



Section 14: Research Projects Using Secondary Data

14.0 Research Studies Using Secondary (Existing) Data Secondary

The term *Secondary (or Existing) Data* is defined as being data collected previously to the research project being proposed. The data may be in the form of data sets, but may also be in the form of interview notes, photographic images, or audio- and/or video-recordings. Data are said to be identifiable if they include direct or indirect information that connects personal information to each participant. Researchers often analyze data that they did not collect. Secondary data may have been collected for research purposes or non-research purposes. Sometimes, data providers require researchers to enter into a data-use agreement or provide other assurances, such as a confidentiality pledge, before they give data to a researcher. Heartland IRB realizes that some research involving existing data sets and archives may not meet the definition of research involving human subjects, and, therefore would not require IRB review. However, some may meet definitions of research and may require IRB review and approval.

14.1 Analysis of De-Identified Data that are Publicly Available

Heartland IRB recognizes that the analysis of de-identified, publicly available data does not constitute human subjects research as defined by 45 CFR 46.102, and that it does not require Institutional Review Board review. Some studies utilize data made available through large data consolidation bureaus and consortiums. To reduce burdens on investigators, Heartland IRB has listed below data holders whose archives include only publicly available, de-identified data. Heartland IRB does not require review of studies involving the analysis of data held by these Organizations, unless a project merges multiple data sets, and in so doing, enables the identification of individuals whose data are analyzed.

- Inter-University Consortium for Political and Social Research (ICPSR)
- U.S. Bureau of the Census
- National Center for Health Statistics
- National Center for Education Statistics
- National Election Studies

14.2 Analysis of Non-Publicly Available Data with Restricted Access to Participant Identifiers (Coded Private Information)

Heartland IRB recognizes that increasing numbers of studies involve the analysis of non-publicly available datasets that include coded private information or that are provided to researchers after the removal of identifying information. Further, a number of data holding organizations (e.g., Bureau of Labor Statistics, Centers for Disease Control, etc.) provide statistical analysis services for investigators that generate aggregated data from data sets with private identifying information. Many of these studies do not involve human subjects.

Data use permissions vary widely across data sets and holders. Because of this, Heartland IRB recommends that investigators conducting studies using coded private information or contracting data holders for statistical analyses of data sets involving private information seek a determination from an Institutional Review Board as to whether the study constitutes human subjects research. This recommendation is made in accordance with guidance from OHRP concerning research involving coded private information.

**14.3 Analysis of publicly available data with private identifiable information
OR Analysis of non-publicly available data with private identifiable information
where researchers will not record individual identifiers**

Research involving the analysis of publicly available data containing private identifiable information or the analysis of non-publicly available data that will not be recorded by the investigator in a manner that allows the direct or indirect identification of individuals qualifies for exempt review (45 CFR 46.101(b)(4)). Research involving these analyses requires the submission of Heartland IRB form: HIRB Waiver for HIPAA Authorization.

Investigators submitting protocols involving these research procedures are asked to provide the following information in their protocol submissions to aid Heartland IRB in making a determination of exemption:

- Description of data set and availability;
- Description of data to be accessed for analysis; and
- Copies of data use agreements required by data holder.

14.4 Analysis of Non-Publicly Available Data Containing Private Identifiable Information

Research involving the analysis of non-publicly available data that contains private identifiable information about living individuals is considered by Heartland IRB to constitute human subjects research that is not exempt from 45 CFR 46 and Heartland IRB review requirements.

Studies involving analysis of this form of data require review by an Institutional Review Board. This review is conducted under expedited or full board review procedures in accordance with the reviewing Institutional Review Board's policies. Investigators submitting protocols involving these research procedures are asked to include the following information in their protocol submissions to aid Heartland IRB in its review.

- Description of data set(s) to be analyzed;
- Copies of data use or security agreements required by data holder; and
- Description of data security and access procedures.

It is important to note that if data are provided to researchers without identifiers, the proposed research does not meet the definition of research with human subjects. If the data are identifiable there is only one protocol application form, regardless of the type of review that will be used.

14.5 Data Held by the Principal Investigator's employing Institution/Organization or by the Principal Investigator from an Earlier Expired Research Study

Research involving the analysis of non-publicly available data that contains private identifiable information about living individuals AND is in the custody of the Principal Investigator or the Principal Investigator's employing institution/organization under a prior, expired Institutional Review Board approval, must still undergo review by an Institutional Review Board.

When submitting an application for a research study using the aforementioned data, the Principal Investigator must include the following documentation with the Heartland IRB application form (*HIRB Secondary Data Protocol Review Form*):

- Approval letter from the Institutional Review Board that reviewed the data's project of origin AND the
- Participant Consent Form used with the original research study.

14.6 Federal Guidance Defining Identifiers

Identifiers include, but are not limited to names, addresses, phone number, social security numbers, geocodes, and images (but not voices).

Federal guidance addresses the question: "Are data 'identifiable' if they include identification codes that could link the data to individuals?" The answer is "yes," unless the researcher and the data provider have a formal agreement stating that the key linking codes with the individual identifiers will never be provided to the researcher.

Restrictions

- If data were collected for research purposes, the original agreement with the subjects must have included permission for identifiable data to be shared with other researchers.

Process

- Protocols for research analyzing existing data can be exempt if identifiable data are publicly available or if the researcher strips any identifiers upon receipt of the data.
- Protocols can be expedited if the only risk is a breach of confidentiality and adequate confidentiality procedures are in place.
- Protocols generally require full review only when it is a requirement of the data provider.

Data-Use Agreements

- A fully-executed data use agreement is one signed by the data supplier and by the Principal Investigator or by the organization's designated signatory.

Other Data Assurances

- Confidentiality pledges and data destruction agreements must also be included in the Data-Use Agreements and are signed by the Principal Investigator or by the organization's designated signatory.