This is an NIH funded project (K01AI168579-01) whose primary goal is to describe social and environmental determinants of disparities in inpatient and emergency visits in US metropolitan areas. We will use two rich encounter-level national claims datasets, available from the Health Care Cost and Utilization Project (HCUP): The State Inpatient Databases (SID) and the State Emergency Department Databases (SEDD). **HCUP data we will use may potentially be in the terabyte scale. As we develop infrastructure for this project, we want to evaluate the high-performance computing at the Drexel URCF as a potential solution for our storage or computing needs. More specifically, we would like to test the Picotte cluster deploying large scale models and explore any storage options; we will not be hosting any of our own clusters.**

# Tech specification questions

(Note I’m being quite vague here, as I’m not sure exact what software we will use or site of models). But I think its enough for them to get a general idea of what we will be doing.

* URL for inclusion on the UCRF? (edited)

Not sure what this means. If they want a lab website? We can just use the UHC or set one up for HCUP.

* Software and Codes: List any software and codes you intend to use on the cluster (edited)
* **R**
* **SAS**
* **Python**
* **Stata**
* Development tools required: List any development tools required (e.g. compilers, libraries)
* **dockers**
* Typical number of processor cores per job

**1-4 cores**

* Typical wall clock time per job

**TBD. Rough estimate 1 -72 hours**

* Use of GPU:   
  **No**
* Expected amount of persistent disk space required (GB)  
   **200 GB**