘11-9032’, ‘11-9033’ and ‘11-9039’ 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 ‘11-9030’ did not match any O*NET codes. ‘11-9030’ is a broad occupation that includes four detailed occupations (‘11-9031’, ‘11-9032’, ‘11-9033’, ‘11-9039’). Therefore, the mean values of scores for SOC codes ‘11-9031’, ‘11-9032’, ‘11-9033’, ‘11-9039’ were used to substitute the scores for SOC code ‘11-9030’.
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 '17-3019' did not match any O*NET codes. The mean values of scores for '17-3011' thru '17-3013' were used to substitute SOC code '17-3019'.
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,' '27-2099' referred to 'Entertainers and Performers, Sports and Related Workers, 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 '21-2090' 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 1 data with the O*NET 2004 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
11-1011	11-1011.01; 11-1011.02	
11-1021	11-1011	Mean substitution are used only when link with Interests and Values data
13-1031	13-1031.01; 13-1031.02	
13-1041	13-1041.01; 13-1041.02; 13-1041.03; 13-1041.04; 13-1041.05; 13-1041.06	
13-1071	13-1071.01; 13-1071.02	
13-2011	13-2011.01; 13-2011.02	
15-1071	15-1071.01	Mean substitution are used only when link with Interests and Values data
17-3023	17-3023.01; 17-3023.02; 17-3023.03	
37-1011	37-1011.01; 37-1011.02	
43-1011	43-1011.01; 43-1011.02	
43-3021	43-3021.01; 43-3021.02; 43-3021.03	
43-4051	43-4051.01; 43-4051.02	Mean substitution are used only when link with Interests and Values data
43-4061	43-4061.01; 43-4061.02	
43-5081	43-5081.01; 43-5081.02; 43-5081.03; 43-5081.04	

Unmatched SOC codes	SOC Codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
43-9051	43-9051.01; 43-9051.02	
43-9071	43-9071.01	
47-1011	47-1011.01; 47-1011.02	
47-2031	47-2031.01; 47-2031.02; 47-2031.03; 47-2031.04; 47-2031.05; 47-2031.06	
47-2073	47-2073.01; 47-2073.02	
47-2152	47-2152.01; 47-2152.02; 47-2152.03	
49-3023	49-3023.01; 49-3023.02	
51-2092	51-2093	
51-4011	51-4011.01	
51-4033	51-4033.01; 51-4033.02	
51-4072	51-4072.01; 51-4072.02; 51-4072.03; 51-4072.04; 51-4072.05	
51-4081	51-4081.01; 51-4081.02	
51-4121	51-4121.01; 51-4121.02; 51-4121.03; 51-4121.04; 51-4121.05	
51-4193	51-4193.01; 51-4193.02; 51-4193.03; 51-4193.04	
51-5022	51-5022.01; 51-5022.02; 51-5022.03; 51-5022.04; 51-5022.05; 51-5022.06; 51-5022.07; 51-5022.08; 51-5022.09; 51-5022.10; 51-5022.11; 51-5022.12; 51-5022.13	
51-5023	51-5023.01; 51-5023.02; 51-5023.03; 51-5023.04; 51-5023.05; 51-5023.06; 51-5023.07; 51-5023.08; 51-5023.09	
51-6011	51-6011.01; 51-6011.02; 51-6011.03	
51-9061	51-9061.01; 51-9061.02; 51-9061.03; 51-9061.04; 51-9061.05	
51-9121	51-9121.01; 51-9121.02	
51-9194	51-9194.01; 51-9194.02; 51-9194.03; 51-9194.04; 51-9194.05; 51-9194.06	
53-3032	53-3032.01; 53-3032.02	
53-7032	53-7032.01; 53-7032.02	
53-7062	53-7062.01; 53-7062.02; 53-7062.03	

Once linked, variables from the O*NET 2004 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 ‘BAABIM1A1A1’ represents the Abilities-Important scores for each SOC code to which it was linked.

Table 3: Variable renaming example

Original O*NET 2004 Variable Name	O*NET-MIDUS Merged Variable Name	O*NET-MIDUS Merged Variable Label
@1.A.1.a.1	BAABIM1A1A1	O*NET 2004 Element ID @1.A.1.a.1: Abilities-Important: Oral Comprehension