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

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
Midlife in the United States

MIDUS
Midlife in Japan
MIDUS: Unique Characteristics

- Population survey
  - Extensive assessments
  - Satellite studies (cog, stress)
  - Longitudinal (9-10 year interval)
- Expanded multi-disciplinary design
  - Aging as integrated bio-psycho-social process
MIDUS: Guiding Conceptual Framework

**Sociodemographic Factors**
- Age
- Gender
- Culture
- Race/Ethnicity
- Marital Status
- Education
- Income
- Occupation

**Health Behaviors**
- Smoking
- Drinking
- Hormone Therapy
- Preventive & Alternative Healthcare
- Exercise
- Drug Abuse

**Psychological & Social Factors**

**Psychological:**
- Personality
- Affect
- Coping
- Health Beliefs
- Religion/Spirituality
- Control
- Goals
- Optimism

**Social:**
- Social Support
- Spousal Support
- Child Abuse
- Parent-Child Ties
- Social Participation
- Job
- Altruism
- Neighborhood

**Life Challenges**

**Daily Stressors**
(e.g., work overload, family arguments, traffic problems)

**Chronic Stressors**
(e.g., caregiving, perceived discrimination, perceived inequalities, work-family spillover, unemployment)

**Acute Events**
(e.g., divorce, remarriage, job change, deaths, relocation)

**Neurobiological Mechanisms**

**Brain:**
- Cerebral Activation Asymmetry
- Prefrontal Function
- Amygdala Activation

**Neuroendocrine**
- Cortisol
- DHEA-S
- Norepinephrine
- Epinephrine

**Cardiovascular**
- Blood Pressure
- Cholesterol
- Fibrinogen
- Glycosyl. Hemog.
- Heart-Rate Var.

**Inflammatory**
- Interleukin-6
- s-IL-6r
- CRP
- ICAM
- E-Selectin

**Integrative Biological Risk**

**Health/Illness**

**Mental:**
- Depression
- Anxiety
- Psychological Well-Being
- Cognitive Functioning

**Physical:**
- Subjective Health
- Health Comparisons
- Chronic Conditions
- Symptoms
- Disability/Functional Limitations
- Mortality
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
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
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

M1 Survey
N=7,108

M2 Survey
N=5,555

Cognition
N=4,512

Biomarker
n=1,255

Daily Stress
N=2,022

Neuroscience
n=332

n=1,152

n=1,011
## MIDUS: M2 Multi-project Participation

<table>
<thead>
<tr>
<th>Projects Completed</th>
<th>Total Number of Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project 1</td>
<td>5555</td>
</tr>
<tr>
<td>Project 1 and 2</td>
<td>2022</td>
</tr>
<tr>
<td>Project 1 and 3</td>
<td>4768</td>
</tr>
<tr>
<td>Project 1 and 4</td>
<td>1255</td>
</tr>
<tr>
<td>Project 1, 2 and 3</td>
<td>1923</td>
</tr>
<tr>
<td>Project 1, 2 and 4</td>
<td>1011</td>
</tr>
<tr>
<td>Project 1, 3 and 4</td>
<td>1152</td>
</tr>
<tr>
<td>Project 1, 4 and 5</td>
<td>331</td>
</tr>
<tr>
<td>Project 1, 2, 3 and 4</td>
<td>960</td>
</tr>
<tr>
<td>Project 1, 2, 4 and 5</td>
<td>238</td>
</tr>
<tr>
<td>Project 1, 3, 4 and 5</td>
<td>295</td>
</tr>
<tr>
<td>Project 1, 2, 3, 4 and 5</td>
<td>221</td>
</tr>
</tbody>
</table>
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**

