



**A NEXT GENERATION,  
REUSABLE, WEB-BASED  
DATA CATALOG**



# Overview of Presentation

- **Background on MIDUS**
  - **Importance of DDI to MIDUS**
    - **Harmonization**
    - **Facilitating discovery and complex analysis**
  - **Current Project Goals**
- **Implementation of Project Goals**
  - **Upgrading MIDUS from DDI 3.1 to 3.2**
  - **Building on the MIDUS-Colectica Portal**
    - **Demonstrating project results so far**

# Background on MIDUS

Midlife in the United States



A National Study of Health & Well-Being



*Advancing Knowledge  
of Factors That  
Promote Positive  
Health and Resilience*

## **MID-LIFE IN THE UNITED STATES** A National Study of Health and Well-Being

### **Baseline: 1995-96**

- **Harvard University**
- **Funded by MacArthur Foundation**
- **N=7,108**
- **Ages 25-74**

# MIDUS: Unique Characteristics

- **Population survey**
  - ▣ **National RDD**
    - **Twin, Sibling subsamples**
- **Extensive assessments**
  - ▣ **Psychological, behavioral, demographic**
- **Multi-disciplinary design**
  - ▣ **Cognitive, daily stress**

# Timelines for MIDUS Longitudinal and Refresher Data Collection

M1

Survey n=7108

Cognitive n=302

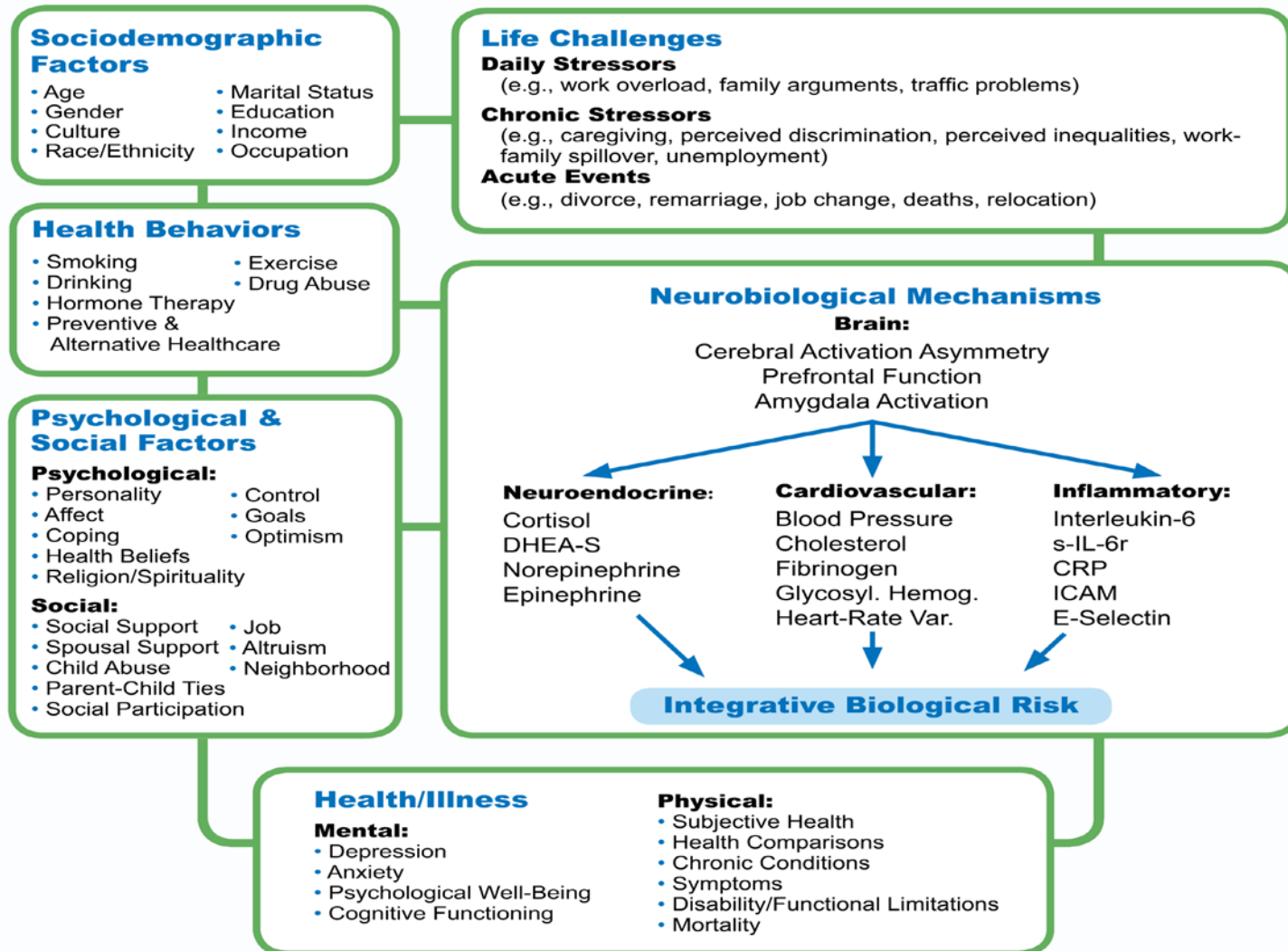
Daily Stress n=1499



# MIDUS: Unique Characteristics

- Population survey
  - ▣ Extensive assessments
  - ▣ Satellite studies (cog, stress)
- **MIDUS 2 (2004)**
  - ▣ **Longitudinal (9-10 year interval)**
- **Expanded multi-disciplinary design**
  - ▣ **Aging as integrated bio-psycho-social process**

