

## Guidance on HRPP/IRB Review of Studies Involving Secondary Data Analysis

This document is intended to provide guidance on Federal and CUNY-wide IRB policies and procedures and to reduce burdens associated with HRPP/IRB review for investigators whose research involves only the analysis of existing data sets and archives. Some research projects involving existing data sets and archives may not meet the definition of “human subjects” research requiring HRPP/IRB review; some may meet definitions of research that is exempt from the federal regulations at 45 CFR part 46; and some may require HRPP/IRB review.

Although projects that only involve secondary data analysis do not involve interactions or interventions with humans, they may still require HRPP/IRB review because the definition of “human subject” at 45 CFR 46.102(f) includes *living individuals about whom an investigator obtains identifiable private information for research purposes*.

**Please note that this guidance is not definitive nor should be perceived as a substitute for appropriate HRPP consultation and HRPP/IRB review. If an investigator has questions about the HRPP/IRB requirements surrounding their proposed project, they should contact the CUNY Graduate School of Public Health and Health Policy (CUNY SPH) HRPP Office at [hrpp@sph.cuny.edu](mailto:hrpp@sph.cuny.edu).**

### 1. When does secondary use of existing data *not* require IRB review?

In general, the secondary analysis of existing data does not require HRPP/IRB review when it does not fall within the regulatory definition of research involving human subjects.

Investigators who are uncertain about whether HRPP/IRB review is required for their proposed project should complete the SPH HRPP Office’s [Human Subjects Research \(HSR\) Assessment Form](#) so the HRPP can advise the investigator on how to proceed.

#### A. Public Use Data Sets

Public use data sets are prepared with the intent of making them available for the public. The data available to the public are not individually identifiable and therefore analysis would not involve human subjects. The IRB recognizes that the analysis of de-identified, publicly available data does not constitute HSR as defined at 45 CFR 46.102 and that it does not require IRB review. The IRB no longer requires the registration or review of studies involving the analysis of public use data sets **unless** a project merges multiple data sets and in so doing enables the identification of individuals whose data is analyzed. An IRB review may be required for a research study that relies exclusively on secondary use of anonymous information BUT records data linkage or disseminates results in such a way that it generates identifiable information.

In addition to being identifiable, existing data must include “private information” in order to constitute research involving human subjects. Private information is defined as information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (e.g., a medical or school record). For example, a study involving only analysis of the published salaries and benefits of university presidents would not need IRB review since this information is not private.

## B. De-identified Data

If a dataset has been stripped of all identifying information and there is no way it could be linked back to the subjects from whom it was originally collected (through a key to a coding system or by other means), its subsequent use by the Principal Investigator or by another researcher would not constitute HSR, since the data is no longer identifiable. “Identifiable” means the identity of the subject is known or may be readily ascertained by the investigator or associated with the information. In general, information is considered to be identifiable when it can be linked to specific individuals by the researcher either directly or indirectly through coding systems, or when characteristics of the information obtained are such that a reasonably knowledgeable person could ascertain the identities of individuals. Even though a dataset has been stripped of direct identifiers (e.g., names, addresses, student ID numbers, etc.), it may still be possible to identify an individual through a combination of other characteristics (e.g., age, gender, ethnicity, place of employment).

**Example:** Many student research projects involve secondary analysis of data that belongs to, or was initially collected by, their faculty advisor or another investigator. If the student is provided with a de-identified, non-coded data set, the use of the data does not constitute research with human subjects because there is no interaction with any individual and no identifiable private information will be used.

**Coded data:** Secondary analysis of coded private information is not considered to be research involving human subjects and would not require IRB review **IF** the investigator(s) cannot readily ascertain the identity of the individuals to whom the coded private information pertains as a result of one of the following circumstances:

1. The investigators and the holder of the key have entered into an agreement prohibiting the release of the key to the investigators under any circumstances, until the individuals are deceased (HHS regulations for HSR do not require the IRB to review and approve this agreement);
2. There are IRB-approved written policies and operating procedures for a repository or data management center that prohibit the release of the key to the investigator under any circumstances, until the individuals are deceased; or
3. There are other legal requirements prohibiting the release of the key to the investigators, until the individuals are deceased.

For more information on when analysis of coded data is or is not HSR, see the HHS Office for Human Research Protections Guidance on Research Involving Coded Private Information or Biological Specimens at <http://www.hhs.gov/ohrp/policy/cdebiol.html>.

**Note:** If a student is analyzing *coded data* from a faculty advisor/sponsor who retains a key, this would be HSR, because the faculty advisor is considered an investigator on the student’s protocol, and can readily ascertain the identity of the subjects since he/she holds the key to the coded data. If the student’s work fits within the scope of the initial protocol from which the dataset originates, the faculty advisor (or investigator who holds the dataset) may wish to consider adding the student and his/her work to the original protocol by means of an amendment application rather than having the student submit a new application for review.

