CIDER Support Information

Clinical Data Extraction for Research and Quality Improvement Initiatives

What Is CIDER?

The Clinical Investigation Data Exploration Repository (CIDER) is a clinical patient data warehouse containing medical information from over 5 million patients within the BJC Healthcare system. Developed by the Center for Biomedical Informatics (CBMI), this HIPAA-compliant software allows for seamless but regulated electronic access to patient data through a powerful querying functionality and a secure, web-based graphical user interface. CIDER provides the opportunity for researchers to efficiently and securely query clinical data, and thus reduces barriers of translational research. CIDER is also a valuable resource in participant recruitment and epidemiological population assessments for process improvement initiatives.
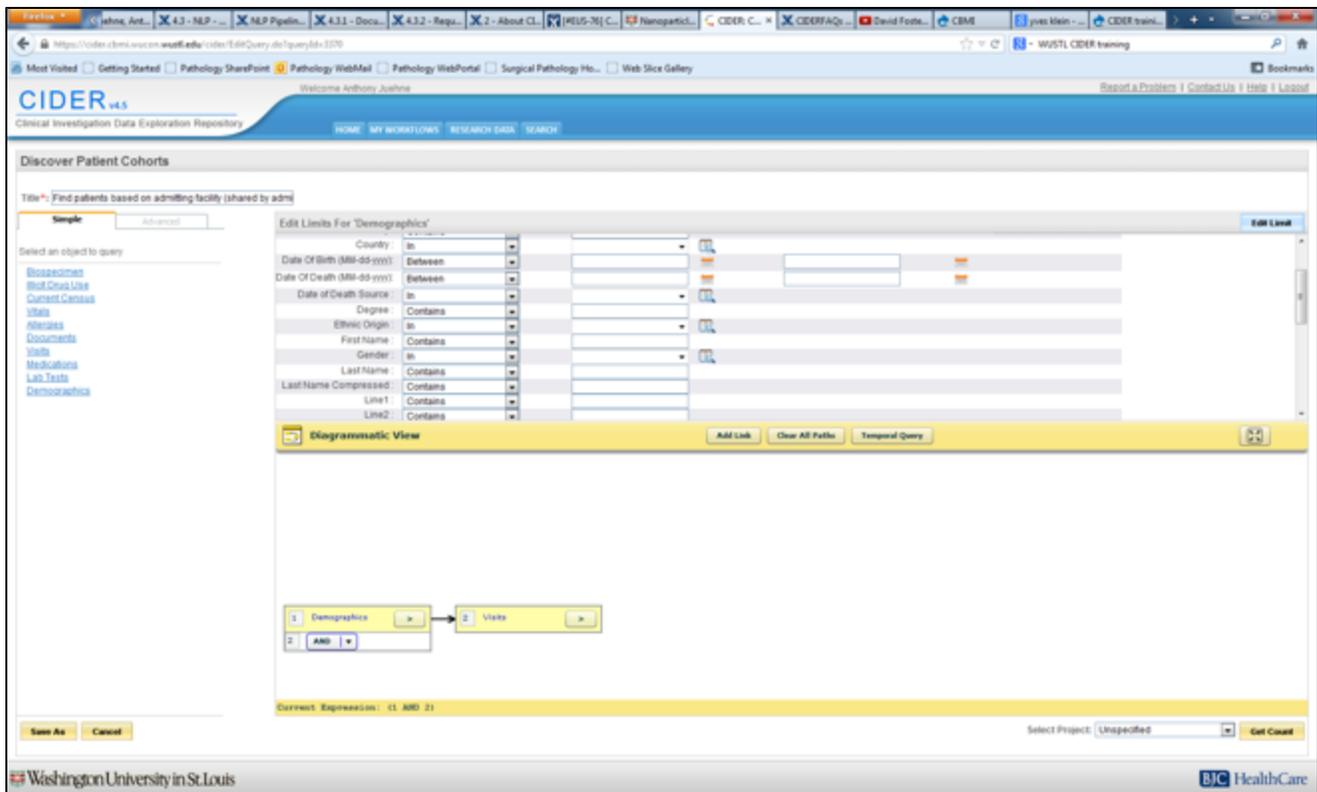
CIDER's web interface is highly intuitive and requires no knowledge of advanced coding or database schematics. All data are refreshed nightly and most queries populate within 5 minutes of execution. Prebuilt queries designed by CIDER administrators are shared with all users to aid in simple queries. However, users can elect to design their own queries to capture a more unique cohort of patient data.

