

Social Science Data: A Data Archive Perspective

George Alter

Director

ICPSR,

Institute for Social Research

University of Michigan



About ICPSR

- Founded in 1962 as a consortium of 21 universities to share the National Election Survey
- Today: 700+ members around the world
- Data dissemination for more than 20 federal and non-government sponsors
- 600,000+ visitors per year

What we do

- Acquire and archive social science data
- Distribute data to researchers
- Preserve data for future generations
- Provide training in quantitative methods

Archive size

- 8,000 data collections, over 60,000 data sets
- Grows by 200+ collections a year

Changes in the information landscape:

- Improved metadata and data discovery
- Linking data and publications
- Emerging sources and types of data
- Sharing confidential data

Improved metadata and data discovery

- Social science data tends to be
 - wide (>400 columns/variables)
 - but not deep (<10,000 rows)
- Data discovery requires variable-level metadata
- Data Documentation Initiative (DDI)
 - XML standard for micro-data in the social sciences

Improved metadata and data discovery

- New search tools
 - Variable level searching
 - Question banks
 - Harmonization tools

“Search/Compare Variables” examines 2.1 million variables in 4,000 data collections

Compare Variables

Compare (5) << Back to search results

Clear All (5)

BL73

P120C

NWIS99D

C0202

T2VOLUNT

[Longitudinal Study of American Youth, 1987-1994, and 2007: LSAY Merged Cohort File \(Base File\)](#)

PARENTS:ANY VOLUNTEER WORK FOR SCHOOL

Question

-

Response Categories

Code	Label	Freq.	%
-99	Not asked	0	0.0
-98	Won't say	1	0.0
-97	Don't know	4	0.1
-95	Did not participate	3463	58.3
1	Yes	564	9.5
2	No	1913	32.2

Time Period
1987--1994 2007

Universe
The methodology and sampling is described in the LSAY user guide.

[more info](#)

[Children of Immigrants Longitudinal Study \(CILS\), 1991-2006](#)

Parent/spouse-school volunteer

Question

Do you and your spouse/partner do any of the following at your child's school? Act as a volunteer in the school?

Response Categories

Code	Label	Freq.	%
1	Yes	989	18.8
2	No	1433	27.2
	Missing Data		
.	-	2840	54.0

Time Period
1991--2006

Universe
Immigrant second generation children born in the United States with at least one foreign-born parent, or children born abroad but brought to the United States at an early age.

[more info](#)

[Span Study of Youth and Social Development, 1992-1997 \(United States\): Teenage Life Questionnaire: Focal Data, Year 1](#)

PARENTS VOLUNTEERED AT SCHOOL

Question

-

Response Categories

Code	Label	Freq.	%
1	NEVER	607	62.1
2	ONCE OR TWICE	131	13.4
3	MORE THAN TWICE	111	11.4
4	DO NOT KNOW	64	6.6
6	MULTIPLE RESPONSE	0	0.0
7	REFUSED	0	0.0
9	MISSING	64	6.6

Time Period
1992--1997

Universe
Adolescents in the United States

[more info](#)

[School Survey on Crime and Safety \(SSOCS\), 2006](#)

PARENT VOLUNTEERS AT SCHOOL

Question

5. What is your best estimate of the percentage of students who had at least one parent or guardian participating in the following events during the 2005-06 school year? d. Volunteered at school or served on a committee

Response Categories

Code	Label	Freq.	%
1	0-25%	-	-
2	26-50%	-	-
3	51-75%	-	-
4	76-100%	-	-
5	School does not offer	-	-

Time Period
2006

Universe
Public elementary and secondary schools in the United States.

