

Policy Number: 41

Title: Research Involving Secondary Use of Existing Data

Date of Last Revision: July 5, 2007

I. Introduction

Data analysis is considered to be a *research* activity and therefore requires IRB oversight; the secondary use of existing data about *human subjects* requires UCLA IRB approval or certification of exemption from UCLA IRB review prior to UCLA investigators' access to and/or provision of the data.

Prospective UCLA IRB review and approval or Certification of Exemption from UCLA IRB review is required for all data analysis activities which are determined to constitute *research* with *human subjects*.¹ Federal regulations define specific Expedited and Exempt Review categories for research analyses of human subject data.

Studies that involve the secondary use of existing data from public or private documents includes review of existing medical records, police reports, vital statistic records, student records, data collected in previous studies, etc. This includes data stored in electronic, digital, paper, audio, or video tape form, whether on computers or equipment that are privately owned, university-owned or maintained, or on removable electronic media.

UCLA Investigators are responsible for obtaining from the UCLA IRB/OPRS a determination of whether proposed secondary analysis of existing data: (1) constitutes research involving human subjects and requires UCLA IRB review and approval or Certification of Exemption from UCLA IRB review, or (2) does not constitute research involving human subjects.²

II. Definitions

- A. **Research:** A systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other

¹ See [OPRS/IRB Policy 3: Human Subject Research Determinations](#)

² UCLA investigators who plan to use "public use" datasets identified in [OPRS/IRB Policy 42: Research Involving Public Use Data Files](#) do not require UCLA IRB review and approval or Certification of Exemption from UCLA IRB review and do not need to submit an application to the UCLA IRB/OPRS, except as described in Section III(C) of the policy.

purposes. For example, some demonstration and service programs may include research activities.

- B. **Human subject** is defined as 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.
1. **Intervention** includes both physical procedures by which data are gathered (for example, venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes.
 2. **Interaction** includes communication or interpersonal contact between investigator (or his/her research staff) and subject (or the subject's identifiable private information).
 3. **Private information** includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be **individually identifiable** (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects. Identifiable private information obtained from a primary participant about a third party may constitute research involving human subjects.
- C. **Existing data** are data that exist at the time the research is proposed. The data may include data sets, interview notes or audio- or video tapes. The data may have been originally collected or created either for research or non-research purposes. Examples of data collected for non-research purposes include: data collected by healthcare providers for clinical care purposes, data collected in the course of internal quality improvement or assessment activities.
- D. **Personal identifiers** within a data set are any data elements that singly or in combination can uniquely identify an individual, such as a social security number, name, address, demographic information (e.g., combining gender, race, job, and location), student identification numbers, or other identifiers (e.g., hospital patient numbers).
- E. **De-identified data** means data that subsequently has been stripped of all elements (including but not limited to personal identifiers) that might enable a reasonably informed and determined person to deduce the identity of the subject.
1. For research that requires that data elements later be linked to an individual's identity, the original data set may be partitioned into two data sets: a *de-identified data set* and an *identity-only data set*. The latter should contain any and all personal identity information absolutely necessary for future conduct of the research. For purposes of

later merging the identity information with other research data, a researcher-assigned identity code (typically a randomly-generated number) that is associated with and unique to each specific individual may be included in both data sets, and later be used to link identity data elements back to the de-identified data set. This identity code should not offer any clue as to the identity of an individual.

