Data Curation for Rehabilitation Research and Related Clinical Trials

University of Florida - Gainesville

May 22, 2019
Program

Speakers

Kenneth J. Ottenbacher, PhD, OTR
Division of Rehabilitation Science, UTMB

Shane Redman, MA
ICPSR, University of Michigan-Ann Arbor

Amol Karmarkar, PhD, MPH
Division of Rehabilitation Science, UTMB

Guest Speaker:
Linda Tickle-Degnen, PhD, OTR
Department of Occupational Therapy, Tufts University
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 8:00 am – 8:30 am | Welcome and Introductions / Overview of the CLDR
|               | K. Ottenbacher                                                         |
| 8:30 am – 9:00 am | Why is Data Sharing Important (even in Clinical Trials research)
|               | A. Karmarkar                                                           |
| 9:00 am – 9:40 am | Overview of ICPSR and Intro to ADDEP
|               | Shane Redman                                                          |
| 9:40 am – 9:50 am | Break                                                                  |
| 9:50 am – 10:35 am | Preparing Data for Archiving (include Identifying and Protecting Confidential Data)
|               | S. Redman                                                            |
| 10:35 am – 11:15 am | Metadata and Transferring Data to a Repository
|               | S. Redman                                                            |
| 11:15 am – 11:25 am | Break                                                                  |
| 11:25 am – 11:45 am | “Why Share Research Data: PI Perspective”
<p>|               | L. Tickle-Degnen                                                      |</p>
<table>
<thead>
<tr>
<th>Time</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>11:45 am – 12:45 pm</td>
<td>Working Lunch – Instructors will divide into groups and make themselves available to discuss related topics (ex: Data Management/Curation, Analysis, and Funding) a. Karmarkar, K. Ottenbacher, S. Redman</td>
</tr>
<tr>
<td>12:45 pm – 1:15 pm</td>
<td>Panel Discussion L. Tickle-Degnen, K. Ottenbacher, A. Karmarkar, &amp; S. Redman</td>
</tr>
<tr>
<td>1:15 pm – 2:00 pm</td>
<td>Data Consultation and Exercises – For those seeking input on data related to planned, ongoing, or completed research projects S. Redman / A. Karmarkar</td>
</tr>
<tr>
<td>2:00 pm – 2:10 pm</td>
<td>Break</td>
</tr>
<tr>
<td>2:10 pm – 2:50 pm</td>
<td>Data Consultation and Exercises continued S. Redman / A. Karmarkar</td>
</tr>
<tr>
<td>2:50 pm – 3:00 pm</td>
<td>Wrap-up</td>
</tr>
<tr>
<td>3:00 pm</td>
<td>End</td>
</tr>
</tbody>
</table>
Welcome and Introductions

Overview of Data Sharing and the CLDR

Kenneth J. Ottenbacher

University of Texas Medical Branch
Learning Objectives

1. Introduction to Data Science and Implications for Disability, Rehabilitation and Recovery Research

2. Review Recent Developments in Data Sharing and Secondary Analysis

3. Identify Resources for Data Sharing and Science in Rehabilitation Research
Data Science

The interdisciplinary field of inquiry in which quantitative and analytical approaches, processes, and systems are developed and used to extract knowledge and insights from increasingly large and/or complex sets of data.

NIH Strategic Plan for Data Science, 2018
Opportunities

NIH Strategic Plan for Data Science
June 4, 2018

NIH releases strategic plan for data science

https://datascience.nih.gov/strategicplan
NIH Big Data to Knowledge (BD2K)

Areas of BD2K

1. Facilitating Broad Use of Data
2. Analysis Methods and Software
3. Enhancing Training
4. Centers of Excellence

https://datascience.nih.gov/bd2k/about
The NIH Strategic Plan for Data Science represents a transition from BD2K to a broader more comprehensive view of Data Science.

The Strategic Plan will address “the broader biomedical research data ecosystem.” The Plan focuses on the following priority areas:

- Data Infrastructure
- Data Modernization
- Data Management
- Workforce Development
- Stewardship/Sustainability
Data Science

The Plan refers to four components of successful Data Science:

• Data are Findable
• Data are Accessible
• Data are Interoperable
• Data are Reusable

*NIH Strategic Plan for Data Science, 2018*

https://datascience.nih.gov/strategicplanrelease
Data Science Opportunities

Important New Datasets are Available
PCORnet, the National Patient-Centered Clinical Research Network, is an initiative of PCORI. It is designed to make it faster, easier, and less costly to conduct clinical research than is now possible by using the power of large health data and patient partnerships.

http://www.pcornet.org/
Opportunities

An Introduction to the

All of Us

Research Program

@AllofUsResearch  #JoinAllofUs

https://allofus.nih.gov/
The mission of the *All of Us* Research Program is to accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us.

- The cornerstone of the larger Precision Medicine Initiative (PMI) – led by the NIH
- One million or more volunteers, reflecting the broad diversity of the U.S.
- Opportunities for volunteers to provide data on an ongoing basis
- Data will inform a variety of research studies

[https://allofus.nih.gov/](https://allofus.nih.gov/)
Data Sharing is Not a New Idea.


https://doi.org/10.17226/2033.
Data Sharing Policy

Federal agencies with more than $100 million in annual research and development (R&D) expenditures must develop plans for increasing public access to the results of the research they support, specifically scholarly publications and digital data.

White House Office of Science and Technology Policy (February 22, 2013)

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
Opportunities

Data Repositories...

BD2K
bioCADDIE
Consortium

NIH Data Sharing Funding Opportunities

1. Resources to help researchers store data
   • Example: R03 mechanism, Archiving and Documenting Child Health and Human Development Data Sets (PAR-16-149)

2. Resources to encourage secondary analysis of available data
   • Example: R21 mechanism to explore secondary analysis of TBI data (RFA-HD-16-001)
Agency for Healthcare Research and Quality (AHRQ) - Public Policy for Access to Federally Funded Research

AHRQ Public Access to Federally Funded Research

Publications and Data

This document is the Agency for Healthcare Research and Quality's (AHRQ's) plan for establishing a policy for public access to scientific publications and scientific data in digital format resulting from AHRQ funding.

https://www.ahrq.gov/funding/policies/publicaccess/index.html#11
“Diabetes now requires authors of original research studies to describe in their manuscripts how readers can access the data and research resources supporting their findings, methods, and conclusions. Submitted manuscripts must provide statements about the availability of data and critical resources supporting the results reported in the article as part of the Research Design and Methods section”.
Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors


### Examples of Data Sharing Statements That Fulfill ICMJE Requirements

<table>
<thead>
<tr>
<th></th>
<th>Example 1</th>
<th>Example 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will individual participant data be available (including data dictionaries)?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>What data in particular will be shared?</td>
<td>All of the individual participant data collected during the trial, after deidentification</td>
<td>Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices)</td>
</tr>
<tr>
<td>What other documents will be available?</td>
<td>Study protocol, statistical analysis plan, informed consent form, clinical study report, analytic code</td>
<td>Study protocol, statistical analysis plan, analytic code</td>
</tr>
<tr>
<td>When will data be available (start and end dates)?</td>
<td>Immediately following publication; no end date</td>
<td>Beginning 3 months and ending 5 years following article publication</td>
</tr>
<tr>
<td>With whom?</td>
<td>Anyone who wishes to access the data</td>
<td>Researchers who provide a methodologically sound proposal</td>
</tr>
<tr>
<td>For what types of analyses?</td>
<td>Any purpose</td>
<td>To achieve aims in the approved proposal</td>
</tr>
<tr>
<td>By what mechanism will data be made available?</td>
<td>Data are available indefinitely at (link to be included)</td>
<td>Proposals should be directed to xxx@yyy To gain access, data requestors will need to sign a data access agreement. Data are available for 5 years at a third-party website (link to be included).</td>
</tr>
</tbody>
</table>

*These examples are meant to illustrate a range of, but not all, data sharing options.*
Resources to Support Data Science

CLDR Center for Large Data Research & Data Sharing in Rehabilitation
### Resources

**Activities supported by the CLDR**

<table>
<thead>
<tr>
<th>Education &amp; Training</th>
<th>Data Directory</th>
<th>Pilot Projects</th>
<th>Visiting Scholars</th>
<th>Data Archiving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workshops, on-line seminars and training modules to develop skills in large data research</td>
<td>Listing of available datasets including purpose, variables, access and contact information</td>
<td>Collaborative projects with CLDR mentors / investigators using large data related to rehabilitation &amp; recovery</td>
<td>Collaborate with CLDR mentors using large rehabilitation datasets. Support for up to six months</td>
<td>Support for linking / merging and archiving data from completed rehabilitation studies to promote secondary data analyses</td>
</tr>
</tbody>
</table>

**Visit us at:** [http://rehabsciences.utmb.edu/cldr](http://rehabsciences.utmb.edu/cldr)

---

**Supported by the National Institutes of Health – the NICHD/NCMRR, the NINDS, and the NIBIB # P2C-HD065702**
Category 1: Traditional Pilot Projects using Large Data

Category 1 projects are pilot studies using secondary data to examine rehabilitation research issues using large datasets.

- Funding up to $25,000
- Applications accepted once per year
- Five to six studies funded each year
Category 2: Data Sharing and Archiving Pilot Studies

Category 2 projects are focused on archiving data from a completed rehabilitation study with support being given to either the PI or a member of the research team.

- Funding up to $10,000
- Applications accepted year-round
Data Management & Curation

ICPSR stores, curates, and provides access to scientific data so others can reuse the data and validate research findings. Curation, from the Latin "to care," is the process that ICPSR uses to add value to data, maximize access, and ensure long-term preservation.

Data curation is akin to work performed by an art or museum curator. Through the curation process, data are organized, described, cleaned, enhanced, and preserved for public use, much like the work done on paintings or rare books to make the works accessible to the public now and in the future. With the modern Web, it's increasingly easy to post and share data. Without curation, however, data can be difficult to find, use, and interpret. Through curation, ICPSR provides meaningful and enduring access to data.
Archive of Data on Disability to Enable Policy and Research (ADDEP)

Disabilities are increasingly recognized as a part of the broad spectrum of diversity and the human experience. Disabilities have important medical, political, social, and economic implications for the global population. Therefore, interest and research on disability has grown in complexity and exposure.

