Building a Harmonized Data Market for Longitudinal Data with MIDUS and DDI (3.2)

European DDI Conference

Overview of Presentation

- Background on MIDUS
  - Importance of DDI for
    - Harmonization
    - Facilitating complex analysis
- Current Project's Goals
- Implementation of Project Goals
  - Creating MIDUS DDI 3.2 Instances
  - Upgrading MIDUS-Colectica Repository/Portal

MIDUS

Advancing Knowledge of Factors That Promote Positive Health and Resilience

Baseline: 1995-96
- Harvard
- MacArthur Found.
- N=7,108
- Ages 25-74
- Twins/Siblings

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

Unique Strengths of the MIDUS Study

In-depth multidisciplinary content *achieved via 5 separate data collection projects*

Wide age range (25–74) *facilitates focus on life course transitions*

MIDUS (Midlife in the U.S.) is a national longitudinal study of how many factors (behavioral, social, psychological, biological, neurological) come together to influence health and well-being as people age from early adulthood into midlife and old age. It was conceived by a multidisciplinary team of scholars interested in understanding aging as an integrative process.
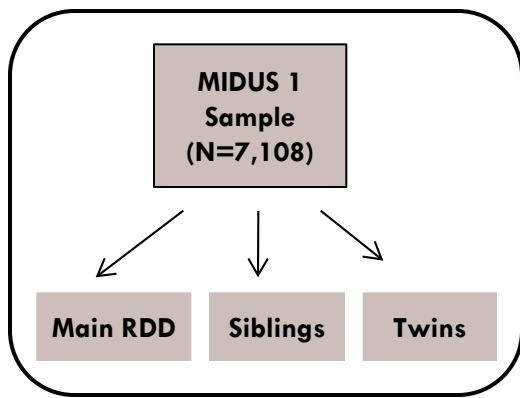
MIDUS Samples

In 1995, MIDUS survey data were collected from a total of 7,108 participants. The baseline sample was comprised of individuals from four subsamples: (1) a national RDD (random digit dialing) sample (n=3,487); (2) oversamples from

In addition, the twin subsample was administered a short screener to assess zygosity and other twin-specific information.

With funding provided by the National Institute on Aging, a longitudinal fol-

MIDUS: Strengths and Complexities

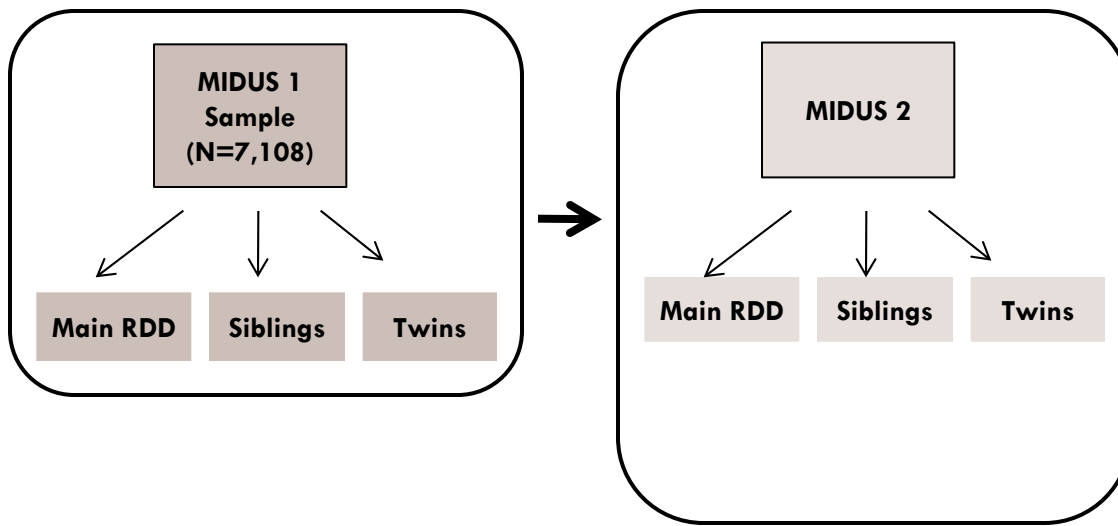☐ **Multiple sample waves (longitudinal)**

MIDUS
Samples
and
Timelines

MIDUS 1
Sample
(N=7,108)