# MIDUS: Guiding Conceptual Framework



## PROJECT 1

### (SURVEY OF A NATIONAL SAMPLE)

Assessed a wide array of psychological constructs (e.g., personality, psychological well-being, positive and negative affect, sense of control, goal orientations) and demographic characteristics (e.g., gender, marital status, socioeconomic standing, employment status), along with extensive health measures (mental and physical).

**MODE:** 30-minute Phone Interview and Two 50-page Self-Administered Questionnaires

## THE MIDUS II PROJECTS

### PROJECT 2

(Daily Diary Study)

**8 days of daily experience obtained via phone interviews.**

(e.g., time use, physical health symptoms and substance use, work productivity, psychological distress)

**4 days of salivary cortisol**

### PROJECT 3

(Cognitive Functioning)

**Phone-based cognitive battery**

(e.g., episodic verbal memory, working memory, verbal ability and speed, fluid intelligence/reasoning, speed of processing, episodic verbal memory/forgetting)

**Face-to-face assessment of cognitive capacities**

### PROJECT 4

(Biomarkers)

**2-Day Clinic Visit:**

**Biomarkers**—neuroendocrine, cardiovascular, immune, bone

Physical exam

Medical history

Medications

Sleep assessments

**Laboratory challenge study**—heart-rate variability, blood pressure, cortisol

### PROJECT 5

(Neuroscience)

**Affective reactivity & recovery:**

- baseline electroencephalography (EEG)
- task-related EEG
- task-related electromyography (EMG; eyeblink startle response, post auricular startle reflex, corrugator supercilli activity)
- structural MRI of neuroanatomy
- task event-related fMRI



# Timelines for MIDUS Longitudinal and Refresher Data Collection

## M1

Survey n=7108

Cognitive n=302

Daily Stress n=1499

## M2

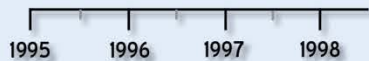
Survey n = 4963

Cognitive n = 4512

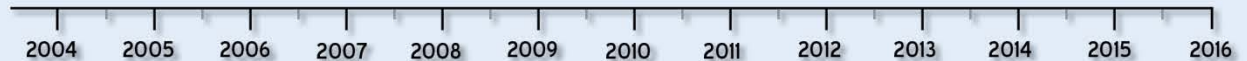
Daily Stress n = 2022

Biomarker n = 1255

Neuroscience n = 331



//



# MIDUS: Unique Characteristics

- Population survey
  - ▣ Extensive assessments
  - ▣ Satellite studies (cog, stress)
- Longitudinal & multi-disciplinary design
  - ▣ Aging as integrated bio-psycho-social process
- **Multiple samples and cohorts**

