

C.S. A1: Analysis of Patient Data from Secondary Sources

C.S. A1.1: Secondary Data with Links to SSNs

Overview

This study uses analysis of patient data from existing VA databases (originally established for patient care and administrative purposes—not research) to compare statistical models of risk adjustment and mortality prediction. There is no direct patient contact, and scrambled patient identifiers are used to link data from various sources. (See additional explanation below under “Data Collection and Confidentiality.”)

Subjects and Sample Size

Data are collected on 5,000 VA patients with ICU admissions. Subjects are to be identified from VA databases using diagnostic criteria.

Data Collection and Confidentiality

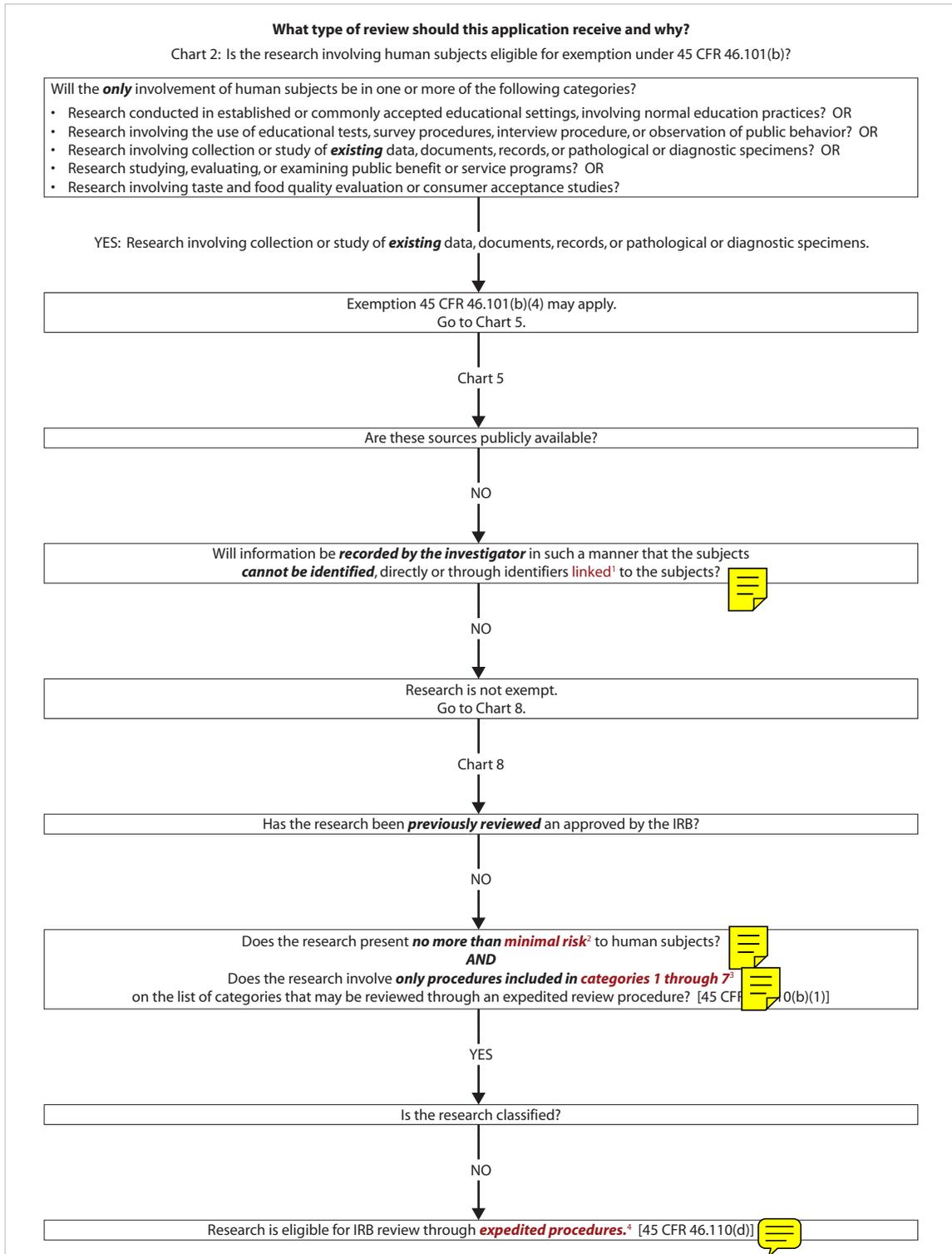
Patient data to be collected include demographic information, date of birth, zip

