Adding Contextual Data to Longitudinal Studies of Aging: Opportunities and Challenges

NIA Longitudinal Studies Workshop

June 5, 2017

HRS | CDR

Health and Retirement Study: Contextual Data Resource

Jennifer Ailshire, PhD

University of Southern California

A joint project of the University of Southern California and the University of Michigan
Sponsored by the National Institute on Aging

Introduction

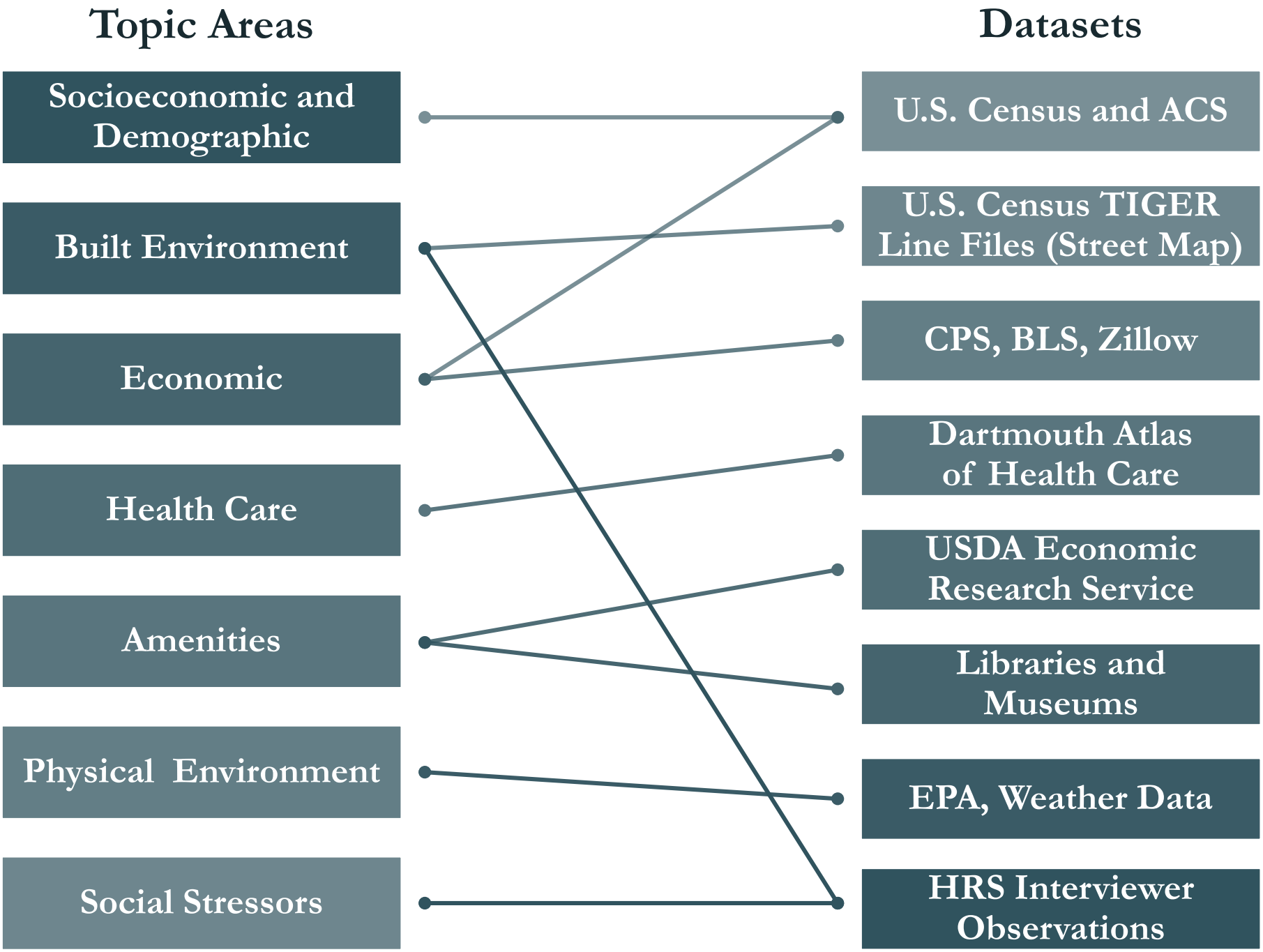
The Health and Retirement Study Contextual Data Resource **HRS-CDR** is a collection of user friendly datasets that enable researchers to study the impact of place on health and well-being among HRS respondents.