# Timelines for MIDUS Longitudinal and Refresher Data Collection

## M1

Survey n=7108

Cognitive n=302

Daily Stress n=1499

## M2

Survey n = 4963

Cognitive n = 4512

Daily Stress n = 2022

Biomarker n = 1255

Neuroscience n = 331

Milwaukee n = 592

MIDJA Survey n=1027

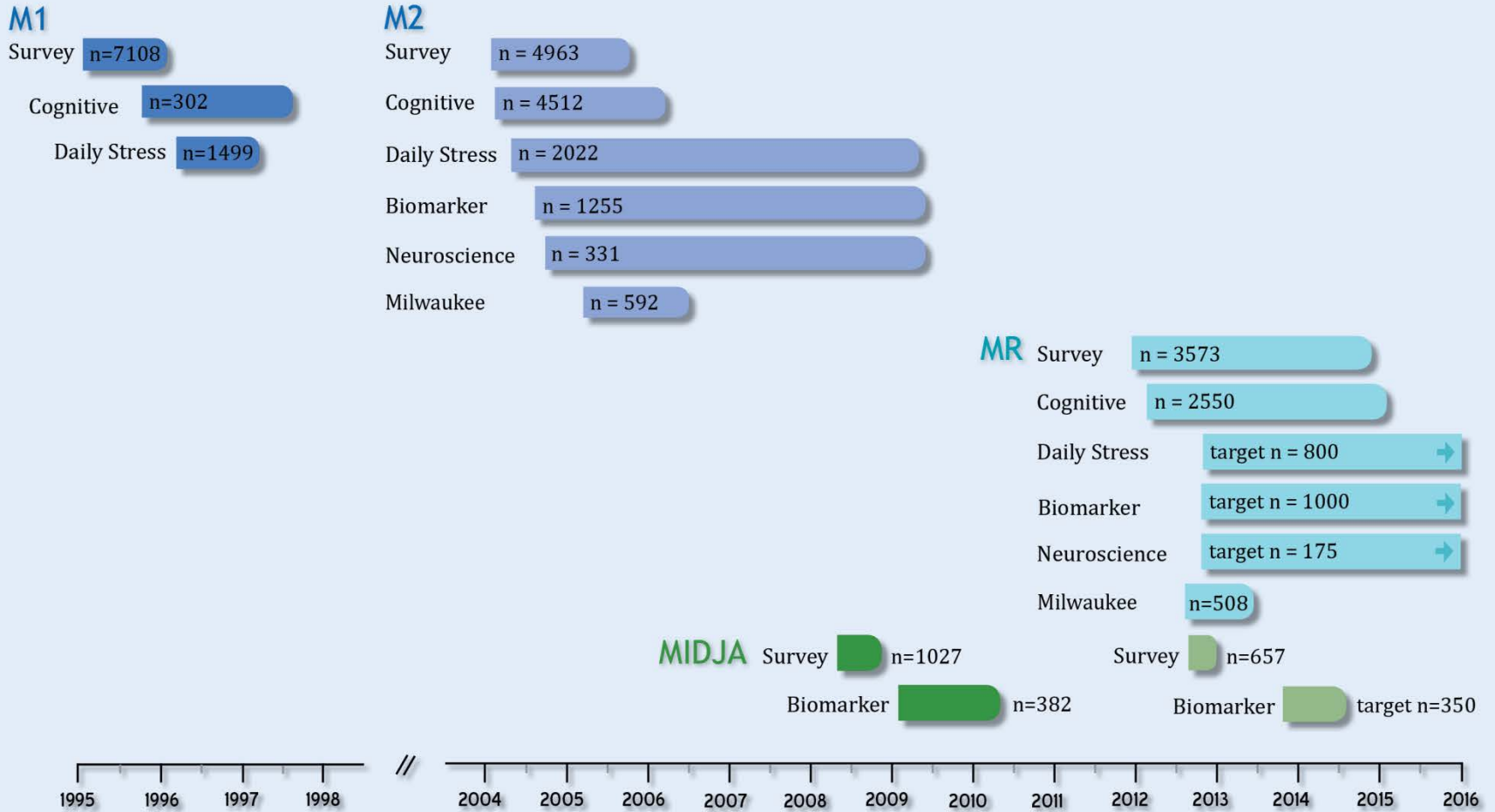
Biomarker n=382



//

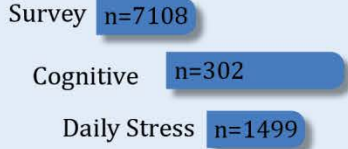


# Timelines for MIDUS Longitudinal and Refresher Data Collection

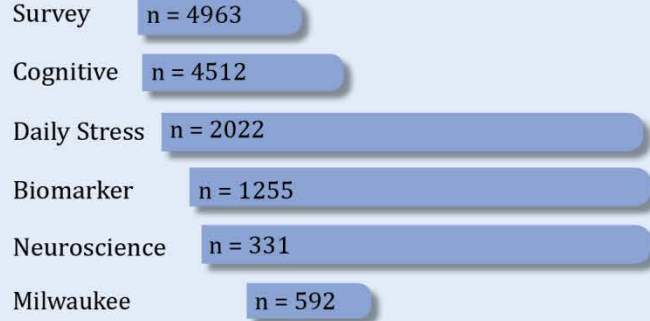


# Timelines for MIDUS Longitudinal and Refresher Data Collection

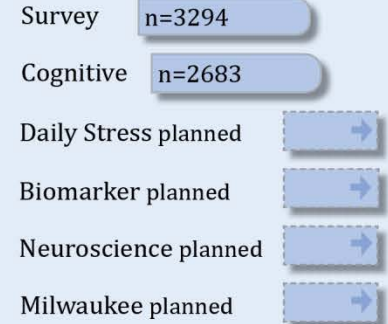
## M1



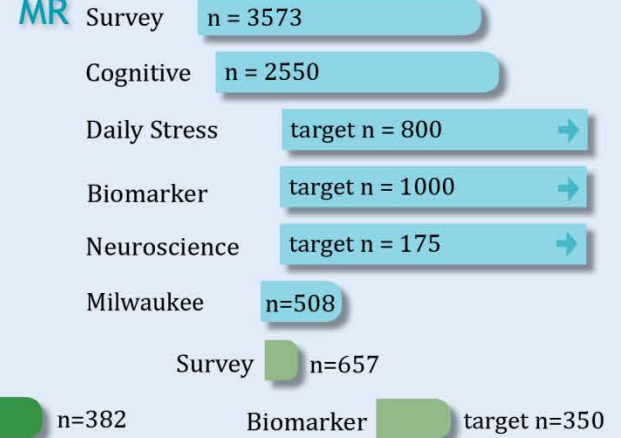