code, gender, ethnicity, ICU admissions, diagnoses, lab results, inpatient treatment information, mortality data, and other outcomes. These data are collected via database search (e.g., Austin data, Pharmacy Benefits Management data, DSS data) and will be used to test and compare risk adjustment methods.

The patient cohort will be obtained from existing VA databases using diagnostic criteria. Some of the databases contain real SSNs, others contain scrambled SSNs. After the study data, including SSNs and scrambled SSNs, are pulled, all real SSNs will be converted to scrambled SSNs, using a file linking scrambled SSNs with real SSNs obtained from a separate Austin database. Thus, all study files with patient data will include only scrambled SSNs. The file linking scrambled SSNs with real SSNs will be maintained by the research team as a separate file, in a password-protected drive that is separate from the drive containing the study data.

C.S. A1.1

[From OHRP Web site: www.hhs.gov/ohrp/humansubjects/guidance/decisioncharts.htm]



Notes for C.S. A1

¹**Definition:** There are identifiers that can be linked to the subject, and the investigator (or investigative team) is maintaining these data.

Discussion: If another person outside the team (e.g., a “data broker”) obtains the data with identifiers, but then removes the identifiers before passing the rest of the data on to the investigator, then the answer to this question is “yes.”

Many IRBs are using the HIPAA definition of de-identified data [from HIPAA Privacy Rule 164.514(a)-(c)] to determine whether or not direct or indirect links exist. According to the rule, de-identified data does not contain the following: name, address (including all geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geo-codes, except for the initial three digits of most zip codes), all elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death, age over 89 and all elements of dates (including year) indicative of age over 89, except that ages over 89 may be aggregated into a single category of “age 90 or older”, telephone and fax number, e-mail address, social security number, medical record number, health plan beneficiary number or account number, certificate/license number, vehicle serial number, URL or IP address, biometric indicators such as finger or voice prints, full face photographic images, any other uniquely identifying characteristic.

²**Definition:** “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” (CFR 46.102(1)).

Discussion: Panel members considered this study to be minimal risk. The *probability and magnitude* of loss of confidentiality, given the safeguards described, are no greater than *that which is encountered in daily life*—e.g., the probability of loss of confidentiality of other, non-research-related health data collected and maintained within VA medical centers and clinics, or by non-VA healthcare providers. The small probability of loss of confidentiality is based on the assumption that the safeguards for maintaining data confidentiality by the investigators are adequate—or, at a minimum, that the procedures are as good as those used elsewhere in the health care facility for ensuring the confidentiality of health-related data. Therefore, sufficient information must be provided by investigators to the IRB committee for them to determine that the procedures for maintaining data confidentiality are acceptable. If there has been a history of problems with maintaining the confidentiality of research data at a particular institution, or if the investigators do not have much experience with the collection and use of data from secondary datasets, then the local IRB may choose full review as a means to more carefully review the procedures and ensure that they are adequate.

³**Definition:** The research involves procedures included in category 5: Research involving materials (data, documents, records, or specimens) that have been collected, or will be collected solely for nonresearch purposes (such as medical treatment or diagnosis). [Return to home page for full list of categories eligible for expedited review under 45 CFR 46.110(b)(1).]

⁴**Discussion:** The research is potentially eligible for expedited review under the assumptions described in the above notes from the panel discussion. A member of the IRB who understands these issues would need to review carefully the proposed research and make this determination.