

DATA DISCOVERABILITY IN THE BIOMEDICAL SCIENCES

Barry T. Radler, PhD

University of Wisconsin-Madison Institute on Aging



Overview



- **MIDUS background:**
 - Study design, unique features, content/measures
- **Research data documentation**
 - Importance of structured metadata
 - Using Data Documentation Initiative (DDI)
 - To facilitate documentation and discovery of longitudinal research data
- **Demonstrate MIDUS Portal**
 - Metadata standards → Search, Harmonization,



MIDUS Background

- Unique Features



- **Multidisciplinary**
 - “Bio-psycho-social” approach
- **Longitudinal**
 - Repeated assessments over 24 years
 - Baseline age range: 25-74
- **Multiple samples**
 - 2 National RDD (Core and Refresher)
 - Sibling, Twins (within Core)
 - 2 African American
 - 1 Japanese



MIDUS Background

- Unique Features



- **Multidisciplinary → multi-project approach:**
 - **Project 1 (Survey): 45 min Phone, 100 page SAQ**
 - **Project 2 (Diary): 8-day phone diary, cortisol**
 - **Project 3 (Cognitive): phone-based cog measures**
 - **Project 4 (Biomarker): 2-day clinic visit**
 - **Project 5 (Neuroscience): MRI affective neuro-imaging**
 - **Project 6 (Gene Expression): RNA**

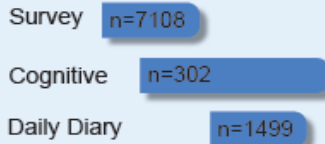


MIDUS Background

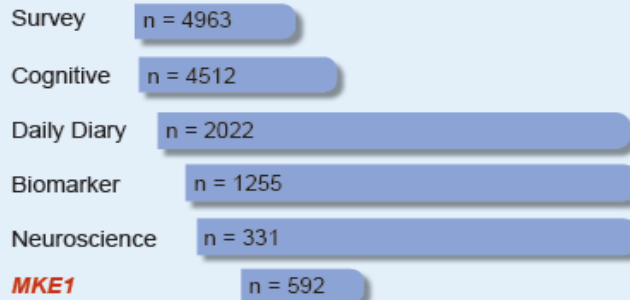
- Unique Features



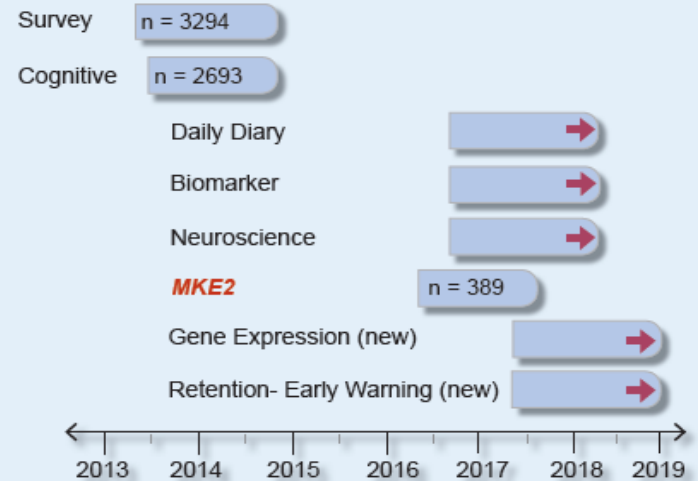
M1



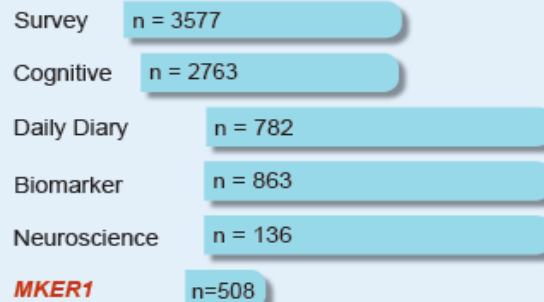
M2



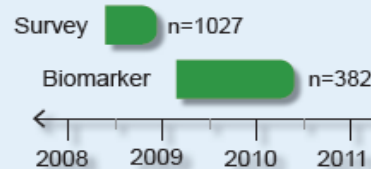
M3



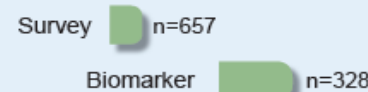
MR1



MIDJA1



MIDJA2



MIDUS Background

- Unique Features



- **23 primary datasets**
 - 31,000 variables
 - $N > 13,000$
- **Open Data**
 - Archived/distributed by ICPSR at University of Michigan
 - Widespread secondary usage since 1999:
 - 121,000 data downloads
 - 17,000 users
 - 1,200 publications



