



Best Practices: EHR Tool Selection and Query Development for Study Recruitment

OCTRI Research Forum

DATE: June 2, 2026 PRESENTED BY: Lindsey Zimmerman, Research Data Program Director, Office of the Chief Research Information Officer and Jiro Inoue, Research Analytics Manager, OCTRI

Agenda

- Cohort Query Development
- Reporting Tools Overview
- More Resources

Why is Accurate Query Development Important:

- It is important to accurately identify potential research participants. Misidentification can:
 - Be upsetting to patients
 - Cause reputational damage to the PI and OHSU
 - Can lead to compliance problems

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or data flow diagram. The lines are vertical and horizontal, with some diagonal connections, and the circles are placed at various points along these lines.

COHORT QUERY DEVELOPMENT

Know Your Data

- Know where and how data is stored in Epic
 - Data...
 - are often difficult to locate
 - are only as good as entered
 - may not mean what you think it means
 - Good data understanding is essential to proper tool selection, query development, correctness checking
 - not all data are available on all platforms
 - some things will be specific to OHSU
 - Include a clinician familiar with the area: If your PI isn't a clinician, a clinical expert can help guide you the the right data and verify correctness for your use-case

Cohort Definition

- Translate from "Clinical" to "Data"
 - Define phenotype
 - Set a timeframe
 - Apply other exclusions
 - Iterate and refine

Running example: Atrial fibrillation patient with low blood pressure on medical management

Patient Phenotype

- How are you defining your cohort?
 - Diagnosis? Procedure? Medication?
 - Hierarchical classifications and codes
 - Individual items or groups
 - Variable precision (broad classes, anatomical locations, combinations, chronic, "unspecified" etc)
 - Coding quality/coverage can also vary
 - Sometimes multiple code vocabularies can be used
 - Registries:
 - Check with clinical expert on details!
 - Other: Clinical note text search? Lab results? Flowsheets?
 - Hard to find in self-serve tools
 - Sometimes you need surrogates or other definitions
 - Watch out for confounders (BMI/pregnancy)

Phenotypes: Diagnosis

- Which coding system?
 - ICD-10-CM, SNOMED
- Which diagnosis?
 - Encounter diagnoses
 - Problem list
 - Claims
 - Chief complaints – not the same!
- Exclusions
 - Secondary (not primary) diagnoses
 - Chronic/prevalent conditions
 - Comorbidities
 - Historical or old diagnoses

Example: First diagnosis of any AF (ICD-10-CM I48.*) - all AF And flutter (13 kinds).

Phenotypes: Setting

- How is the patient interacting with OHSU?
- Encounter type
 - Inpatient
 - Outpatient
 - Emergency
 - Telehealth
 - By clinic
 - By provider
- How many visits over time?
- Referrals, orders, non-face-to-face encounters

Example: OP visits in KCVI arrhythmia clinic every year

Phenotypes: Procedure

- CPT codes codes are typically used
 - But OHSU also has proprietary codes
 - Clinical expertise, text search
- Procedures can cover a lot of different things - sometimes you need to check specific categories. Depends on the data model!
 - Imaging
 - Surgery
 - Radiation therapy

Example: No CV procedure

Phenotypes: Medications

- Epic has a built-in hierarchy for medications
- Some systems can use RxNorm, AHFS etc
- Hospital administered or prescribed?
 - Prescribed doesn't mean taken!
 - No medication doesn't mean not taken!
- Can be useful instead of diagnosis

Example: Prescribed Amiodarone

Phenotypes: Results/Values

- Some results are really complicated (e.g. microbiology!)
- Lab results
 - Lots of dates: ordering, performed, resulting etc.
- Flowsheet
 - Could be anything from 1/0 flag to text
 - Find these in Epic Hyperspace!
- Vitals
 - Usually standard flowsheet values
- Freetext note search
 - These can be complicated. AI/LLM tool development in progress!

Example: Blood pressure below 90/60

Timeframe Scope

- Most cohorts need a timeframe scope
- Index event
 - Phenotype events (dx, proc, result etc.)
 - Administrative events (admit, discharge etc.)
 - How to deal with multiples?
- Timeframe
 - Beginning/End dates
 - Follow-up periods?
 - Did something change over timeframe?

Example: Jan 1, 2023 and Dec 31, 2025. Index event is first recorded diagnosis.

Other Exclusions

- Many other exclusions are possible:
 - Demographics: age, sex, language, race, ethnicity
 - Location: in-state, zip, RUCA, FIPS
 - Active patients (encounter within 2 years)
 - Deceased patients
 - In Epic, this is POSITIVE ONLY
 - Too many/severe comorbidities
- Not everything can be checked at the query development stage!

