WSU HRPP Supplemental Policy 3.2A: Public Data Sets Version 1.0, 09/21/2022

Policy

This policy is supplemental to the WSU Human Research Protection Program (HRPP) Policies and Procedures Manual, Section 3.2 entitled "Human Subjects Research Determination" which states: "The investigator is responsible for initial determination of whether an activity constitutes human subjects research. The investigator should make this determination based on the definitions of "human subject" and "research" in Sec 18 (or by using the "Human Subject Research Determination" form). Since Washington State University will hold them responsible if the determination is not correct, investigators are urged to request a confirmation that an activity does not constitute human subjects research from the HRPP Office if they have any doubt about the determination. Determinations as to whether an activity constitutes human subjects research will be made by the HRPP according to the definitions in Sec 18 using either the Human Subject Research Determination form, or via the electronic submission system. Determinations regarding activities that either clearly are or clearly are not human subjects research may be made by experience HRPP staff, the ADRA or the Chair. Determinations regarding less clear activities will be referred to the ADRA and/or IRB Chair, who may make the determination or refer the matter to the convened IRB.

Documentation of all determinations made through the HRPP Office will be recorded and maintained in the IRB document management system. Formal submissions will be responded to in writing and a copy of the submitted materials and determination letter/email will be kept on file."

The HRPP and Institutional Review Board (IRB) consider the research use of certain publicly available data sets to not involve "human subjects" as defined by federal regulations (see 45CFR46 and 21CFR 50 and 56). The data contained within these specific data sets are neither identifiable nor private, nor do they involve individuals who become participants in research either as a recipient of a test article or control, and thus do not meet the federal definition of "human subject" under either HHS or FDA regulations. Therefore, the HRPP does not require these research projects to be reviewed and approved by the IRB.

The HRPP does not consider controlled access data sets, including those from NIH dbGaP and similar data repositories, to be public data sets. Prospective IRB approval or a determination of exempt status is required for any controlled or limited access data sets.

Definitions

Public Data Sets are data (or specimens) that are accessible to anyone in the general public, without the need for special qualifications, permissions, or privileges.

- The majority of public data sets will contain data that is not individually identifiable or in a readily identifiable form.

- Data sets available to any researcher with an academic or other research affiliation may still be considered public data sets, however any qualification requirement beyond this would result in the data set no longer being considered public.
- The federal definition of human subject requires that data be both private and identifiable to require IRB review, therefore publicly available data may or may not be individually identified or in a readily identifiable form.
- The merging of de-identified public data sets that allows for the re-identification of an individual (or their private information) would result in the merged data set no longer meeting the definition of public data.
- Public data files include data files that have been reviewed under the jurisdiction of an IRB with the intent of making them available for public use. In the case of federal statistical data collections, the federal government is responsible for that review.

Restricted Use Data Sets are not publicly available and are not covered by this policy.

- Restricted use data are special files distributed by government agencies, research organizations, and others upon which use restrictions are imposed.
- These files often contain data such as Social Security numbers, names, or extensive life history markers that might enable an unauthorized user to identify an individual.
- The use restrictions vary but may include secure (locked) data storage and password-protected computers, prohibiting storage of data on systems that may be accessed through a network.
- The use agreements may limit the types of analyses that can be performed.

Publicly Available means that the data are widely available (available to anyone in the general public). Data may still be considered publicly available 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 means that the recipient of the data cannot readily ascertain the identity of the subject or associate identity with the information, and that the data set does not include 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.

Exceptions to this policy/procedure

The following uses of public data sets are not covered by this policy and require prior IRB approval or a determination of exempt status:

- The merging of multiple de-identified public data sets, or a combination of publicly available de-identified and identifiable data, that enables re-identification of an individual and their private information.
- Enhancing or merging a public data set with identifiable, or potentially identifiable, non-public data.

The data host may require the researcher or the researcher's institution to sign a **Data Use Agreement** that explicitly requires IRB approval or a determination of exempt status. If so, then the Data Use Agreement prevails over this policy/procedure and over federal guidance about the definition of "human subjects."

Researchers whose research project also uses non-public data and/or interacts or intervenes with human subjects must first submit an application for IRB approval or a determination of exempt status for those activities.

Documentation for research sponsors or funding agencies

Publishers, research sponsors or funding agencies may require documentation that proposed research activities using public data sets do not require IRB approval or a determination of exempt status. In some instances, providing a copy of this policy with the publicly available data set included in the list below may satisfy this requirement. When it does not, a researcher will need to submit a request for documentation to the HRPP that includes any relevant details listed below.

1. Written communication (e.g., an email or memo), requesting documentation that the proposed research activities using a public data set does not require IRB approval or a determination of exempt status. The communbication should include:
    - A human subject research determination application or;
    - The name, department, and contact information of the lead researcher and;
    - The name of the data set, which must be on the pre-approved list below and;
    - A description of the research activities

2. Funding information, (if applicable and available) including:
    - The type of support (grant, contract, subcontract, fellowship, gift, etc.)
    - Name of funding agency or sponsor
    - ORSO # (if known) or;
    - if ORSO # is not known:
        - Title of award or contract
        - Name of person listed as principal investigator on the funding proposal
        - Start and end date of the funding

3. Answers to the following questions:
    - Will the research involve merging any of the data sets in such a way that individuals might be identified?
    - Will any data sets be used that are considered restricted?
    - Does the research activity involve adding other data to the data sets?

