Requirements for Research Projects Involving Secondary Analysis of Existing Data

Ascension St. John Hospital

July 2018

Background:

There are many local, regional and national databases that can be used to answer research questions. Local and regional databases are sometimes publicly available, free of charge. Many national databases (for example the National Inpatient Survey, the National Readmission Database) require that each potential user signs and submits a data use agreement. There is also a fee for each year of the database requested.

Other databases have more stringent requirements. The SEER-Medicare linkage database, for example, is available to investigators for research purposes. Although the personal identifiers for each patient are removed, there is the remote risk of re-identification. Thus, these data are not public-use data and investigators are required to obtain approval for specific research questions in order to purchase the data. SEER requires that a letter of IRB approval be submitted along with the proposed study design.

National databases provide the best opportunity to answer difficult questions, particularly when the disease or risk factor is rare. These datasets are typically designed with multi-level sampling so that the data are representative of the US population. Therefore, the statistical analysis is complex and requires that the data analyst account for the study/sample design. It is imperative that the researcher (i.e. resident or fellow) understand how the database was designed, how the sample was drawn, what elements of data exist and how the data are analyzed, given the complex sampling design.

Requirements:

If a resident or fellow wishes to do a research project that employs secondary analysis of existing data, the following requirements must be met:

1. The resident or fellow will write a protocol, following the standard protocol design.
    a. ***Background:*** The review of the relevant literature should be an in-depth review of what is known about the topic.

b. ***Research question/hypotheses:*** The research question and hypotheses should be precisely stated.

c. ***Methods:*** The methods sections should include the methods that were used to collect the data, i.e. the design of the survey (for example, NHANES) or database (example, NRD). The resident should be able to describe: the purpose of the database, the history of the database, the types of variables that are available, the sampling technique used to collect the data, any noted limitations of the data or of a particular year of the data, whether the data can be linked across years and whether the data, once weighted, are representative of the US population.

   This section should include how informed consent was obtained for the original data collection, if applicable.

   The resident or fellow is required to read the documentation that accompanies the database and choose the specific variables that should be employed in the analysis, including ICD-9 and ICD-10 diagnosis and procedure codes, SNOMED codes, CPT codes, CCS codes, demographic factors and other variables unique to the dataset. These codes and variables should be described and listed in the methods section.

   The resident or fellow should also clearly define the study population of interest (e.g. males over 50 or all adults 18-90 years etc.)

d. ***Sample size/Power analysis:*** Because these databases are often very large, a sample size or power analysis may not always be required. For hospital, local or regional databases, however, a sample size calculation is still required to ensure that there are enough patients to reject the null hypothesis if false.

e. ***Statistical analysis:*** As usual, the statistical analysis should be written by the medical researcher, unless the resident or fellow has advanced training in statistics. Given that not all medical researchers are familiar with these types of analyses, this section should be written by Dr. Szpunar or, for General Surgery residents, by Dr. Hayward.

The resident should be able to demonstrate to the GME Research Department medical researcher that they understand the basic elements of the data analysis, particularly the design effects.

f. ***Patient confidentiality/HIPAA:*** For many of these databases, the data are already de-identified; however, for hospital-level databases, patient identifiers may be present. In these situations, the resident or fellow should describe what measures will be taken to protect the data.

g. ***Predicted results:*** In this section, the resident or fellow will describe what they expect to find and the impact on the practice of medicine.

2. **Funding:** Many databases must be purchased. The databases can be purchased by the GME Research Department, after approval by the Director of Research. In some cases, the project may be referred to the GME Research Committee. Once purchased, the databases can be used for additional analyses by other residents, given all data use requirements are met. Each resident or fellow must write his or her own protocol.

3. IRB Review

a. All protocols will be submitted to the IRB for review.

b. National databases will usually qualify for an exemption; however, per current Federal guidelines, only the IRB can make this decision.

c. Local or hospital-level databases that include patient identifiers will typically require expedited review.

d. No data analysis should be done until IRB approval/exemption is granted.

e. Per standard policy, for data that include patient identifiers, only the individuals on the approved team list (IRB Form A-2) can have access to the data. For national databases, all elements of the data use agreement required by the database owner will be followed.

4. **Presentation and publication**
    a. To be eligible to present the work at the annual Ascension St. John Research Day, all the requirements in this document should be met and the resident should understand and be able to answer questions related to study design and analysis.
    b. For any publication, all requirements of the database owner must be followed (i.e. how the database is referenced etc.) In some instances, the database owner (for example, Blue Cross/Blue Shield) may wish to see any paper prior to submission to a journal.

Per current policy, all resident and fellow projects are submitted to the IRB by the ***GME medical researcher*** who is working with the resident on the protocol.

For any questions, please contact Dr. Szpunar at 313-343-7838 or susan.szpunar@ascension.org