[more info](#)

[Early Childhood Longitudinal Study \(United States\): Kindergarten Class of 1998-1999, Kindergarten-Eighth Grade Full Sample: ECLS-K Kindergarten-Eighth Grade Full Sample Public-Use File](#)

T2 Q11D PARENT VOLUNTEERED IN SCHOOL

Question

-

Response Categories

Code	Label	Freq.	%
1	YES	8274	38.6
2	NO	8258	38.6
3	NOT APPLICABLE/NOT OFFERED	1453	6.8
	Missing Data		
-9	NOT ASCERTAINED	1310	6.1
-8	DON'T KNOW	0	0.0
-7	REFUSED	0	0.0
.	-	2114	9.9

Time Period
1998--2007

Universe
Children and their families, teachers, and schools in the United States.

[more info](#)

Improved metadata and data discovery

- Federated search tools
 - Data Preservation Alliance for the Social Sciences (Data-PASS) combined catalog searches:
 - Institute for Quantitative Social Science, Harvard University
 - Odum Institute, University of North Carolina-Chapel Hill
 - ICPSR, University of Michigan
 - Electronic and Special Media Records Service Division, National Archives and Records Administration
 - Roper Center, University of Connecticut
 - Social Science Data Archive, University of California, Los Angeles (UCLA)

Linking data and publications

- Benefits of linking data and publications
 - Encourages research transparency
 - Promotes re-use of existing data
 - Links researchers across disciplines
- ICPSR *Bibliography of Data-related Literature*
 - 60,000+ citations
 - Challenge of poor citation practices
- DataCite: DOIs for data
- *Data Citation Index*

All Databases

<< Back to results list

Record 9 of 66 >>

MGet It Mirlyn

(0) | | Save to: [ENDNOTE® WEB](#) [ENDNOTE®](#) [RefWorks](#) [I Wrote These Publications](#) [R](#) more options

Perceived age discrimination and mental health

Author(s): Yuan, ASV (Yuan, Anastasia S. Vogt)

Source: SOCIAL FORCES **Volume:** 86 **Issue:** 1 **Pages:** 291-311 **DOI:** 10.1353/sof.2007.0113 **Published:** SEP 2007

Times Cited: 5 (from Web of Science)

Cited References: 40 [[view related records](#)] [Citation Map](#)

Abstract: Although perceived discrimination (especially due to race-ethnicity) decreases mental health, the influence of perceived discrimination due to other reasons on mental health needs to be explored. This study examines the relationship between perceived age discrimination and mental health and determines whether psychosocial resources explain or buffer (i.e., reduce the strength of) this relationship. Using a nationally-representative sample of persons aged 25-74 from the Midlife Development in the United States survey, this study finds that perceived age discrimination is associated with higher psychological distress and lower positive well-being. Perceived age discrimination is more negatively associated with women's mental health than men's. Although sense of control buffers the relationship between perceived age discrimination and psychological distress, perceived age discrimination decreases sense of control and social support.

Accession Number: WOS:000251031500013

Document Type: Article

Language: English

Keywords Plus: PSYCHOLOGICAL DISTRESS; RACIAL-DISCRIMINATION; SOCIAL SUPPORT; STRESS PROCESS; OLDER-ADULTS; DEPRESSION; VULNERABILITY; EPIDEMIOLOGY; COMPETENCE; STEREOTYPE

Reprint Address: Yuan, ASV (reprint author), Virginia Polytech Inst & State Univ, Dept Sociol, 560 McBryde Hall 0137, Blacksburg, VA 24061 USA.

Addresses:

1. Virginia Tech, Blacksburg, VA 24061 USA

E-mail Address: avy@vt.edu

Publisher: UNIV NORTH CAROLINA PRESS, BOX 2288, JOURNALS DEPT, CHAPEL HILL, NC 27515-2288 USA

Web of Science Categories: Sociology

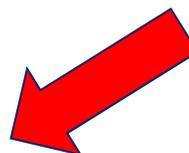
Research Areas: Sociology

IDS Number: 232PL

ISSN: 0037-7732

Associated Data Citation Index Records: [[View All](#)]

National Survey of Midlife Development in the United States (MIDUS), 1995-1996.	Data study
---	------------



Save to: [ENDNOTE® WEB](#) [ENDNOTE®](#) [RefWorks](#) [I Wrote These Publications](#) [R](#) [more options](#)