Mission

The Archive of Data on Disability to Enable Policy and research (ADDEP) is a new ICPSR initiative to build a repository of quantitative and qualitative data on disability. The mission of ADDEP is to improve and enable further research on disability for researchers, policymakers, and practitioners by acquiring, enhancing, preserving, and sharing data. ADDEP data can be used to better understand and inform the implementation of the Americans with Disabilities Act and other disability policies with the goal of improving the lives and conditions of people with disabilities. In addition to providing quality data for research and analysis, ADDEP strives to offer user support and community outreach to enable data to be optimally used by the widest audience.

Primary Aims/Services

1. Build disability data sharing capacity
2. Disseminate disability data
3. Provide technical assistance for data users
4. Build a connection between research data, community outreach, and activism

https://www.icpsr.umich.edu/icpsrweb/content/ADDEP/
Visiting Scholars Program

Investigators can be supported for periods of up to six months at one of the consortium institutions to work with experienced investigators using large datasets.
Pilot Projects and Visiting Scholars

**Pilot Projects**
- CRRLD 2010-2015
  - Boston University
  - Cedars-Sinai Medical Center
  - Colorado State University
  - Craig Hospital
  - Duke University
  - Kessler Foundation Research Center
  - McMaster University
  - Mount Sinai School of Medicine
  - Rehabilitation Institute of Chicago
  - Toronto Rehabilitation Institute
  - University of Buffalo
  - University of California - Davis
  - University of Michigan
  - Pittsburgh
  - Texas Medical Branch

**Visiting Scholars**
- CRRLD 2010-2015
  - Bedford VA Medical Center
  - BiOM (Bedford, MA)
  - Boston University
  - Colorado State University
  - Duke University
  - LSU Health Science Center
  - Path Clinical Research Institute
  - University of Oklahoma
  - University of Southern California
  - UTHSCSA
  - University of Texas Medical Branch

**Pilot Projects**
- CLDR 2015-2020
  - Boston University
  - Brown University
  - Craig Hospital
  - Northwestern University
  - Duke University
  - University of Illinois
  - University of Maryland
  - University of Michigan
  - University of Pittsburgh
  - University of Southern California
  - University of Texas Medical Branch
  - University of Utah
  - University of Washington

**Visiting Scholars**
- CLDR 2015-2020
  - Aegis Therapies and Golden Living
  - Brown University
  - Colorado State University
  - MGH Institute of Health Professions
  - Temple University
  - University of Michigan
  - University of Pittsburgh

*All partnerships not shown include visiting scholars from King Saud University in Saudi Arabia and Chun Shan Medical University in Taiwan.*
Making an Impact

Searching Google for “Images for Big Data”

https://www.google.com/search?q=images+of+large+data
Research Infrastructure

Centralized research infrastructure in specific domains:

- Analysis of large data sets
- Biomechanics and modelling of movement
- Technology assessment and product development
- Clinical trial design
- Regenerative Medicine
- Neuromodulation: clinical applications

Sites offer:

- State-of-the-art research facilities
- Courses and workshops
- Mentorship and consultations
- Pilot grants
- Other collaborative opportunities

WWW.NCMRR.ORG

The CLDR is funded by the NIH - National Institute of Child Health and Human Development, through the National Center for Medical Rehabilitation Research, the National Institute for Neurological Disorders and Stroke, and the National Institute of Biomedical Imaging and Bioengineering. (P2CHD065702).
Why is Data Sharing Important? (even in Clinical Trials research)

Amol M. Karmarkar
University of Texas Medical Branch
&
James E. Graham
Colorado State University/CLDR
Learning Objectives

• **Why**: Why should researchers share data?
  - Mandates & Policies.
  - Combining Data from Small Trials (Observational & Experimental).
  - Return of Investment (ROI).
  - Foster Multidisciplinary Collaboration.
  - Education/training & Capacity building.
  - Professional Recognition.

• **Where**: Resources that are currently available.
  - Data Access.
  - Data Repositories.

• **Barriers & Challenges**

• **How**: Process for data sharing.
Combining Experimental/Observational Studies

- Rehabilitation Studies: Individual studies are smaller in nature.
- Cause & Effect Relationship is not always strong.
- Evidence is limited & mixed.
Pooled Sample Analysis (Example)

Gait Speed and Survival in Older Adults

Stephanie Studenski, MD, MPH
Subashan Perera, PhD
Kushang Patel, PhD
Caterina Rosano, MD, PhD
Kimberly Faulkner, PhD
Marco Inzitari, MD, PhD
Jennifer Brach, PhD
Julie Chandler, PhD
Peggy Cawthon, PhD
Elizabeth Barrett Connor, MD
Michael Nevitt, PhD
Marjolein Visser, PhD
Stephen Kritchevsky, PhD
Stefania Badinelli, MD
Tamara Harris, MD
Anne B. Newman, MD
Jane Cauley, PhD
Luigi Ferrucci, MD, PhD
Jack Guralnik, MD, PhD

Context Survival estimates help individualize goals of care for geriatric patients, but life tables fail to account for the great variability in survival. Physical performance measures, such as gait speed, might help account for variability, allowing clinicians to make more individualized estimates.

Objective To evaluate the relationship between gait speed and survival.

Design, Setting, and Participants Pooled analysis of 9 cohort studies (collected between 1986 and 2000), using individual data from 34 485 community-dwelling older adults aged 65 years or older with baseline gait speed data, followed up for 6 to 21 years. Participants were a mean (SD) age of 73.5 (5.9) years; 59.6%, women; and 79.8%, white; and had a mean (SD) gait speed of 0.92 (0.27) m/s.

Main Outcome Measures Survival rates and life expectancy.

Results There were 17 528 deaths; the overall 5-year survival rate was 84.8% (confidence interval [CI], 79.6%-88.8%) and 10-year survival rate was 59.7% (95% CI, 46.5%-70.6%). Gait speed was associated with survival in all studies (pooled hazard ratio per 0.1 m/s, 0.88; 95% CI, 0.87-0.90; P < .001). Survival increased across the full range of gait speeds, with significant increments per 0.1 m/s. At age 75, predicted 10-year survival across the range of gait speeds ranged from 19% to 87% in men and from 35% to 91% in women. Predicted survival based on age, sex, and gait speed was as accurate as predicted based on age, sex, use of mobility aids, and self-reported function or age, sex, chronic conditions, smoking history, blood pressure, body mass index, and hospitalization.

Conclusion In this pooled analysis of individual data from 9 selected cohorts, gait speed was associated with survival in older adults.

JAMA. 2011;305(1):50-58

https://jamanetwork.com/journals/jama/fullarticle/644554
Pooled Sample Analysis (Example)

Figure 2. Predicted Median Life Expectancy by Age and Gait Speed

https://jamanetwork.com/journals/jama/fullarticle/644554
Pooled Sample Analysis (Example)

<table>
<thead>
<tr>
<th>Study</th>
<th>No. of Deaths</th>
<th>Total Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular Health Study, 1991</td>
<td>3851</td>
<td>5801</td>
</tr>
<tr>
<td>Established Populations for the Epidemiologies Study of the Elderly, 1985</td>
<td>1955</td>
<td>2128</td>
</tr>
<tr>
<td>Health, Aging, and Body Composition Study, 2009, 2005</td>
<td>848</td>
<td>3048</td>
</tr>
<tr>
<td>Hispanic Established Populations for Epidemiological Study of the Elderly, 1999</td>
<td>972</td>
<td>1905</td>
</tr>
<tr>
<td>Invecchiare in Chianti, 2000</td>
<td>187</td>
<td>972</td>
</tr>
<tr>
<td>Osteoporotic Fractures in Men, 2005</td>
<td>1073</td>
<td>5833</td>
</tr>
<tr>
<td>Third National Health and Nutrition Examination Study, 2004</td>
<td>2837</td>
<td>3958</td>
</tr>
<tr>
<td>Predicting Elderly Performance, 2003</td>
<td>293</td>
<td>491</td>
</tr>
<tr>
<td>Study of Osteoporotic Fractures, 1990</td>
<td>5512</td>
<td>10349</td>
</tr>
</tbody>
</table>

[https://jamanetwork.com/journals/jama/fullarticle/644554](https://jamanetwork.com/journals/jama/fullarticle/644554)
RAND Data Harmonization

Overview

HEALTH & RETIREMENT STUDIES: CORE CONTENT AREAS

Demographic
- Education, marital status, age, resident, birth year, birth month
- Cognition, disease, depression, injury, physical functioning, physical measures, health behaviors

Health
- Insurance: utilization, expenditure, out-of-pocket spending

Health Services
- Employment status/history, labor force, earnings, disability, retirement, type of work, pension

Work & Employment
- Earnings, asset income, government transfers, pension, financial assets, housing, non-financial assets

Economic Status
- Parents' information, household structure, family exchange, family support, social participation

Family Structure & Social Network

What's New

- Register for an upcoming webinar
- Harmonized SHARE Life History now available
- Harmonized ELSA Life History now available
- Harmonized SHARE Ver.D.5 now available
- New: Harmonized HRS End of Life
- New resources for Stress research
- New Open Position: Research Programmer II

https://g2aging.org/
Rehabilitation Research Cross-Dataset Variable Catalog

Browse or search for detailed variable level rehabilitation relevant information across 6 major datasets. The catalog provides variable names, labels, survey questions, response categories and other related variables that can be exported into an excel spreadsheet for your use.

Note: this tool is designed to provide an overview of rehabilitation related information available across multiple datasets - always use the dataset's codebook/dictionary to guide your actual analysis.

