

## Columbia University Institutional Review Board- Morningside Heights

### Policy for Using Human Subjects Data Sets Without Identifiers: Conditions for Exemption from Full IRB Review

The Columbia University Morningside Campus Institutional Review Board (IRB) recently reviewed their current policies and procedures for the review of projects that conform to the requirements of the exempt category 45 CFR 46.101b(4). At the June 2002 meeting the IRB unanimously approved the following policy governing the use of human subjects data sets without identifiers.

Exemption from full IRB review under this category is specifically for human subjects data sets **without identifiers**, i.e., "Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects."

Included in the Columbia IRB policy are the recommendations on public use data sets of the National Human Subjects Protections Advisory Committee (NHRPAC) in its "Recommendations on Public Use Data Files – April 2002" <http://ohrp.osophs.dhhs.gov/nhrpac/documents/dataltr.pdf> and the conditions identified in the OHRP document "Issues to Consider in the Research Use of Stored Data or Tissues" - November 1997 <http://ohrp.osophs.dhhs.gov/humansubjects/guidance/reposit.htm>. It must be clearly understood that, pursuant to Federal regulations, the "Exempt" category does **not** mean that the human subjects protocol does not need to be reviewed; it means that the protocol can be reviewed (and approved) by the Chair of the IRB.

#### Public Use Data Files

Public use data files are those prepared by investigators or data suppliers with the intent of making them available for public use. The data available to the public are not individually identified or maintained in a readily identifiable form. Producers or suppliers of public use data files are responsible for having public use data files appropriately reviewed by an IRB before making them available to the public. In the case of Federal statistical data collections, the Federal Government has the responsibility for that review.

The Columbia University IRB requires that the use of these files be registered.

***Users seeking to merge public use data files or enhance a public use data file with identifiable or potentially identifiable data need to obtain IRB review and approval.***

#### Public Data Sets Currently Approved for Use Under this Policy

Columbia University investigators who wish to access/use the data sets available through these sources will need to submit the protocol to the IRB via the RASCAL Human

Subjects Module (<<http://www.rascal.columbia.edu>>) as with any other request for approval. The IRB chair will review the protocol and determine whether this policy applies to your project.

Currently only the following sources of public use data files have been approved by the IRB as qualifying under this policy:

1. Inter-University Consortium for Political and Social Research (ICPSR), CU IRB #01/02-832AEX  
URL: <http://www.icpsr.umich.edu/index.html>
2. U.S. Bureau of the Census, CU IRB#01/02-833AEX  
URL: <http://www.census.gov/>
3. National Center for Health Statistics, CU IRB#01/02-834AEX  
URL: <http://www.cdc.gov/nchs/National Center for Education Statistics CU IRB#01/02-835AEX> URL: <http://nces.ed.gov/>
4. National Election Studies, CU IRB#01/02-836AEX  
URL: <http://www.umich.edu/~nes/>

#### Investigator Guidelines for Registering a Data Set Without Identifiers

Data sets that may qualify for inclusion on Columbia University's list of approved data sources include:

- 1) Public use data sets posted on the Internet that include a responsible use statement or other confidentiality agreement for authors to protect human subjects.
- 2) Protocols submitted by investigators or research teams that include plans for making data publicly available. In reviewing such data sets for public use, the Investigator needs to provide evidence that either:
  - (a) The original data collection was gathered in an anonymous form or on unknown persons, or
  - (b) The original data collection was gathered on identified subjects but the data file has been stripped of direct identifiers and indirect identifiers that may risk disclosure of subjects' identity.

When the IRB is satisfied that a protocol involves the collection of data in anonymous form or on unknown persons, these data may be classified as public use data files. When the IRB is asked to authorize public data files from data originally collected with identifiers, the following factors will be considered by the IRB to be certain that the data files have been effectively de-identified for analysis by secondary users:

- (a) Removal of any identifiers of a human subject or of persons named by a human subject;
- (b) Removal of any variables that by definition would serve as surrogates for the identity of a human subject;
- (c) Collapse or combine categories of a variable to remove the possibility of identification due to a human subject being in a small set of persons with specific

- attributes regarding a variable (for example, due to the infrequency of subjects in a lower or upper range);
- (d) Collapse or combine variables to provide summary measures to mask what otherwise would be identifiable information;
  - (e) Use of statistical methods, where necessary, to add random variation with variables otherwise impossible to mask;
  - (f) Removal of any variables that could be linked to identifiers by secondary users.

Questions with respect to this policy and/or registration of additional data sets on the Morningside Campus should be directed to: Alison Dewhurst, IRB Administrator, 401B Low Memorial Library, Mail Code 4330 Telephone: 212/854-1324; Fax: 212/854-2577; E-Mail: [ad23@columbia.edu](mailto:ad23@columbia.edu) or Matthew Stafford, Assistant IRB Administrator, 401B Low Memorial Library, Mail Code 4330, Telephone: 212/854-0382, Fax: 212/854-2577; E-Mail: [mads35@columbia.edu](mailto:mads35@columbia.edu).