1. Title: **National Survey of Midlife Development in the United States (MIDUS), 1995-1996.**
Author(s): Brim, Orville G; Baltes, Paul B; Bumpass, Larry R; et al.
Source: Inter-university Consortium for Political and Social Research
DOI: <http://dx.doi.org/10.3886/ICPSR02760> Version: v7
Document Type: Data study Times Cited: 66 (from All Databases) [[View abstract](#)]
2. Title: **Midlife Development in the United States (MIDUS): Boston Study of Management Processes, 1995-1997.**
Author(s): Lachman, Margie E
Source: Inter-university Consortium for Political and Social Research
DOI: <http://dx.doi.org/10.3886/ICPSR03596> Version: v1
Document Type: Data study Times Cited: 3 (from All Databases) [[View abstract](#)]
3. Title: **National Survey of Midlife Development in the United States (MIDUS II), 2004-2006.**
Author(s): Ryff, Carol; Almeida, David M; Ayanian, John S; et al.
Source: Inter-university Consortium for Political and Social Research
DOI: <http://dx.doi.org/10.3886/ICPSR04652> Version: v2
Document Type: Data study Times Cited: 1 (from All Databases) [[View abstract](#)]
4. Title: **Social Responsibility at Midlife.**
Author(s): Colby, Anne
Source: Institute for Quantitative Social Science
Source URL: <http://hdl.handle.net/1902.1/00022> Version: 2 1
Document Type: Data study Times Cited: 0 (from All Databases) [[View abstract](#)]

Emerging sources and types of data

- Geo-spatial
- Video
- Administrative data
- Online text
- Transactions
- Clicks
- Sensors

Safety Pilot Project

- University of Michigan Transportation Research Institute under a USDOT contract
- Vehicle-to-Vehicle and Vehicle-to-Infrastructure devices in volunteer's cars
- 2800 cars, trucks, and buses with Vehicle Awareness Devices, sensors, and video
- ~ 1 Petabyte of data per year

Adaptive education software records every click.



Casinos capture data too...



SECTION 97.

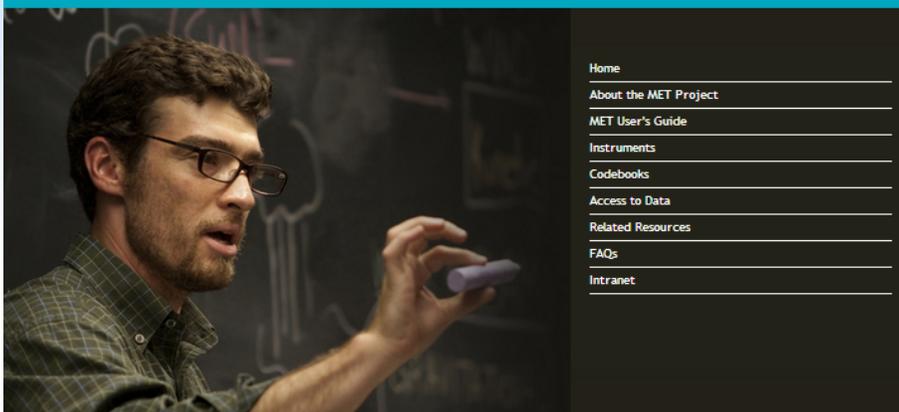
..., a gaming establishment shall supply the Massachusetts gaming Commission with customer tracking data collected or generated by loyalty programs, player tracking software, player card systems, online gambling transactions or any other information system. ... The Commission shall convey the anonymized data to a research facility which shall make the data available to qualified researchers ...

Chapter 194 of the Acts of 2011: An Act Establishing Expanded Gaming in the Commonwealth of Massachusetts, signed into law November 22, 2011.

<http://www.malegislature.gov/Laws/SessionLaws/Acts/2011/Chapter194>

MET Longitudinal Database

MEASURES of EFFECTIVE TEACHING
Longitudinal Database



How do you **QUANTIFY**
EXCELLENCE in **TEACHING**?

