



University of California
San Francisco

Overview: Using clinical data for research

Leslie Yuan
Clinical and Translational Science Institute (CTSI)



Image: chiefmartec.com

Five common scenarios



- 1. Counts**
- 2. De-identified data**
- 3. EHR data, including real dates and zip codes**
- 4. EHR, including personal identifiers**
- 5. Recruit**

1

I need **counts** of patients

- 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?



Image: readingcraze.com

2

I need **de-identified data** for secondary data analysis

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



Image: ketogenic-diet-resource.com

3

I need **EHR data**, including real dates & zips

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



Image: globe-views.com

4

I need data from the **EHR**, including **personal identifiers**

- 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



Image: physiciansweekly.com

5

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

<http://recruit.ucsf.edu>

- CTSI mailing service available



Image: usmailinghouse.com

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-ided 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

Have a plan!



Image: advancedlifekills.com