Screen Reader Friendly Version

Show/Hide Help

Search by Topic, Data Set, or Keyword

Select Topic

Select Data Set

Keyword Search:  

Search

Open All Close All

Demographics

- Age
- Education
- Gender
- Marital Status
- Race/Ethnicity

Disability & Medical Conditions

Employment

Health Behaviors

Healthcare Access & Utilization

Healthcare Costs & Health Insurance

HRS MEPS NHANES NLTCS NSCH SIPP

Search
Return on Investment (ROI)

- Typical clinical trial workflow:

1. Study design/planning phase
2. Secure funding
3. Obtain IRB/other approvals
4. Recruit participants → Collect data
5. Data analysis
6. Dissemination
7. Close study protocol
Return on Investment (ROI)

“The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data”

Amy M. Pienta, George Alter, Jared Lyle
Inter-university Consortium for Political and Social Research, Institute for Social Research, University of Michigan

“Multivariate models of count of publications suggest that data sharing, especially sharing data through an archive, leads to many more times the publications than not sharing data. This finding is robust even when the models are adjusted for PI characteristics, grant award features, and institutional characteristics”.

https://deepblue.lib.umich.edu/bitstream/handle/2027.42/78307/pientaAlterLyle_100331.pdf?sequence=1&isAllowed=y
Multidisciplinary Collaboration

• Dataset for secondary analysis purposes:
  ➢ Can be Agnostic: If the research question being asked is different than the original reason/s behind data collection.
  ➢ Opportunity for every team member to bring their skills sets/expertise on the table.
Category 3 Travel Award for Secondary Analysis of Archived Studies is an opportunity for early career researchers, postdoctoral fellows and doctoral students to perform secondary analysis of data from completed studies in the Archive of Data on Disability to Enable Policy and research (ADDEP) repository. Award recipients will participate in the CLDR Symposium at the American Congress of Rehabilitation Medicine (ACRM) annual conference. Funds, up to $2,500, will support conference registration, travel, and accommodations. Note: International travel is not covered by this award.

We are only targeting the following datasets for the current competition:

- Well Elderly 2, Los Angeles, California, 2004-2008 (ICPSR 33641)
- Hispanic Established Populations for Epidemiologic Studies of the Elderly (EPESE) Frailty Study: 2006-2009 (ICPSR 36321)
- The Anatomical Tracings of Lesions after Stroke (ATLAS) Dataset
- WHO Study on Global AGEing and Adult Health (SAGE): Wave 0, 2002-2004 (ICPSR 28502)
- WHO Study on Global AGEing and Adult Health (SAGE): Wave 1, 2007-2010 (ICPSR 31381)
- Early Intervention Colorado (EI-CO) Participant Characteristics, Service Use, and Outcomes, Colorado, 2014-2016 (ICPSR 37131)
- Medical University of South Carolina Stroke Data (ARRA) (ICPSR 37122)
Education/Training & Capacity Building

• Develop a teachable process that can be integrated in the existing curriculum.

• Process: Study design, Data management, Data analysis.

• Products (abstracts poster/presentations, manuscripts) coming out from class-related projects.
Professional Recognition

Utilization for Hispanic Established Populations for Epidemiologic Studies of the Elderly (EPESE) Frailty Study: 2006-2009

Using ICPSR Metrics

**General Information**
- Initial release date: 2016-03-29
- Last updated: 2016-03-29
- Related publications: 2

**Unique Users**

<table>
<thead>
<tr>
<th></th>
<th>Total Unique Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Datasets/Files</td>
<td>150</td>
</tr>
<tr>
<td>DS0: Study-Level Files</td>
<td>99</td>
</tr>
<tr>
<td>DS1: Frailty 1</td>
<td>102</td>
</tr>
<tr>
<td>DS2: Frailty 2</td>
<td>71</td>
</tr>
</tbody>
</table>

**Unique Sessions**

<table>
<thead>
<tr>
<th></th>
<th>Total Data/Documentation Sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any Datasets/Files</td>
<td>374</td>
</tr>
<tr>
<td>DS0: Study-Level Files</td>
<td>178</td>
</tr>
<tr>
<td>DS1: Frailty 1</td>
<td>155</td>
</tr>
<tr>
<td>DS2: Frailty 2</td>
<td>119</td>
</tr>
</tbody>
</table>

**Downloads by Member Institutions**
- University of Alabama at Birmingham
- University of Chicago
- University of Kansas
- University of Michigan
- University of Southern California
- Washington University in St. Louis
Availability of Resources

- Data Access: Rehabilitation Data Directory developed by CLDR
- Data Repositories
American Community Survey (ACS) 2008-2017
- Prevalence
- Employment Rate
- Not Working but Actively Looking for Work
- Full-Time / Full-Year Employment
- Annual Earnings
- Annual Household Income
- Poverty
- Supplemental Security Income (SSI)
- Educational Attainment
- Veterans Service-Connected Disability
- Health Insurance Coverage (and Type)

- Prevalence
- Labor Market Activity
- Employment
- Household Income
- Poverty

EEOC Charge Data 2006-2014
Report Dashboard
- State Level Reports on Common Issues and Bases
- Compare the Most Common Bases Cited Across States
- Compare the Most Common Issues Cited Across States
- Charge Rate Comparison by State
- Charge Rate Comparison by Employment Discrimination Area

Ask Our Researchers
- Can't find a statistic or have a question about disability data or datasets?
  E-mail your question to our researchers at disabilitystatistics@cornell.edu
# Rehabilitation Dataset Directory: Dataset Profile

**Dataset: Spinal Cord Injury Rehab (SCIRehab)**

## Basic Information

<table>
<thead>
<tr>
<th>Dataset Full Name</th>
<th>Spinal Cord Injury Rehab</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset Acronym</td>
<td>SCIRehab</td>
</tr>
</tbody>
</table>

## Summary

The SCIRehab is a 5-year, multi-center study based on Practice Based Evidence (PBE) approach which collects data related to inpatient rehabilitation process of care and outcomes for individuals with Spinal Cord Injury (SCI). The SCIRehab project collects detail information on specific intervention, duration of therapy, patient participation, caregiver role, social participation, and patient-centered outcomes including functional status and quality of life.

The data includes detailed information on discipline-specific rehabilitation interventions which are not recorded adequately by the clinicians in normal documentation during inpatient rehabilitation process. Six major US inpatient rehabilitation hospitals specializing in providing rehabilitation services for individuals with SCI were involved in collaborative research. They report point of care data of a patient with SCI and rehabilitation service outcomes to Craig Hospital in Englewood, Colorado.

## Key Terms

- Spinal Cord Injury
- Rehabilitation Outcomes
- Practice Based Evidence
- Nursing education
- Occupational therapy
- Physical therapy
- Speech therapy
- Therapeutic Recreation
- Psychology
- Social work/case management
- Physiatry

## Study Design

Longitudinal

## Data Type(s)

- Administrative
- Clinical
- Survey

## Sponsoring Agency/Entity

Department of Health and Human Services (HHS): National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR)

## Health Conditions/Disability Measures

<table>
<thead>
<tr>
<th>Health Condition(s)</th>
<th>Spinal cord injury (SCI),</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disability Measures</td>
<td>Functional limitations (ADLs and/or IADLs)</td>
</tr>
</tbody>
</table>

## Measures/Outcomes of Interest

### Topics

- Patient and injury characteristics. Detailed information about the rehabilitation process and intensity of therapy during inpatient rehabilitation by each discipline: physical therapy, occupational therapy, therapeutic recreation, speech-language pathology, psychology, nursing, and social work/case management. Individual and Group Therapy Co-treatment with other disciplines, Duration of therapy. Family/Caregiver involvement. Missed therapy and factors that impact rehabilitation sessions. Patient participation in Physical Therapy and Occupational therapy sessions.

- The level of assistance needed in performing activities of daily living. Functional Independence Measure (FIM). A detailed description of functional status by a subset of the skills. Detailed information on the burden of care by distinguishing “total assistance of more than one person” from “total assistance of one person.”

- Admission and discharge function, Discharge destination, rehospitalization in the first year, variables related to community participation, work or school attendance, depression, the presence of pressure ulcers, usage of assistive technology, and job satisfaction at one year post-injury.
<table>
<thead>
<tr>
<th>Data Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Website</strong></td>
</tr>
<tr>
<td>ICPSR site:</td>
</tr>
<tr>
<td><a href="https://www.icpsr.umich.edu/icpsrweb/ADDEP/studies/36724/summary">https://www.icpsr.umich.edu/icpsrweb/ADDEP/studies/36724/summary</a></td>
</tr>
<tr>
<td><a href="http://www.scirehab.net/">http://www.scirehab.net/</a></td>
</tr>
<tr>
<td><strong>Data Access</strong></td>
</tr>
<tr>
<td><a href="http://www.scirehab.net/scirehabdataform.aspx">http://www.scirehab.net/scirehabdataform.aspx</a></td>
</tr>
<tr>
<td><strong>Data Access Requirements</strong></td>
</tr>
<tr>
<td>Data Use agreement, No cost</td>
</tr>
<tr>
<td><strong>Summary Tables/reports</strong></td>
</tr>
<tr>
<td>N/A</td>
</tr>
<tr>
<td><strong>Data Components</strong></td>
</tr>
<tr>
<td>8 Files:</td>
</tr>
<tr>
<td>- Patient Level Data</td>
</tr>
<tr>
<td>- Physical Therapy</td>
</tr>
<tr>
<td>- Occupational Therapy-1</td>
</tr>
<tr>
<td>- Occupational Therapy-2</td>
</tr>
<tr>
<td>- Speech Language Pathology</td>
</tr>
<tr>
<td>- Therapeutic Recreation</td>
</tr>
<tr>
<td>- Psychology</td>
</tr>
<tr>
<td>- Social Work-Case Management</td>
</tr>
<tr>
<td>- Nursing</td>
</tr>
<tr>
<td><strong>Selected papers</strong></td>
</tr>
<tr>
<td><strong>Other Papers</strong></td>
</tr>
</tbody>
</table>
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2716826/ |
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2716827/ |
| **Technical**                                   |
| Data Documentation (ICPSR):                     |
| https://www.icpsr.umich.edu/icpsrweb/ADDEP/studies/36724/datadocumentation# |
| SciRehab papers:                                |

**Ask Our Researchers**

Have a question about disability data or datasets?  
E-mail your question to our researchers at disability.statistics@cornell.edu
### Spinal Cord Injury Rehabilitation Study, United States, 2007-2010 (ICPSR 36724)

**Version Date:** May 3, 2018  
**Principal Investigator(s):** Gale G. Whiteneck, Craig Hospital (Englewood, Colo.)

https://doi.org/10.3886/ICPSR36724.v1

**Version V1**

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Preview</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS0 Study-Level Files</td>
<td>0 bytes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS1 Patient Level Data</td>
<td>143 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS2 Physical Therapy</td>
<td>635 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS3 Occupational Therapy-1</td>
<td>483 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS4 Occupational Therapy-2</td>
<td>462 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS5 Speech Language Pathology</td>
<td>35 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS6 Therapeutic Recreation</td>
<td>85 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS7 Psychology</td>
<td>109 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS8 Social Work Case Management</td>
<td>226 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS9 Nursing</td>
<td>78 MB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes**

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.
Availability of Resources

• Over 2,000 research data repositories listed on re3data.org
• Lots of open access to research data out there – but are the data usable for the purposes of further research? Will it increase research impact?
## NIH Data Sharing Repositories

This table lists NIH-supported data repositories that make data accessible for reuse. Most accept submissions of appropriate data from NIH-funded investigators (and others), but some restrict data submission to only those researchers involved in a specific research network. Also included are resources that aggregate information about biomedical data and information sharing systems. The table can be sorted according by name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites of the individual repositories. Are we missing a data sharing repository? [Contact us](#).