- F. **Minimal risk** means that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests. [45 CFR 46.102(i)]
- G. **Publicly available** means that the general public can obtain the data. Data are not considered “publicly available” if access to the data is limited to researchers.
- H. **Individually identifiable data** means data that are still attached to a participant identifier (such as full name, address and other contact information, social security number, identifiable photographic images, medical record number, etc.) such that the identity of the subject is or may readily be ascertained by the investigator or associated with the information.
- I. **Practicably** means that it is feasible to obtain a subject's consent to participate in research without affecting the scientific integrity of the study.
- J. **Readily identifiable** means that the identity of the subject could be ascertained by the investigator or associated with the data without requiring time or special effort.
- K. **Coded data** means that collected data are unidentified for research purposes by use of a random or arbitrary alphanumeric code but the data may still be linked to their sources through use of a key to the code available to an investigator or collaborator.
- L. **Unlinked data** means human data that were initially collected with identifiers but, prior to research use, have been irreversibly stripped of all identifiers by use of an arbitrary or random alphanumeric code and the key to the code is destroyed, thus making it impossible for anyone to link the samples to the sources. This process does not preclude linkage with existing clinical, pathological, and demographic information before participant identifiers are removed (a.k.a., **anonymized data**).
- M. **Anonymous data** means that the data were collected without identifiers of any kind. Data may retain demographic or diagnostic information and still be considered anonymous if such information cannot be used to reveal the identity of the source. (a.k.a., **unidentified data**)

III. Examples of Secondary Data Analysis Scenarios, and Related IRB Requirements

The following scenarios detail the materials required for IRB/OPRS review of different types of secondary data analysis, and illustrate factors which determine the type of review required for each.

A. Change in Data Analysis Plan for Currently Approved Research

When analysis prompts changes to the data analysis plan for currently approved research, for example if investigators make a significant change to their data analysis plan and/or add new investigators to a UCLA IRB-approved protocol or a protocol that has been certified exempt from UCLA IRB review, the investigator must submit an application to obtain UCLA IRB approval or certification of exemption of the proposed modifications to the original protocol. Please refer to [OPRS/IRB Policy 11: IRB Review of Modifications to Previously Approved Research](#) for details.

B. Analysis of Data from Another Investigator's Currently Approved Research

When a UCLA investigator proposes to analyze data collected by *another* investigator, the UCLA investigator proposing analysis must submit a new protocol to the IRB/OPRS describing the analysis plan. The application should explain whether the analysis includes data that remain identifiable, and should include a copy of the IRB approval and IRB-approved consent document for the original collection of the data. The investigator who originally collected the data may need to amend his/her protocol to allow for sharing.

C. Analysis of Data from Previously Approved Research

1. If the original UCLA IRB-approved protocol has been closed, and the *original investigators* now wish to re-analyze or otherwise examine data from that protocol - i.e., conduct a 'secondary analysis' of their own data, the original UCLA investigators must submit a new protocol to the UCLA IRB/OPRS describing the analysis plan and requesting UCLA IRB review and approval or Certification of Exemption from UCLA IRB review.
2. If the original UCLA IRB-approved protocol has been closed, and the original investigators now wish to *share their data with a UCLA student or colleague*, the UCLA investigator who proposes to access the data must submit a new protocol to the UCLA IRB/OPRS describing the analysis plan, and requesting UCLA IRB review and approval or Certification of Exemption from UCLA IRB review.
3. Applications proposing to analyze data from previously-approved research must:
 - a. explain whether the analysis includes data that remain identifiable, and
 - b. include a copy of the IRB approval and IRB-approved consent document for the original collection of the data.

D. Analysis of Data from Currently Approved Research Which Does Not Include a Process for Obtaining Subjects' Consent for Their Identifiable Data to be Used for Research Purposes Beyond the Original Protocol

If the original IRB-approved protocol for the collection of the data does not include: (1) an IRB-approved consent process wherein subjects are asked to consent to the storage and sharing of their identifiable data for research purposes beyond the original protocol, and/or (2) IRB-approved procedures to store and share identifiable human subjects data with persons outside the original research team, the following is required:

1. The Principal Investigator of the original protocol must either:
 - a. obtain IRB approval for a modification to the protocol to include storage and sharing of data and a consent process wherein subjects are re-consented and asked to consent to the storage and sharing of their identifiable data for research purposes beyond the original protocol, or
 - b. obtain IRB approval for the storage and sharing of subjects' identifiable data including IRB approval of a waiver of the requirement for subjects' informed consent for their identifiable data to be stored and shared for research purposes beyond the original protocol.
2. The UCLA investigator proposing secondary analysis of the data must include the following in the application for either IRB review or certification of exemption:
 - a. explain whether the analysis includes data that are identifiable,
 - b. include a copy of the IRB approval and IRB-approved consent document for the original collection of the data, and
 - c. provide a written ethical and scientific justification for a waiver of the requirement for informed consent, explaining why the proposed secondary analysis: (1) involves no more than minimal risk to the subjects; (2) will not adversely affect the rights and welfare of the subjects; and (3) could not practicably be carried out without the waiver or alteration.

