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

**Baseline:** 1995-96

- Harvard University
- Funded by MacArthur Foundation
- N=7,108
- Ages 25-74
- Satellite studies of stress, cognition

*Advancing Knowledge of Factors That Promote Positive Health and Resilience*
Timelines for MIDUS Longitudinal and Refresher Data Collection

1. Survey: n=7108
   Cognitive: n=302
   Daily Stress: n=1499

2. Survey: n=4693
   Cognitive: n=4512
   Daily Stress: n=2022
   Biomarker: n=1255
   Neuroscience: n=331
   Milwaukee: n=592

3. Survey: target n=3500
   Cognitive: target n=2280
   Daily Stress: target n=800
   Biomarker: target n=1000
   Neuroscience: target n=125
   Milwaukee: n=500

4. Survey: target n=3300
   Cognitive: target n=2680
   Daily Stress planned
   Biomarker planned
   Neuroscience planned
   Milwaukee planned

5. Survey: target n=350
   Biomarker: target n=350
### 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

<table>
<thead>
<tr>
<th>PROJECT 2</th>
<th>PROJECT 3</th>
<th>PROJECT 4</th>
<th>PROJECT 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Daily Diary Study)</td>
<td>(Cognitive Functioning)</td>
<td>(Biomarkers)</td>
<td>(Neuroscience)</td>
</tr>
</tbody>
</table>
| 8 days of daily experience obtained via phone interviews. (e.g., time use, physical health symptoms and substance use, work productivity, psychological distress) | **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) | **2-Day Clinic Visit: Biomarkers—neuroendocrine, cardiovascular, immune, bone** | **Affective reactivity & recovery:**
  - baseline electroencephalography (EEG)
  - task-related EEG
  - task-related electromyography (EMG; eyelink startle response, post auricular startling reflex, corrugator supercilli activity)
  - structural MRI of neuroanatomy
  - task event-related fMRI |
| 4 days of salivary cortisol | **Face-to-face assessment of cognitive capacities** | **Physical exam** | **Laboratory challenge study—heart-rate variability, blood pressure, cortisol** |

---
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/
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