<table>
<thead>
<tr>
<th>Repository ID</th>
<th>Repository Name</th>
<th>Repository Description</th>
<th>Data Submission Policy</th>
<th>Access to Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI</td>
<td>Cancer Nanotechnology Laboratory (cancanolab)</td>
<td>canNanLab is a data sharing portal designed to facilitate information sharing in the biomedical nanotechnology research community to expedite and validate the use of nanotechnology in biomedicine. canNanLab provides support for the annotation of nanomaterials with characteristics resulting from physico-chemical, in vitro, and in vivo studies, and the sharing of these characterization and associated nanotechnology protocols in a secure fashion.</td>
<td><a href="#">How to submit your data to canNanLab</a></td>
<td><a href="#">How to access canNanLab data</a></td>
</tr>
<tr>
<td>NCI</td>
<td>The Cancer Imaging Archive (TCIA)</td>
<td>The image data in The Cancer Imaging Archive (TCIA) is organized into purpose-built collections of subjects. The subjects typically have a cancer type and/or anatomical site (lung, brain, etc.) in common.</td>
<td><a href="#">How to submit data to TCIA</a></td>
<td><a href="#">How to access TCIA data</a></td>
</tr>
<tr>
<td>NCI (NHGRI, NIGMS)</td>
<td>PeptideAtlas</td>
<td>PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometry output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences.</td>
<td><a href="#">How to submit data to PeptideAtlas</a></td>
<td><a href="#">How to access PeptideAtlas data</a></td>
</tr>
<tr>
<td>NEI</td>
<td>EyeGENE®</td>
<td>The EyeGENE® Biorepository and corresponding Database contain family history and clinical eye exam data from subjects enrolled in EyeGENE® Program coupled to clinical grade DNA samples. This data and samples are submitted by collaborators throughout the US and Canada and the data is available on a controlled access basis to researchers worldwide.</td>
<td><a href="#">How to submit data to EyeGENE®</a></td>
<td><a href="#">How to access EyeGENE® data</a></td>
</tr>
<tr>
<td>NHGRI</td>
<td>FlyBase: A Drosophila Genomic and Genetic Database</td>
<td>Drosophila genomic and genetic database that includes proteomics data, microarrays and Tiling ERAs.</td>
<td><a href="#">How to submit data to Flybase</a></td>
<td><a href="#">How to access Flybase data</a></td>
</tr>
<tr>
<td>NHGRI</td>
<td>The Zebrafish Model Organism Database (ZFIN)</td>
<td>ZFIN serves as the zebrafish model organism database. It aims to: a) be the community database resource for the laboratory use of zebrafish, b) develop and support integrated zebrafish genetic, genomic and developmental information, c) maintain the definitive reference data sets of zebrafish research information, d) to link this information extensively to corresponding data in other model organism and human databases, e) facilitate the use of zebrafish as a model for human biology, and f) serve the needs of the research community.</td>
<td><a href="#">How to submit data to ZFIN</a></td>
<td><a href="#">How to access ZFIN data</a></td>
</tr>
<tr>
<td>NHGRI</td>
<td>WormBase</td>
<td>WormBase is an international consortium of biologists and computer scientists dedicated to providing the research community with accurate, current, accessible information concerning the genetics, genomics and biochemistry of C. elegans and related nematodes.</td>
<td><a href="#">How to submit data to WormBase</a></td>
<td><a href="#">How to access WormBase data</a></td>
</tr>
<tr>
<td>NHGRI/NIGMS</td>
<td>The Universal Protein Resource (UniProt)</td>
<td>The Universal Protein Resource (UniProt) is a comprehensive resource for protein sequence and annotation data. The UniProt databases are the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (Uniparc).</td>
<td><a href="#">How to submit data to UniProt</a></td>
<td><a href="#">How to access UniProt data</a></td>
</tr>
</tbody>
</table>
## Availability of Resources

<table>
<thead>
<tr>
<th>NICHD</th>
<th>Repository Name</th>
<th>Repository Description</th>
<th>Data Submission Policy</th>
<th>Access to Data</th>
<th>Current NIH funding support</th>
<th>Open data submission</th>
<th>Open data access</th>
<th>Open time frame for data deposit</th>
<th>Sustained support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data and Specimen Hub (DASH)</td>
<td>NICHD DASH is a centralized resource for researchers to store and access de-identified data from NICHD funded research studies for the purposes of secondary research use. It serves as a mechanism for NICHD-funded extramural and intramural investigators to share research data from studies in accordance with the NIH Data Sharing Policy and the NIH Genomic Data Sharing Policy.</td>
<td>How to submit data to DASH</td>
<td>How to access DASH data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Data Sharing for Demographic Research (DSDR)</td>
<td>DSDR is a data sharing project providing curation and archiving services for the demographic and population sciences community. It focuses on data collected through funding from the NICHD Population Dynamics Branch (PDB), but also provides these services for other data sets that fall within the scientific mission of PDB.</td>
<td>How to submit data to DSDR</td>
<td>How to access DSDR data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>National Children’s Study (NCS) Archive</td>
<td>The NCS Archive, a data and sample repository for the National Children’s Study, provides access to data and samples collected from over 5,600 U.S. birth families to study environmental influences on child health and development. Data and biological and environmental samples are freely available, with an approved request, for scientific research.</td>
<td>How to submit data to NCS</td>
<td>How to access NCS data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>PhonBank</td>
<td>PhonBank is an open database for the study of early phonological development using the Phon program.</td>
<td>How to submit data to PhonBank</td>
<td>How to access PhonBank data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Xencase</td>
<td>Xencase is a Xenopus laevis and Xenopus tropicalis biology and genomics resource.</td>
<td>How to submit data to Xencase</td>
<td>How to access Xencase data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>NICHD (NINDS, &amp; NIBIB)</td>
<td>Archive of Data on Disability to Enable Policy (ADDEP)</td>
<td>ADDEP provides access to data including a wide range of topics related to disability. ADDEP data can be used to better understand and inform the implementation of the Americans with Disabilities Act and other disability policies.</td>
<td>How to submit data to ADDEP</td>
<td>How to access ADDEP data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>NIH (NIA, NICHD, NIDA)</td>
<td>Inter-university Consortium for Political and Social Research (ICPSR)</td>
<td>ICPSR provides leadership and training in data access, curation, and methods of analysis for the social science research community. ICPSR maintains a data archive of more than 250,000 files of research in the social and behavioral sciences. ICPSR collaborates with a number of funders, including U.S. statistical agencies and foundations, to create thematic data collections and data stewardship and research projects. NICHD (DSDR), NIA (NAACDA), and NIDA (NAHDAP) currently support specialty archives that use the ICPSR infrastructure.</td>
<td>How to submit data to ICPSR</td>
<td>How to access ICPSR data</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
OpenICPSR

Share your behavioral health and social science research data in three easy steps:

- Name your project
- Upload and describe files
- Publish

GET STARTED NOW

family services

47,009
file uploads
in the last 12 months

591
projects

2,583,027
downloads
in the last 12 months

Free to share!
Get cited for sharing data
Control your own updates
Challenges

• **Infrastructure**: insufficient platforms to store & manage clinical trial data under a variety of access models.
• **Technological**: existing platforms are not well-known, intuitive, or standardized.
• **Workforce**: lacks the skills & knowledge to manage the operational & technical aspects of data sharing.
• **Sustainability**: costs exclusive to original sponsors & investigators; need to be distributed equitably across both data generators & users.
Thank You!
Archive of Data on Disability to Enable Policy and Research

Shane Redman
icpsr-addep@umich.edu

Advancing Research on Disability
Overview

• What is ICPSR?
• What is ADDEP?
• How to use ICPSR/ADDEP
ICPSR

- Inter-university Consortium for Political and Social Research
  - Within Institute for Social Research at the University of Michigan
- 750+ Member institutions
- ~1,000 Summer training participants annually
- Over 11,000 studies archived at ICPSR
- 80,000 citations in the searchable bibliography
- Over 5 million variables tagged
What is ADDEP?