<table>
<thead>
<tr>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M1 Variable Name</strong></td>
<td><strong>M2 Variable Name</strong></td>
<td><strong>MKE Variable Name</strong></td>
<td><strong>MR Variable Name</strong></td>
<td><strong>MKER Variable Name</strong></td>
<td><strong>M3 Variable Name</strong></td>
<td><strong>M1 Variable Label</strong></td>
</tr>
<tr>
<td>67</td>
<td>HEALTH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>A1PA4</td>
<td>B1PA1</td>
<td>BACA1</td>
<td>RA1PA1</td>
<td>RAACA1</td>
<td>C1PA1</td>
</tr>
<tr>
<td>69</td>
<td>A1PA5</td>
<td>B1PA2</td>
<td>BACA2</td>
<td>RA1PA2</td>
<td>RAACA2</td>
<td>C1PA2</td>
</tr>
<tr>
<td>70</td>
<td>A1PA6</td>
<td>B1PA3</td>
<td>BACA3</td>
<td>RA1PA3</td>
<td>RAACA3</td>
<td>C1PA3</td>
</tr>
<tr>
<td>71</td>
<td>A1PA7</td>
<td>B1PA4</td>
<td>BACA4</td>
<td>RA1PA4</td>
<td>RAACA4</td>
<td>C1PA4</td>
</tr>
<tr>
<td>72</td>
<td>A1PA7A</td>
<td>B1PA4A</td>
<td>BACA4A</td>
<td>RA1PA4A</td>
<td>RAACA4A</td>
<td>C1PA4A</td>
</tr>
<tr>
<td>73</td>
<td>A1PA7BA</td>
<td>B1PA4BA</td>
<td>BACA4BA</td>
<td>RA1PA4BA</td>
<td>RAACA4BA</td>
<td>C1PA4BA</td>
</tr>
<tr>
<td>74</td>
<td>A1PA7BB</td>
<td>B1PA4BB</td>
<td>BACA4BB</td>
<td>RA1PA4BB</td>
<td>RAACA4BB</td>
<td>C1PA4BB</td>
</tr>
<tr>
<td>75</td>
<td>A1PA7BC</td>
<td>B1PA4BC</td>
<td>BACA4BC</td>
<td>RA1PA4BC</td>
<td>RAACA4BC</td>
<td>C1PA4BC</td>
</tr>
<tr>
<td>76</td>
<td>A1PA8</td>
<td>B1PA5</td>
<td>BACA5</td>
<td>RA1PA5</td>
<td>RAACA5</td>
<td>C1PA5</td>
</tr>
<tr>
<td>77</td>
<td>A1PA8A</td>
<td>B1PA5A</td>
<td>BACA5A</td>
<td>RA1PA5A</td>
<td>RAACA5A</td>
<td>C1PA5A</td>
</tr>
<tr>
<td>78</td>
<td>A1PA8B</td>
<td>B1PA5BA</td>
<td>BACA5BA</td>
<td>RA1PA5BA</td>
<td>RAACA5BA</td>
<td>C1PA5BA</td>
</tr>
<tr>
<td>79</td>
<td>A1PA8BB</td>
<td>B1PA5BB</td>
<td>BACA5BB</td>
<td>RA1PA5BB</td>
<td>RAACA5BB</td>
<td>C1PA5BB</td>
</tr>
<tr>
<td>80</td>
<td>A1PA9</td>
<td>B1PA5BC</td>
<td>BACA5BC</td>
<td>RA1PA5BC</td>
<td>RAACA5BC</td>
<td>C1PA5BC</td>
</tr>
<tr>
<td>81</td>
<td>A1PA9</td>
<td>B1PA5BC</td>
<td>BACA5BC</td>
<td>RA1PA5BC</td>
<td>RAACA5BC</td>
<td>C1PA5BC</td>
</tr>
<tr>
<td>82</td>
<td>A1PA10</td>
<td>B1PA6A</td>
<td>BACA6A</td>
<td>RA1PA6A</td>
<td>RAACA6A</td>
<td>C1PA6A</td>
</tr>
<tr>
<td>83</td>
<td>A1PA8</td>
<td>B1PA6B</td>
<td>BACA6B</td>
<td>RA1PA6B</td>
<td>RAACA6B</td>
<td>C1PA6B</td>
</tr>
<tr>
<td>84</td>
<td>A1PA6C</td>
<td>BACA6C</td>
<td>RA1PA6C</td>
<td>RAACA6C</td>
<td>C1PA6C</td>
<td>History of secondary stroke</td>
</tr>
<tr>
<td>85</td>
<td>A1PA8D</td>
<td>B1PA6D</td>
<td>BACA6D</td>
<td>RA1PA6D</td>
<td>RAACA6D</td>
<td>C1PA6D</td>
</tr>
<tr>
<td>86</td>
<td>A1PA11</td>
<td>B1PA7</td>
<td>BACA7</td>
<td>RA1PA7</td>
<td>RAACA7</td>
<td>C1PA7</td>
</tr>
<tr>
<td>87</td>
<td>A1PA12</td>
<td>B1PA7A</td>
<td>BACA7A</td>
<td>RA1PA7A</td>
<td>RAACA7A</td>
<td>C1PA7A</td>
</tr>
<tr>
<td>88</td>
<td>A1PA11B</td>
<td>B1PA7BA</td>
<td>BACA7BA</td>
<td>RA1PA7BA</td>
<td>RAACA7BA</td>
<td>C1PA7BA</td>
</tr>
<tr>
<td>89</td>
<td>A1PA11BB</td>
<td>B1PA7BB</td>
<td>BACA7BB</td>
<td>RA1PA7BB</td>
<td>RAACA7BB</td>
<td>C1PA7BB</td>
</tr>
<tr>
<td>90</td>
<td>A1PA11BC</td>
<td>B1PA7BC</td>
<td>BACA7BC</td>
<td>RA1PA7BC</td>
<td>RAACA7BC</td>
<td>C1PA7BC</td>
</tr>
<tr>
<td>91</td>
<td>A1PA11BD</td>
<td>B1PA7BD</td>
<td>BACA7BD</td>
<td>RA1PA7BD</td>
<td>RAACA7BD</td>
<td>C1PA7BD</td>
</tr>
<tr>
<td>92</td>
<td>A1PA11BE</td>
<td>B1PA7BE</td>
<td>BACA7BE</td>
<td>RA1PA7BE</td>
<td>RAACA7BE</td>
<td>C1PA7BE</td>
</tr>
<tr>
<td>93</td>
<td>A1PA11BF</td>
<td>B1PA7BF</td>
<td>BACA7BF</td>
<td>RA1PA7BF</td>
<td>RAACA7BF</td>
<td>C1PA7BF</td>
</tr>
<tr>
<td>94</td>
<td>A1PA11BG</td>
<td>B1PA7BG</td>
<td>BACA7BG</td>
<td>RA1PA7BG</td>
<td>RAACA7BG</td>
<td>C1PA7BG</td>
</tr>
<tr>
<td>95</td>
<td>A1PA11BH</td>
<td>B1PA7BH</td>
<td>BACA7BH</td>
<td>RA1PA7BH</td>
<td>RAACA7BH</td>
<td>C1PA7BH</td>
</tr>
<tr>
<td>96</td>
<td>A1PA11RH</td>
<td>B1PA7RH</td>
<td>BACA7RH</td>
<td>RA1PA7RH</td>
<td>RAACA7RH</td>
<td>C1PA7RH</td>
</tr>
</tbody>
</table>
### Harmonization - reconciling