HRS | CDR

What is the HRS Contextual Data Resource?

- Compilation of datasets containing contextual measures, with complete documentation, merged with HRS geographic identifiers.
- Datasets at multiple geographic levels (e.g., census tract, county/metro area/hospital service area).
  - Buffer and distance-based measures based on household location.
- Temporal coverage over much of the survey period (i.e., 1990-2010)
  - Most measures are annual/semi-annual depending on source

HRS | CDR


Topic Areas
Datasets
Socioeconomic and Demographic
Built Environment
Economic
Health Care
Amenities
Physical Environment
Social Stressors
U.S. Census and ACS
U.S. Census TIGER Line Files (Street Map)
CPS, BLS, Zillow
Dartmouth Atlas of Health Care
USDA Economic Research Service
Libraries and Museums
EPA, Weather Data
HRS Interviewer Observations

Why Create a CDR for the HRS?

- Linkage of longitudinal survey data with longitudinal contextual data creates valuable new research avenues
  - E.g., Exploring geographic differences in health, mechanisms underlying health disparities, person-environment interactions, aging in place/aging in changing places
- But, data linkage can create user burden
  - Most geographic ids only available in restricted data
  - Requires time, knowledge, resources
- Due to data confidentiality, users cannot create GIS-based measures

HRS | CDR

Data Development Process

- Simultaneous and coordinated work conducted at USC and UM
  - Maximizes user-end data product while preserving respondent confidentiality.

USC CDR

Data Creation Node

HRS | CDR

Data Linkage Node

USC

1. Download Datasets
2. Convert to Consistent, Usable Format
3. Geocode Data and Interpolate
4. Construct Summary Measures and Scales

CDR

...could be linked with other surveys.

HRS

1. Geocode Respondent Addresses
2. Link to Administrative Boundaries
3. Merge CDR Data with HRS Respondents
4. Construct Respondent Proximity Measures

HRS | CDR

Data Access

The HRS-CDR is a Restricted Data Product that can be accessed through the MiCDA Enclave Virtual Desktop Infrastructure (VDI) with an approved restricted data application. http://hrsonline.isr.umich.edu/rda/

HRS | HEALTH AND RETIREMENT STUDY
A Longitudinal Study of Health, Retirement, and Aging
Sponsored by the National Institute on Aging

Home | Introduction | Documentation | Data Products | Publications | Search

Restricted Data Overview » HRS - MiCDA Enclave Application Information

Step-By-Step: Accessing HRS restricted data through the MiCDA Enclave Virtual Desktop Infrastructure (VDI)

1. Review the **application checklist**
2. The application should include:
   - Letter from Department chair (Students only)
   - **Research Proposal:** Applicants for HRS Restricted Data must provide to HRS staff a short (1-3 page) research proposal, that includes a synopsis (or a full statement, if necessary) of your research goals, and specifies:
     - the types of variables from HRS Restricted Datasets you intend to use in your research; and
     - why you believe the unrestricted versions of those variables, if any, are not adequate for your research purposes.

     For each research project proposed, applicants must provide:
     - Project Title
     - Project Executive Summary (one paragraph abstract of research goals)
     - Study Team Details - for each study team member, defined as anyone who will have access to the restricted data, provide Name, Role on Project, Contact Information (Complete business street address, Email, Telephone)
   - **Data order form**
   - **MiCDA VDI Data Security Plan**
   - **MICDA Data Enclave Acceptable Use Policy**
   - **ISR Pledge to Safeguard Respondent Privacy**
3. Accessing restricted data as a visitor to the MiCDA Enclave: Certain data merges can only be performed by visiting the Enclave in person. In addition to the above, visitors to the Enclave should complete the **Confidentiality Agreement Restricting Disclosure and Use of Data from the MiCDA Enclave**

Please note that a separate, complete application package is required for each participating institution. A single application may include multiple users. The completed application package should be emailed to HRS Restricted Data Applications Processing (**hrsrdaapplication@umich.edu**).