## M2



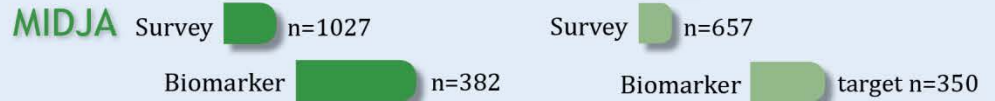
## M3



## MR



## MIDJA



### KEY

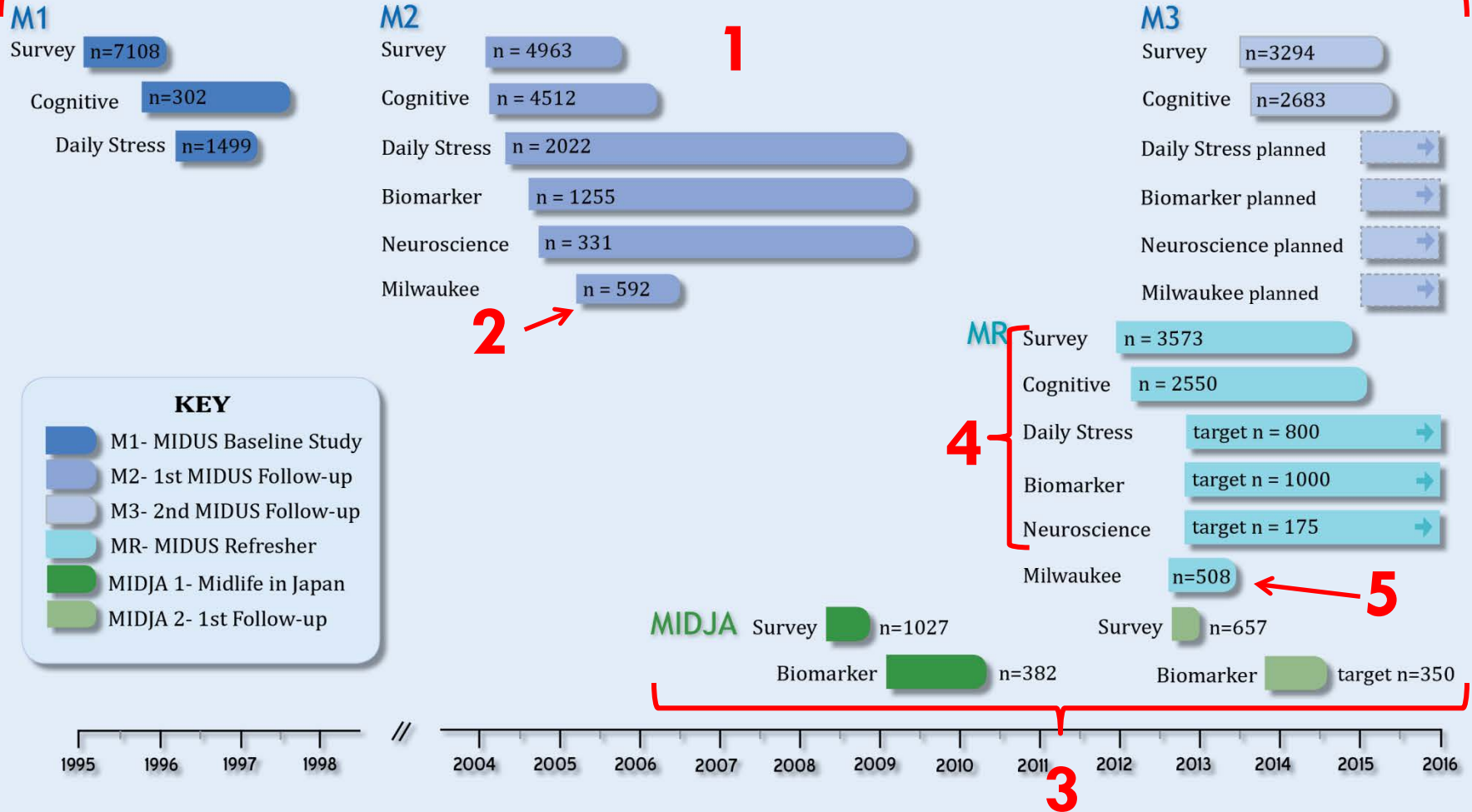
- M1- MIDUS Baseline Study
- M2- 1st MIDUS Follow-up
- M3- 2nd MIDUS Follow-up
- MR- MIDUS Refresher
- MIDJA 1- Midlife in Japan
- MIDJA 2- 1st Follow-up



//



# Timelines for MIDUS Longitudinal and Refresher Data Collection



# MIDUS: Unique Characteristics

- Population survey
  - ▣ Extensive assessments
- Longitudinal & multi-disciplinary design
  - ▣ Aging as integrated bio-psycho-social process
- Multiple samples and cohorts
- **Increasing complexity**
  - ▣ **Produced research products:**
    - 25,000 variables
    - N > 13,000

