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

Documentation of Data Linkage between MIDUS 2 (M2) Survey and O*NET 2004 (7.0) Database

This document provides details on the data linkage performed between the MIDUS 2 (M2) Survey sample and the 2004 O*NET 7.0 database that produced the following standalone dataset:

M2_ONET2004_N3362_20220222.sav

Specifically, this dataset was derived by linking M2 cases' survey responses (collected between 2004-05) 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 2 sample used SOC codes from 3,362 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 3,362 MIDUS 2 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 directyly matched with MIDUS 2 sample (out of the 3,362 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 M2 Sample	% of Cases Matched with M2 Sample
Abilities (IM & LV)	916	2,402	71%
Interests	900	2,312	69%
Values	900	2,312	69%
Styles (IM)	280	n.a.	n.a.
Skills (IM & LV)	916	2,402	71%
Knowledge (IM & LV)	916	2,402	71%
Activities (IM & LV)	916	2,402	71%
Context (CX & CT)	916	2,402	71%

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 2 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
11-1031	11-1011; 11-1021	When link with Interests and Value data, only 11-1011 is used for substitution
11-2031	11-2011; 11-2021; 11-2022	Mean substitution are used only when link with Interests and Values data
11-3031	11-3031.01; 11-3031.02	
11-3049	11-3040; 11-3041; 11-3042	
11-3071	11-3071.01; 11-3071.02	
11-9011	11-9011.01; 11-9011.02; 11-9011.03	
11-9039	11-9031; 11-9032; 11-9033	
11-9199	11-9011; 11-9012; 11-9021; 11-9031; 11-9032; 11-9033; 11-9041; 11-9051; 11-9061; 11-9071; 11-9081; 11-9111; 11-9121; 11-9131; 11-9141; 11-9151	
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	

Unmatched SOC codes	SOC codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
13-1061	13-1011; 13-1021; 13-1022; 13-1023; 13-1031; 13-1032; 13-1041; 13-1051; 13-1071; 13-1072; 13-1073	
13-1071	13-1071.01; 13-1071.02	
13-1079	13-1071; 13-1072; 13-1073	
13-1081	13-1011; 13-1021; 13-1022; 13-1023; 13-1031; 13-1032; 13-1041; 13-1051; 13-1071; 13-1072; 13-1073	
13-2011	13-2011.01; 13-2011.02	
13-2021	13-2021.01; 13-2021.02	
15-1011	15-1021; 15-1031; 15-1032; 15-1041; 15-1051; 15-1061; 15-1071; 15-1081	
15-1071	15-1071.01	Mean substitution are used only when link with Interests and Values data
17-2081	17-2011; 17-2021; 17-2041; 17-2051; 17-2061; 17-2071; 17-2072	Mean substitution are used only when link with Interests and Values data
17-2111	17-2111.01; 17-2111.02; 17-2111.03	
17-2121	17-2121.01; 17-2121.02	
17-3011	17-3011.01; 17-3011.02	
17-3012	17-3012.01; 17-3012.02	
17-3023	17-3023.01; 17-3023.02; 17-3023.03	
17-3025	17-3021; 17-3022; 17-3023; 17-3024; 17-3026; 17-3027	Mean substitution are used only when link with Interests and Values data
17-3031	17-3031.01; 17-3031.02	
19-1013	19-1013.01; 19-1013.02	
19-3031	19-3031.01; 19-3031.02; 19-3031.03	
19-3091	19-3091.01; 19-3091.02	
19-4051	19-4051.01; 19-4051.02	
19-4093	19-4091; 19-4092	
23-2091	23-2092; 23-2093	
25-1000	25-1021; 25-1022; 25-1032; 25-1041; 25-1042; 25-1043; 25-1052; 25-1054; 25-1061; 25-1062; 25-1063; 25-1065; 25-1066; 25-1067; 25-1071; 25-1072; 25-1121; 25-1123; 25-1124; 25-1125; 25-1191; 25-1194	
25-1011	25-1021; 25-1022; 25-1032; 25-1041; 25-1042; 25-1043; 25-1052; 25-1054; 25-1061; 25-1062; 25-1063; 25-1065; 25-1066; 25-	