MIDUS Documentation

- Metadata are critical



- **Data Documentation Initiative:**

- **DDI is a metadata standard to describe social science data and make it...**
 - Machine-actionable
 - Independently understandable

- **University of Michigan (ICPSR):**

- **Uses DDI Codebook**

- **University of Wisconsin (IOA):**

- **Uses DDI Lifecycle**



MIDUS Documentation

- Metadata are critical



- **MIDUS Portal uses Colectica**

- **Based on DDI Lifecycle**

- **Harmonization (internal)**

- Clarifies the related nature of versions of longitudinal and cross-cohort variables

- **Search**

- Searches various fields: variable name, label, question text, assigned concepts

- **Data extract**

- Researchers can focus on variables of interest
 - Facilitate accurate merges across numerous datasets
 - Ease data management burden





The MIDUS Colectica Portal

http://midus.colectica.org

← → ↻ 🏠 midus.colectica.org


Apps Personal Work DDI All files and folders - My Drive - Google Drive UW-Madison NetID MIDUS Colectica repo DTD Portal MIDUS 3 MySQL MIDUS - N

MIDUS Search Explore Basket 2 Admin


Home

Welcome

Welcome to the MIDUS Colectica Portal. This website provides users access to publicly available data and richly-structured documentation for the MIDUS study.



The Portal is based on the DDI metadata standard and is powered by Colectica software.



Midlife in the United States

Midlife in the United States (MIDUS) is a national longitudinal study of health and well-being. It was conceived by a multidisciplinary team of scholars interested in understanding the health and well-being of middle-aged and older Americans, such that it now includes data from over 12,000 individuals, comprising thousands of variables in different scientific areas among distinct cohorts. This website is a one-stop portal that provides MIDUS data and metadata (information about data) for exploration and analysis. As of 2016 the portal supports the following features:

- **Home:** The home page still allows users to browse variables, read abstracts, download instruments and documentation, and search for variables within projects.
- **Search:** improved and expanded search capabilities allow searches across ALL publicly-available MIDUS datasets, by variable name, label, or question text.
 - Search results show all variable-level metadata and descriptive statistics.
 - Search results can be shared by copying and pasting the URL from the search results window.
- **Explore:** harmonization and concordance information is available for all SURVEY variables. This information can be found by clicking on the Explore menu and browsing the "Variables" associated with each survey variable.
 - Under Explore, click on the "graph" icon to produce visualizations across concordance variables.
 - Click on the "grid" icon to examine code list comparability across concordance variables.
 - Any threats to comparability among concordance variables will be detailed in two fields called "Comparability notes" and "Comparability class."
- **Basket (Custom Download):** variables displayed under the Search tab can be added to a variable "basket" by clicking on the shopping cart icon next to them. Users can then download their basket in several formats:
 - A custom CSV or SPSS data file (custom data downloads from multiple datasets are presented in a "wide" format, i.e. the same measures from different datasets are presented in the same column).
 - A custom PDF codebook. We strongly suggest that you download a codebook with your custom dataset – the codebooks include important information about the variables in your basket.
 - The DDI-compliant XML code.

The MIDUS study and this portal are supported by multiple grants from the National Institute on Aging (5R37AG027343, 5P01AG020166, 1R03AG046312) and by the National Institute of Health (5R01AG020166, 5P01AG020166, 1R03AG046312).

MIDUS 1

Project 1 (Survey)

In 1994/95, the MacArthur Midlife Research Network carried out a national survey of over 7,000 Americans aged 25 to 74.

MIDUS Portal – DDI Lifecycle

- Search



Variables

smoking



Sort by: Relevance

Syntax

Item types: Variables

Query: smoking

Results 1 to 20 of 125 (0.205 seconds)

C1PA41

Ever tried to quit smoking

Since we last interviewed you in [M2PhoneMO], [M2PhoneYR] have you tried to quit smoking?