E. Analysis of Data from Previously Approved Research Which Did Not Include a Process for Obtaining Subjects' Consent for Their Data to be Used for Research Purposes Beyond the Original Protocol

If the original IRB-approved protocol for the collection of the data did not include: (1) an IRB-approved consent process wherein subjects were asked to consent to the storage and sharing of their data for future research purposes, and/or (2) IRB-approved procedures to store and share the data with persons outside the original research team, the UCLA investigator's application for UCLA IRB review and approval of the proposed secondary analysis of the data must:

1. explain whether the analysis includes data that remain identifiable,
2. include a copy of the IRB approval and IRB-approved consent document for the original collection of the data, and
3. provide a written ethical and scientific justification explaining why the proposed secondary analysis: (1) involves no more than minimal risk to the subjects; (2) will not adversely affect the rights and welfare of the subjects; and (3) could not practicably be carried out without the waiver or alteration.

F. Long-Term Storage of Data

As an increasing number of investigators plan to share data with colleagues and students, the UCLA IRB strongly recommends that when data sets are maintained beyond the end of a research project, investigators routinely archive those data for long-term storage, and do this by stripping all identifiers (e.g., name, birth date, social security number, phone number, address). Such procedures to deidentify stored data serve to minimize risk to subjects and facilitate future IRB reviews of proposals for secondary data analysis. Please refer to [OPRS/IRB Policy 40: Data Repositories](#) for details.

IV. “Public Use” Data

- A. Generally, public use data sets have been prepared by data-suppliers for wide scale dissemination. Data suppliers may have both publicly available de-identified and restricted identifiable data sets. Examples of public use data include portions of the U.S. Census data, and the Health and Retirement Survey. Data shared informally among colleagues does not constitute public use data.
- B. UCLA investigators who seek to use “public use” datasets identified in [OPRS/IRB Policy 42: Research Involving Public Use Data Files](#) do not require UCLA IRB review and approval or Certification of Exemption from UCLA IRB review.
- C. UCLA Investigators who are importing a dataset to UCLA that may meet a “public use” standard should contact the UCLA IRB/OPRS for guidance. Please see [OPRS/IRB Policy 42: Research Involving Public Use Data Files](#) for details.
- D. UCLA investigators whose proposed research includes activities in addition to access to “public use” data must submit an application to the UCLA IRB/OPRS to obtain prospective UCLA IRB review and approval or Certification of Exemption from UCLA IRB review of the activities to be conducted in addition to the access to “public use” data.

V. **“Restricted Use” Data**

- A. A number of federal agencies and research organizations distribute special files to investigators on which they impose use restrictions. These files generally contain data fields, such as social security numbers, names, or extensive life history markers that might enable an unauthorized user to identify a participant. The use restrictions vary, but they typically involve secure (locked) data storage and password protected computers, and forbid the storage of data on computer hard drives that may be accessed through a computer network connection. These agreements may also limit the types of analyses that are done by the investigator.
- B. UCLA investigators who propose to access “restricted use” data must submit an application to the UCLA IRB/OPRS to obtain UCLA IRB review and approval prior to access to the data.

VI. **Data Enclave**

- A. Some organizations create secure environments where investigators can conduct analyses on “unexpurgated”³ data; these environments are referred to as “data enclaves” or “secure data enclaves”. Investigators have to file a research plan and meet strict criteria for what can be printed, saved, and removed from the site.