# MIDUS: Unique Characteristics

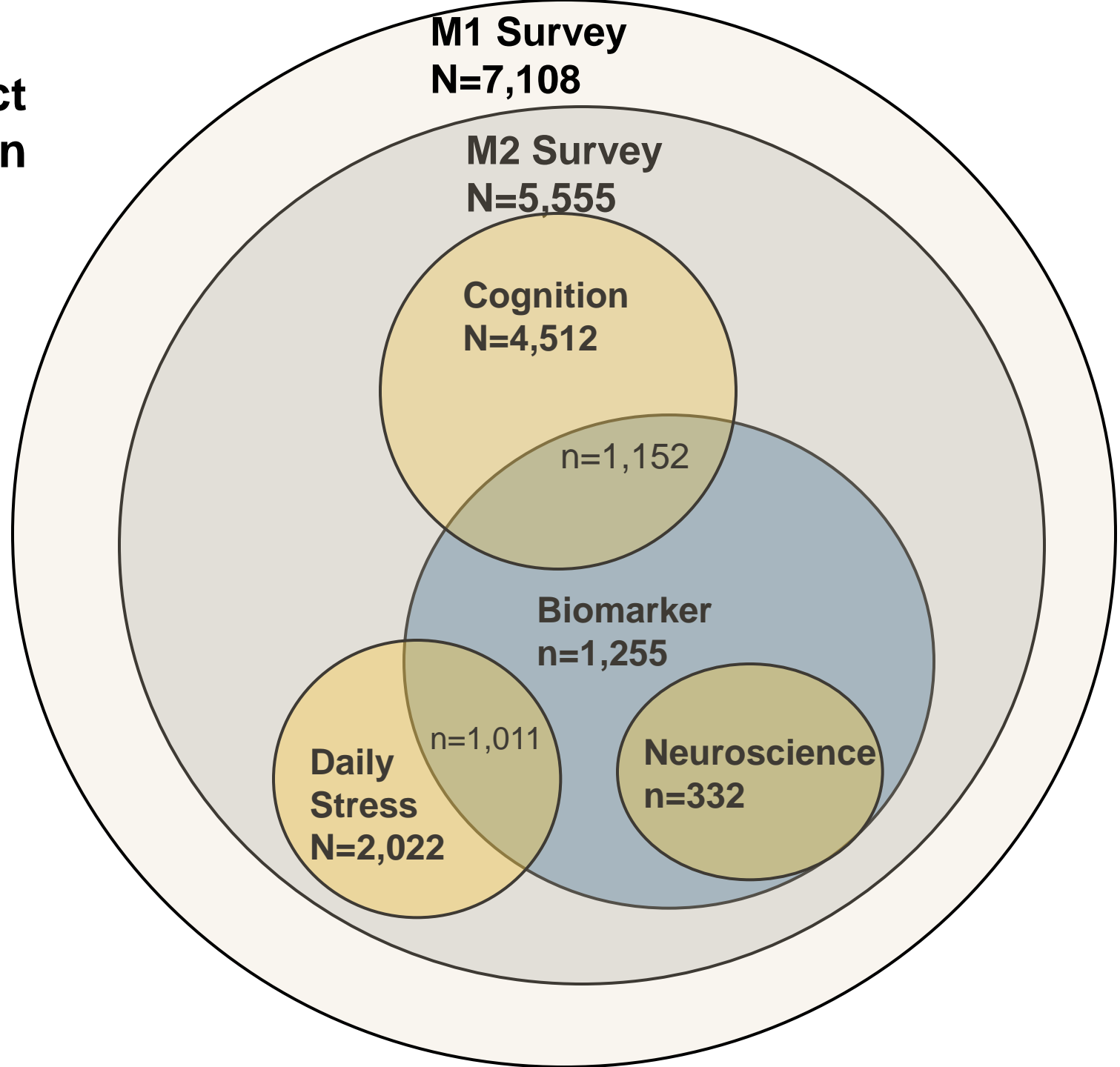
- Population survey
- Longitudinal & multi-disciplinary design
- Multiple samples and cohorts
- Increasing complexity
- **Extensive use of MIDUS (via ICPSR)**
  - ▣ **60k+ downloads; 20k+ users**
    - **Top 10 data download at ICPSR**
  - ▣ ***600 publications***



# MIDUS DDI 2.x Codebooks

- <http://midus1-project1.ssc.wisc.edu/>
- <http://midus2-project1.ssc.wisc.edu/>
- <http://midus2-project2.ssc.wisc.edu/>
- <http://midus2-project3.ssc.wisc.edu/>
- <http://midus2-project4.ssc.wisc.edu/>
- <http://midus2-project5.ssc.wisc.edu/>
- <http://midus2-project1.ssc.wisc.edu/milwaukee/>
- <http://midus2-project1.ssc.wisc.edu/midja/>