Unmatched SOC codes	SOC codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
	1067; 25-1071; 25-1072; 25-1121; 25-1123; 25-1124; 25-1125; 25-1191; 25-1194	
25-1112	25-1021; 25-1022; 25-1032; 25-1041; 25-1042; 25-1043; 25-1052; 25-1054; 25-1061; 25-1062; 25-1063; 25-1065; 25-1066; 25-1067; 25-1071; 25-1072; 25-1121; 25-1123; 25-1124; 25-1125; 25-1191; 25-1194	
25-1122	25-1121; 25-1123; 25-1124; 25-1125	
25-1126	25-1121; 25-1123; 25-1124; 25-1125	
25-1193	25-1191; 25-1194	
25-2020	25-2021; 25-2022; 25-2023	
25-3000	25-3011; 25-3021	
27-1012	27-1011; 27-1013	
27-1013	27-1013.01; 27-1013.02; 27-1013.03; 27-1013.04	
27-1027	27-1027.01; 27-1027.02	
27-2012	27-2012.01; 27-2012.02; 27-2012.03; 27-2012.04; 27-2012.05	
27-2041	27-2041.01; 27-2041.02; 27-2041.03	
27-2042	27-2042.01; 27-2042.02	
27-3043	27-3043.01; 27-3043.02; 27-3043.03; 27-3043.04	
27-4021	27-4021.01; 27-4021.02	
29-2034	29-2034.01; 29-2034.02	
31-9011	31-9091; 31-9092; 31-9093; 31-9095; 31-9096	Code 31-9095 was not available when link with Interests and Value data
31-9094	31-9091; 31-9092; 31-9093; 31-9095; 31-9096	Code 31-9095 was not available when link with Interests and Value data
31-9095	31-9091; 31-9092; 31-9093; 31-9096	Mean substitution are used only when link with Interests and Values data
33-1011	33-1012	
33-2011	33-2011.01; 33-2011.02	
33-3021	33-3021.01; 33-3021.02; 33-3021.03; 33-3021.04; 33-3021.05	
33-3051	33-3051.01; 33-3051.02; 33-3051.03	
35-2013	35-2011; 35-2012; 35-2014; 35-2015	
37-1011	37-1011.01; 37-1011.02	
37-1012	37-1012.01; 37-1012.02	

Unmatched SOC codes	SOC codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
39-5094	39-5091; 39-5092	
41-3031	41-3031.01; 41-3031.02	
41-4011	41-4011.01; 41-4011.02; 41-4011.03; 41-4011.04; 41-4011.05; 41-4011.06	
41-9021	41-9022	Mean substitution are used only when link with Interests and Values data
43-1011	43-1011.01; 43-1011.02	
43-2021	43-2021.01; 43-2021.02	
43-3021	43-3021.01; 43-3021.02; 43-3021.03	
43-4031	43-4031.01; 43-4031.02; 43-4031.03	
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	
43-4181	43-4181.01; 43-4181.02	
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	
43-9051	43-9051.01; 43-9051.02	
43-9071	43-9071.01	
45-1011	45-1011.01; 45-1011.02; 45-1011.03; 45-1011.04; 45-1011.05; 45-1011.06	
45-2092	45-2092.01; 45-2092.02	
45-4022	45-4022.01	
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	
47-5021	47-5021.01; 47-5021.02	
49-2011	49-2011.01; 49-2011.02; 49-2011.03	
49-2022	49-2022.01; 49-2022.02; 49-2022.03; 49-2022.04; 49-2022.05	
49-3011	49-3011.01; 49-3011.02; 49-3011.03	
49-3023	49-3023.01; 49-3023.02	
49-9012	49-9012.01; 49-9012.02; 49-9012.03	
49-9021	49-9021.01; 49-9021.02	
49-9031	49-9031.01; 49-9031.02	

Unmatched SOC codes	SOC codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
49-9063	49-9063.01; 49-9063.02; 49-9063.03; 49-9063.04	
51-2092	51-2093	
51-3011	51-3011.01; 51-3011.02	
51-4011	51-4011.01	
51-4031	51-4031.01; 51-4031.02; 51-4031.03; 51-4031.04	
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	
51-4122	51-4122.01; 51-4122.02; 51-4122.03; 51-4122.04;51-4122.05	
51-4191	51-4191.01; 51-4191.02; 51-4191.03	
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	
51-6031	51-6031.01; 51-6031.02	
51-6052	51-6052.01; 51-6052.02	
51-7041	51-7041.01; 51-7041.02	
51-7042	51-7042.01; 51-7042.02	
51-8013	51-8013.01; 51-8013.02	
51-8021	51-8021.01; 51-8021.02	
51-9041	51-9041.01; 51-9041.02	
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	
51-9195	51-9195.01; 51-9195.02; 51-9195.03; 51-9195.04; 51-9195.05; 51-9195.06; 51-9195.07	
51-9198	51-9198.01; 51-9198.02	
53-2022	53-2021	
53-3032	53-3032.01; 53-3032.02	

Unmatched SOC codes	SOC codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
53-5011	53-5011.01; 53-5011.02	
53-5021	53-5021.01; 53-5021.02; 53-5021.03	
53-6051	53-6051.01; 53-6051.02; 53-6051.03; 53-6051.04; 53-6051.05; 53-6051.06	
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 'B1ABIM1A1A1' 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	B1ABIM1A1A1	O*NET 2004 Element ID @1.A.1.a.1: Abilities-Important: Oral Comprehension