Example of a data enclave at UCLA: The California Center for Population Research Secure Data Enclave (<http://www.ccpr.ucla.edu/asp/sde.asp>).

- B. UCLA investigators planning to access data in a “secure data enclave” must submit an application to the UCLA IRB/OPRS to obtain UCLA IRB review and approval or Certification of Exemption from UCLA IRB review prior to access to the data.
- C. Enclaves that do not collect, bank, and share data, but rather provide a venue where investigators “work” with specific data sets obtained for their protocol do not require UCLA IRB review and approval or Certification of Exemption from UCLA IRB review. UCLA investigators who collect, bank, and share data which is accessed in the enclave must obtain UCLA IRB approval for the activities related to the collection, banking, and sharing of the data.
- D. Enclaves that do collect, bank, and share data must obtain prospective UCLA IRB review and approval of all activities related to the collection, banking, and sharing of data to be stored in the enclave. Please refer to [OPRS/IRB Policy 40: Data Repositories](#) for details.

³ Unexpurgated is defined as complete, whole, or not having material deleted.

VII. Determining Type of Review Required

A. Certification of Exemption from IRB Review

According to federal regulations, secondary analysis of existing data may be considered for Certification of Exemption from IRB review under either of the two following conditions:

1. The data are publicly available. OR
2. The investigator records the data in such a way that subjects cannot be identified.

B. IRB Review by Expedited Review Procedures or Convened Board

1. If an investigator records pre-existing data with identifiers, secondary analysis of the data must be reviewed through expedited or full procedures.
2. Expedited review procedures may be used when the only risk to the identified subjects is a breach of confidentiality and adequate confidentiality procedures are in place.
3. If the purpose of the secondary data analysis is found to differ significantly from the purpose of the original study, the application for secondary data analysis is likely to require expedited or full committee review by the UCLA IRB. The UCLA IRB will determine whether or not secondary analysis of the data increases risk to participants and whether or not participants were adequately informed during the original study about the use, confidentiality, and destruction of their identifiable data. Based on these findings, the UCLA IRB may require that informed consent for secondary analysis be obtained from subjects.
4. If the content of the data to be analyzed could put the participants at risk, the application for secondary data analysis is likely to undergo review by the Convened Board. The UCLA IRB will determine whether or not secondary analysis of the identifiable data increases risk to participants and whether or not participants were adequately informed during the original study about the use, confidentiality, and destruction of their identifiable data. Based on these findings, the UCLA IRB may require that informed consent for secondary analysis be obtained from subjects.
5. Additional examples of secondary data analysis that must be reviewed through the expedited or full review process include:
 - a. Analysis that involves merging sets of unidentifiable data creating an identifiable data set.
 - b. Analysis of restricted portions of public use data sets.
 - c. Analysis of video and audio recordings when investigators propose to analyze the material in a manner not specified in the original agreement with the subjects.

VIII. Investigator Responsibilities

- A. UCLA Investigators are responsible for obtaining from the UCLA IRB/OPRS a determination of whether proposed secondary analysis of existing data:
1. constitutes research involving human subjects and requires UCLA IRB review and approval or Certification of Exemption from UCLA IRB review, or
 2. does not constitute research involving human subjects.

Investigators may not make this determination; the determination must be made by the UCLA IRB/OPRS.⁴ Please refer to [OPRS/IRB Policy 3: Human Subject Research Determinations](#) for additional details.

- B. UCLA Investigators who obtain data from other researchers must obtain UCLA IRB review and approval or Certification of Exemption from UCLA IRB review prior to receiving the data from the other researcher. The researcher providing the data must have approval from his/her respective institution, allowing him/her to share the data with the UCLA investigator.
- C. UCLA Investigators who are conducting secondary analysis of existing data that constitutes human subjects research must submit an application for UCLA IRB review and approval or Certification of Exemption from UCLA IRB review.