RDA links

RDA Overview
Product List
Application Materials
Extending Acccess
Annual Renewals
Termination
Audits
FAQs
CMS Research Data

HRS Order Form
CMS Order Form

Contact Information

MiCDA Enclave

http://hrsonline.isr.umich.edu/index.php?p=micdahrsapply&jumpfrom=RR

HRS | CDR

Major Challenges

- Large data dimensionality
- Balancing access and data availability against data confidentiality and administrative burden
- Periodicity of data resource updates
- Adding CDRs to other NIA studies

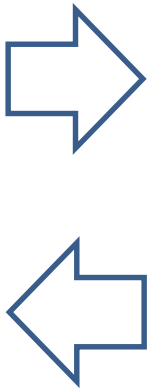
Data Can Quickly Become Unwieldy

- National level contextual data at micro-scale creates data processing burden for GIS-based measures (e.g., walkability measures within 1 mile buffer)
- Many data files. HRS-CDR is organized by:
  - Dataset (e.g., Dartmouth Atlas)
    - Geography (e.g., state, HRR, HSA)
      - Topic (e.g., post-discharge events, primary care access, prescription drug utilization)
- Longitudinal data on individuals and place creates a large data set

HRS | CDR

Individual-level data are long format
Contextual data are wide format

Respondent-Level File		
hhid	year	tract
4002	1992	19921
4002	1994	19921
4002	1996	19921
5636	1992	40601
5636	1994	40601
5636	1996	40601
7042	1992	15283
7042	1994	15283
7042	1996	15283
7042	1998	15283
7053	2000	15283
7053	2002	15283
7053	2004	15283
7053	2006	15283
7053	2008	15283
7053	2010	15283
10181	1992	1021
10181	1994	1021
10181	1996	1021
10181	1998	1021
10181	2000	1021
10181	2002	1021
10181	2004	1021
10181	2006	1021
10181	2008	1021
10181	2010	1021



Contextual Data File: Dartmouth Atlas Post-Discharge Events (by State)					
d100st	d102st	d333st2004	d333st2008	d333st2009	d333st2010
1	AL	18.86	19.01	19.73	17.44
2	AK		16.34	12.72	15.25
3	AZ	19.77	17.72	17.75	18.79
4	AR	18.31	17.64	18.86	18.60
5	CA	18.10	18.38	17.83	17.40
6	CO	15.71	15.18	17.07	14.35
7	CT	19.93	20.41	19.05	18.58
8	DE	21.35	18.21	16.79	17.59
9	DC	26.33	19.96	20.44	22.11
10	FL	20.10	19.34	18.51	17.99
11	GA	18.99	18.45	17.34	17.39
12	HI	19.80	18.80	20.25	18.06
13	ID	14.80	13.24	14.04	12.22
14	IL	20.38	20.07	19.48	19.01
15	IN	18.08	17.09	16.69	16.23
16	IA	17.29	16.70	15.65	15.02
17	KS	17.26	15.71	18.10	16.44
18	KY	21.02	19.50	20.16	18.96
19	LA	20.77	19.21	18.12	17.94
20	ME	19.86	20.01	18.26	15.97

HRS | CDR

Maximizing User Product & Access

- Providing data through a virtual enclave increases access and increases administrative control over restricted data
- But users limited to data in the enclave and linkages by user may not be feasible
- Users may not have access to linked data of their choice

HRS | CDR

Updating the CDR

- Respondents move and places change
- Some administrative data are updated annually (e.g., CMS, EPA), or every few years
- Some studies are conducted annually (NHATS) or every couple years (HRS/PSID)

Extending CDR Linkages

- Administrative staff requirements
  - Need data managers to conduct data linkages
  - Need staff with GIS capability to create certain measures
- Data confidentiality concerns
  - Data linkages should conform to study-specific requirements for maintaining respondent confidentiality (e.g., identifiability)
- User access
  - How will users gain access to the data?

HRS | CDR

Potential Strategies

- Data Warehousing
  - A place where geographic data are housed for individuals and studies to download and link to the individual-level data they're working with
  - Means to distribute NIA funded national and longitudinal data contextual data resources
- File Sharing Infrastructure
  - Creating infrastructure to allow individual studies to input respondent household (and workplace) coordinates and get spatially linked data in return

HRS | CDR

Acknowledgments

- Grant funding: **NIA R21 AG045625**
- Project personnel & collaborators
  - University of Southern California
    - Jennifer Ailshire, Eileen Crimmins, Sarah Mawhorter, Matthew Young, Catherine Perez, Hyewon Kang
  - University of Michigan
    - Philippa Clarke, Jessica Faul, Robert Melendez, Daniel Awad

HRS | CDR