• Archive of Data on Disability to Enable Policy and research
• A joint initiative of the Center for Large Data Research and Data Sharing in Rehabilitation (CLDR) and the Inter-university Consortium for Political and Social Research (ICPSR)
• Mission: Improve and enable research among researchers, policymakers, and practitioners by *acquiring* disability and rehabilitation data and *providing access* to these data
ADDEP: A Topical Archive

- Topical data archive providing curated disability and rehabilitation data for secondary analysis
  - Offers wide range of resources including data analysis tools, bibliography of related publications, and data resources for various audiences
ADDEP

www.icpsr.umich.edu/ADDEP
Primary ADDEP Activities

• Acquire, prepare, and disseminate data
  • Well-documented datasets in multiple formats
  • Study-level and variable-level metadata for each study
• Online analysis
• Bibliography of related publications
Prepare and Disseminate Data (Curation Services)

• Disclosure Review
• Missing Data Standardization
• Outliers, Wild Codes
• Study-level and Variable-level Metadata
• ICPSR Codebook
• Quality Checks
• Dissemination in SAS, SPSS, Stata, R, ASCII
Primary ADDEP Activities

• Educate and train users
  • Webinars, workshops, and presentations
    • Best practices for data sharing and preparation
    • Data analysis tools
  • Online guides
    • How to prepare data for archiving
    • How to access and analyze archived data
• User support
Educate and Train Users

Education and Training

ADDEP and the [CLDR](http://cldr.org) will provide education, training, and learning experiences about both archiving data and using archived data in the disability and rehabilitation medicine fields. Future training includes workshops, webinars, written guides, and other training tailored to fit the needs of the disability and rehabilitation research communities to strengthen and support the infrastructure around shared data.

ICPSR offers education and training materials as well as other forms of technical assistance for data depositors and data users.

Data Management and Curation Resources available at ICPSR:

- [Guide to Social Science Data Preparation and Archiving](http://www.icpsr.umich.edu/icpsrweb/guide/)
- [Data Management & Curation](http://www.icpsr.umich.edu/icpsrweb/guide/)
- [Confidentiality Preservation](http://www.icpsr.umich.edu/icpsrweb/guide/)
- [Citing Data](http://www.icpsr.umich.edu/icpsrweb/guide/)
- [Depositing Data](http://www.icpsr.umich.edu/icpsrweb/guide/)

Teaching Resources available at ICPSR:

- [TeachingWithData.org](http://www.teachingwithdata.org)
- [ICPSR Data-Driven Learning Guides (DDLGs)](http://www.icpsr.umich.edu/icpsrweb/daa/)
- [Instructor’s Guide to DDLGs](http://www.icpsr.umich.edu/icpsrweb/daa/)

Webinars and Workshops Presented by ADDEP:

- [Introduction to ADDEP](http://www.addep.org)
- [Archiving Data with ADDEP](http://www.addep.org)
- [Data Curation for Disability and Rehabilitation Outcomes Research Workshop](http://www.addep.org/)

Researchers can find additional details about data preparation, data management, and other relevant topics on the [ICPSR YouTube Channel](https://www.youtube.com канал).
ADDEP’s Value to the Research Community

• For researchers and depositors:
  • Increases discoverability and shareability of data
  • Enhances quality of data for analysis
  • Technical assistance with data preparation
  • Data citation – those who share data receive credit

• For secondary users:
  • Access to high-quality, curated data on a wide range of topics
  • Multiple resources, links to publications, and tools for secondary data analysis and data exploration
Current ADDEP Holdings

- 119 studies total discoverable on ADDEP website
  - 12 Series of datasets
  - Over 2,500 associated publications
  - Over 80,000 variables

- Recent releases:
  - Boston Rehabilitative Impairment Study of the Elderly (Boston RISE)
  - Anatomical Tracings of Lesions after Stroke (ATLAS) – Release 1.2
  - Medical University of South Carolina Stroke Data (ARRA)
ICPSR can help!

• Collecting data?
• Writing new study proposals?
• Ideas for interesting projects?

• Contact us!
• Depositing data with us and user support is always free.

icpsr-addep@umich.edu
Using ICPSR/ADDEP Website

www.icpsr.umich.edu
### Results from Keyword Search

<table>
<thead>
<tr>
<th>Study Title/Investigator</th>
<th>Released/Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rehabilitation Measures Database</strong> (<strong>ICPSR 36587</strong>)</td>
<td>2016-11-28</td>
</tr>
<tr>
<td>Center for Rehabilitation Outcomes Research (CROR), Rehabilitation Institute of Chicago; Department of Medical Social Sciences Informatics, Northwestern University Feinberg School of Medicine (NU M55)</td>
<td></td>
</tr>
<tr>
<td><strong>Public Support for Rehabilitation in Ohio, 1996</strong> (<strong>ICPSR 2543</strong>)</td>
<td>2006-03-30</td>
</tr>
<tr>
<td>Applegate, Brandon K.</td>
<td></td>
</tr>
<tr>
<td><strong>Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014</strong> (<strong>ICPSR 37045</strong>)</td>
<td>2018-05-31</td>
</tr>
<tr>
<td>Bean, Jonathan</td>
<td></td>
</tr>
<tr>
<td><strong>Spinal Cord Injury Rehabilitation Study, United States, 2007-2010</strong> (<strong>ICPSR 36724</strong>)</td>
<td>2018-05-03</td>
</tr>
<tr>
<td>Whiteneck, Gale G.</td>
<td></td>
</tr>
<tr>
<td>Pullen, Suzanne; English, Kim</td>
<td></td>
</tr>
<tr>
<td>Pearson, Frank S.</td>
<td></td>
</tr>
<tr>
<td><strong>Functional Independence in Children at a Pediatric Clinic in Guanajuato, Mexico, 2004-2013</strong> (<strong>ICPSR 37068</strong>)</td>
<td>2018-07-09</td>
</tr>
<tr>
<td>Bucior, Samuel; Toro Hernández, Maria Luisa; Pearlman, Jon; Dausey, David J.</td>
<td></td>
</tr>
<tr>
<td><strong>Matching Treatment and Offender: North Carolina, 1980-1982</strong> (<strong>ICPSR 8515</strong>)</td>
<td>1992-02-16</td>
</tr>
<tr>
<td>Marsden, Mary Ellen; Orsagh, Thomas</td>
<td></td>
</tr>
</tbody>
</table>

- Studies
- Variables
- Series
- Publications
- ICPSR Website (news, announcements, etc.)
### Results from Keyword Search

#### Variables (3,861)

<table>
<thead>
<tr>
<th>Var. Name</th>
<th>Label/Question Text</th>
<th>Var. Type</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUQ088C</td>
<td>Rehabilitation facility</td>
<td>numeric</td>
<td>DS223 - Questionnaire: Hospital Utilization</td>
</tr>
<tr>
<td>HUQ088C</td>
<td>Rehabilitation facility</td>
<td>numeric</td>
<td>DS224 - Questionnaire: Hospital Utilization</td>
</tr>
<tr>
<td>SF470</td>
<td>REHABILITATION PROGRAM</td>
<td>numeric</td>
<td>DS1 - Survivor Data File</td>
</tr>
<tr>
<td>SF1125</td>
<td>REHABILITATION PROGRAM</td>
<td>numeric</td>
<td>DS1 - Survivor Data File</td>
</tr>
<tr>
<td>SF1160</td>
<td>REHABILITATION PROGRAM</td>
<td>numeric</td>
<td>DS1 - Survivor Data File</td>
</tr>
</tbody>
</table>

- Studies
- Variables
- Series
- Publications
- ICPSR Website (news, announcements, etc.)
# Results from Keyword Search

<table>
<thead>
<tr>
<th></th>
<th>Studies (2,303)</th>
<th>Variables (3,861)</th>
<th>Series (132)</th>
<th>Publications (198)</th>
<th>ICPSR Website (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>ABC News/Washington Post Poll Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Adult Education Surveys Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Afrobarometer Survey Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Americans' Use of Time Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>American Community Survey (ACS) Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>American Housing Survey Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>American National Election Study (ANES) Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>American Time Use Survey (ATUS) Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td>Annual Parole Survey Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td>Annual Probation Survey Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td>Annual Survey of Governments Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td>Annual Survey of Jails Data Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Annual Survey of Jails in Indian Country Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Canadian National Elections Study (CNES) Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Census of Governments Series</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Studies
- Variables
- Series
- Publications
- ICPSR Website (news, announcements, etc.)
# Results from Keyword Search

<table>
<thead>
<tr>
<th>Studies (2,303)</th>
<th>Variables (3,861)</th>
<th>Series (132)</th>
<th><strong>Publications (198)</strong></th>
<th>ICPSR Website (6)</th>
</tr>
</thead>
</table>

**Per page:** 50  
**Sort by:** Title A-Z

<table>
<thead>
<tr>
<th>Pub. Type</th>
<th>Pub. Year</th>
<th>Citation</th>
</tr>
</thead>
</table>
**Full Text Options:** DOI, WorldCat, Google Scholar  
**Export Options:** RIS, EndNote |

|           | 2011      | Lindsay, Brianna, Albrecht, Jennifer, Terplan, Mishka *Against professional advice: Treatment attrition among pregnant methamphetamine users. Substance Abuse and Rehabilitation*, 2, (1), 189-195.  
**Full Text Options:** DOI, WorldCat, Google Scholar  
**Export Options:** RIS, EndNote |

**Full Text Options:** WorldCat, Google Scholar  
**Export Options:** RIS, EndNote |

**Full Text Options:** DOI, WorldCat, Google Scholar  
**Export Options:** RIS, EndNote |

**Export Options:** RIS, EndNote |

**Export Options:** RIS, EndNote |

- Studies
- Variables
- Series
- Publications
- ICPSR Website (news, announcements, etc.)
## Results from Keyword Search

