LEVERAGING DDI 3.2 TO POWER A HARMONIZED DATA EXTRACTION TOOL FOR MIDUS

EDDI 2014 - London

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**

Background on MIDUS

Midlife in the United States

A National Study of Health & Well-Being

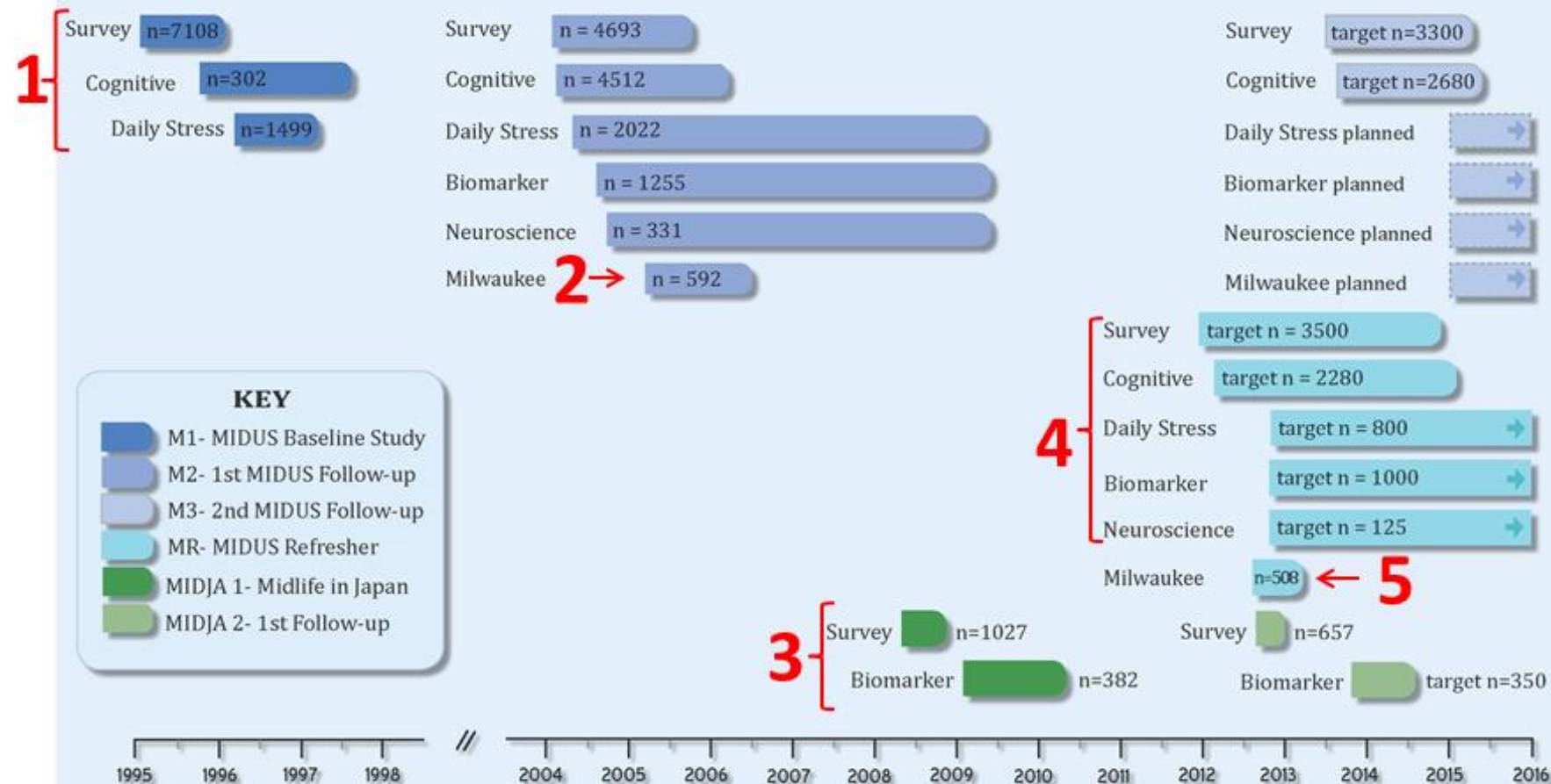Advancing Knowledge of Factors That Promote Positive Health and Resilience

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
- Satellite studies of stress, cognition

Timelines for MIDUS Longitudinal and Refresher Data Collection

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 electro-encephalography (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

MIDUS: Unique Characteristics

- **Longitudinal (9-10 year interval)**
- **Multiple samples/cohorts**
- **Multidisciplinary design**
  - **Aging as integrated bio-psycho-social process**
- **Result**
  - **25,000 variables**
  - **N > 13,000**

MIDUS: Unique Characteristics

- **Multiple waves (9-10 year interval)**
- **Multiple samples/cohorts**
- **Multidisciplinary design**
- **Many datasets, cases**
- **Extensive use of MIDUS – Open Data philosophy**
  - **Top 10 data download at ICPSR**
  - ***600 publications***

Status of Current DDI Efforts

MIDUS Metadata Portal

http://midus.colectica.org/

Moving Forward

- **2012 – National Institute on Aging RFA:**

   "Secondary Analyses and Archiving of Social and Behavioral Datasets in Aging"

- **MIDUS project funded in 2013!**

   Acknowledgement:

   NIA grant (R03-AG046312)

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

Harmonization – cross-referencing

	B	C	D	E	F	G	H	
	M1 Variable Name	M2 Variable Name	MKE Variable Name	MR Variable Name	MKER Variable Name	M3 Variable Name	M1 Variable Label	
67		HEALTH						
68	A1PA4	B1PA1	BACA1	RA1PA1	RAACA1	C1PA1	Physical health	Physical healt
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 una
74	A1PA7BB	B1PA4BB	BACA4BB	RA1PA4BB	RAACA4BB	C1PA4BB	# of Days mental	Num days una
75	A1PA7BC	B1PA4BC	BACA4BC	RA1PA4BC	RAACA4BC	C1PA4BC	# of Days combination	Num days una
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	A1PA11BH	B1PA7BH	BACA7BH	RA1PA7BH	RAACA7BH	C1PA7BH	Heart murmur	Diagnosis - H

Phone + SAQ | Comparability Class

Harmonization – reconciling

	N M1 Variable Name	O Comparability Notes (among MIDUS Project 1)	P Comparability Class	Q 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

Phone + SAQ / Comparability Class

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 `RepresentedVariable` for each row
- Store these in the repository

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
Upcoming	Web-based Discoverability
Upcoming	Data Extraction

NADDI 2015 – April 8 - 10

www.ssc.wisc.edu/naddi2015/

University of Wisconsin - Madison

Thank you

Jeremy Iverson – Colectica (jeremy@colectica.com)

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

Dan Smith – Colectica (dan@colectica.com)

midus.wisc.edu

www.colectica.com

colectica