**MIDUS**  
**Multi-project**  
**participation**



# MIDUS: M2 Multi-project Participation

<b>Projects Completed</b>	<b>Total Number of Respondents</b>
<b>Project 1</b>	<b>5555</b>
<b>Project 1 and 2</b>	<b>2022</b>
<b>Project 1 and 3</b>	<b>4768</b>
<b>Project 1 and 4</b>	<b>1255</b>
<b>Project 1, 2 and 3</b>	<b>1923</b>
<b>Project 1, 2 and 4</b>	<b>1011</b>
<b>Project 1, 3 and 4</b>	<b>1152</b>
<b>Project 1, 4 and 5</b>	<b>331</b>
<b>Project 1, 2, 3 and 4</b>	<b>960</b>
<b>Project 1, 2, 4 and 5</b>	<b>238</b>
<b>Project 1, 3, 4 and 5</b>	<b>295</b>
<b>Project 1, 2, 3, 4 and 5</b>	<b>221</b>

# MIDUS DDI 3 - Portal

---

## **MIDUS Metadata Portal**

<http://midus.colectica.org/>

# MIDUS and DDI: Moving Forward

- **2012 - National Institute on Aging RFA:**
  - “Secondary Analyses and Archiving of Social and Behavioral Datasets in Aging”
- **Proposal funded in 2013:**
  - “Facilitating Secondary Analyses and Archiving of MIDUS through DDI”

# Current Project goals

## Under a DDI 3.2 rubric...

- **1. Harmonization (internal, post-hoc)**
  - ▣ Clarify related nature of longitudinal and cross-cohort *survey* variables (RepresentedVariable)
  - ▣ Provide information/procedures for reconciliation
- **2. Custom Data Extract (CDE)**
  - ▣ Allow researchers to focus on variables of interest
  - ▣ Facilitate accurate merges across numerous datasets

# Harmonization

## □ **Concordance table**

- Cross-referenced record of each variable
- Includes “Comparability notes” and “Comparability class”
- Future plans:
  - Provide code or procedures
    - Reconciliation or transformation of incompatible versions
    - Constructed variables

# MIDUS Harmonization

## Cross-walking and Cross-referencing

	B	C	D	E	F	G	H	
1	M1 Variable Name	M2 Variable Name	MKE Variable Name	MR Variable Name	MKER Variable Name	M3 Variable Name	M1 Variable Label	
67		<b>HEALTH</b>						
68	A1PA4	B1PA1	BACA1	RA1PA1	RAACA1	C1PA1	Physical health	Physical health
69	A1PA5	B1PA2	BACA2	RA1PA2	RAACA2	C1PA2	Mental or emotional health	Mental/emotio
70	A1PA6	B1PA3	BACA3	RA1PA3	RAACA3	C1PA3	Self-evaluated health	Health compa
71	A1PA7	B1PA4	BACA4	RA1PA4	RAACA4	C1PA4	Days work limited by health	Days unable t
72	A1PA7A	B1PA4A	BACA4A	RA1PA4A	RAACA4A	C1PA4A	Unable to work, physical, mental	Reason unabl
73	A1PA7BA	B1PA4BA	BACA4BA	RA1PA4BA	RAACA4BA	C1PA4BA	# of Days physical	Num days unc
74	A1PA7BB	B1PA4BB	BACA4BB	RA1PA4BB	RAACA4BB	C1PA4BB	# of Days mental	Num days unc
75	A1PA7BC	B1PA4BC	BACA4BC	RA1PA4BC	RAACA4BC	C1PA4BC	# of Days combination	Num days unc
76	A1PA8	B1PA5	BACA5	RA1PA5	RAACA5	C1PA5	# Days cut back on work due health	Days cut back
77	A1PA8A	B1PA5A	BACA5A	RA1PA5A	RAACA5A	C1PA5A	Physical, mental or both	Reason cut ba
78	A1PA8BA	B1PA5BA	BACA5BA	RA1PA5BA	RAACA5BA	C1PA5BA	# of Days physical	Num days cut
79	A1PA8BB	B1PA5BB	BACA5BB	RA1PA5BB	RAACA5BB	C1PA5BB	# of Days mental	Num days cut
80	A1PA8BC	B1PA5BC	BACA5BC	RA1PA5BC	RAACA5BC	C1PA5BC	# of Days combination	Number days
81	A1PA9						Physical health at 16	
82	A1PA10						Mental health at 16	
83		B1PA6A	BACA6A	RA1PA6A	RAACA6A	C1PA6A		History of stro
84		B1PA6B	BACA6B	RA1PA6B	RAACA6B	C1PA6B		History of seri
85		B1PA6C	BACA6C	RA1PA6C	RAACA6C	C1PA6C		History of Par
86		B1PA6D	BACA6D	RA1PA6D	RAACA6D	C1PA6D		History of othe
87	A1PA11	B1PA7	BACA7	RA1PA7	RAACA7	C1PA7	Heart problems ever	Heart trouble :
88	A1PA11A	B1PA7A	BACA7A	RA1PA7A	RAACA7A	C1PA7A	Age of heart problem	Age doctor tol
89	A1PA11BA	B1PA7BA	BACA7BA	RA1PA7BA	RAACA7BA	C1PA7BA	Heart attack	Diagnosis - H
90	A1PA11BB	B1PA7BB	BACA7BB	RA1PA7BB	RAACA7BB	C1PA7BB	Angina	Diagnosis - A
91	A1PA11BC	B1PA7BC	BACA7BC	RA1PA7BC	RAACA7BC	C1PA7BC	High blood pressure	Diagnosis - H
92	A1PA11BD	B1PA7BD	BACA7BD	RA1PA7BD	RAACA7BD	C1PA7BD	Valve disease/mitrovalve prolap	Diagnosis - V
93	A1PA11BE	B1PA7BE	BACA7BE	RA1PA7BE	RAACA7BE	C1PA7BE	Hole in heart/atrial septal dfct	Diagnosis - H
94	A1PA11BF	B1PA7BF	BACA7BF	RA1PA7BF	RAACA7BF	C1PA7BF	Blocked/closed artery/corony art	Diagnosis - B
95	A1PA11BG	B1PA7BG	BACA7BG	RA1PA7BG	RAACA7BG	C1PA7BG	Irregular/fast heart beat/arrhyt	Diagnosis - Ir
96	A1PA11RH	B1PA7RH	BACA7RH	RA1PA7RH	RAACA7RH	C1PA7RH	Heart murmur	Diagnosis - H