<table>
<thead>
<tr>
<th>Page Link</th>
<th>Last Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Featured Data: Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)</strong> &lt;br&gt;<a href="https://www.icpsr.umich.edu/icpsrweb/ICPSR/news/1079">https://www.icpsr.umich.edu/icpsrweb/ICPSR/news/1079</a> &lt;br&gt;The Boston Rehabilitative Impairment Study of the Elderly (Boston RISE) was a prospective cohort study of older primary care patients aged 65 years and older, who were at risk for declining mobility and disability at baseline. The study was designed to investigate which combinations of neuromuscular impairments are most responsible for mobility decline and disability over 2 years of follow up.</td>
<td>2018-08-06</td>
</tr>
<tr>
<td>2. <strong>Featured Data: Impact of WHO's 8-step Wheelchair Service Provision in Indonesia, 2013-2014 (ICPSR 37093)</strong> &lt;br&gt;<a href="https://www.icpsr.umich.edu/icpsrweb/ICPSR/news/1130">https://www.icpsr.umich.edu/icpsrweb/ICPSR/news/1130</a> &lt;br&gt;The Impact of the World Health Organization's (WHO) 8-step Wheelchair Service Provision in Indonesia, 2013-2014 includes data collected in Indonesia by United Cerebral Palsy (UCP) Wheels for Humanity and the University of Pittsburgh. The purpose of this study was to investigate how wheelchairs provided to individuals with mobility impairments related to mobility, participation in society, quality of life, wheelchair skills, wheelchair maintenance, and satisfaction with mobility as compared to a control group.</td>
<td>2018-08-20</td>
</tr>
<tr>
<td>3. <strong>OR Meeting: Program and Workshops</strong> &lt;br&gt;<a href="https://www.icpsr.umich.edu/icpsrweb/content/membership/or/ormeet/workshops.html">https://www.icpsr.umich.edu/icpsrweb/content/membership/or/ormeet/workshops.html</a></td>
<td>2017 Biennial ICPSR Meeting</td>
</tr>
<tr>
<td>4. <strong>Thematic Data Collections</strong> &lt;br&gt;<a href="https://www.icpsr.umich.edu/icpsrweb/content/about/thematic-collections.html">https://www.icpsr.umich.edu/icpsrweb/content/about/thematic-collections.html</a></td>
<td></td>
</tr>
</tbody>
</table>
Filter Search Results

<table>
<thead>
<tr>
<th>Study Title/Investigator</th>
<th>Released/Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Rehabilitation Measures Database (ICPSR 36587)</td>
<td>2016-11-28</td>
</tr>
<tr>
<td>Center for Rehabilitation Outcomes Research (CROIR), Rehabilitation Institute of Chicago Department of Medical Social Sciences Informatics, Northwestern University Feinberg School of Medicine (NU MSS)</td>
<td></td>
</tr>
<tr>
<td>2. Public Support for Rehabilitation in Ohio, 1996 (ICPSR 2543)</td>
<td>2006-03-30</td>
</tr>
<tr>
<td>Applegate, Brandon K.</td>
<td></td>
</tr>
<tr>
<td>Bean, Jonathan</td>
<td></td>
</tr>
<tr>
<td>Whitener, Gale G.</td>
<td></td>
</tr>
<tr>
<td>Pullen, Suzanne; English, Kim</td>
<td></td>
</tr>
<tr>
<td>Pearson, Frank S.</td>
<td></td>
</tr>
<tr>
<td>7. Functional Independence in Children at a Pediatric Clinic in Guanajuato, Mexico, 2004-2013 (ICPSR 37068)</td>
<td>2018-07-09</td>
</tr>
<tr>
<td>Buicic, Samuel; Toro Hernandez, Maria Luisa; Pearlman, Jon; Deasy, David J.</td>
<td></td>
</tr>
<tr>
<td>Marsden, Mary Ellen; Orsagh, Thomas</td>
<td></td>
</tr>
<tr>
<td>Odii, Glenn; Oobido, Kenneth; Kuo, Yong Fang</td>
<td></td>
</tr>
<tr>
<td>10. The Traumatic Brain Injury Model Systems National Data and Statistical Center (ICPSR 36589)</td>
<td>2016-11-29</td>
</tr>
<tr>
<td>Traumatic Brain Injury Model Systems National Data and Statistical Center; Craig Hospital Research Department</td>
<td></td>
</tr>
<tr>
<td>National Data and Statistical Center for the Burn Model System</td>
<td></td>
</tr>
</tbody>
</table>
Variable Search
Compare Variables
## Compare Variables

<table>
<thead>
<tr>
<th>NAME</th>
<th>LABEL</th>
<th>QUESTION</th>
<th>RESPONSES</th>
<th>STUDY</th>
<th>TIME PERIOD</th>
<th>UNIVERSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HU0888C</td>
<td>Rehabilitation facility</td>
<td>Target: B(1 Yrs. to 150 Yrs.) English Text: What was the type of facility? English instructions: CODE ALL THAT APPLY</td>
<td>Value</td>
<td>Label</td>
<td>Unweighted Frequency</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>Rehabilitation facility</td>
<td>66</td>
<td>0.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Missing Values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>Total</td>
<td>10056</td>
<td>99.3%</td>
</tr>
<tr>
<td>HU0888C</td>
<td>Rehabilitation facility</td>
<td>Target: B(1 Yrs. to 150 Yrs.) English Text: What was the type of facility? English instructions: CODE ALL THAT APPLY</td>
<td>Value</td>
<td>Label</td>
<td>Unweighted Frequency</td>
<td>%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>Rehabilitation facility</td>
<td>63</td>
<td>0.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Missing Values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td>Total</td>
<td>10976</td>
<td>99.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Every Study has a Homepage

Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)

Published: May 31, 2018  Cite this Study | Share this study

Principal Investigator(s):  
Jonathan Bean, New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S.); Harvard Medical School Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

https://doi.org/10.3886/IACPSR37045.v1

Version V1

Project Description

Summary

The Boston Rehabilitative Impairment Study of the Elderly (Boston RISE) was a prospective cohort study of older primary care patients, aged 65 years and older, who were at risk for declining mobility and disability at baseline. The study was designed to investigate which combinations of neuromuscular impairments are most responsible for mobility decline and disability over 2 years of follow up.
Study Homepage Information

• Study title, PI, date published, doi

• Project description
  • Summary, citation, funding source, subject terms, geographic coverage, access information

• Scope of Project
  • Time period(s), date of collection, data collection notes, etc.

• Methodology
  • Study purpose, design, sample, units of observation, data type, mode of data collection, response rate, etc.

• Version

• Analysis Information
Data & Documentation

Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)

Published: May 31, 2018  
Principal Investigator(s):  
Jonathan Bean. New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S.); Harvard Medical School. Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

https://doi.org/10.3886/ICPSR37045.v1

Version V1

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Preview</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS0 Study-Level Files</td>
<td>3 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS1 Main Data File</td>
<td>13 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS2 Falls and Rehabilitation Follow-Up</td>
<td>4 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS3 Medications</td>
<td>11 MB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DS4 Tracking</td>
<td>3 MB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.
- **One or more files in this data collection have special restrictions.** Restricted data files are not available for direct download from the website; click on the Restricted Data button to learn more.
- The citation of this study may have changed due to the new version control system that has been implemented.
# Variables

**Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)**

Published: May 31, 2018  
Cite this Study | Share this study

Principal Investigator(s):  
Jonathan Bean, New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S); Harvard Medical School, Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

https://doi.org/10.3886/ICPSR37045.v1

Version V1

<table>
<thead>
<tr>
<th>Name</th>
<th>Label/Question Text</th>
<th>Type</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>Activities Specific Balance Confidence Scale (ABC) total score</td>
<td>numeric</td>
<td>DS1</td>
</tr>
<tr>
<td></td>
<td>Taken from: Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>Age</td>
<td>numeric</td>
<td>DS1</td>
</tr>
<tr>
<td></td>
<td>Taken from: Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHR</td>
<td>Are you currently being treated by a doctor for: Abnormal heart beat</td>
<td>numeric</td>
<td>DS1</td>
</tr>
<tr>
<td></td>
<td>Taken from: Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes**

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.
- One or more files in this data collection have special restrictions. Restricted data files are not available for direct download from the website; click on the Restricted Data button to learn more.
- The citation of this study may have changed due to the new version control system that has been implemented.

This study is maintained and distributed by Advancing Research on Disability.
Publications

Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)

Published: May 31, 2018
Principal Investigator(s): Jonathan Bean, New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S.); Harvard Medical School. Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

https://doi.org/10.3886/ICPSR37045.v1

Version V1

Notes

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.

- One or more files in this data collection have special restrictions. Restricted data files are not available for direct download from the website; click on the Restricted Data button to learn more.

- The citation of this study may have changed due to the new version control system that has been implemented.

This study is maintained and distributed by Advancing Research on Disability.
Accessing Data

• Access to all ADDEP studies is free
• However, free does not mean openly accessible
• If no disclosure risk, files are public-use
  • Users can access without restriction beyond agreeing to basic ICPSR Terms of Use
• If data contain sensitive information, access is restricted
  • Users must request access and meet specified requirements (researcher credentials, IRB approval, data security plan, etc.)
• Study documentation (e.g., codebook) is public-use even if data files are restricted
Accessing Data

Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)

Published: May 31, 2018

Principal Investigator(s): Jonathan Bean, New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S.); Harvard Medical School. Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

https://doi.org/10.3886/ICPSR37045.v1

Version V1

Download | Analyze Online | Access Restricted Data

At A Glance | Data & Documentation | Variables | Publications | Export Metadata | Staff

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Preview</th>
<th>Download</th>
</tr>
</thead>
<tbody>
<tr>
<td>D50 Study-Level Files</td>
<td>3 MB</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>D51 Main Data File</td>
<td>13 MB</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>D52 Falls and Rehabilitation Follow-Up</td>
<td>4 MB</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>D53 Medications</td>
<td>11 MB</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
<tr>
<td>D54 Tracking</td>
<td>3 MB</td>
<td>![Icon]</td>
<td>![Icon]</td>
</tr>
</tbody>
</table>

Documentation [Excel 2007 spreadsheet] | PI Codebook

Questionnaire [PDF]

Notes

- The public-use data files in this collection are available for access by the general public. Access requires an institutional subscription or individual citation of the ICPSR member institution.
- One or more files in this data collection have special restrictions. Restricted data files are not available for direct download from the website; click on the Restricted Data button to learn more.
- The citation of this study may have changed due to the new version control system that has been implemented.