<table>
<thead>
<tr>
<th></th>
<th>B M1 Variable Name</th>
<th>C M2 Variable Name</th>
<th>D MKE Variable Name</th>
<th>E MR Variable Name</th>
<th>F MKER Variable Name</th>
<th>G M3 Variable Name</th>
<th>H M1 Variable Label</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>HEALTH</td>
<td></td>
<td></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>
<td>Physical health</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>A1PA5</td>
<td>B1PA2</td>
<td>BACA2</td>
<td>RA1PA2</td>
<td>RAACA2</td>
<td>C1PA2</td>
<td>Mental or emotional health</td>
<td>Mental/mental</td>
</tr>
<tr>
<td>70</td>
<td>A1PA6</td>
<td>B1PA3</td>
<td>BACA3</td>
<td>RA1PA3</td>
<td>RAACA3</td>
<td>C1PA3</td>
<td>Self-evaluated health</td>
<td>Health complaints</td>
</tr>
<tr>
<td>71</td>
<td>A1PA7</td>
<td>B1PA4</td>
<td>BACA4</td>
<td>RA1PA4</td>
<td>RAACA4</td>
<td>C1PA4</td>
<td>Days work limited by health</td>
<td>Days unable to</td>
</tr>
<tr>
<td>72</td>
<td>A1PA4A</td>
<td>B1PA4A</td>
<td>BACA4A</td>
<td>RA1PA4A</td>
<td>RAACA4A</td>
<td>C1PA4A</td>
<td>Unable to work, physical, mental</td>
<td>Reason unable</td>
</tr>
<tr>
<td>73</td>
<td>A1PA4BA</td>
<td>B1PA4BA</td>
<td>BACA4BA</td>
<td>RA1PA4BA</td>
<td>RAACA4BA</td>
<td>C1PA4BA</td>
<td># of Days physical</td>
<td>Num days unit</td>
</tr>
<tr>
<td>74</td>
<td>A1PA4BB</td>
<td>B1PA4BB</td>
<td>BACA4BB</td>
<td>RA1PA4BB</td>
<td>RAACA4BB</td>
<td>C1PA4BB</td>
<td># of Days mental</td>
<td>Num days unit</td>
</tr>
<tr>
<td>75</td>
<td>A1PA4BC</td>
<td>B1PA4BC</td>
<td>BACA4BC</td>
<td>RA1PA4BC</td>
<td>RAACA4BC</td>
<td>C1PA4BC</td>
<td># of Days combination</td>
<td>Num days unit</td>
</tr>
<tr>
<td>76</td>
<td>A1PA8</td>
<td>B1PA5</td>
<td>BACA5</td>
<td>RA1PA5</td>
<td>RAACA5</td>
<td>C1PA5</td>
<td># Days cut back on work due health</td>
<td>Days cut back</td>
</tr>
<tr>
<td>77</td>
<td>A1PA3A</td>
<td>B1PA5A</td>
<td>BACA5A</td>
<td>RA1PA5A</td>
<td>RAACA5A</td>
<td>C1PA5A</td>
<td>Physical, mental or both</td>
<td>Reason cut back</td>
</tr>
<tr>
<td>78</td>
<td>A1PA8BA</td>
<td>B1PA5BA</td>
<td>BACA5BA</td>
<td>RA1PA5BA</td>
<td>RAACA5BA</td>
<td>C1PA5BA</td>
<td># of Days physical</td>
<td>Num days unit</td>
</tr>
<tr>
<td>79</td>
<td>A1PA8BB</td>
<td>B1PA5BB</td>
<td>BACA5BB</td>
<td>RA1PA5BB</td>
<td>RAACA5BB</td>
<td>C1PA5BB</td>
<td># of Days mental</td>
<td>Num days unit</td>
</tr>
<tr>
<td>80</td>
<td>A1PA8BC</td>
<td>B1PA5BC</td>
<td>BACA5BC</td>
<td>RA1PA5BC</td>
<td>RAACA5BC</td>
<td>C1PA5BC</td>
<td># of Days combination</td>
<td>Number days</td>
</tr>
<tr>
<td>81</td>
<td>A1PA9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Physical health at 16</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>A1PA10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mental health at 18</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>A1PA6A</td>
<td>B1PA6A</td>
<td>BACA6A</td>
<td>RA1PA6A</td>
<td>RAACA6A</td>
<td>C1PA6A</td>
<td>History of stroke</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>A1PA6B</td>
<td>B1PA6B</td>
<td>BACA6B</td>
<td>RA1PA6B</td>
<td>RAACA6B</td>
<td>C1PA6B</td>
<td>History of spondylosis</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>A1PA6C</td>
<td>B1PA6C</td>
<td>BACA6C</td>
<td>RA1PA6C</td>
<td>RAACA6C</td>
<td>C1PA6C</td>
<td>History of Pericarditis</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td>A1PA6D</td>
<td>B1PA6D</td>
<td>BACA6D</td>
<td>RA1PA6D</td>
<td>RAACA6D</td>
<td>C1PA6D</td>
<td>History of other</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>A1PA11</td>
<td>B1PA7</td>
<td>BACA7</td>
<td>RA1PA7</td>
<td>RAACA7</td>
<td>C1PA7</td>
<td>Heart problems ever</td>
<td>Heart troubles</td>
</tr>
<tr>
<td>88</td>
<td>A1PA11A</td>
<td>B1PA7A</td>
<td>BACA7A</td>
<td>RA1PA7A</td>
<td>RAACA7A</td>
<td>C1PA7A</td>
<td>Age of heart problem</td>
<td>Age doctor told</td>
</tr>
<tr>
<td>89</td>
<td>A1PA11BA</td>
<td>B1PA7BA</td>
<td>BACA7BA</td>
<td>RA1PA7BA</td>
<td>RAACA7BA</td>
<td>C1PA7BA</td>
<td>Heart attack</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>90</td>
<td>A1PA11BB</td>
<td>B1PA7BB</td>
<td>BACA7BB</td>
<td>RA1PA7BB</td>
<td>RAACA7BB</td>
<td>C1PA7BB</td>
<td>Diagnosis - A</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>91</td>
<td>A1PA11BC</td>
<td>B1PA7BC</td>
<td>BACA7BC</td>
<td>RA1PA7BC</td>
<td>RAACA7BC</td>
<td>C1PA7BC</td>
<td>Angina</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>92</td>
<td>A1PA11BD</td>
<td>B1PA7BD</td>
<td>BACA7BD</td>
<td>RA1PA7BD</td>
<td>RAACA7BD</td>
<td>C1PA7BD</td>
<td>High blood pressure</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>93</td>
<td>A1PA11BE</td>
<td>B1PA7BE</td>
<td>BACA7BE</td>
<td>RA1PA7BE</td>
<td>RAACA7BE</td>
<td>C1PA7BE</td>
<td>Diagnosis - V</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>94</td>
<td>A1PA11BF</td>
<td>B1PA7BF</td>
<td>BACA7BF</td>
<td>RA1PA7BF</td>
<td>RAACA7BF</td>
<td>C1PA7BF</td>
<td>Diagnosis - B</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>95</td>
<td>A1PA11BG</td>
<td>B1PA7BG</td>
<td>BACA7BG</td>
<td>RA1PA7BG</td>
<td>RAACA7BG</td>
<td>C1PA7BG</td>
<td>Diagnosis - Ir</td>
<td>Diagnosis - H</td>
</tr>
<tr>
<td>96</td>
<td>A1PA11RH</td>
<td>B1PA7RH</td>
<td>BACA7RH</td>
<td>RA1PA7RH</td>
<td>RAACA7RH</td>
<td>C1PA7RH</td>
<td>Heart murmur</td>
<td></td>
</tr>
<tr>
<td>M1 Variable Name</td>
<td>Comparability Notes (among MIDUS Project 1)</td>
<td>Comparability Class</td>
<td>Question Description/Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>148 A1PA29CI</td>
<td></td>
<td></td>
<td>History of Severe Chest Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>147 A1PA29CJ</td>
<td></td>
<td></td>
<td>History of Severe Chest Pain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>148 A1PA29CK</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>149</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>Not in M1.</td>
<td>General Item Availability</td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>152 A1PA33</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>153 A1PA34</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>154</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>155</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>156</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>157</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>158</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>159</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>160 A1PA30</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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>161 A1PA30</td>
<td></td>
<td></td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>162 A1PA31</td>
<td></td>
<td></td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>163 A1PA32S</td>
<td></td>
<td></td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>164 A1PA32D</td>
<td></td>
<td></td>
<td>History of Blood Pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>165 A1PA36</td>
<td></td>
<td></td>
<td>Ever Had Cancer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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