<table>
<thead>
<tr>
<th>N</th>
<th>M1 Variable Name</th>
<th>Comparability Notes (among MIDUS Project 1)</th>
<th>Comparability Class</th>
<th>Question Description/Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>148</td>
<td>A1PA29CI</td>
<td>In M1 only: M1 has a separate variable (A1PA29CK) to indicate 'None', while 'None' is a separate response category in other waves.</td>
<td>Response Category Availability</td>
<td>History of Severe Chest Pain</td>
</tr>
<tr>
<td>147</td>
<td>A1PA29CJ</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Severe Chest Pain</td>
</tr>
<tr>
<td>146</td>
<td>A1PA29CK</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>149</td>
<td>A1PA33</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>151</td>
<td>A1PA34</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>154</td>
<td>A1PA30</td>
<td>Not in M1: Since this question was open-ended in M1, response was not coded.</td>
<td>Format Difference</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>155</td>
<td>A1PA31</td>
<td>Not in M1: Since this question was open-ended in M1, response was not coded.</td>
<td>Format Difference</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>156</td>
<td>A1PA32S</td>
<td>Not in M1: Since this question was open-ended in M1, response was not coded.</td>
<td>Format Difference</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>157</td>
<td>A1PA32D</td>
<td>Not in M1: Since this question was open-ended in M1, response was not coded.</td>
<td>Format Difference</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>158</td>
<td></td>
<td>Not in M1: Since this question was open-ended in M1, response was not coded.</td>
<td>Format Difference</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>159</td>
<td></td>
<td>Not in M1: Since this question was open-ended in M1, response was not coded.</td>
<td>Format Difference</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>160</td>
<td></td>
<td>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.</td>
<td>Coding Scheme</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>161</td>
<td></td>
<td></td>
<td>Coding Scheme</td>
<td>History of Blood Pressure</td>
</tr>
<tr>
<td>162</td>
<td></td>
<td></td>
<td>History of Blood Pressure</td>
<td>Ever Had Cancer</td>
</tr>
<tr>
<td>163</td>
<td></td>
<td></td>
<td>History of Blood Pressure</td>
<td>Ever Had Cancer</td>
</tr>
</tbody>
</table>
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           |
| ✓ 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
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
www.colectica.com