4. The HRPP Office will review this information and provide a determination letter.

Nomination process: Adding another data set to the pre-approved list

1. Anyone may nominate a data set for inclusion on the Public Data Sets list at the end of this document by providing the HRPP with the following information:
    - Name of the data set
    - Description of the data set (list of variables, or URL/web address for the list)
    - Description of the identifiability of the data (are the data coded, etc.)
    - Description of how the data were collected, including: from/about whom (subject populations), by whom, and when
    - Description of how the data were obtained by the host (if different from above)
    - Name and brief description of the host archive or institution
    - Description of how the data set may be obtained by researchers
    - Description of any conditions or restrictions about the use of the data
    - A copy of any agreement or statement that a researcher or institution must sign in order to gain access to the data (or URL for the agreement)
    - Any URLs that are relevant

    WSU researchers who propose to provide public use data files are responsible for having those files appropriately reviewed by the WSU IRB (or another qualified IRB with approval of the WSU IRB) before making the data available to the public.

2. Nominations are reviewed by HRPP staff. In the event of any uncertainty or dispute involving the public status of a data set, the Assistant Director or Director of the Office of Research Assurances, who have responsibility for administration of the WSU HRPP, will make the final decision.

    - Documentation.

a. The decision to add a public data set to the pre-approved list will be documented in writing (email or formal notification) by HRPP staff
b. Written notification of the review outcome is provided by the HRPP to the person who nominated the data set.
c. The HRPP Office will create a file to document nominations, reviews, and outcomes. The file contains:
    i. The nomination (e.g., print of an email request)
    ii. A written summary of the review and determination
    iii. Copy or URL of the Public Data Set(s)
    iv. Any other relevant materials

List of Data Sets Considered Public (includes but is not limited to)

NOTE: Research involving only the analysis of data from the following public data sets does not require WSU IRB approval or a determination of exempt status. Researchers may conduct this research without submitting an application or other materials for review. A copy of this policy may be used in lieu of an approval or determination letter for publication or other purposes following analysis of publicly available data.

- American College of Surgeons National Trauma Data Bank (NTDB)
- American Hospital Association Annual Survey
- Behavioral Risk Factor Surveillance System (BRFSS; public data only)
- Comprehensive Hospital Abstract Reporting System (CHARS; public data only)
- Demographic Health Survey (DHS), Standard and Interim Surveys
- European Union Open Data Portal
- Fatality Analysis Reporting System (FARS)
- Google Public Data Explorer
- Healthcare Cost and Utilization Project (H-CUP) healthcare databases
    - The Nationwide Inpatient Sample (NIS)
    - The Kids' Inpatient Database (KID)
    - The State Inpatient Databases (SID)
    - The State Ambulatory Surgery Databases (SASD)
    - The State Emergency Department Databases (SEDD)
- HIV Prevention Trials Network D01: Vaccine Preparedness Study/Uninfected Protocol Cohort – 4 files
- Hospital Compare
- Inter-University Consortium for Political and Social Research (ICPSR) – public use data only
    - National Latino and Asian American Study (NLAAS) – public use data only
- Medical Expenditure Panel Survey (MEPS)
    - Household Component Full-Year files
    - Household Component Event files
    - Household Component Point-in-time files

- Pooled Linkage files
- Medicare Healthcare Cost Report Information System (HCRIS)
- National Automotive Sampling System (NASS) General Estimates System (GES)
- National Center for Health Statistics
  - Life Tables
  - LSOAs: Longitudinal Studies of Aging
  - NHANES: National Health and Nutrition Examination Survey
  - NHCS: National Health Care Survey
  - NHIS: National Health Interview Survey
  - NIS: National Immunization Survey
  - NSFG: National Survey of Family Growth
  - SLAITS: State & Local Area Integrated Telephone Survey
  - Vital Statistics: National Vital Statistics System
- National Center for Education Statistics
- National Election Studies
- National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)-Wave 1 & Wave 2
- National Hospital Ambulatory Medical Care Survey (NHAMCS)
- National Longitudinal Survey (NLSY)
  - National Longitudinal Survey of Youth 1997 (NLSY97)
  - National Longitudinal Survey of Youth 1979 (NLSY79)
  - NLSY79 Children and Young Adults
  - National Longitudinal Survey of Young Women and Mature Women
  - National Longitudinal Survey of Young Men and Mature Men
- National Survey of Children's Health (public version)
- National Survey of Children with Special Health Care Needs (public version)
- National Trauma Data Bank
- Organ Procurement and Transplantation Network (OPTN)
- Roper Center for Public Opinion Research
- Survey of Consumer Finances (SCF)
- U.S. Bureau of the Census
- U.S. Bureau of Labor Statistics
- World health Organization (WHO) Open Data Repository

REGULATORY AND OTHER REFERENCES

45 CFR 46
32 CFR 219.102
21 CFR 50 and 56