MCW Office of Research
Standard Operating Procedure

PUBLIC DATA SETS

Unit: Human Research Protections Program (HRPP), Office of Research
Applies to: MCW/FH Faculty and Staff involved in human research

PURPOSE:
Per 45 CFR 46.101, research using certain publicly available data sets does not involve "human subjects". The data contained within these specific data sets are neither identifiable nor private and thus do not meet the federal definition of "human subject" as defined in 45 CFR 46.102. Therefore these research projects do not need to be reviewed and approved by the Institutional Review Board, (IRB).

DEFINITIONS:
Public Data Sets: files prepared by investigators or data suppliers with the intent of making them available for public use.
- Are not individually identified or in a readily identifiable form.
- Include data files that have been reviewed under the jurisdiction of an IRB with the intent of making them available for public use. The federal government has responsibility for the review of federal statistical data collections.
- Does not include data shared informally among colleagues.
- May not include all data from a data supplier. Data suppliers may have (1) publicly available de-identified data as well as (2) restricted use data from the same data set.

Restricted Use Data Sets: not considered publicly available and not covered by this SOP.
- Special files upon which use restrictions are imposed. These may be distributed by governmental agencies, research organizations, and others.
- Often contain data such as Social Security numbers, names, or extensive life history markers that might enable an unauthorized user to identify a subject.
- Use restrictions vary, but typically involve secure (locked) data storage and password-protected computers, and forbid the storage of data on systems that may be accessed through a computer network connection.
- Use agreements for this type of data may limit the types of analyses that can be performed.

Publicly Available: data are widely available, even when:
- A fee is charged for obtaining the data.
- Access to the data is limited to researchers, if any researcher with a standard academic or research affiliation has access.

De-identified: the recipient of the data cannot readily ascertain the identity of the subject or associate identity with the information, and the data set does not include any
one of the 18 HIPAA direct identifiers. If there is a code that links the data to direct identifiers, the code: (1) may not be derived from or related to the information about the individual, and (2) could not be translated to identify the individual.

HIPAA Identifiers:
- Names
- All geographic subdivisions smaller than a state
- All elements of dates
- Telephone numbers
- Fax numbers
- E-mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers
- Device identifiers and serial numbers
- Web universal resource locators (URLs)
- Internet protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images
- Any other unique identifying number, characteristic or code

PROCEDURE:
**MCW List of Data Sets Considered Public**
The IRB Office will maintain on its website a current list of data sets it considers to be public which do not require MCW/FH IRB approval or a determination of exempt status. Researchers may conduct this research without submitting an IRB application.

IRB Review
1. This policy does not apply to the following activities and an IRB application must be submitted in eBridge:
   - Merging public data sets in such a way that individuals may be identified
   - Enhancing a public data set with identifiable, or potentially identifiable, data
   - Using non-publicly available datasets along with publicly available data sets
   - Interacting or intervening with human subjects along with publicly available data sets
   - Using controlled access data sets from the NIH dbGaP data repository
   - Files prepared by MCW investigators with the intent of making them available for public use

2. The data host may require the researcher or researcher's institution to sign a Data Use Agreement that explicitly requires IRB approval or a determination of exempt status. If IRB review is required:
   - Complete and submit an IRB application in eBridge
   - Upload a copy of the Data Use Agreement in section 52.1 of the eBridge Study SmartForm. An unexecuted copy is acceptable.
   - Submit a copy of the Data Use Agreement to the MCW Grants and Contracts Office for review and execution.

3. If the Data Use Agreement does not require IRB review and the database is listed on the **MCW List of Data Sets Considered Public**: 
- Submit a copy of the Data Use Agreement to the MCW Grants and Contracts Office for review and execution.

4. If the proposed data set is not on the *MCW List of Data Sets Considered Public*, the investigator should submit an IRB application in eBridge or follow the procedure outlined below to request that the IRB add a new data set to the list.

**Obtaining documentation from the IRB for research sponsors or funding agencies**

1. If documentation from the IRB is needed confirming that the use of specified data sets do not involve human subjects, complete the *Application for Activity that is Not Human Subjects Research* found on the IRB website.

2. Provide a memo that describes the following:
   - Name, department, and contact information of lead researcher
   - Name of data set, which must be on the *MCW List of Data Sets Considered Public* available on the IRB website
   - A description of research activities including:
     - If the research will involve merging any of the data sets in such a way that individuals might be identified
     - If any data sets to be used are considered restricted
     - If the research activity involves adding other data to the data sets
   - Funding information including:
     - Type of support (grant, contract, subcontract, etc.)
     - Name of funding agency or sponsor
     - Title of award or contract
     - Name of person listed as principal investigator on the funding proposal
     - Start and end date of the funding
     - Name of the office that processed the funding protocol (e.g. UW-Milwaukee Office of Sponsored Programs)

3. Provide a complete copy of the grant application or contract, if applicable.

4. Submit packet to the MCW/FH IRB Committee #5, IRB Office.

5. The IRB will review the submission and provide a determination letter to the researcher.

**Adding to the list of data sets considered public**

1. A data set may be nominated for inclusion on the *MCW List of Data Sets Considered Public* by providing the MCW/FH IRB with the following information:
   - Name of data set
   - Description of data set (list of variables or URL/web address for the list)
   - Whether the data are identifiable through direct identifiers, use of a code or other mechanism
   - How data were collected, including: from/about whom (subject populations), by whom, and when
   - How the data were obtained by the data supplier (if different from above)
   - Name and brief description of the data supplier or institution
   - How the data set may be obtained by researchers
   - Description of any conditions or restrictions about the use of the data
   - A copy of the agreement or statement that a researcher or institution must sign in order to gain access to the data (or URL for the agreement)
   - Any other URLs that are relevant

2. The nominated data set and documentation provided to the MCW/FH IRB will be reviewed by the IRB Office.

3. Written notification of the review outcome will be provided to the individual who submitted the nomination request.
4. If the data set is found eligible, the data set will be added to the *MCW List of Data Sets Considered Public*.

**REFERENCES:**
45 CFR 46.102(f)
32 CFR 219.102(f)
21 CFR 50.3(g)
45 CFR 160 and 164
OHRP Guidance on Research Involving Coded Private Information or Biological Specimens: October 16, 2008
National Human Subjects Protection Advisory Committee (NHRPAC), "Recommendations on Public Use Data Files"; January 29-29, 2002 meeting

**SUPPORTING DOCUMENTS:**
*Application for Activity that is Not Human Subjects Research*
*MCW List of Public Data Sets Considered Public*
*IRB Form: Public Data Set Nomination Checklist*
*IRB Staff: Nominated Data Set Letter*

---

Effective Date: 02/04/2013
Version number: 1.0
Previous Version/date: N/A
Responsible Office: HRPP Office
Approval Date: 02/04/2013

Approved By
HRPP Authorized Official: 
David Clark, PhD, Assistant Dean for Clinical Research and Director, HRPP
Human Research Protections Program (HRPP)
Office of Research
Medical College of Wisconsin

IRB SOP PUBLIC DATA SETS
Page 4 of 4