Sight OUtcomes Research CollaborativE (SOURCE) Ophthalmology

Acceptable IRB Approval Types for analysis of de-identified data:

No IRB review of each site is required by UM, however should be obtained if required by the site's institution. An appropriate Unfunded Agreement (UFA)/Data Use Agreement (DUA) will be acquired.

- Note, for UFAs, the PI's HUM (not the REP), will be utilized in order to work within the U of M system utilized for agreements.
    - Use "HUM00134954 U-M Ophthalmology Data Repository: Characteristics of Patients with Ocular Disease"; this study is linked to the REP.

Describe the purpose, study question, study objectives or aims for this project:

The SOURCE Ophthalmology Data Repository (originated at the University of Michigan) contains clinical data from a consortium of academic ophthalmology departments and some private practices contributing de-identified electronic health records data on patients with ocular diseases seen at their institutions/practices.

This large dataset holds the unique potential to study a) risk factors of ocular disease b) utilization of eye care services, c) outcomes of eye care, d) sources of variation in eye care service use, and e) disparities and inequities in eye care. By combining point of care clinical information among multiple institutions, it is anticipated that one can more fully address questions that were previously unanswerable using traditional health care claims databases, though to date, researchers have never successfully tapped into this new data source to try to assess whether it can effectively be used to study patients with ocular diseases.

Describe how the results will be used including any plans for presentation or publication:

The SOURCE database will be cited in publications. Any analyses done by any consortium researcher will utilize de-identified information only, thus not meeting the definition of human subjects research. There will be no ability to re-identify patients whose data have been contributed to SOURCE.

Describe the data/information that will be collected for the study

The following information from the Electronic health records of patients with ocular diseases will be included:

- Patient Demographics
- Structured Data from Clinical Visits (Vision, IOP)
- Billing and Administrative data
- Patient Reported Outcomes Data
- Pathology Results
- Laboratory Test Results
- Medications

- Interventions/Treatment
- Data from Operative Reports
- Diagnostic Data

Only those who are contributing data will be able to retrieve data; this will be regulated by a Research Committee that will include a member from each site.

Describe if the data will include individually identifying information (e.g. name, DOB, MRN, email address, other codes; etc.):

The SOURCE repository will not include any individually identifying information. The data coming from each institution/private practice will be scrubbed of identifiers before being sent to Michigan Medicine, and the data that UM provides will be de-identified as well.

Data will be stored on Turbo and Yotabyte servers (UM servers) that are HIPAA-Compliant. Datavant de-identification software will be utilized, which converts HIPAA identifiers or removes them as necessary.

Will the data be linked to individual identifiers? How will you avoid duplicates? Will there be ongoing data feeds?

None of the individual sites, including UM, will have access to HIPAA identifiers once they are added to the SOURCE data set. The data will be completely de-identified once your IT runs it through the Datavant software.

Once the data is in SOURCE, there should be no link to identifiers: dates will be shifted, and the data will be tokenized only so that newly imputed data in subsequent feeds can be connected to the appropriate existing individual using the software (this is the mechanism of the software and is purposeful to make sure that we can say that the data is truly de-identified and to avoid having duplicate individuals).

In order to get your site data into SOURCE, a member of the local IT team at each site who will be responsible for extracting the data from EPIC/your medical record system will have access to data containing identifiers. Once they run this data through the Datavant de-identification software, all identifiers will be removed and replaced with 44 character encrypted hashed tokens. It is the tokenized data that will get sent in to SOURCE. Once the data is in SOURCE, only the software will be able to link to individuals; no individuals will be able to do so.

If your site wishes to maintain a copy of your site's extracted data containing the identifiers (before it runs through the Datavant software), so that you can use it for analyses specific to your site, you or someone at your site will be responsible for ensuring that any proper IRB approvals as needed are in place to work with it, and this would be considered a separate activity than what we cover in SOURCE.