Study: [MIDUS 3 Project 1](#) / Data File: [MIDUS 3](#) / Variable Group: [Phone](#) / Variable Group: [Phone A: Health](#)

A2DNOSMO

Never smokes

R never smokes/ R may smoke

Study: [MIDUS 1 Project 2](#) / Data File: [M1_P2_DAILY DAIRY_N=1499_11-01-12](#) / Variable Group: [MIDUS 1 Project 2](#) / Variable Group: [Personal Background](#)

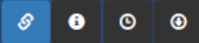
MIDUS Portal – DDI Lifecycle

- Search



MIDUS Search Explore IFDTC 2017 Demo 7 Admin

Help Feedback bradler@wisc.edu



Appears Within

- MIDUS 3 Project 1
 - MIDUS 3
- MIDUS 3 Project 1
 - MIDUS 3
- M3_P1_AGGREGATE_N=3294_20150608
 - Phone
 - Phone A: Health

Ever tried to quit smoking

MIDUS 3 Project 1 MIDUS 3 Phone 2

Variable Details Lineage Concordance Variables

Name C1PA41

Label Ever tried to quit smoking

Question Text Since we last interviewed you in [M2PhoneMO], [M2PhoneYR] have you tried to quit smoking?

Backward Skip (C1PA37 = NEVER) OR (C1PA38A = NO, DK, OR REFUSED) OR (C1PA39 = NO, DK, OR REFUSED)

Forward Skip GO TO C1PA44.

Dataset M3_P1_AGGREGATE_N=3294_20150608

Comparability Class Wording in Question - Difference exists among waves because there were wording changes in question just at particular waves.

Comparability Notes Wording change in question text: 'Since we last interviewed you in [MONTH], [YEAR]' phrase was deleted from MR and MKER baseline instruments.

Value	Label	Frequency
1	YES	203
2	NO	103
9	INAPP	2988

Valid	Invalid	Min	Max
306	2988	1	2

MIDUS Portal – DDI Lifecycle

- Harmonization



MIDUS Search **Explore** Basket 7 Admin

Help Feedback bradler@wisc.edu

Survey

- Administration
- Recession Experience
- Health
 - Self-Evaluated Health
 - Physical/Mental Health Limits Work
 - Health at Age 16
 - History of Neurological Disorder
 - History of Heart Trouble
 - Cardiovascular Health
 - Angina
 - Chest Pain
 - Blood Pressure
 - Cancer
 - Hypertension/Stroke Removal
 - History of Smoking**
 - History of Drinking
 - Depression - Depressed Affect
 - Depression - Anhedonia
 - Depression - Depressed Affect and Anhedonia
 - Anxiety/Worry/Panic
- Education, Occupation, and Marital Status
- Household Roster and Children
- Caregiving
- Living Arrangements
- Race and Ethnicity
- Life Satisfaction

	M1P1	M2P1	M3P1	MKE1	MKE2	MRP1	MKER
Age smoked first cigarette	A1PA41	B1PA37	C1PA37	BACA36	CACA36	RA1PA37	RAACA36
Smoked cigarettes ever	A1PA40	B1PA38A	C1PA38A	BACA37	CACA37	RA1PA38A	RAACA37
Age begin smoke regularly	A1PA42	B1PA38B	C1PA38B	BACA38	CACA38	RA1PA38B	RAACA38
Smoke cigarettes regularly now	A1PA43	B1PA39	C1PA39	BACA39	CACA39	RA1PA39	RAACA39
Ave # cigs per day_cur smoker	A1PA44	B1PA40	C1PA40	BACA40	CACA40	RA1PA40	RAACA40
Ever tried to quit smoking	A1PA45	B1PA41	C1PA41	BACA41	CACA41	RA1PA41	RAACA41
Age when last smoked regularly	A1PA46	B1PA42	C1PA42	BACA42	CACA42	RA1PA42	RAACA42
Ave # cigs per day_ex-smoker	A1PA47	B1PA43	C1PA43	BACA43	CACA43	RA1PA43	RAACA43
Ever used pipe/cigars/snuff/chew		B1PA44	C1PA44	BACA44	CACA44	RA1PA44	RAACA44
Who smoked - Father		B1PA45A		BACA45A		RA1PA45A	RAACA45A
Who smoked - Mother		B1PA45B		BACA45B		RA1PA45B	RAACA45B
Who smoked - Other		B1PA45C		BACA45C		RA1PA45C	RAACA45C
In home anyone smoke/use tobacco current		B1PA46	C1PA46	BACA46	CACA46	RA1PA46	RAACA46
At job anyone smoke/use tobacco ever		B1PA47	C1PA48	BACA47	CACA47	RA1PA47	RAACA47
At job anyone smoke/use tobacco current		B1PA48	C1PA47	BACA48	CACA48	RA1PA48	RAACA48