This study is maintained and distributed by Advancing Research on Disability.
Documentation (Codebook)

ICPSR 37045
Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014

Jonathan Bean
New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S.); Harvard Medical School. Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

Codebook for Dataset 1: Main Data File

<table>
<thead>
<tr>
<th>Race</th>
<th>Unweighted</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>225</td>
<td>17.4%</td>
</tr>
<tr>
<td>1</td>
<td>1025</td>
<td>82.6%</td>
</tr>
<tr>
<td>Total</td>
<td>1250</td>
<td>100%</td>
</tr>
</tbody>
</table>

Based upon 1,250 valid cases out of 1,250 total cases.
* Minimum: 0.00
* Maximum: 1.00
Location: 022-002 (width: 1, decimal: 0)
Variable Type: numeric

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>Unweighted</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Less than 12th grade</td>
<td>119</td>
<td>2.2%</td>
</tr>
<tr>
<td>1</td>
<td>12th grade or General Educational Development (GED)</td>
<td>269</td>
<td>22.4%</td>
</tr>
<tr>
<td>2</td>
<td>Any College or Vocational or Technical School</td>
<td>370</td>
<td>37.8%</td>
</tr>
<tr>
<td>3</td>
<td>Graduate or Professional School</td>
<td>203</td>
<td>21.6%</td>
</tr>
<tr>
<td>-9</td>
<td>Missing Data</td>
<td>240</td>
<td>10.5%</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1250</td>
<td>100%</td>
</tr>
</tbody>
</table>

Based upon 1,060 valid cases out of 1,220 total cases.
* Minimum: 0.00
* Maximum: 3.00
Location: 414-2 (width: 2, decimal: 0)
Variable Type: numeric
Range of Missing Values: -9

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>Unweighted</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>MARITAL: What is your marital status?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

Boston Rehabilitative Impairment Study of the Elderly (Boston RISE), 2009-2014 (ICPSR 37045)

Published: May 31, 2018

Principal Investigator(s):
Jonathan Bean, New England Geriatric Research Education and Clinical Center (GRECC), VA Boston Healthcare System (U.S.); Harvard Medical School. Department of Physical Medicine and Rehabilitation; Spaulding Rehabilitation Hospital Cambridge, MA

https://doi.org/10.3886/ICPSR37045.v1

Version V1

- **DS0 Study-Level Files**: 3 MB
- **DS1 Main Data File**: 13 MB
- **DS2 Falls and Rehabilitation Follow-Up**: 4 MB
- **DS3 Medications**: 11 MB
- **DS4 Tracking**: 3 MB

**Notes**

- The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution.

- **One or more files in this data collection have special restrictions.** Restricted data files are not available for direct download from the website; click on the Restricted Data button to learn more.

- The citation of this study may have changed due to the new version control system that has been implemented.
MyData Account

Log into ICPSR

Use MyData

Email:

Password:

Log In

Request a new password

You must set your browser to accept cookies in order for login to succeed. Your login session will last for nine hours (or until you log out).

New User?

By creating a MyData account, you can:

- Download data and use online analysis tools
- Download data available only to ICPSR members, if you are from a member institution
- Generate utilization reports and download statistics
- Update your account profile and settings
- Access your past searches, downloads, and deposits (forthcoming)
- Access Summer Program registration and history of courses taken (forthcoming)

Or one of these:

Google
LinkedIn
Facebook
ORCID
Restricted Data Application

My Application  Request #29909

To be able to complete the application you must complete multiple sections which are listed below. Please feel free to complete the sections in any order. Once you complete a section, the system will mark that section as complete. When your application is complete, click the Submit to ICPSR button.

- Investigator Information  (Incomplete)
- Research Staff Information  (Incomplete)
- Research Description  (Incomplete)
- Data Selection  (Incomplete)
- Data Format  (Incomplete)
- Confidential Data Security Plan  (Incomplete)
- IRB Review Approval  (Incomplete)
- Final Signatures  (Incomplete)

Submit to ICPSR  Please complete the above sections before submitting your application for approval.
Questions?

Contact:
Shane Redman
icpsr-addep@umich.edu
Preparing Data for Archiving

Shane Redman
icpsr-addey@umich.edu
Overview

• How to Prepare Data & Documentation for Sharing
  • Variable-Level
  • Derivative/Constructed Data Sets
  • File-Level
  • Study-Level

• How Restricted/Sensitive Data are Handled
  • Disclosure Risk Review
  • Personally Identifiable Information
  • How to De-Identify your Data
  • ICPSR’s Approach
Why is this important?

• A well-prepared data collection “contains information intended to be complete and self-explanatory” for future users.
  • “How should I set up the data so secondary users can independently understand the data collection?”
• How you prepare your data can affect how others will use it.
What Should My Deposit Include?

- Data
- Documentation
- Descriptions (metadata)
Structuring Data for Sharing

• How will your data be organized?
• Think of data at three levels:
  • Variable level
  • File level
  • Study level
Variable-Level Structure

• Variable Naming Conventions
• Variable Labels
• Value Labels
• Missing Data
• Documentation
# Variable-Level Details

<table>
<thead>
<tr>
<th>Var Name</th>
<th>Var Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>V20</td>
<td>Q10B WERE PRI FELONY CONV FOR HOMICIDE?</td>
</tr>
</tbody>
</table>

**Var Type:** numeric

**Location:** 41-42 (width: 2; decimal: 0)

**Range of Valid Data Values:** 1 thru 9

**Range of Missing Values (M):** 99

**Question:** If Q. 10a coded 1, Q. 10b. Were any of these convictions for criminal homicide?

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Murder</td>
</tr>
<tr>
<td>2</td>
<td>Invol manslauter</td>
</tr>
<tr>
<td>3</td>
<td>Vehic manslauter</td>
</tr>
<tr>
<td>4</td>
<td>Attempted murder</td>
</tr>
<tr>
<td>5</td>
<td>Yes - Other</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>Unknown (fr Qst)</td>
</tr>
<tr>
<td>8</td>
<td>Manslaughter</td>
</tr>
<tr>
<td>9</td>
<td>Vol manslauter</td>
</tr>
<tr>
<td>99 (M)</td>
<td>Unknown/NA</td>
</tr>
</tbody>
</table>
Variable Naming Conventions

- One-up numbers (V1, V2)
- Question numbers (Q1, Q2)
- Mnemonic names (age, race)
- Prefix, root, suffix systems (FAED, MOED)
Variable Labels

• Item/Question number
• Variable description
• Indicate if variable was constructed

• Example 1 (original variable):
  • Rhealth: Q14 Self-Assessment of Respondent’s Health

• Example 2 (constructed variable):
  • Rhealth_C: Index of Respondent’s Health (constructed)
Value Labels

- Mutually exclusive, exhaustive, and defined
- **R_Employ**: Respondent’s Employment Status
  - Unemployed (1)
  - Self-employed (2)
  - Employed, not by self (3)
- Are there weight variables? Are they documented?
Missing Data

• Are there missing data?
• Are missing data labeled?

• 77 = Inapplicable
• 88 = Don’t Know
• 99 = Refused to Answer
Documentation

• Files that explain data collection
• Materials to independently understand data
• Examples:
  • Codebooks, Data Dictionary, User Guide
  • Data Collection Instruments/Questionnaires, Protocols
• Format: MS Word, PDF, ASCII, DDI XML
Derivative/Constructed Data

• Sharing of data not collected by you depends on level of access to original data set

• If original data set cannot be shared, you can share derived/constructed data
  • Statistical code and documentation that describes how data can be accessed and how new data set was constructed can be shared
  • In other words, give the “recipe” and instructions to secondary users
During the process, ask yourself…

• Is the data collection complete, accurate, and well-documented?
  • Do the data match the documentation? Are values and/or labels listed in one but not in the other?
  • Are all codes in the data valid (documented) according to the data collection instrument or PI's codebook?
  • Are there duplicate records?
  • Does the spelling look okay?
## Data Files

### Table 1: Sample Data

<table>
<thead>
<tr>
<th>Count</th>
<th>Card</th>
<th>Height</th>
<th>Gender</th>
<th>Age</th>
<th>Eye color</th>
<th>Hand</th>
<th>Siblings</th>
<th>Coffees</th>
<th>TV time</th>
<th>Die 1</th>
<th>Die 2</th>
<th>Satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>177.8</td>
<td>male</td>
<td>65</td>
<td>green</td>
<td>right</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>170.2</td>
<td>female</td>
<td>65</td>
<td>blue</td>
<td>left</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>185.4</td>
<td>male</td>
<td>65</td>
<td>brown</td>
<td>right</td>
<td>1</td>
<td>3</td>
<td>1.0</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>185.4</td>
<td>male</td>
<td>65</td>
<td>brown</td>
<td>right</td>
<td>2</td>
<td>1</td>
<td>0.0</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

### Additional comments/questions for Q8: For each of the following statements, please rate...

<table>
<thead>
<tr>
<th>ID</th>
<th>Org Structure</th>
<th>Focus of Org</th>
<th>Q10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Community-based org</td>
<td>Regional</td>
<td>The state of was on a strong track of championing, implementing, and building health and wellness initiatives. The political will, local support, and financial for these efforts has reduced drastically over the past 3 years. We are doing the best we can in a “do more with less” atmosphere and optimistic the tide will change in our favor.</td>
</tr>
<tr>
<td>2</td>
<td>School</td>
<td>Regional</td>
<td>Our community is in the process of developing many policies and procedures to make healthy, local food and gardens a part of every school. We have a tremendous amount of community support for their work.</td>
</tr>
<tr>
<td>3</td>
<td>Community-based org</td>
<td>County</td>
<td>Being rural and getting schools to step up are two of our biggest challenges. “Old agriculture” is strong here, but we need to get the holders of power and influence to link our agriculture with healthy food.</td>
</tr>
<tr>
<td>4</td>
<td>Community-based org</td>
<td>Local school/district</td>
<td>There are pockets of support for healthy food but most of the community needs a great deal of education on this topic.</td>
</tr>
<tr>
<td>5</td>
<td>College</td>
<td>Statewide</td>
<td>It is a case by case basis when working with schools. Also, working statewide, schools are only one part and each regionally has its own efforts. There is not a lot of momentum in connecting efforts. No true convergence.</td>
</tr>
</tbody>
</table>
Common Statistical Data File Formats

• SPSS, SAS, Stata
• ASCII + setups
• Access, Excel
Non-Statistical File Formats

Text
- Text
- Rich Text
- PDF (+OCR)
- MS Word

Image
- TIFF
- JPEG2000
- PNG
- JPEG/JFIF
- GIF

Audio
- MPEG
- WAV
- AIFF

Video
- JPEG2000
- MOV
- AVI
File-Level Organization

• Possible file structures:
  • One file with all data in same location
  • One file with multiple sheets
  • Several files in different folders/subfolders