Main RDD   Siblings   Twins
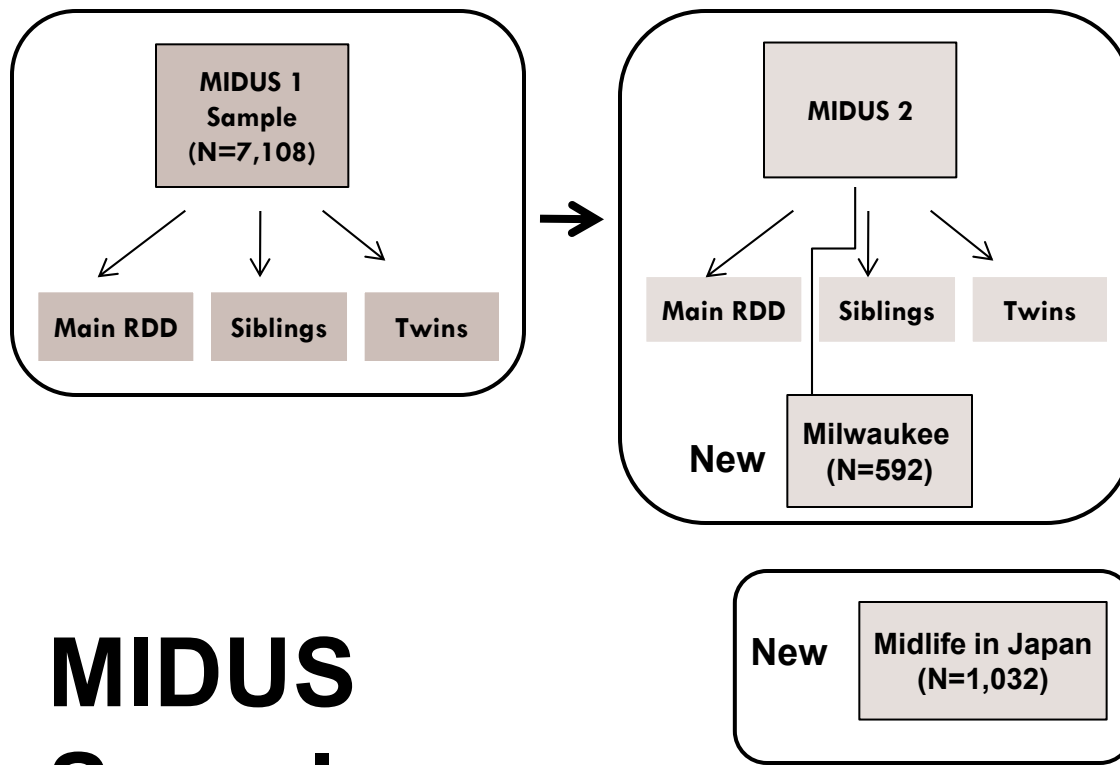
1995   2005   2015