MIDUS Portal – DDI Lifecycle

- Harmonization



MIDUS Search Explore Basket 5 Admin

Help Feedback

	MIDUS 1	MIDUS 2	MIDUS 3	MIDUS 2 Milwaukee	MIDUS Refresher	Refresher Survey
Age smoked first cigarette	A1PA41	B1PA37	C1PA37	BACA36	RA1PA37	RAACA36
Smoked cigarettes ever	A1PA40	B1PA38A	C1PA38A	BACA37	RA1PA38A	RAACA37
Age begin smoke regularly	A1PA42	B1PA38B	C1PA38B	BACA38	RA1PA38B	RAACA38
Smoke cigarettes regularly now	A1PA43	B1PA39	C1PA39	BACA39	RA1PA39	RAACA39
Ave # cigs per day_cur smoker	A1PA44	B1PA40	C1PA40	BACA40	RA1PA40	RAACA40
Ever tried to quit smoking	A1PA45	B1PA41	C1PA41	BACA41	RA1PA41	RAACA41
Age when last smoked regularly	A1PA46	B1PA42	C1PA42	BACA42	RA1PA42	RAACA42
Ave # cigs per day_ex-smoker	A1PA47	B1PA43	C1PA43	BACA43	RA1PA43	RAACA43
Ever used pipe/cigars/snuff/chew		B1PA44	C1PA44	BACA44	RA1PA44	RAACA44
Who smoked - Father		B1PA45A		BACA45A	RA1PA45A	RAACA45A
Who smoked - Mother		B1PA45B		BACA45B	RA1PA45B	RAACA45B
Who smoked - Other		B1PA45C		BACA45C	RA1PA45C	RAACA45C

Include missing values

Who smoked - Mother

[Details...](#)

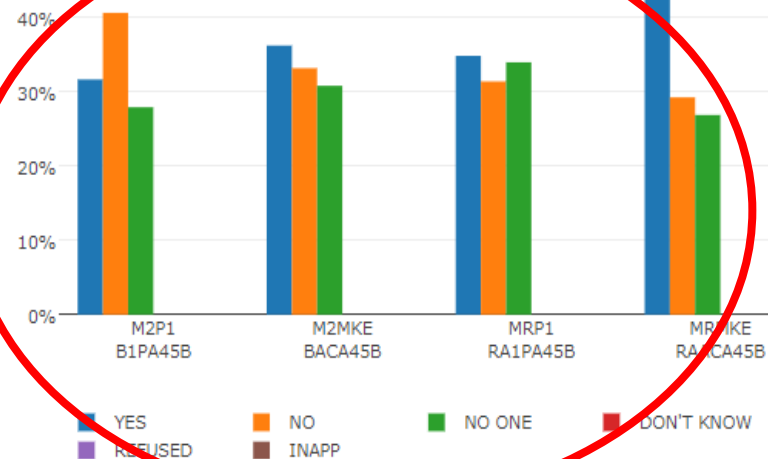
Comparability Class

General Item Availability; Baseline Item

Comparability Notes

Not in M1 and M3; was introduced at M2 as a new baseline item.

Who smoked - Mother



MIDUS Portal – DDI Lifecycle

- Harmonization



MIDUS Search Explore Basket 1 Admin Help Feedback bradler@wisc.edu

Survey

- Administration
- Recession Experience
- Health
 - Self-Evaluated Health**
 - Physical/Mental Health Limits Work
 - Health at Age 16
 - History of Neurological Disorder
 - History of Heart Trouble
 - Cardiovascular Health
 - Angina
 - Chest Pain
 - Blood Pressure
 - Cancer
 - Hysterectomy/Ovary Removal
 - History of Smoking
 - History of Drinking
 - Depression - Depressed Affect
 - Depression - Anhedonia
 - Depression - Depressed Affect and Anhedonia
 - Anxiety/Worry/Panic
- Education, Occupation, and Marital Status
- Household Roster and Children
- Caregiving
- Living Arrangements
- Race and Ethnicity
- Life Satisfaction
- Your Health
- Health Questions for Women

Physical health

	M1P1	M2P1	M3P1	MKE1	MKE2	MRP1	MKER
Physical health	A1PA4	B1PA1	C1PA1	BACA1	CACA1	RA1PA1	RAACA1
Mental or emotional health	A1PA5	B1PA2	C1PA2	BACA2	CACA2	RA1PA2	RAACA2
Self-evaluated health	A1PA6	B1PA3	C1PA3	BACA3		RA1PA3	RAACA3

Label Physical health

Concept Self-Rated Physical Health

Comparability Class Coding Scheme - Difference exists among waves because responses were coded differently.