Begin with the Measures of Effective Teaching Longitudinal Database. Found within the collection are over 20,000 hours of rich video data – active classrooms, real teachers, current practices. The collection also includes a diverse set of quantitative data ranging from standardized instrument scores to teacher and student surveys to student assessments.

Be a part of designing fairer, more powerful, and more reliable measures of teacher effectiveness.

[Contact Us](#) [Privacy Policy](#) [Log In/Create Account](#)

BILL & MELINDA
GATES foundation ICPSR

- Variety of indicators
 - 5 instruments for scoring classroom observation
 - Student surveys
 - teacher evaluations
 - effort
 - enjoyment
 - “Value-added” computed from test scores
 - State math and English language tests
 - Other standardized tests
 - Other instruments and surveys

MET Longitudinal Database

MEASURES of EFFECTIVE TEACHING

Account Admin Help Logout

Options...

Home
Data & Documentation
Virtual Data Enclave
Contact Us
Help
Related Links

Join our Email list
Get notifications about updates to our studies.
Email Address:

Video Documentation

Context Files

Date: Monday, July 12, 2010
Time: 11:15 A.M.
Teacher:
Location:
Grade: 5th grade
Subject: Mathematics
Topic: Fractions

Notes Instruments

Notes

BILL & MELINDA GATES foundation

ICPSR | A PARTNER IN SOCIAL SCIENCE RESEARCH

[Privacy Policy](#)

Scale

- 2 academic years
- 6 large school districts
- 6 grade levels
- 3,000 teachers
- 44,500 students
- 24,000 videos
- 22,500 observation sessions
- 900+ observers trained by ETS to score videos
- **~12GB of quantitative data**
- **~10TB of video**

Metadata: MET LDB

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 34309

Measures of Effective Teaching: Section-Level Analytical Files, 2009-2011

Bill and Melinda Gates Foundation

Year 1 Section Level Analytical File 4th-8th Grade Codebook

Inter-university Consortium for Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Sample Codebook

- 864 variables
- 423 pages
- 1 of 30+ data files in the MET LDB collection
- ICPSR codebooks are generated from DDI.

- Study 34309 -
Measures of Effective Teaching: Section-Level Analytical Files, 2009-2010
Year 1 Section Level Analytical File 4th-8th Grade

Variables

GRADE_LEVEL: Grade Level
Location: 1-1 (width: 1, decimal: 0)
Variable Type: numeric
Notes: frequencies are available
Notes: summary statistics are not available

Value	Unweighted Frequency	%
4	502	15.6 %
5	515	16.0 %
6	819	25.5 %
7	730	22.7 %
8	647	20.1 %

Based upon 3,213 valid cases out of 3,213 total cases.

SCF_SUBJ: Subject
Location: 2-64 (width: 63, decimal: 0)
Variable Type: character
Notes: frequencies are available
Notes: summary statistics are not available

Value	Unweighted Frequency	%
Elementary ELA	402	12.5 %
Elementary ELA + math	642	20.0 %
Elementary math	334	10.4 %
Math	16	0.5 %
Middle school ELA	959	29.8 %
Middle school math	800	26.8 %

Based upon 3,213 valid cases out of 3,213 total cases.

