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

C.S. A1.3: Use of SSNs to Obtain Secondary Data

Overview

This study uses analysis of patient data from existing VA databases (originally established for patient care and administrative purposes—not research) and from electronic medical records 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 Austin databases using diagnostic criteria.

Data Collection and Confidentiality

Austin databases are searched by diagnostic code to determine eligible patients. The SSNs of these eligible patients are obtained, and are used to identify the patients' electronic medical records, from which relevant clinical data for the study are obtained. These clinical data are entered into the study database, which does not contain any SSNs or linkages to patient identifiers. A file containing the list of SSNs of eligible patients—but no linkages and no other data—is temporarily maintained until all of the medical record data are obtained.

C.S. A1.3

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

What type of review should this application receive and why?

Chart 2: Is the research involving human subjects eligible for exemption under 45 CFR 46.101(b)?

Will the ***only*** involvement of human subjects be in one or more of the following categories?

- Research conducted in established or commonly accepted educational settings, involving normal education practices? OR
- Research involving the use of educational tests, survey procedures, interview procedure, or observation of public behavior? OR
- Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens? OR
- Research studying, evaluating, or examining public benefit or service programs? OR
- Research involving taste and food quality evaluation or consumer acceptance studies?

↓

YES: Research involving collection or study of ***existing*** data, documents, records, or pathological or diagnostic specimens.

↓

Exemption 45 CFR 46.101(b)(4) may apply.
Go to Chart 5.

↓

Chart 5

↓

Are these sources publicly available?

↓

NO

↓

Will information be ***recorded by the investigator*** in such a manner that the subjects ***cannot be identified***, directly or through identifiers linked[1] to the subjects?

↓

NO

↓

Research is not exempt.
Go to Chart 8.

↓

Chart 8

↓

Has the research been ***previously reviewed*** an approved by the IRB?

↓

NO

↓

Does the research present ***no more than minimal risk***[2] to human subjects?
AND
Does the research involve ***only procedures included in categories 1 through 7***[3]
on the list of categories that may be reviewed through an expedited review procedure? [45 CFR 46.110(b)(1)]

↓

YES

↓

Is the research classified?

↓

NO

↓

Research is eligible for IRB review through ***expedited procedures.***[4] [45 CFR 46.110(d)]

Notes for C.S. A1.3

[1]**Discussion:** The first part of the study—the identification of eligible subjects--definitely involves the use of identifiers (SSNs). Therefore, the majority of the panel responded "no" to this question, because subjects can be identified. Therefore, this part of the study is not exempt from IRB review.

However, the second part of the study, after the SSNs are destroyed, might not involve the use of identifiable data, and could possibly be considered exempt. It depends on what kinds of data are obtained from the medical records. If the medical record data can be used indirectly to identify subjects, then the second part of the study is also not exempt from IRB review. To be considered exempt, the investigator would need to provide information showing that none of the data obtained from the medical records could be used to identify individuals.

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.

Notes for C.S. A1.3 ***(cont.)***

[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.