What Data is Available?

CIDER's querying capabilities allow for the acquisition of data within eleven domains: Demographics, Visits, Lab Tests, Medication Orders, Vitals, Documents, Illicit Drug Use, Current Census, Allergies, Tumor Registry, and Biospecimens. CIDER contains information dating back to 1992 collected during inpatient and some outpatient clinical visits at 13 area hospitals under the umbrella of the BJC, Washington University, and other

healthcare institutes (early data is more limited, as the use of electronic records evolved slowly across the systems). Anonymous, de-identified, limited- identified, or fully- identified data can be exported depending on user privileges and IRB approval status (for research projects) or BJC approval (for QA/QI projects).

Natural Language Processing Capabilities

CIDER is fitted with Natural Language Processing software based upon IBM's LanguageWare which can access over one-thousand types of medical documents to extract and structure valuable information stored as unstructured free text. At this time, the NLP capabilities have been used to develop a structured data terminology for querying one commonly used domain for research that is not discretely stored within the EMR: illicit drug use.

In addition, the CBMI support team can assist researchers with the selective extraction of data from a group of documents. Use of NLP can give clinical researchers access to meaningful clinical information that previously required laborious extraction if it were to be recovered at all.

End User Support & Associated Costs

The CBMI offers three levels of support for the use of CIDER:

Level 1 Support (no charge)

- The researcher uses CIDER independently after completing training
- The CBMI Help Desk assists with technical issues
- The CBMI CIDER Data Support Team provides guidance on (maximum one hour of consultation):
    - building count/data queries
    - basic project feasibility - is the cohort discoverable or data available in CIDER

Level 2 Support (hourly charge)

- The researcher needs assistance with CIDER for more than one hour (4 hours minimum) for a short-term data project with defined requirements. The support includes:
    - Query development
    - Personalized training
    - Basic and/or specialized data activities (e.g. natural language processing to extract data from documents, limited data restructuring)
        - Users who require more assistance with data management activities can consult with the following resources available at WU/BJC:
            - Research Design and Biostatistics Group
        - IMPORTANT: Availability of data services depends on the availability of resources at the CBMI/HSIL at the time of request. At some times, we may only be able to provide raw data to users. Please contact our support team to inquire regarding our availability for data work.
- Current CBMI/I2 rates:

Funding Type	Member Type	Hourly Rate
Departmental funds or NIH/internal grant funds	Full rate	86.00
	ICTS membership	62.00
	SCC membership (project must be cancer-related to qualify)	46.00
Sponsored project	N/A	118.25

Getting Your Project Started

Prior to gaining CIDER access or data, this list of requirements must be fulfilled, depending on the level of support needed:

1. Level 1 Support – Researcher Uses CIDER

- User must be WU or BJC employee
- PI requirements
    - ICTS membership (free)
    - WU/BJC faculty member
    - Signed data confidentiality agreement
- General CIDER training (free)
- Completion of WU/BJC HIPAA training

- Signed data confidentiality agreement for user (if not PI)
- Access to either WUCON or BJC Carenet secure network (VPN may be needed for some users, as well as for use outside of WU/BJC networks)
- Project description
- Identified data access (i.e., access to PHI) requires the following (see 'Understanding Project Privileges' for further explanation):
    - IRB (if research) **or** BJC (if QA/QI) approval
    - IRB application must describe use of CIDER
    - SFTP server set-up (via Cyberduck) for secure download of data

2. Level 2 Support - CIDER Support Team runs queries for project (minimum of four hours billed)

- All requirements of Level 1 Support (#1 above)
- Standardized data request form (see Data Request Process below)
- JIT grant or other funding

Citing CIDER in Research Publications

1. Be sure to cite the **NIH CTSA Grant # UL1 TR000448** when your research was supported by ICTS/CTSA funds or any ICTS core service, including services at the CBMI.
2. **CIDER citation and DOI**: Washington University in St. Louis Center for Biomedical Informatics, 2010, Clinical Investigation Data Exploration Repository, project number: [fill in your project number], created at: [fill in the date of your query], doi:10.7936/B6CR0000.
3. If CIDER is used to identify your study cohort and/or to obtain data for your research, also **describe this in the Research Methods section** of your manuscript. For example: "Cases were selected using a searchable electronic medical record database (Clinical Investigation Data Exploration Repository, CIDER) maintained by Washington University's Center for Biomedical Informatics..."