Comparability Notes The M1 version of this variable is reverse coded

Statistics Code Comparison Correspondence tree

	M1P1 A1PA4	M2P1 B1PA1	M3P1 C1PA1	MKE1 BACA1	MKE2 CACA1	MRP1 RA1PA1	MKER RAACA1
POOR	1	5	5	5	5	5	5
FAIR	2	4	4	4	4	4	4
GOOD	3	3	3	3	3	3	3
VERY GOOD	4	2	2	2	2	2	2
EXCELLENT	5	1	1	1	1	1	1
DON'T KNOW	7	7	7	7	7	7	7
REFUSED	8	8	8	8	8	8	8
INAPP	9	9		9			

MIDUS Portal – DDI Lifecycle

- Data Extract



MIDUS Search Explore **Basket 7** Admin

Help Feedback bradler@wisc.edu

- Survey
 - Administration
 - Recession Experience
 - Health
 - Self-Evaluated Health
 - Physical/Mental Health Limits Work
 - Health at Age 16
 - History of Neurological Disorder
 - History of Heart Trouble
 - Cardiovascular Health
 - Angina
 - Chest Pain
 - Blood Pressure
 - Cancer
 - Hysterectomy/Ovary Removal
 - History of Smoking**
 - History of Drinking
 - Depression - Depressed Affect
 - Depression - Anhedonia
 - Depression - Depressed Affect and Anhedonia
 - Anxiety/Worry/Panic
 - Education, Occupation, and Marital Status
 - Household Roster and Children
 - Caregiving
 - Living Arrangements
 - Race and Ethnicity

	M1P1	M2P1	M3P1	MKE1	MKE2	MRP1	MKER
Age smoked first cigarette	A1PA41	B1PA37	C1PA37	BACA36	CACA36	RA1PA37	RAACA36
Smoked cigarettes ever	A1PA40	B1PA38A	C1PA38A	BACA37	CACA37	RA1PA38A	RAACA37
Age begin smoke regularly	A1PA42	B1PA38B	C1PA38B	BACA38	CACA38	RA1PA38B	RAACA38
Smoke cigarettes regularly now	A1PA43	B1PA39	C1PA39	BACA39	CACA39	RA1PA39	RAACA39
Ave # cigs per day_cur smoker	A1PA44	B1PA40	C1PA40	BACA40	CACA40	RA1PA40	RAACA40
Ever tried to quit smoking	A1PA45	B1PA41	C1PA41	BACA41	CACA41	RA1PA41	RAACA41
Age when last smoked regularly	A1PA46	B1PA42	C1PA42	BACA42	CACA42	RA1PA42	RAACA42
Ave # cigs per day_ex-smoker	A1PA47	B1PA43	C1PA43	BACA43	CACA43	RA1PA43	RAACA43
Ever used pipe/cigars/snuff/chew		B1PA44	C1PA44	BACA44	CACA44	RA1PA44	RAACA44
Who smoked - Father		B1PA45A		BACA45A		RA1PA45A	RAACA45A
Who smoked - Mother		B1PA45B		BACA45B		RA1PA45B	RAACA45B
Who smoked - Other		B1PA45C		BACA45C		RA1PA45C	RAACA45C
In home anyone smoke/use tobacco current		B1PA46	C1PA46	BACA46	CACA46	RA1PA46	RAACA46
At job anyone smoke/use tobacco ever		B1PA47	C1PA48	BACA47	CACA47	RA1PA47	RAACA47
At job anyone smoke/use tobacco current		B1PA48	C1PA47	BACA48	CACA48	RA1PA48	RAACA48

midus.colectica.org/explore/concordance/midus.wisc.edu/c2c30710-8ddc-40d9-a14c-184176fe725e#

MIDUS Portal – DDI Lifecycle

- Data Extract



MIDUS Search Explore **Basket 7** Admin

MANAGE
Create a Basket +

BASKETS
Basket

Basket

This is your active basket.

Documentation Downloads

This basket contains 7 items.

