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

Documentation of Data Linkage between MIDUS Refresher 1 (MR1) Survey and O*NET 2012 (17.0) Database

This document provides details on the data linkage performed between the baseline MIDUS Refesher (MR1) Survey sample and the 2012 O*NET 17.0 database that produced the following standalone dataset:

MR1_ONET2012_N2348_20210914.sav

Specifically, this dataset was derived by linking MR1 cases' survey responses (collected between 2012-14) 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 Refresher sample used SOC codes from 2,348 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 2,348 MIDUS 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 2,348 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 MR1 Sample	% of Cases Matched with MR1 Sample
Abilities (IM & LV)	903	1,889	80%
Interests	897	1,888	81%
Values	974	1,914	81%
Styles (IM)	902	1,889	80%
Skills (IM & LV)	903	1,889	80%
Knowledge (IM & LV)	903	1,889	80%
Activities (IM & LV)	903	1,889	80%
Context (CX & CT)	898/902	1,888	80%

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.' '11-9030' is a broad occupation code, while '11-9031' and '11-9032' 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 available SOC codes '11-9031', '11-9032', '11-9033' 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 '31-1014' did not match any O*NET codes. The mean values of scores for '31-1011' and '31-1013' were used to substitute SOC code '31-1014'.

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, '15-1199' referred to 'Computer occupations, all other,''19-1099' referred to 'Life scientists, 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 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
11-3031	11-3031.01; 11-3031.02	
11-3071	11-3071.01; 11-3071.02; 11-3071.03	
11-9013	11-9013.01; 11-9013.03	
11-9030	11-9031; 11-9032; 11-9033	
11-9039	11-9031; 11-9032; 11-9033	When the link with Value data, used mean scores of 11-9039.01 and 11-9039.02 to substitute
11-9199	11-9199.01; 11-9199.02; 11-9199.04; 11-9199.08	
13-1031	13-1031.01; 13-1031.02	
13-1041	13-1041.01; 13-1041.02; 13-1041.03; 13-1041.04; 13-1041.06; 13-1041.07	
13-1199	13-1199.01; 13-1199.03; 13-1199.04	
13-2011	13-2011.01; 13-2011.02	
13-2021	13-2021.01; 13-2021.02	
13-2071	13-2071.01	No substitution when the link with Value data
15-1152	15-1151	No substitution when link with Value data
15-1199	15-1199.01; 15-1199.02; 15-1199.03; 15-1199.04; 15-1199.05; 15-1199.08; 15-1199.09; 15-1199.11	
17-2199	17-2199.01; 17-2199.02; 17-2199.03; 17-2199.04; 17-2199.05; 17-2199.07; 17-2199.08; 17-2199.10	
17-3023	17-3023.01; 17-3023.03	

Unmatched SOC codes	SOC Codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
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	
17-3031	17-3031.01; 17-3031.02	
19-3031	19-3031.01; 19-3031.02; 19-3031.03	
19-3039	19-3039.01	
19-3099	19-3099.01	
19-4011	19-4011.01; 19-4011.02	
19-4041	19-4041.01; 19-4041.02	
19-4099	19-4099.01; 19-4099.02; 19-4099.03	
21-1019	21-1011; 21-1012; 21-1013; 21-1014; 21-1015	
21-1029	21-1021; 21-1022; 21-1023	
21-2099	21-2011; 21-2021	
23-2099	23-2091; 23-2093	
25-1199	25-1191; 25-1192; 25-1193; 25-1194	
25-2020	25-2021; 25-2022; 25-2023	
25-2051	25-2053; 25-2054; 25-2059	No substitution when link with Value data
25-2052	25-2053; 25-2054; 25-2059	No substitution when link with Value data
25-3099	25-3011; 25-3021	When link with Value data, scores for 25-3099.02 are used for substitution
27-1029	27-1021; 27-1022; 27-1023; 27-1024; 27-1025; 27-1026; 27-1027	
27-2012	27-2012.01; 27-2012.02; 27-2012.03; 27-2012.04; 27-2012.05	
27-2020	27-2021; 27-2022; 27-2023	
27-2041	27-2041.01; 27-2041.04	
27-2042	27-2042.01; 27-2042.02	
27-2099	27-2011; 27-2012; 27-2021; 27-2022; 27-2023; 27-2031; 27-2032; 27-2041; 27-2042	
27-3043	27-3043.04; 27-3043.05	
29-1069	29-1069.01; 29-1069.02; 29-1069.03; 29-1069.04; 29-1069.05; 29-1069.06; 29-1069.07; 29-1069.08; 29-1069.09; 29-1069.10; 29-1069.11; 29-1069.12	
29-1129	29-1122; 29-1123; 29-1124; 29-1125; 29-1126; 29-1127	Code 29-1128 is added to compute mean scores when link with Value data
29-1199	29-1199.01; 29-1199.04; 29-1199.05	
29-2099	29-2099.01; 29-2099.06	
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

Unmatched SOC codes	SOC Codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
31-9099	31-9099.01; 31-9099.02	
33-1021	33-1021.01; 33-1021.02	
33-1099	33-1011; 33-1012; 33-1021	
33-2011	33-2011.01; 33-2011.02	
33-2021	33-2021.01; 33-2021.02	
33-3021	33-3021.01; 33-3021.02; 33-3021.03; 33-3021.05; 33-3021.06	
33-3051	33-3051.01; 33-3051.03	
33-9099	33-9091; 33-9092; 33-9093	When link with Value data and Interests data, scores for 33-9099.02 are used for substitution
39-3019	39-3011; 39-3012	
39-5090	39-5091; 39-5092; 39-5093; 39-5094	
39-9099	39-9011; 39-9021; 39-9031; 39-9032; 39-9041	
41-3031	41-3031.01; 41-3031.02	
41-3099	41-3011; 41-3021; 41-3031; 41-3041	When link with Value data and Interests data, scores for 41-3099.01 are used for substitution
41-9099	41-9091	
43-3021	43-3021.01; 43-3021.02	
43-4031	43-4031.01; 43-4031.02; 43-4031.03	
43-4041	43-4041.01; 43-4041.02	
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	
45-2092	45-2092.01; 45-2092.02	
45-4029	45-4021; 45-4022; 45-4023	
47-2031	47-2031.01; 47-2031.02	
47-2152	47-2152.01; 47-2152.02	
49-3023	49-3023.01; 49-3023.02	
49-9021	49-9021.01; 49-9021.02	
49-9099	49-9099.01	When link with Interests data, mean scores of codes from 49-9091 to 49-9098 are used for substitution
51-2099	51-2091; 51-2092; 51-2093	
51-4121	51-4121.06; 51-4121.07	
51-4199	51-4191; 51-4192; 51-4193; 51-4194	
51-8099	51-8099.03	When link with Interests data, mean scores of codes from 51-8091 to 51-8093 are used for substitution
51-9199	51-9199.01	When link with Interests data, mean scores of codes from 51-

Unmatched SOC codes	SOC Codes Used to Compute Mean Substitution Scores	Substitution Variation Across Datasets
		9191 to 51-9198 are used for substitution
53-3099	53-3011; 53-3021; 53-3022; 53-3031; 53-3032; 53-3033; 53-3041	
53-5021	53-5021.01; 53-5021.02; 53-5021.03	
53-6051	53-6051.01; 53-6051.07; 53-6051.08	
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 'RA1ABIM1A1A1' 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	RA1ABIM1A1A1	O*NET 2012 Element ID @1.A.1.a.1: Abilities-Important: Oral Comprehension