Data Request Process (Level 2 Support)

The Standardized Data Request Form was developed to assist users with 1) identifying the patient cohort/population from whom they would like to obtain data, and 2) identifying the data required to accurately test study hypotheses or explore relationships among the variables.

While the form is a good place to start, frequently this is an iterative process when users engage with us for Level 2 support. The clinical data available is complex, both in terms of how clinical occurrences can be defined, as well as how it is stored and can be extracted. We also rely on users' clinical expertise for guidance on how to find the necessary data. Please understand that your engagement throughout this process is vital and expected for the successful completion of your data request.

Here are some tips for the process (most from http://ps.columbia.edu/CERS/how-work-programmer):

- Misunderstandings can and do occur. Try to be as clear as possible in your request.
    - E.g., "BJC" generally refers to the entire BJC Healthcare system and its affiliated hospitals, and should not be used to indicate from which facility you'd like to obtain data. Specify from exactly which location(s) you would like data, such as "BJH, WUSM, and SLCH."
- Identify where/how the data you need is stored in the pertinent clinical system.
    - E.g., data are contained in the "Emergency Department Physician Discharge Summary" document, or listed as discrete lab values in ClinDesk.
- Be realistic with turnaround times for data requests. At times we are busy with numerous existing projects and may not be able to prioritize your request.
- Obtaining data is an iterative process:
    - Multiple discussions with the investigator may be necessary to clarify the data request.
    - Once the data is explored, specific variable needs may need to be re-evaluated (e.g., a special lab test is found to be a "send out" and not stored discretely in the data).
    - You may wish to obtain more variables after the data is delivered to you, which may not be quick and easy to accomplish, depending on the activities required to pull your specific data. Please be patient and polite should this need occur.
- Note that your idea of how the data are structured may not be correct.
    - In most cases, it will be best to provide the data generated for you in separate files for each variable domain (e.g., one spreadsheet with all patient lab values, another for all patient diagnoses). This often makes the data less cumbersome to deal with compared to receiving it all in one large flat file, but requires the data to be merged carefully for analyses.
    - Data are frequently stored in a "normalized" fashion. This means there may be multiple rows of data for one patient, such as a row for each individual diagnosis, surgical procedure, lab test, etc., associated with a specific visit.
    - Be prepared to spend time reviewing the data to understand the variables and values delivered.
    - Allot time for data checking, cleaning, and reformatting once it is received. Data entry errors do occur in the clinical systems that feed into CIDER, and some data may have been collected using varying processes over the years.
- Users will be responsible for their own data management activities, although we can provide some consultation if needed. Other assistance can be obtained through resources such as the WUSM Biostatistics Core Resources center.

Disclaimer on Data Quality and Structure

Please be aware that our team cannot correct for missing data, data inaccuracies, or formatting issues within data as it originates within electronic medical record systems across BJC Healthcare and Washington University. If data is entered into the medical record inaccurately or altered when transferred across systems, these issues must be reported directly to BJC. Given these limitations, our core's services involve providing the data

in the form they are received by our applications. Advances are being made across BJC Healthcare and Washington University to ensure clean and reliable collection and storage of medical data; however, this move towards meaningful use is a fluid and ongoing process. Thus, we ask for your understanding that data can be a sometimes "messy" or "untidy" resource given the complexity of care and medical record systems provided across area hospitals. We strive to provide efficient and effective methods of extracting the most reliable medical data collected across hospital systems and ask for your understanding in facing any remaining constraints and barriers.

Contact Us

For more information, please contact us at help@bmi.wustl.edu.