[Download PDF Codebook](#) [Download DDI 3.2 XML](#)

Variables (7)

[Download CSV](#) [Download SPSS](#)

- BACA36**
Age had first cigarette
- A1PA41**
Age had first cigarette
- B1PA37**
Age had first cigarette
- CACA36**
Age had first cigarette

MIDUS Portal – DDI Lifecycle

- Data Extract



*Basket (12).sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help



	Name	Label	Values	Missing	Type	Width	Decimals	Columns
1	MIDUSID	Master ID created by concatenating M2ID, MRID, MIDJA_IDs	None	None	Numeric	8	0	8
2	SAMPLMAJ	Major sample identification (aka Sample)	{1, MAIN RDD}...	None	Numeric	8	0	8
3	M2ID	MIDUS 2 ID number	None	None	Numeric	8	0	8
4	MRID	MIDUS Refresher ID number	None	None	Numeric	8	0	8
5	A1PA41	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	97, 98, 99	Numeric	8	0	8
6	B1PA37	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	96, 97, 98	Numeric	8	0	8
7	C1PA37	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	97, 98, 96	Numeric	8	0	8
8	BACA36	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	97, 98, 99	Numeric	8	0	8
9	CACA36	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	96, 97, 98	Numeric	8	0	8
10	RA1PA37	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	97, 98, 97	Numeric	8	0	8
11	RAACA36	Age had first cigarette	{96, NEVER HAD A CIGARETTE}...	97, 98, 97	Numeric	8	0	8

*Basket (12).sav [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help



	MIDUSID	SAMPLMAJ	M2ID	MRID	A1PA41	B1PA37	C1PA37	BACA36	CACA36	RA1PA37	RAACA36	var	var	var	var
1	10005	3	10005	.	36	35				
2	10014	1	10014	.	12	12				
3	10018	4	10018	.	9	7				
4	10020	1	10020	.	11	96	96				
5	10029	13	10029	96	13	.	.				
6	10039	1	10039	.	13	12				
7	10047	1	10047	.	96	96	96				
8	10063	3	10063	.	10	13	13				

Visible:

MIDUS Portal – DDI Lifecycle

- Data Extract



Midlife in the United States (MIDUS)

GSA 2019 Demo

Generated on Saturday, November 09, 2019 7:04 PM

Data File

Variable Count

23

SAMPLMAJ - Major sample identification

Type Code

Dataset

MR_MKE_AGGREGATE DATA_N508_20160914.sav

Comparability Class

General Item Availability

Comparability Notes

Variable specific to M1, M2, M3, M4, and M5.

Value	Label	Frequency	%
21	MILWAUKEE REFRESHER	508	100.0%

Valid	Invalid	Minimum	Maximum
508	0	21	21

MRID - MIDUS Refresher ID

Type Numeric (Long)

Dataset

MR_MKE_AGGREGATE DATA_N508_20160914.sav

Comparability Notes

The M2ID case identification system is used for the original MIDUS core sample first assessed in 1995. The MRID case identification system is used for the MIDUS Refresher sample.

2 of 14

Midlife in the United States (MIDUS)

Valid	Invalid	Minimum	Maximum
508	0	30005	39991

SAMPLMAJ - Major sample identification

Type Code

Dataset

MR_P1_AGGREGATE_N3577_20160824.sav

Comparability Class

Constructed Item - Difference exists among waves because some items were constructed only in particular waves.

Comparability Notes

Constructed in MKE1, MKE2, and MKE3 only.

Value	Label	Frequency	%
20	MIDUS REFRESHER	3,577	100.0%

Valid	Invalid	Minimum	Maximum	Mean	StdDev
3577	0	20	20	20	0

MRID - MIDUS REFRESHER ID - PUBLIC USE

Type Numeric (Long)

Dataset

MR_P1_AGGREGATE_N3577_20160824.sav

Comparability Notes

The M2ID case identification system is used for the original MIDUS core sample first assessed in 1995. The MRID case identification system is used for the MIDUS Refresher sample.

Valid	Invalid	Minimum	Maximum
3577	0	30000	39999

SAMPLMAJ - Major sample identification

Type Code

3 of 14

MIDUS and DDI

- Looking ahead



- **DDI is question- and survey-centric**
 - Expand capabilities to describe non-survey data
 - Bake details into DDI fields for biomarker, laboratory, and derived (constructed) variables
- **Explore harmonization with other DDI-compliant studies**
 - CLOSER, NSHAP, AddHealth, NHATS, WLS
 - Requires harmonization of concepts, then measures
 - Leverage controlled vocabularies?





<http://midus.wisc.edu/>



Barry T. Radler, PhD

University of Wisconsin-Madison Institute on Aging

This research is supported by the following grants
from the National Institute on Aging:
R03-AG046312, 5P01-AG020166, R01-AG047154