Example: Age 18+ on first diagnosis

Example case

Recently diagnosed, adult atrial fibrillation patient with low blood pressure on medical management being treated at KCVI

- Age 18+
- BP under 90/60
- First diagnosis (I48.*) between Jan 1, 2023 and Dec 31, 2025
- Yearly visits KCVI arrhythmia clinic since diagnosis
- Prescribed amiodarone
- No cardiovascular procedures on record

Iterate and Refine

- Does the count make sense? How many can you actually recruit? Is this enough for your study?
- Most self-serve tools will allow adjustments
- Adjusting counts:
 - Increase/reduce timeframe
 - Loosen/refine diagnosis hierarchy
 - Remove/add criteria or evidence (lab test etc)

This cohort is NOT your final contact list!

Additional Considerations

- Familiarize yourself with the recruitment plan in your IRB approved protocol
 - Appropriate tool
 - Restrictions/requirements for the use of PHI
- Before you contact anybody...
 - Using MyChart? - Check for MyChart recruitment opt out
 - Data could have changed (even from yesterday)
 - Verify information in Epic Hyperpace before you call people
 - [Epic Access Request](#)

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or data flow diagram. The lines are vertical and horizontal, with some diagonal connections, and the circles are placed at various points along these lines.

REPORTING TOOLS OVERVIEW



Research Reporting Options:

- The focus today will be on the main tools to help identify patient cohorts:
 - Research Data Services
 - Self-service tools

Self-Service Tool Comparison

	TriNetX (OCTRI)	SlicerDicer (ITG)	Reporting Workbench (RWB) (ITG)	Business Objects ITG)
Best For	Multisite Research, Cohort discovery	Quick insights in Epic	Actionable operational reporting	System-wide analytics
Access	Access is via a web app.	Accessed via Epic Hyperspace.	Accessed via Epic Hyperspace.	Access via the BI launch pad portal.
Training	Training is required (sign up via Compass)	Training is not required but training materials are available.	Training is not required but training materials are available.	Training is required (sign up via Compass)
Cost	No cost for self-service aggregate numbers, MRNs or LDS (w/out MRNs) – cost for full-service custom data set	No cost	No cost	No cost
Delay	Data updated weekly	Data updated daily from Epic Clarity (data source is Epic Caboodle Data Warehouse)	Main source is Chronicles, so can include real-time information	Data updated daily from Epic Clarity (data source is Epic Caboodle Data Warehouse)
Access to PHI	PHI can be accessed via informatics team	PHI can be accessed automatically via the tool.	PHI can be accessed automatically via the tool.	PHI can be accessed automatically via the tool.
Data Omissions	No genetic opt-out patients, no affiliate visit data, data that cannot be mapped to standards	Permissions based (e.g., no affiliate-only data w/out permissions)	Permissions based (e.g., no affiliate-only data w/out permissions)	Permissions based (e.g., no affiliate-only data w/out permissions)
Support	The informatics team is available to help with queries or extended data needs	Demonstrations and troubleshooting	Demonstrations and troubleshooting	No central research support for queries.
Complexity	Low-moderate	Low	Moderate	High

TriNetX:

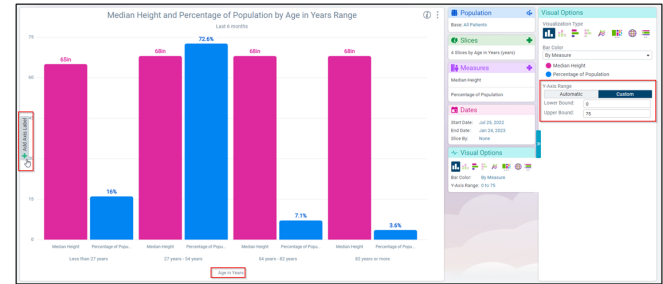
- Pros
 - Quickly generate cohort counts and subgroup counts
 - Request MRN lists for free
- Cons
 - TriNetX data model may not reflect OHSU organization (e.g. specific clinics)
 - Genetic opt-out patients are excluded
- Use Case examples:
 - Self-serve complex cohort development
 - Request MRN list for the same cohort
 - Request customized data set from OCTRI RDS for cohort data including MRN and other clinical data

RDS:

- Pros
 - Customized recruitment lists based on nuanced criteria with OCTRI data analyst support
 - Most data in the EHR can be used for inclusion criteria
- Cons
 - Fee charged based on analyst time
- Use Case examples:
 - My inclusion criteria are only stored in clinical notes
 - I have complex inclusion/exclusion criteria across multiple domains
 - I need help defining my cohort because the criteria are not well captured in standard data fields
 - I want to exclude patients based on OHA death data

SlicerDicer and Reporting Workbench

- Main display
 - SlicerDicer: graph visualizations
 - Reporting Workbench: patient lists
- Pros
 - Quick and Easy
 - Access to PHI (w/ IRB approval)
 - Good chart review workflows
- Cons
 - Epic access required
 - No lists within lists (generally)
 - Not good for complex criteria
 - Not good for unstructured data
 - Can be difficult to identify which data fields to use