• If the structure of your data files is complex, consider creating a user guide that explains the organization of files
Importance of File Organization for Non-Statistical Data

• File naming to understand how files are related - good to use descriptive information
  • e.g., female_1, female_2
• Consider using folders and sub-folders to organize files and relationships
• Data listing is a document that helps users identify a subset of files with relevant characteristics (e.g., male participants at first interview who were age 45 and older)
Study-Level Metadata

• Study description that explains the data, collection methods, sample, etc.
• Focus on data collection efforts
• Data collection summary is not simply a publication abstract
## Study-Level Metadata

<table>
<thead>
<tr>
<th>Who</th>
<th>PIs, Participants, Funders</th>
</tr>
</thead>
<tbody>
<tr>
<td>What</td>
<td>Summary/Project Description</td>
</tr>
<tr>
<td>Where</td>
<td>Geography</td>
</tr>
<tr>
<td>When</td>
<td>Timeframe</td>
</tr>
<tr>
<td>Why</td>
<td>Purpose, Goals, Aims</td>
</tr>
<tr>
<td>How</td>
<td>Study Design/Methodology, Sampling, Weighting</td>
</tr>
</tbody>
</table>
Identifying and Protecting Confidential Data
Data Confidentiality

• Starts prior to data collection with informed consent
  • What to include in informed consent document:
    • Assurance of confidentiality, including as much about process of data handling as possible -- e.g., all names and other identifying information will be stripped when presenting results or sharing data with the scientific community
    • Don’t promise such a tight process that you will not be able to share data with sponsors (who care about respondent safety) or other researchers
Example Language

“The information in this study will only be used in ways that will not reveal who you are. You will not be identified in any publication from this study or in any data files shared with other researchers. Your participation in this study is confidential. Federal or state laws may require us to show information to university or government officials [or sponsors], who are responsible for monitoring the safety of this study.”

Source: Pienta and Marz example at Qualitative Data Repository
Disclosure Risk

• The potential identification of an individual based on information in the data record
  • Detailed geography
  • Exact date of birth or other exact dates
  • Exact occupations held
  • Knowing an individual participated and using online data to “find” that person
  • Combination of variables
Disclosure vs Sensitivity

Data with Disclosure Risk (potential to identify)

- Detailed geography (i.e., state, county, or census tract of residence)
- Exact date of birth
- Exact occupations held
- Exact dates of events
- Detailed income
- Combinations of variables

Data with Sensitive Personal Information

- Health information
- Drug use
- Previous or current illegal behavior
# Defining Sensitive Data, Sensitively

## Potentially Sensitive
- Psychological well-being/mental health
- Use of alcohol or drugs
- Illegal behavior
- Behavior damaging to individual’s financial standing, employability, or reputation
- Medical information that could lead to discrimination, stigmatization

## Probably Not Sensitive
- Consumer product usage
- Type of car you drive
- Past voting behavior
- Political party affiliation
- Your ‘real’ age
- Social media rants
- Data records where respondents are long deceased
Key Concern

• Allows information about the respondent to be revealed that would not otherwise be known
Direct Identifiers

• Point directly to the individual
  • Names
  • Photos, videos
  • Addresses or phone numbers
  • Social Security Numbers
  • Driver’s license numbers, employee ID numbers, or other unique numbers
Indirect Identifiers

• Used with other information to identify individual respondents
  • Geographic information (from state-level down, including zip and area codes)
  • Employment or social organizations
  • Education histories with school/year
  • Detailed income
Do you have identifiable data?

• Prior to sharing data, ensure the data set is de-identified
• Identifying variables may be specific based on whether your data are regulated by HIPAA, FERPA, or other law
  • Example: HIPAA 18 identifiers that must be removed
    • Name, telephone number, SSN, IP address, exact age if over 89, etc.
How to De-Identify your Data

- Variables that include PII should be removed/redacted or recoded
  - Remove all names, SSNs, or other direct identifiers completely
  - Recode age variable so that all respondents 89 and older are a group
- Images and video typically cannot be de-identified without affecting use
  - Share de-identified transcripts
ICPSR Approach

- Data processors must be certified through disclosure risk training (supervisors receive extra training)
- Evaluation of deposited data using decision tree
  - Defined steps and protocols depending on level of disclosure risk
Data Dissemination

• May require recoding/collapsing or perturbing
  • Pay attention to sub-populations
  • Changing exact dates to years only, grouping income into categories, top coding, swapping cases

• Some projects require specific methods – e.g., Safe Harbor method of de-identification

• But sometimes the data lose too much utility when all specificity is removed, so restricted-use versions created
Sharing Data with Disclosure Risk or Sensitivity Issues

• Data sharing via:
  – Secure Download
  – Virtual Data Enclave
  – Physical Enclave

• Users of restricted-use data committed to applications and approvals for data access, data security measures, and communication

• Repositories equipped with technology, legal counsel, and staff
Additional Resources

- ICPSR, “What is a codebook?”
- Amsterdam Public Health Research Institute Quality Handbook
- Princeton University Data and Statistical Services, “How to Use a Codebook”
- UCLA Social Science Data Archive, “About Codebooks”
- UK Data Archive, “Data-level Documentation”
- ICPSR Guide to Social Science Data Preparation and Archiving: Phase 3: Data Collection and File Creation
Additional Resources

- American Statistical Assoc. Data Access and Personal Privacy: Appropriate Methods of Disclosure Control
- Census Bureau’s Approach to Statistical Disclosure Control
- Disclosure Limitation and Confidentiality Protection in Linked Data (US Census)
- Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPPA) Privacy Rule (HHS.gov)
- ICPSR’s Approach to Confidentiality
Transferring Data to a Repository

Shane Redman
icpsr-addep@umich.edu
Overview

• Choosing a home for your data
• ICPSR deposit options
• Deposit process at ICPSR
• What happens after deposit?
Choosing a Destination

• Several options when selecting home for data
Options

• Domain (subject/disciplinary) repositories
• Institutional repositories
• Personal Website
• FigShare, Zenodo, etc.
Things to Consider

• Where do people look for data like yours?
• Does the repository have a history (long-term preservation)?
• What information will others need to reuse your data and will that be made available? What about user support?
• Requirement of funder, university, or journal?
• Will the repository help you show impact of your work?
Elements of a High-Quality Repository

• FAIR Data Principles
  • Findable, Accessible, Interoperable, and Reusable
• Open Archival Information System (OAIS) Reference Model
• CoreTrustSeal Certification (formerly Data Seal of Approval)
  • 16 Requirements
Options Available at ICPSR

• ICPSR/ADDEP repository
  • Curated collection means people can find and use data easily
  • Collection of similar topics attracts relevant researchers
  • Ability to handle data with disclosure risk

• openICPSR for self-deposit
  • Great for replication datasets
  • Quick DOI
  • Can pay for curation, then data freely available at ICPSR or another topical archive
Preparing Data for Deposit

• Should start data collection with sharing in mind
  • Document everything (questionnaire, interviewer instructions, sampling strategy, etc.)

• Ensure that your data files have:
  • Meaningful labels (variable and value)
  • Missing data codes
  • No direct identifiers (or work with repository to handle this)
  • For data with confidentiality issues, ICPSR staff can help with creating a file for restricted-use deposit

• Start by reviewing ICPSR’s Guide to Social Science Data Preparation and Archiving:
  http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/
How to Deposit Data with ICPSR

• Data submitted to ICPSR through online data deposit manager
  • [http://www.icpsr.umich.edu](http://www.icpsr.umich.edu)
  • Follow links to start sharing data/deposit data

• Create MyData account
• Follow online steps to begin new deposit
Deposit Form

• Save and return at any time
• Shareable with colleagues or ICPSR staff
• Have prepared:
  • Final version of each data file
  • Codebook
  • Information about study (summary, sample, etc.)
  • Copy of data collection instrument(s)
  • Any other helpful documentation
• Distinct data sets or studies should be deposited on separate deposit forms
Deposit Agreement

• Ensure you have rights to make data publicly available
  • IRB approval
• Grant permission to ICPSR to:
  • Share data
  • Promote/advertise data
  • Preserve data
• Confirm data are de-identified
• Release date (embargo period)
What Happens After Deposit?

• Data transferred to secure server where they stay throughout curation
• Checking documentation, variables, values (quality control)
• Questions back to PI as needed
• Disclosure risk review
  • Dissemination options: public-use, encrypted download, Virtual Data Enclave (VDE), physical enclave
• File formats for major statistical packages and preservation copy created
• Collect citations
Data Citation

• Each study receives DOI
• Secondary users must agree to terms of use
• Terms include agreement to reference data collection citation in any publication
  • Also required to send citations of published work to be included in bibliography
Tracking Usage

• Each study homepage contains usage report

# of Views
# of Downloads
# of Related Publications
Deposit Data
Our online deposit form provides a secure upload of files and establishes the terms and conditions of the data transfer.

Start Deposit

Instructions for Depositing Data
To complete a data deposit, click on the “Start Deposit” button above. Please make sure that “ADDEP” appears in the “Archive” drop-down menu so that your data is archived with us.

Most fields are self-explanatory, but contact us at icpsr-addep@umich.edu if you have any questions.

Information about how ICPSR prepares data for public release is available.

The following files are necessary for deposit:

1. Final version of each dataset generated during the project, including scale or other derived variables created for published analyses
2. Codebook listing the variable names, variable labels, value labels, and missing value designations (an SPSS dictionary with these elements can suffice)

The following files are very helpful to the archiving work of ADDEP:
Make Sure to Choose “ADDEP”
• File-level management – can document the study/project and each of the files therein
• Can upload most file types
• Automatic versioning if edits needed
Dashboard

• Each user who creates a deposit will have a dashboard where they can find the completed or in progress deposits

• Do not need to complete deposit in one session
Questions?

Shane Redman
icpsr-addep@umich.edu
www.icpsr.umich.edu/ADDEP