MIDUS
Samples
and
Timelines

| 1995 | 2005 | 2015 |

MIDUS: Strengths and Complexities

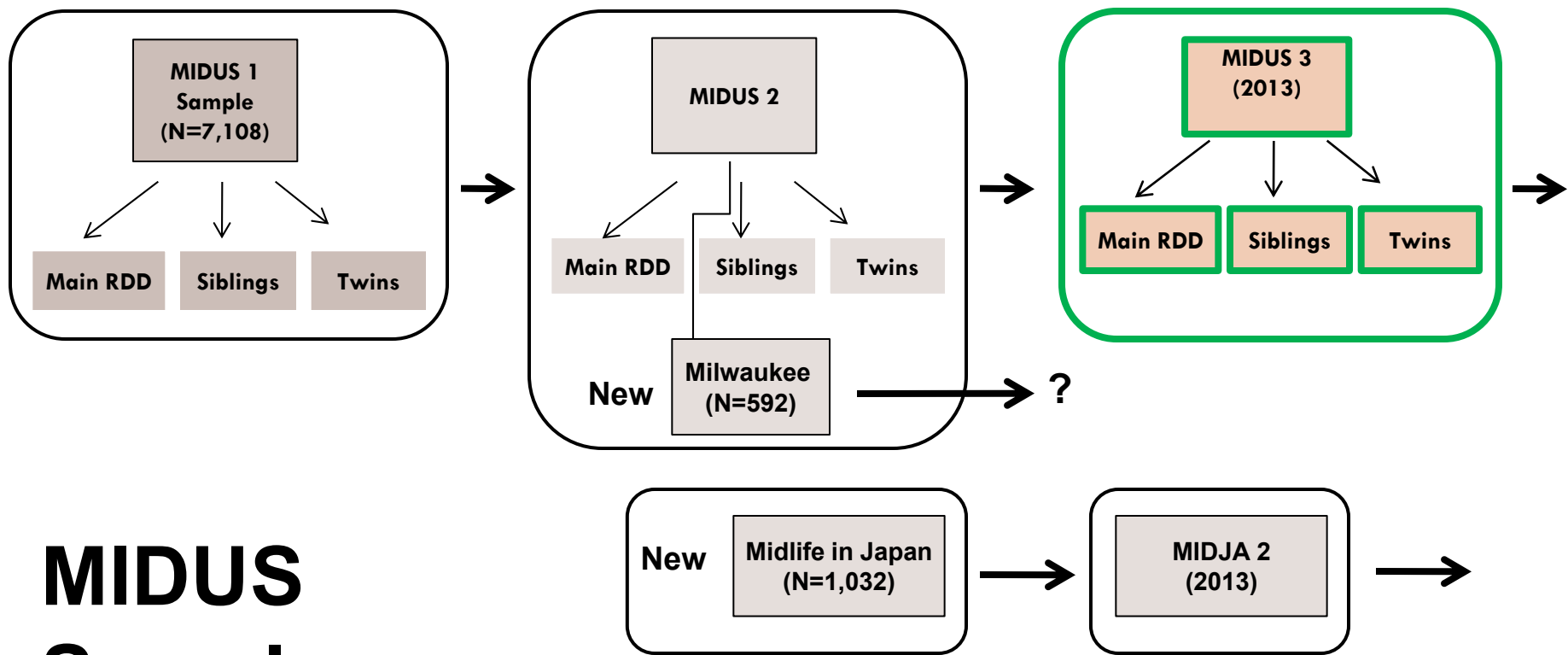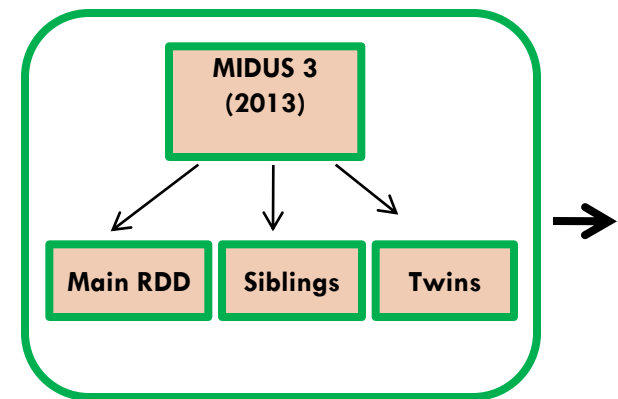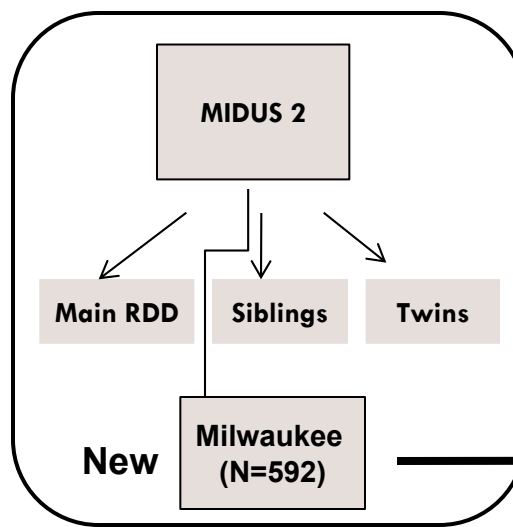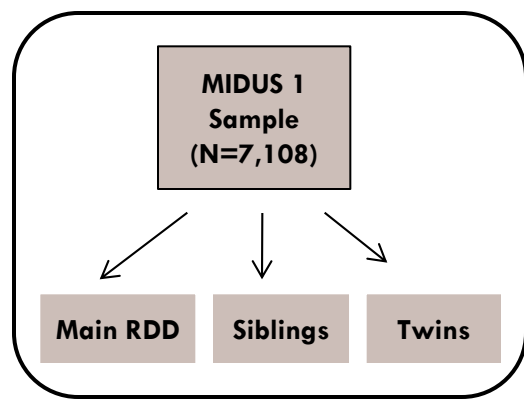- **Multiple sample waves (longitudinal)**
- **Multiple cohorts**