# Harmonization - reconciling

	N	O	P	Q
1	M1 Variable Name	Comparability Notes (among MIDUS Project 1)	Comparability Class	Question Description/Concept
146	A1PA29CI			History of Severe Chest Pain
147	A1PA29CJ			History of Severe Chest Pain
148	A1PA29CK	In M1 only: M1 has a separate variable (A1PA29CK) to indicate 'None', while 'None' is a separate response category in other waves.	Response Category Availability	History of Severe Chest Pain
149		Not in M1.	General Item Availability	History of Blood Pressure
150		Not in M1.	General Item Availability	History of Blood Pressure
151		Not in M1.	General Item Availability	History of Blood Pressure
152	A1PA33			History of Blood Pressure
153	A1PA34			History of Blood Pressure
154		Not in M1: Since this question was open-ended in M1, response was not coded.	Format Difference	History of Blood Pressure
155		Not in M1: Since this question was open-ended in M1, response was not coded.	Format Difference	History of Blood Pressure
156		Not in M1: Since this question was open-ended in M1, response was not coded.	Format Difference	History of Blood Pressure
157		Not in M1: Since this question was open-ended in M1, response was not coded.	Format Difference	History of Blood Pressure
158		Not in M1: Since this question was open-ended in M1, response was not coded.	Format Difference	History of Blood Pressure
159		Not in M1: Since this question was open-ended in M1, response was not coded.	Format Difference	History of Blood Pressure
160	A1PA30		Coding Scheme	History of Blood Pressure
161	A1PA30	M1 is not directly comparable with M2, MKE, MR, MKER, M3: M1 responses were coded as number of months, while other waves broke out number and unit separately.	Coding Scheme	History of Blood Pressure
162	A1PA31			History of Blood Pressure
163	A1PA32S			History of Blood Pressure
164	A1PA32D			History of Blood Pressure
165	A1PA36			Ever Had Cancer

# Custom Data Extract

- **Creating customized MIDUS datasets**
  - A Researcher favorite
  - Search variables, use shopping basket
  - Include variables from across all MIDUS projects
    - Merge different datasets
    - Provide different formats (csv, SPSS, SAS, Stata)
    - Associated DDI codebook
  - More efficient, cleaner, comprehensive use of dataset

# Development Milestones



1. Metadata Quality Report
2. Harmonization
3. Web-based Discoverability
4. Data Extraction

# Step 1. Metadata Quality Report

- Compare the harmonization spreadsheet to the Repository
- Check for:
  - ▣ Missing information
  - ▣ Inconsistent labels
  - ▣ Inconsistent data types
- Update the metadata to improve quality

# Step 2. Harmonization

- Use the harmonization spreadsheet
- Create a Represented Variable for each row
- Store these in the repository

The screenshot shows a Google Spreadsheet with the following table structure:

	A	B	C	D	E	F	G	H
1	M1 Variable Name	M2 Variable Name	MKE Variable Name	MR Variable Name	MKR Variable Name	M3 Variable Name	J1 Variable Name	Comparability Notes (between MIDJA and MIDUS)
398				RA1PB2DN	RAACB2DN	C1PB2DN		Question not included in MIDJA instrument.
399				RA1PB2DU	RAACB2DU	C1PB2DU		Question not included in MIDJA instrument.
400	A1PB3A	B1PB3A	BACB3A	RA1PB3A	RAACB3A	C1PB3A		Question not included in MIDJA instrument.
401	A1PB3B	B1PB3B	BACB3B	RA1PB3B	RAACB3B	C1PB3B		Question not included in MIDJA instrument.
402	A1PB3C	B1PB3C	BACB3C	RA1PB3C	RAACB3C	C1PB3C		Question not included in MIDJA instrument.
403	A1PB3D	B1PB3D	BACB3D	RA1PB3D	RAACB3D	C1PB3D		Question not included in MIDJA instrument.
404	A1PB3E	B1PB3E	BACB3E	RA1PB3E	RAACB3E	C1PB3E		Question not included in MIDJA instrument.
405	A1PB3F	B1PB3F	BACB3F	RA1PB3F	RAACB3F	C1PB3F		Question not included in MIDJA instrument.
406	A1PB3G	B1PB3G	BACB3G	RA1PB3G	RAACB3G	C1PB3G		Question not included in MIDJA instrument.
407	A1PB3H	B1PB3H	BACB3H	RA1PB3H	RAACB3H	C1PB3H		Question not included in MIDJA instrument.
408	A1PB3I	B1PB3I	BACB3I	RA1PB3I	RAACB3I	C1PB3I		Question not included in MIDJA instrument.
409	A1PB3J	B1PB3J	BACB3J	RA1PB3J	RAACB3J	C1PB3J		Question not included in MIDJA instrument.
410	A1PB3K	B1PB3K	BACB3K	RA1PB3K	RAACB3K	C1PB3K		Question not included in MIDJA instrument.
411		B1PBWORK		RA1PB3WK	RAACB3WK	C1PB3WK		Question not included in MIDJA instrument.
412	A1PB5	B1PB4N	BACB4N	RA1PB4N	RAACB4N	C1PB4N		Question not included in MIDJA instrument.
413	A1PB5	B1PB4M	BACB4M	RA1PB4U	RAACB4M	C1PB4U		Question not included in MIDJA instrument.
414		B1PB5	BACB5	RA1PB5	RAACB5	C1PB5		Question not included in MIDJA instrument.
415	A1PB4A	B1PB5	BACB5	RA1PB5A	RAACB5A	C1PB5A		Question not included in MIDJA instrument.
416	A1PB4B	B1PB5	BACB5	RA1PB5B	RAACB5B	C1PB5B		Question not included in MIDJA instrument.
417	A1PB4C	B1PB5	BACB5	RA1PB5C	RAACB5C	C1PB5C		Question not included in MIDJA instrument.
418	A1PB4D	B1PB5	BACB5	RA1PB5D	RAACB5D	C1PB5D		Question not included in MIDJA instrument.
419	A1PB4E	B1PB5	BACB5	RA1PB5E	RAACB5E	C1PB5E		Question not included in MIDJA instrument.
420	A1PB4F	B1PB5	BACB5	RA1PB5F	RAACB5F	C1PB5F		Question not included in MIDJA instrument.
421	A1PB4G	B1PB5	BACB5	RA1PB5G	RAACB5G	C1PB5G		Question not included in MIDJA instrument.
422				RA1PB5AB	RAACB5AB	C1PB5AB		Question not included in MIDJA instrument.
423	A1PB6	B1PB6	BACB6	RA1PB6	RAACB6	C1PB6	J1SE1	
424	A1PB8_2	B1PB7	BACB7	RA1PB7	RAACB7	C1PB7		Question not included in MIDJA instrument.
425	A1PB8_3	B1PB7A	BACB7A	RA1PB7A	RAACB7A	C1PB7A		Question not included in MIDJA instrument.

# Step 3. Web-based Discoverability

- Build on top of Colectica Portal
  - ▣ Searching and information retrieval out-of-the-box
- Add cross-reference tables for easy discoverability
- Choose variables or groups of variables to include in the data extract

# Step 4. Data Extraction

- Store master data in Colectica Repository
- Based on a user's selected variables, generate:
  - ▣ Datasets
    - CSV, R, SAS, SPSS, Stata
  - ▣ HTML and PDF codebooks
  - ▣ DDI XML

# Progress

✓ Complete	Metadata Quality Report
✓ Complete	Harmonization
✓ Complete	Web-based Discoverability
✓ Complete	Data Extraction
Upcoming	More and better discoverability





# Demonstration



# The Benefits of a Data Documentation Standard



- Share tools
- Share funding
- Benefit from other organizations' investments

# CLOSER



- 9 independent studies
- Data from 1930 – 2014
- 200,000+ variables
- Purpose: Maximize the use, value, and impact of the UK's longitudinal studies
- Document and harmonize all variables and questionnaires



# Demonstration



# Acknowledgement

- **This research project is supported by a grant from the National Institute on Aging (R03-AG046312).**
- **CLOSER funding is from the UK Economic and Social Research Council and Medical Research Council**

# Thank you

Jeremy Iverson – Colectica (jeremy@colectica.com)

Barry Radler – UW-Madison (bradler@wisc.edu)

Dan Smith – Colectica (dan@colectica.com)

[midus.wisc.edu](http://midus.wisc.edu)

[www.colectica.com](http://www.colectica.com)