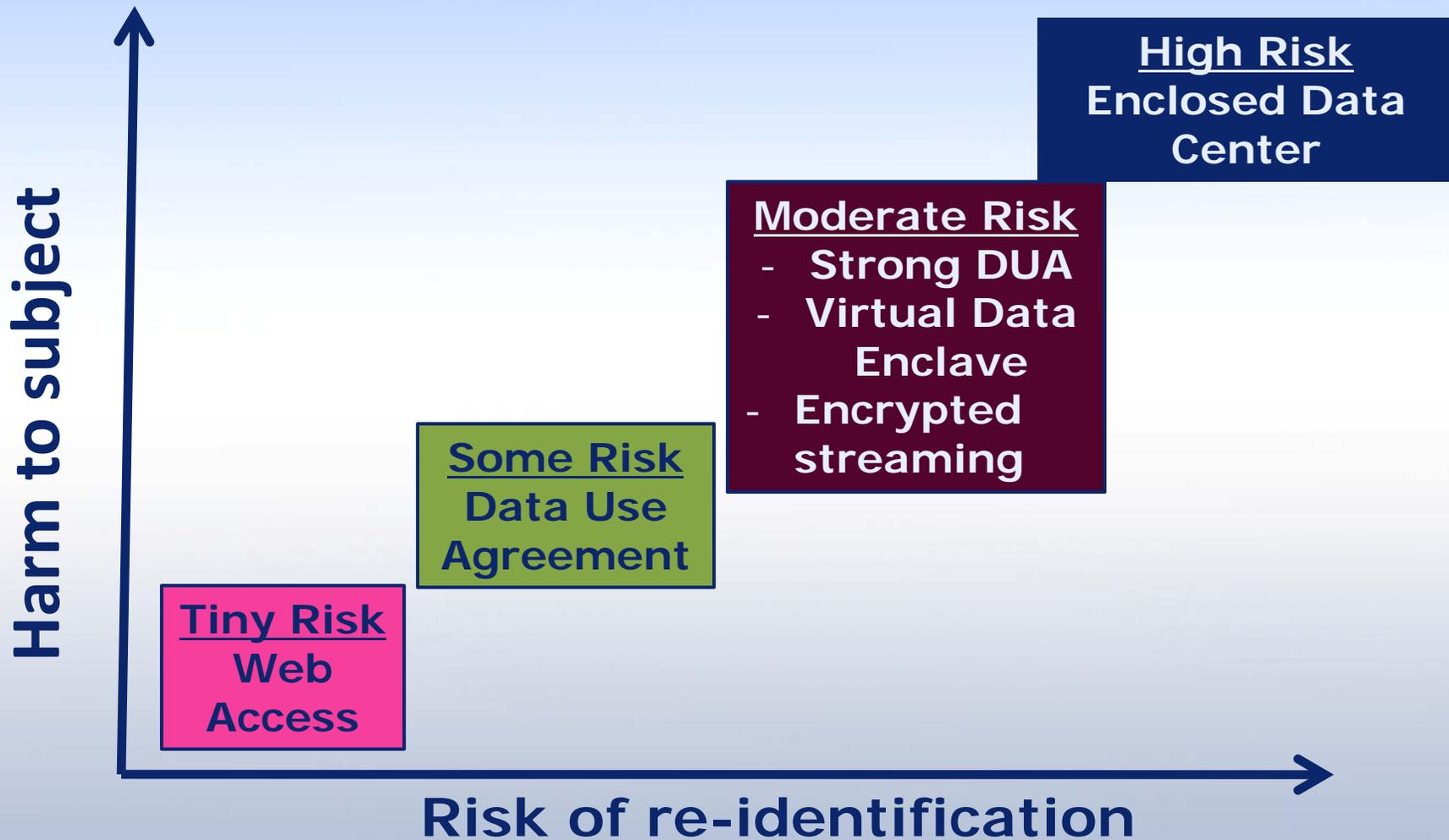
SD_HISPANIC: District Administrative Data: Proportion of Hispanic
Location: 65-68 (width: 4, decimal: 2)
Variable Type: numeric
Based upon 3,213 valid cases out of 3,213 total cases.

- Mean: 0.33
- Median: 0.23
- Mode: 0.00
- Minimum: 0

Sharing confidential data

- **Safe data:** Modify the data to reduce the risk of re-identification
- **Safe places:** Physical isolation and secure technologies
- **Safe people:** Training and Data use agreements

Confidentiality: Harm, Risk & Restriction



Challenges

- Secure access to confidential information
 - Partnerships between public/private data centers and research communities
 - New access models
- Providing accurate and complete metadata, especially provenance
 - Automated metadata capture
- Improving data management and citation practices

Thank you.

George Alter
altergc@umich.edu