MIDUS 1 Sample (N=7,108) → Main RDD, Siblings, Twins

MIDUS 2 → Main RDD, Siblings, Twins

New Milwaukee (N=592)

New Midlife in Japan (N=1,032)

MIDUS Samples and Timelines

1995   2005   2015

MIDUS Samples and Timelines

MIDUS 1 Sample (N=7,108)
Main RDD | Siblings | Twins

MIDUS 2
Main RDD | Siblings | Twins
New Milwaukee (N=592) → ?

MIDUS 3 (2013)
Main RDD | Siblings | Twins

New Midlife in Japan (N=1,032) → MIDJA 2 (2013)

1995 | 2005 | 2015

MIDUS Samples and Timelines

MIDUS 1 Sample (N=7,108)
- Main RDD
- Siblings
- Twins

MIDUS 2
- Main RDD
- Siblings
- Twins
- New Milwaukee (N=592) → ?

MIDUS 3 (2013)
- Main RDD
- Siblings
- Twins

New Midlife in Japan (N=1,032) → MIDJA 2 (2013)

New Refresher Sample (N≈2,659)
- RDD (N≈2,152)
- Milwaukee (N≈507)
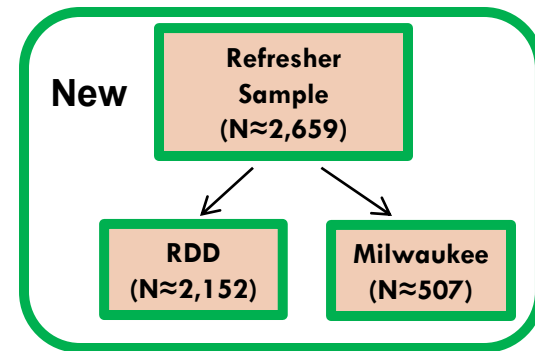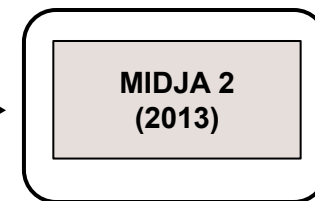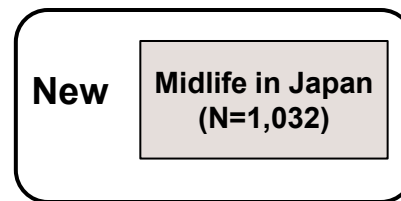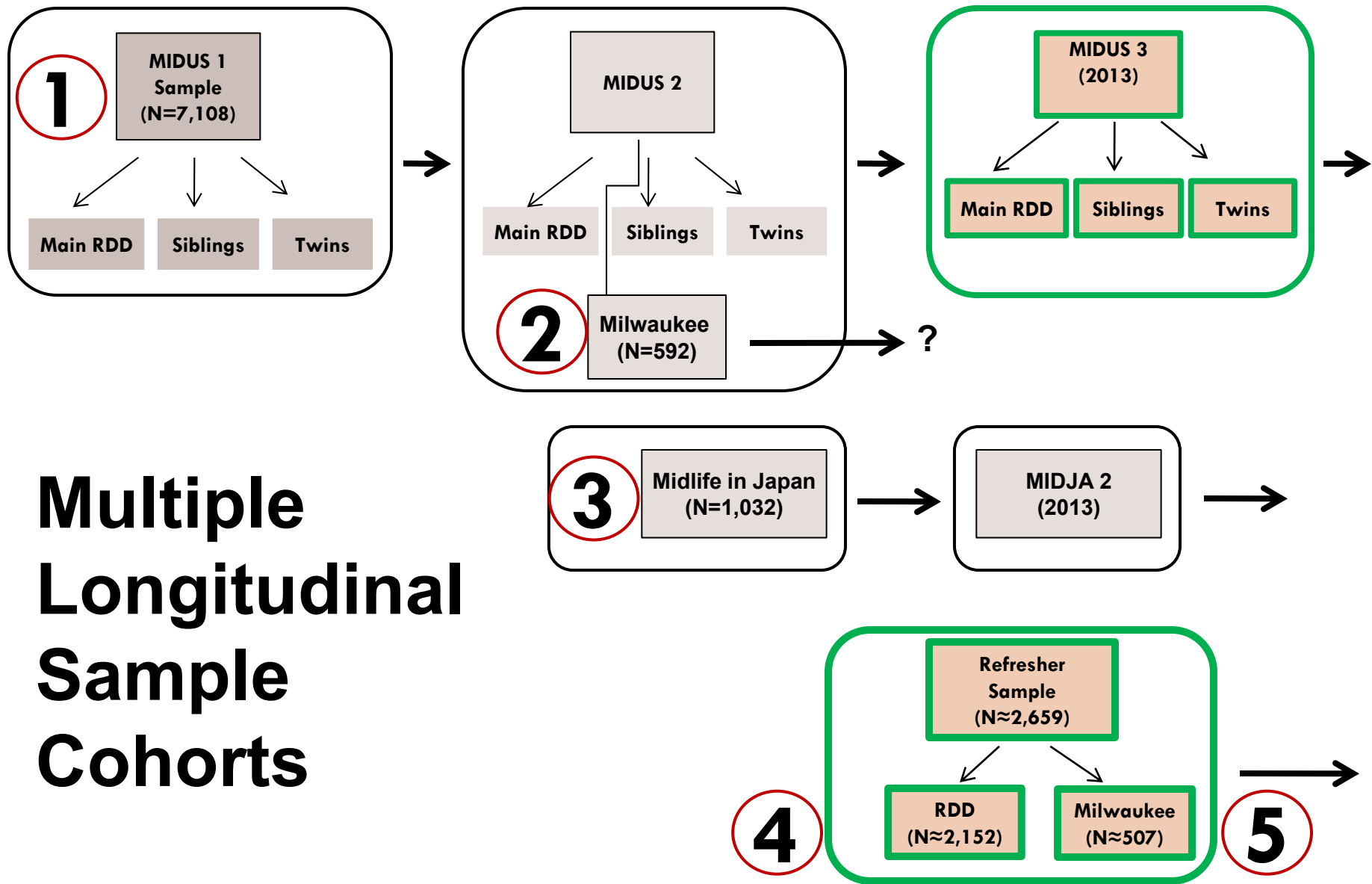
1995   2005   2015