Patient	MRN	DOB	Current Medications	PCP
John Smith	1234567891	05/05/3035	Aspirin, Tylenol	Dr. Smith

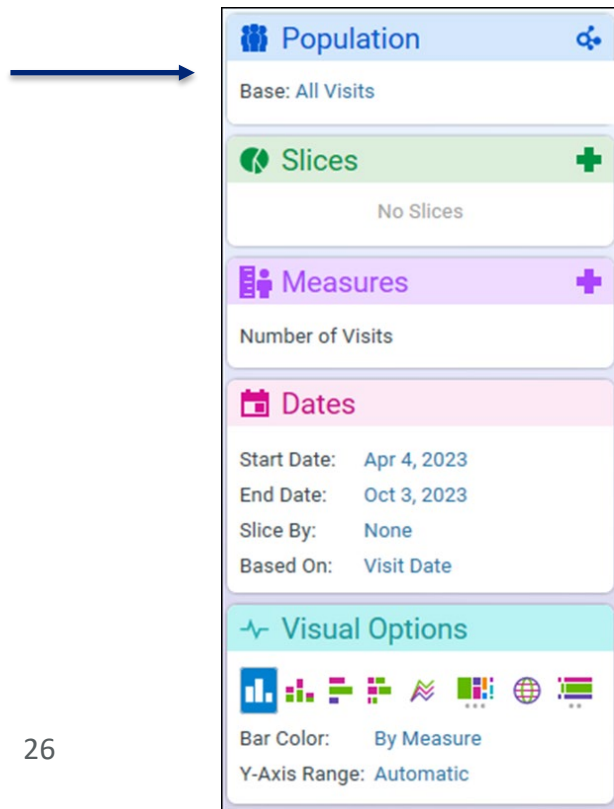
SlicerDicer:


- Pros
 - Can be used for data exploration as default view is de-identified
 - Sequential criteria
- Cons
 - Limited display criteria at the patient level
- Use Case examples:
 - I am recruiting adults with poorly controlled type 2 diabetes seen in the last 12 months
 - SlicerDicer can answer if there are enough patients, what clinics see the most patients, how tight eligibility criteria can be made before recruitment becomes unrealistic
 - SlicerDicer can also produce a list of patients with email and phone number

Reporting Workbench:


- Pros
 - Many templates to choose from (find patients, labs, appointments, etc.)
 - Many display criteria to choose from
- Cons
 - You may not find everything you are looking for with one template
- Use Case examples:
 - I would like to find all patients scheduled for an appointment with Dr. Smith over the next week w/ contact info
 - I would like to find all patients who are between the ages of 18 and 25 and have a diagnosis of sarcoma on their health record and show their last blood pressure, next appointment date and contact information

Main Locations for Cohort Criteria and Display:




Population 

Base: All Visits

Slices 

No Slices

Measures 

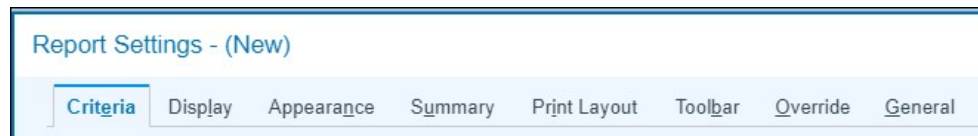
Number of Visits

Dates

Start Date: Apr 4, 2023
End Date: Oct 3, 2023
Slice By: None
Based On: Visit Date

Visual Options

Bar Color: By Measure
Y-Axis Range: Automatic



Report Settings - (New)

Criteria Display Appearance Summary Print Layout Toolbar Override General

*Note that options may change based on model or template, the following are examples of options to use for specific best practices

	SlicerDicer	Reporting Workbench (with IRB approval)
Validation (with IRB approval)	Chart search – right click on patient in Detail Table and Crosstab	Chart search – right click on patient in report results
Deceased individuals	Patient Demographics – Patient Status	Patient Status (alive/deceased)
Comorbidities	Each criteria will have a settings button (⚙️) that you can use to exclude the criterion, in addition to ways to link criteria	Reporting workbench has customizable logic that you can use to exclude criteria (e.g., 1 and 2 and not 3)

A decorative graphic on the left side of the slide, consisting of a complex network of light blue lines and circles. The lines are vertical and horizontal, with some branching out, resembling a circuit board or a data network. The circles are small and white, connected to the lines at various points.

MORE RESOURCES



- inouej@ohsu.edu, octri@ohsu.edu, Zimmerli@ohsu.edu, researchdata@ohsu.edu
- [OCTRI Informatics Page](#)
- [Research Data Website](#), [Epic data for research page](#)
- [Previous OCTRI forums](#):
 - Respectful First Contact: Best practices for initial communications with potential research participants
 - Transforming Your Grant into a Protocol
- The [Biostatistics and Design Program](#) is also a good resource for discussing study design prior to beginning recruitment.



Thank You