An application submitted for UCLA IRB review and approval must include the following:

1. Form HS-1 “UCLA Application to Involve Human Subjects in Research”. Sections I, II, III, and VII must be completed.
2. Section VII of the HS-1 form must describe how subjects’ privacy and confidentiality will be protected, including:
 - a. Who will have access to the data
 - b. How and when personal information will be separated from other data
 - c. Whether the data will be retained at the conclusion of the study
3. Section VII of the HS-1 form must include a written ethical and scientific justification explaining why the proposed secondary analysis: (1) involves no more than minimal risk to the subjects; (2) will not adversely affect the rights and welfare of the subjects; and (3) could not practicably be carried out without the waiver or alteration.

⁴ UCLA investigators who plan to use “public use” datasets identified in [OPRS/IRB Policy 42: Research Involving Public Use Data Files](#) do not require UCLA IRB review and approval or Certification of Exemption from IRB review and do not need to submit an application to the UCLA IRB/OPRS, except as described in Section III(C) of the policy.

4. Copies of relevant IRB approval notices and IRB-approved consent documents, as described in Sections III-VI above.
5. A copy of the original protocol's IRB-approved modification to the consent process or storage/sharing procedures, as described in Section III(D)(1).

An application submitted for Certification of Exemption from IRB review must include the following:

1. Form HS-7 "Claim of Exemption From Institutional Review Board (IRB) Review."
 2. Section I of the HS-7 form must include a written ethical and scientific justification explaining why the proposed secondary analysis: (1) involves no more than minimal risk to the subjects; (2) will not adversely affect the rights and welfare of the subjects; and (3) could not practicably be carried out without the waiver or alteration.
 3. Copies of relevant IRB approval notices and IRB-approved consent documents, as described in Sections III-VI above.
 4. A copy of the original protocol's IRB-approved modification to the consent process or storage/sharing procedures, as described in Section III(D)(1).
- D. UCLA investigators who seek to add company, community, census tract, or geographical data to any existing or new survey data set must apply to the UCLA IRB for approval to merge these data, and provide a plan for the secure analyses of these data. The addition of such data to individual records increases the probability that an unauthorized user of the dataset may be able to identify an individual participant.
- E. The Investigator is responsible for complying with all applicable federal and state laws regarding the confidentiality of information (such as the California Information Practices Act as described below).
1. The release of identifiable data held by the University of California and the receipt of identifiable data from another state agency both fall under the terms of the California Civil Code 1798.24, as amended in 2005. Unless subjects have provided authorization as required by the law, or another exception exists as outlined in the law, the release of identifiable information to or by UC requires review by the Committee for the Protection of Human Subjects of the California Health and Human Services Agency.⁵

IX. IRB/OPRS Responsibilities

- A. Upon review of applications for secondary data analysis, or which in part include analysis of secondary data, the UCLA IRB will specifically determine:

⁵ <http://www.oshpd.ca.gov/CPHS/index.htm>

1. whether or not secondary analysis of the data increases risk to participants,
 2. whether or not participants were adequately informed during the original study about the use, confidentiality, and destruction of their identifiable data, and
 3. whether or not informed consent for secondary analysis must be obtained from participants.
- B. The UCLA IRB will conduct initial and continuing review in accordance with its standard policies as described in the [OPRS/IRB Policy 9: IRB Review Process - Full Committee Review](#) and [OPRS/IRB Policy 8: IRB Review Process - Expedited Review](#).
- C. The UCLA OPRS will certify applicable submissions Exempt from UCLA IRB review as described in the [OPRS/IRB Policy 5: OPRS Review Process - Certification of Exemption from IRB Review](#).

Regulations:

45 CFR 46.101
45 CFR 46.102(d)
45 CFR 46.102(f)
45 CFR 46.110
45 CFF 46.111
45 CFR 46.116(d)

References:

OHRP Guidance on Research Involving Coded Private Information or Biological Specimens, August 10, 2004. (<http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.pdf>)

U.S. Office for Human Research Protections' (OHRP, formerly OPRR) January 26, 1999 Memorandum, "Engagement of Institutions in Research". (<http://www.hhs.gov/ohrp/humansubjects/assurance/engage.htm>)