Multiple Longitudinal Sample Cohorts

1. MIDUS 1 Sample (N=7,108) → Main RDD, Siblings, Twins

2. MIDUS 2 → Main RDD, Siblings, Twins; Milwaukee (N=592) → ?

MIDUS 3 (2013) → Main RDD, Siblings, Twins

3. Midlife in Japan (N=1,032) → MIDJA 2 (2013)

4. Refresher Sample (N≈2,659) → RDD (N≈2,152), Milwaukee (N≈507) 5.

1995   2005   2015

MIDUS: Strengths and Complexities

- **Multiple sample waves (longitudinal)**
- **Multiple cohorts**
- **Multidisciplinary design**
  - **Aging as integrated bio-psycho-social process**

THE MIDUS II PROJECTS

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

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

MIDUS Core
-
Longitudinal
Multi-project
Participation
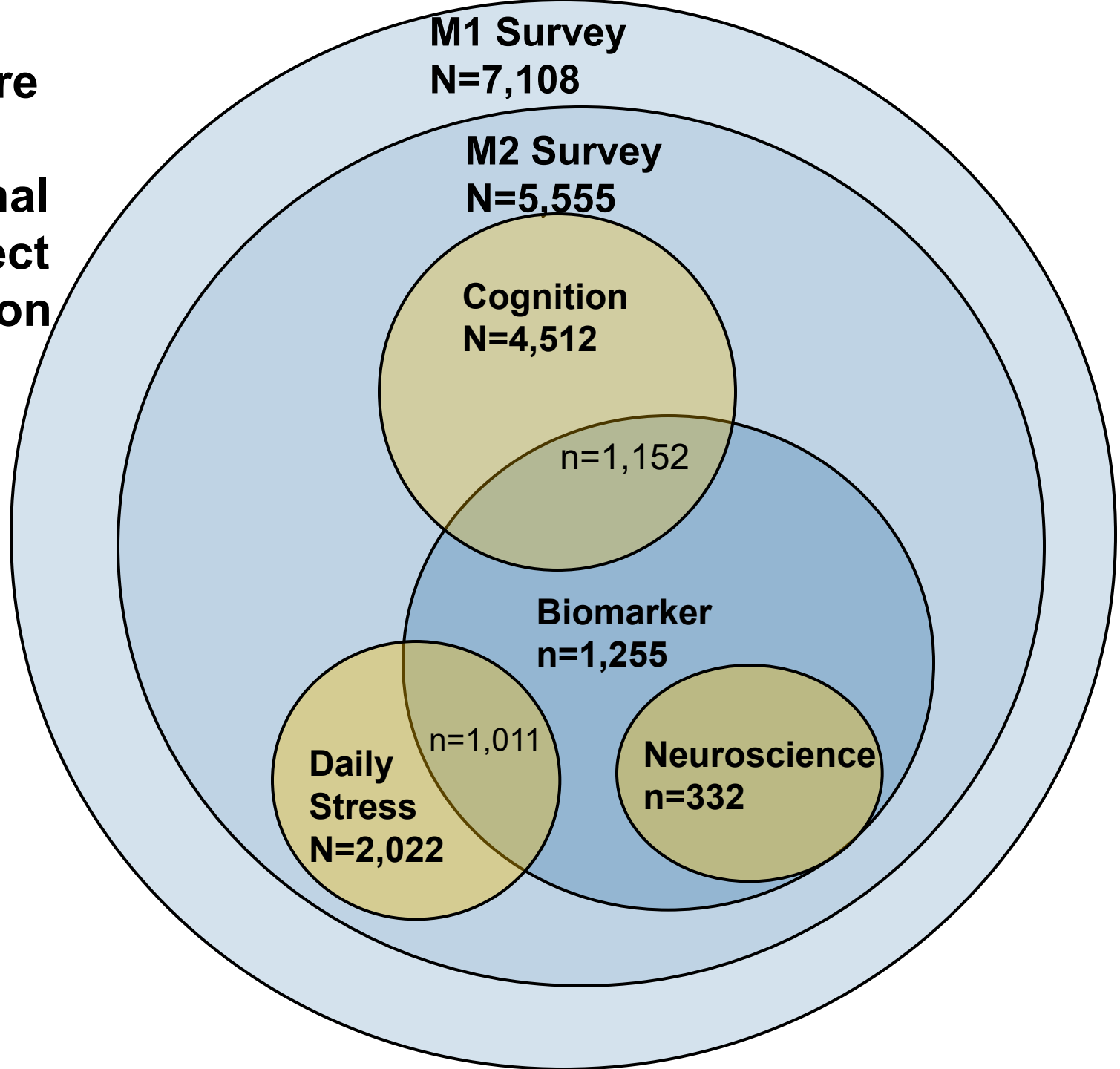
M1 Survey
N=7,108

M2 Survey
N=5,555

Cognition
N=4,512

n=1,152

Biomarker
n=1,255

Daily
Stress
N=2,022

n=1,011

Neuroscience
n=332

MIDUS: Strengths and Complexities

- Multiple sample waves (longitudinal)
- Multiple cohorts
- Multidisciplinary design
  - Aging as integrated bio-psycho-social process
- Wide use of MIDUS – Open Data philosophy
  - #1 data download at NACDA
  - Top 10 data download at ICPSR
  - 500 publications

Current DDI Efforts

MIDUS Metadata Repository/Portal

http://midus.colectica.org/

Current Project goals

Under a DDI rubric…

1. **Harmonization (internal)**
   - Clarify related nature of longitudinal and cross-cohort variables (improving search function)
   - Provide information/procedures for reconciliation

2. **Customized Data Extract (CDE)**
   - Allow researchers to focus on variables of interest
   - Facilitate accurate merges across numerous datasets

Harmonization

Harmonization

- Concordance table
  - Includes "Comparability notes" and "Comparability class"
  - Example: Variable A1PA30 "time since last BP test"
    - "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."
    - Offer code/algorithm for reconcilation

Custom Data Extract

- Customized dataset
  - Search variables, use shopping basket
  - Allow variables from across MIDUS
    - Merge different datasets
    - Different formats (csv, SPSS, SAS, Stata)
    - Associated DDI codebook

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
In Progress	Harmonization
Upcoming	Web-based Discoverability
Upcoming	Data Extraction

Thank you

Barry Radler – UW-Madison (bradler@wisc.edu)
Jeremy Iverson – Colectica (jeremy@colectica.com)
Dan Smith-Colectica (dan@colectica.com)

midus.wisc.edu

www.colectica.com

colectica