

Columbia University Institutional Review Board Guidance

Use of Publicly Available Datasets for Research

I. SCOPE:

This guidance identifies a specific set of conditions under which research involving the analysis of de-identified data in publicly available datasets is considered to **not** qualify as “research” with “human subjects” per applicable federal regulation and therefore does not need to be submitted to the IRB for review.

II. EFFECTIVE DATE: November 30, 2007

III. BACKGROUND:

Columbia University IRB policy, which is based on applicable federal regulations for the protection of human subjects, requires that all “research” with “human subjects” be submitted to the Institutional Review Board (“IRB”) for review.

The purpose of the IRB review is to:

- (1) ensure that the requirements of the regulations (e.g., informed consent will be sought and documented, risks are minimized) are met, or
- (2) determine that, although the project involves research with human subjects, all procedures fall into one or more of the “exempt” categories defined in the regulations.

Some research involves the analysis of data about humans for which the regulatory definition of “human subject” is not met. One example is research that involves only the analysis of de-identified data contained within publicly available datasets.

For the purposes of this guidance, “publicly available datasets” are limited to those that

- (1) are available to any one regardless of occupation, purpose, or affiliation, and*
- (2) have legitimately attained such status, i.e., those individuals who are responsible for posting the dataset had legitimate access to the data and have employed the necessary mechanisms to ensure the privacy and confidentiality of the individuals about whom the data were collected.*

While the activity described above meets the regulatory definition of research, the definition of human subjects is not met because data about a living person is not obtained through interaction or intervention, and no private, identifiable information about a living individual is obtained.

Therefore, protocols for research activities that involve **only** the analysis of de-identified data within a publicly available dataset need not be submitted to the IRB for review or for a determination that the project falls into an “exempt” category.

It is important to note that this guidance does not apply to research projects that merge or append datasets, unless each individual dataset is publicly available and contains only de-identified data.

IV. DEFINITIONS:

“Research”: a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.

“Human subject”: a living individual about whom an investigator (whether professional or student) conducting research obtains

- (1) Data through intervention or interaction with the individual, or
- (2) Identifiable private information.

“De-identified” - the identity of the subject cannot be ascertained by the investigator, research team or any individual collaborating in the research.

IV. GUIDANCE:

The IRB has determined that research projects involving only the analysis of de-identified datasets available from the following institutions do not qualify as “research” with “human subjects” and therefore, need not be submitted to the IRB for review or a determination of exemption.

1. Columbia University Electronic Data Service (EDS): <http://www.columbia.edu/acis/eds/>
2. **Inter-University Consortium for Political and Social Research (ICPSR)**: <http://www.icpsr.umich.edu/index.html>
3. **U.S. Bureau of the Census**: <http://www.census.gov/>
4. **National Center for Health Statistics**: <http://www.cdc.gov/nchs/>
5. **National Center for Education Statistics**: <http://nces.ed.gov/>
6. **National Election Studies**: <http://www.electionstudies.org/>

This list may be amended as the IRB determines the status of other data sources. For an up-to-date list of qualifying datasets, access the IRB website at: *[link to IRB website]*. Suggestions for additions to the list should be forwarded to the IRB via the email link on the website.

The IRB strongly recommends that investigators conducting research using data from any other source seek a determination from the IRB regarding the need for review. Failure to do so could result in regulatory non-compliance, for which investigators would be held responsible.