**Example:** Researcher B plans to examine the relationships between attention deficit hyperactivity disorder (ADHD), oppositional defiance disorder, and teen drug abuse using data collected by Agencies I, II, and III that work with “at risk” youth. The data will be coded and the agencies have entered into an

agreement prohibiting release of the key to the researcher that could connect the data with identifiers. The use of the data would not constitute research with human subjects.

If the HRPP determines that the project does not constitute HSR, the investigator will be notified. If the HRPP determines that the project does involve HSR, the investigator will be asked to submit a protocol for consideration by the IRB.

## 2. When is the secondary use of existing data exempt from IRB review?

There are six categories of research activities involving human subjects that may be exempt from the requirements of the federal regulations on HSR protections (45 CFR part 46). However, only one exemption category (Category 4) applies specifically to existing data. If research is found to be exempt, it need not receive full or expedited review.

A determination of exemption can only be made by the SPH HRPP Office following review of an IRB application in IRB Manager.

Research involving collection or study of existing data, documents, and records can be exempted under Category 4 of the federal regulations if: (i) the sources of such data are publicly available; or (ii) the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

The latter condition of this category applies in cases where the investigators initially have access to identifiable private information but abstract the data needed for the research in such a way that the information can no longer be connected to the identity of the subjects. This means that the abstracted data set does not include **direct identifiers** (names, social security numbers, addresses, phone numbers, etc.) or **indirect identifiers** (codes or pseudonyms that are linked to the subject's identity). Furthermore, it must not be possible to identify subjects by combining a number of characteristics (e.g., date of birth, gender, position, and place of employment). This is especially relevant in smaller datasets, where the population is confined to a limited subject pool.

The following do not qualify for exemption: Research involving prisoners, and FDA-regulated research.

**Example:** Student A will be given access to data from her faculty advisor's health survey research project. The data consists of coded survey responses, and the advisor will retain a key that would link the data to identifiers. The student will extract the information she needs for her project without including any identifying information and without retaining the code. The use of the data does constitute research with human subjects because the initial data set is identifiable (albeit through a coding system); however, it would qualify for exempt status.

## 3. When does the secondary use of existing data require expedited or full-board IRB review?

If secondary analysis of existing data does involve research with human subjects and does not qualify for exempt status as explained above, the project must be reviewed either through expedited procedures or by the full (convened) IRB, and an application must be submitted for HRPP/IRB review in IRB Manager.

**Consent:** Researchers using data previously collected under another study should consider whether the currently proposed research is a “compatible use” with what subjects agreed to in the original consent form. For non-exempt projects, a consent process description or justification for a waiver must be included in the research protocol.

The IRB may require that informed consent for secondary analysis be obtained from subjects whose data will be accessed.

Alternatively, the IRB can consider a request for a waiver of one or more elements of informed consent under 45 CFR 46.116(d). In order to approve such waiver, the IRB must first be satisfied that the research:

1. presents minimal risk (no risks of harm, considering probability and magnitude, greater than those ordinarily encountered in daily life or during the performance of routine examinations or tests); and
2. the waiver or alteration will not adversely affect the rights and welfare of the subjects; and
3. the research could not practicably be carried out without the waiver or alteration; and
4. whenever appropriate, the subjects will be provided with additional pertinent information after participation.

**“Restricted Use Data”:** Certain agencies and research organizations release files to researchers with specific restrictions regarding their use and storage. The records frequently contain identifiers or extensive variables that combined might enable identification, even though this is not the intent of the researcher. Research using these data sets most often requires non-exempt level review.

**Examples:**

1) Student C will be given access to coded mental health assessments from his faculty advisor’s research project. The student plans to analyze the data with a code attached to each record, and the advisor will retain a key to the code that would link the data to identifiers. The use of the data does constitute research with human subjects and does not qualify for exempt status since subjects can be identified. This student project would require an application to be submitted for non-exempt review by the IRB.

**Note:** As previously noted, if the student’s work fits within the scope of the initial protocol from which the dataset originates, the faculty advisor (or investigator who holds the dataset) may wish to consider adding the student and his/her work to the original protocol by means of an amendment application rather than having the student submit a new application for non-exempt review.

2) Student D is applying to the National Center for Health Statistics for use of data from the National Health and Nutrition Examination Survey that includes geographic identifiers and date of examination. The analysis of this restricted use data would require non-exempt review by the IRB.

The SPH HRPP Office is available to answer any questions regarding this guidance by email at [hrpp@sph.cuny.edu](mailto:hrpp@sph.cuny.edu).

*The CUNY SPH HRPP Office acknowledges the excellent guidance document prepared by the University of California-Berkeley Committee for Protection of Human Subjects, which served as the model for this Guidance.*