SECONDARY ANALYSIS OF EXISTING DATASETS

Background

In many areas of the social sciences, one of the most common methods of research is secondary analysis of publicly available files of data collected from surveys. One of the main sources of such data is the federal government (e.g., Bureau of Labor Statistics). Increasingly, federal agencies require that social science investigators make data collected with assistance of federal funds publicly available. In addition, social science professional organizations and journals are establishing the norm that investigators make data sets used for the production of published scholarly papers accessible in public use data files to encourage scholarly replication of research.

In response, the University of Utah Institutional Review Board (IRB) has adopted a policy developed by the University of Wisconsin-Madison that notes research projects involving analysis of secondary data from pre-approved datasets/repositories will NOT require prior IRB approval.

The University of Utah Human Research Protection Program (HRPP) wishes to express our sincere gratitude to officials of the University of Wisconsin-Madison for allowing us to adopt and disseminate this policy.

Copyright © 2008 University of Wisconsin System Board of Regents

Description

Research projects involving analysis of secondary data from the following datasets/repositories will NOT require prior IRB approval:

- Better Access to Data for Global Interdisciplinary Research (BADGIR)
- Bureau of Labor Statistics
- Center for Research on Child Wellbeing
- Inter-University Consortium for Political and Social Research (ICPSR)
- Luxembourg Income Study (LIS)
- Measure Demographic Health Surveys (DHS)
- Medical Expenditure Panel Survey (MEPS)
- National Center for Health Statistics
- National Center for Education Statistics
- National Election Studies
- National Opinion Research Center
- National Center for Statistics and Analysis (NCSA), of the National Highway Traffic Safety Administration
- Panel Study of Income Dynamics (PSID)
- Roper Center for Public Opinion Research
- Social Science Computing Cooperative
- Surveillance, Epidemiology and End Results Program (SEER) Research Data
- University of Wisconsin-Madison Data and Information Services Center (DISC)
- U.S. Bureau of the Census

The University of Utah IRB has determined that data in the above listed datasets has been stripped of identifiers and is publicly available. As a result, research using this data does not rise to the level of "human subjects research," under the federal Common Rule, 45 CFR Part 46 and University policies, and, therefore, does not require IRB review. See, *Definition of Research Involving Human Subjects Policy.*

Note: Research projects that merge more than one dataset in such a way that individuals may be identified are not covered by this policy, and require prior IRB approval. Some of the databases listed above have restricted-use components that include indicia, and use of these components must be approved by the University of Utah IRB before the research begins.

Submitting a Dataset for Pre-approval

Datasets that may qualify for inclusion on the University of Utah list of approved data sources include:

1. Public use datasets posted on the Internet that include a responsible use statement or other confidentiality agreement for authors to protect human subjects (for an example, see the ICPSR's responsible use statement)
2. Survey data distributed by University of Utah principal investigators who can certify that:
    a. the data collection procedures were approved by an IRB that satisfies the Common Rule criteria for an IRB, and
    b. the dataset and documentation as distributed do not contain information that could be used to identify individual research participants.

Investigators must submit the following information on potentially eligible datasets to the University of Utah IRB office prior to conducting research:

1. Name of dataset.
2. URL of the dataset or other information on how to obtain the dataset.
3. Abstract (one page maximum) describing the content of the dataset and its potential use.