# A Research Centered Approach to Sharing Patient Data While Protecting Patient Privacy

Qingbo Wang, Gaurav Luthria

# Goals

**Objective:** Identify a solution for sharing patient data while maintaining patient privacy

**Our Approach (Two Fold)**

- **Advanced Querying Language**
  - Group_by / aggregation / custom filtering / statistical test
  - Custom threshold to exclude outliers
  - Keeps the raw data secure in the cloud

- **Realistic Synthetic Data Generation from the Query**
  - Robust
  - Produce Realistic data
  - Suppress Outliers (outliers are the most easily identifiable)
  - Does not require model complexities or parameters
  - Work with smaller datasets

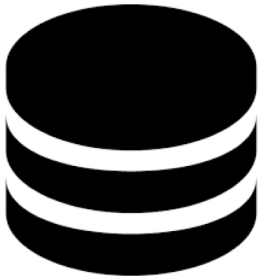
## Query the Original Data

```
d = d[d.REGION_C==1]
d.groupby("EVENT").agg("AVGDOS_N").
describe()
```

```
case = np.array(d[d.MALE]["EVENT"])
oth = np.array(d[~d.MALE]["EVENT"])
stats.ttest_ind(case, oth)
```

```
case = np.array(d[d["AGE"]<70]
["TRTN"])
oth = np.array(d[d["AGE"]>70]
["TRTN"])
tst = stats.chisquare(case, oth)
```

(Cloud)



Real Data

Efficacy Sensitivity Analysis

Generate Synthetic Data from Query

Downstream Analysis Use Vivli Platform to validate analysis on real data

| MASK_ID | | E_TEXT | TT_EVENT | EVENT |
|---|---|---|---|---|
| 0 | | RFS back date recurrence to midpoint | 805 | 0.0 |
| 0 | | | | |
| 0 | | | | |

| MASK_ID | PARNAM1C | PARNAM1A | PARUNT1C | LABRSL1N |
|---|---|---|---|---|
| 0 | CREA | Creatinine | mg/dL | 10.666667 |
| 0 | PLAT2 | Platelet Count | thous/mm3 | 216.666667 |
| 0 | SGOT | AST | U/L | 83.333333 |
| 0 | SGPT | ALT | U/L | 131.666667 |

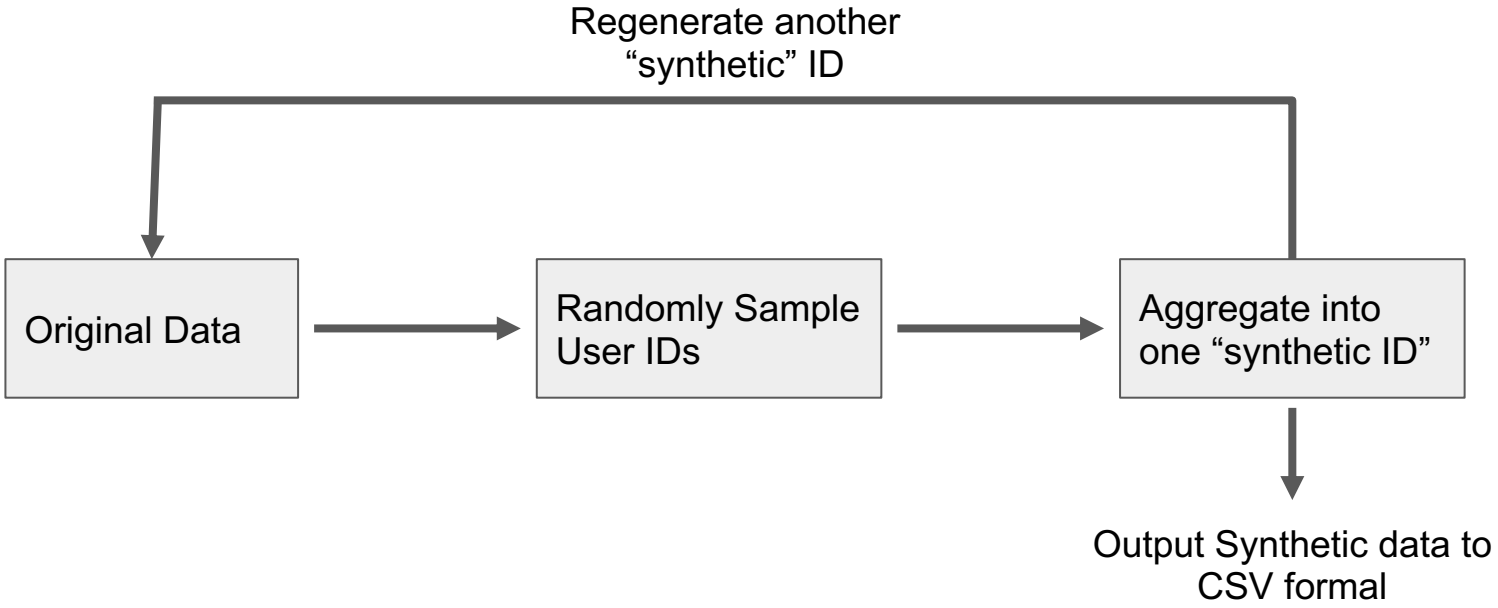
Lab Results

| MASK_ID | TERMREAS | AVGDOS_N | AVGDOS_C |
|---|---|---|---|
| 0 | 4.0 | 292.623333 | 1.0 |
| 1 | 4.0 | 256.606667 | 1.0 |
| 2 | 4.0 | 294.456667 | 1.0 |
| 3 | 1.0 | 298.210000 | 1.0 |

Drug Exposure Data

# Generating Synthetic Data

Want to Generate Synthetic Data that is Robust, Realistic, Suppresses Outliers
**Method: Permutation Test**



# Demo

https://drive.google.com/file/d/1nPdZNeYrEvshGEoLx7ntUHco3-gpT0kq/view?usp=sharing