

2017 UCSF Clinical Data Colloquium Notes: Overview

May 10, 2017

Five common scenarios

1. I need **counts** of patients, I don't need a dataset yet
2. I need **de-identified data** for secondary data analysis
3. I need **EHR data, including real dates and zip codes**; don't need names or MRNs
4. I need data from the **EHR, including identifiers**, such as names, MRNs
5. I want to **recruit** using data from the EHR

1. I need **counts** of patients

Example scenarios:

- Obtain counts for a grant, sample size calculation, feasibility, hypothesis generation
- How many patients have been discharged from the hospital by year with COPD as one of the discharge diagnoses?

Relevant Notes:

- Several self-serve/DIY tools available (even across UC)
- Consultants available for recharge for custom counts

2. I need **de-identified data**, for secondary data analysis

Example scenario:

- I am developing a predictive model to identify patients at high risk for hypoglycemia. I need 20 covariates and outcome data.

Relevant notes:

- De'ided data available; may use without IRB and DIY
- Typically very large files; need a plan for data management, tools & storage
- Consultants available for recharge to use these files, for custom data extractions, data management etc.

NOTE: Scenarios 3-5 cover more sensitive data, i.e., includes identifiers and more. As such, there are compliance, regulatory and security considerations that apply to these scenarios, such as required IRB approval, use of Quick Disclosure system, secured data storage, encryption in transit.

And there are further requirements if you are planning to share the data outside of UCSF.

3. I need **EHR data, including real dates & zips. (a Limited Dataset)**

Example scenario:

- I am interested to study the seasonal impact of the flu and location (rural versus urban environment)

Relevant notes:

- This is called a **Limited Dataset**
- Use UCSF consultant to extract data
- Multiple requirements e.g., IRB, Data Use Agreement, other regulatory requirements

4. I need data from the **EHR, including identifiers**

Example scenarios:

- I have MRNs and need data for these patients
- I need *ongoing and updated* data on consented patients for my study
- I want to mine “notes” fields from the EHR

Relevant notes:

- Use UCSF consultant to extract data
- Multiple requirements e.g., IRB, secure environment, use of Quick Disclosure system etc.
- Notes fields in the EHR are always considered “Identified data”

5. I want to **recruit** using data from the EHR

Relevant notes:

- CTSI has a service to act as Honest Broker and mail letters for you.
<http://recruit.ucsf.edu>

Summary

Scenario	Self-serve (Free)	Consult Required (May have recharge)	IRB needed	Requires MyResearch account or other secure environment	Includes clinical notes	UC Health data available in addition to UCSF data
Counts	Yes	No	No	No	No	Yes
De- <u>ided</u> data	Yes	No	No	No	No	Yes *
Limited data	No	Yes	Yes	Yes	No	Yes *
Id'ed data	No	Yes	Yes	Yes	Yes	Yes *
Recruitment	No	Yes	Yes	Yes	No	Yes *

* Process to obtain data from UC Health sites can be very lengthy

Overarching Considerations - Have a Plan!

For all data:

- Have a plan to use it
 - Receiving large files
 - Secure, ample storage
 - Data analysis tools

For limited and/or identified data:

- Have a plan for completing & complying with regulatory and security requirements
- All regulatory/compliance requirements must be in place before a programmer will work on your request