

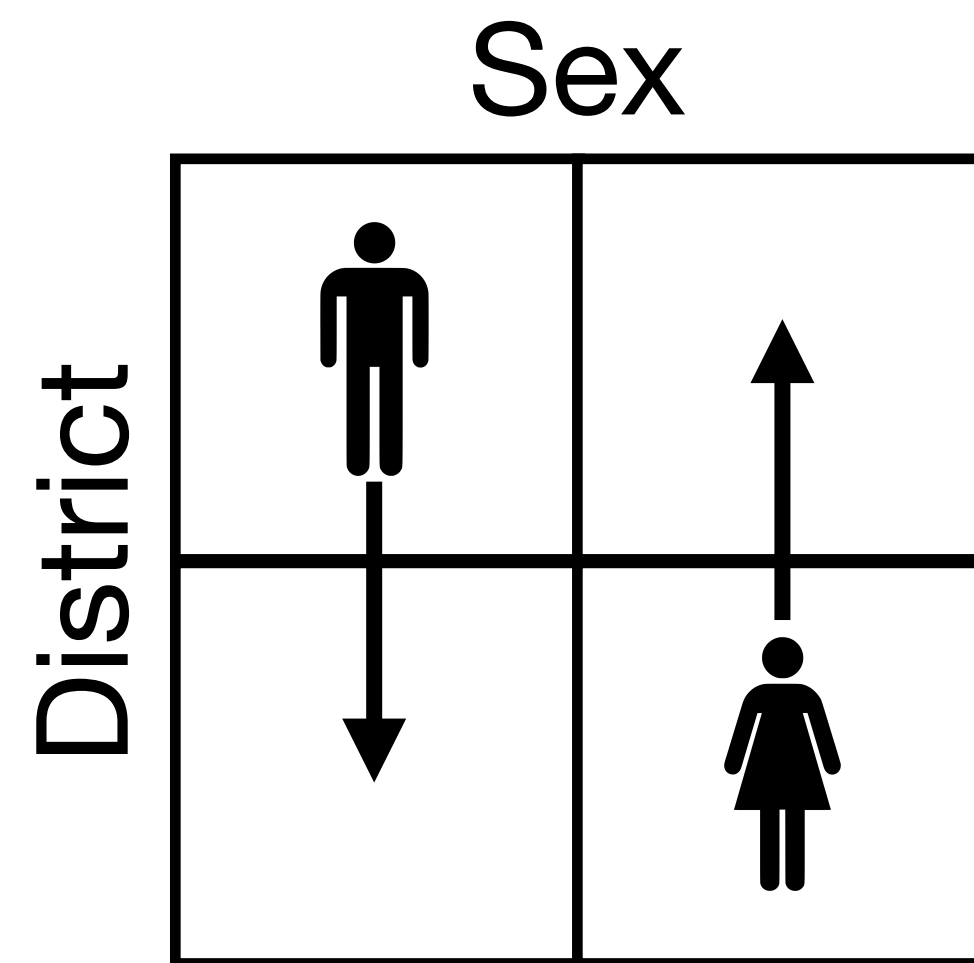
Valid Inference under Data Swapping

Kevin Eng, Ruobin Gong, Minge Xie

Data Swapping

What is it?

- Data swapping is a **statistical disclosure control** method used to protect confidentiality in contingency tables.



Data Swap

161	37	256
95	107	144
256	144	

→

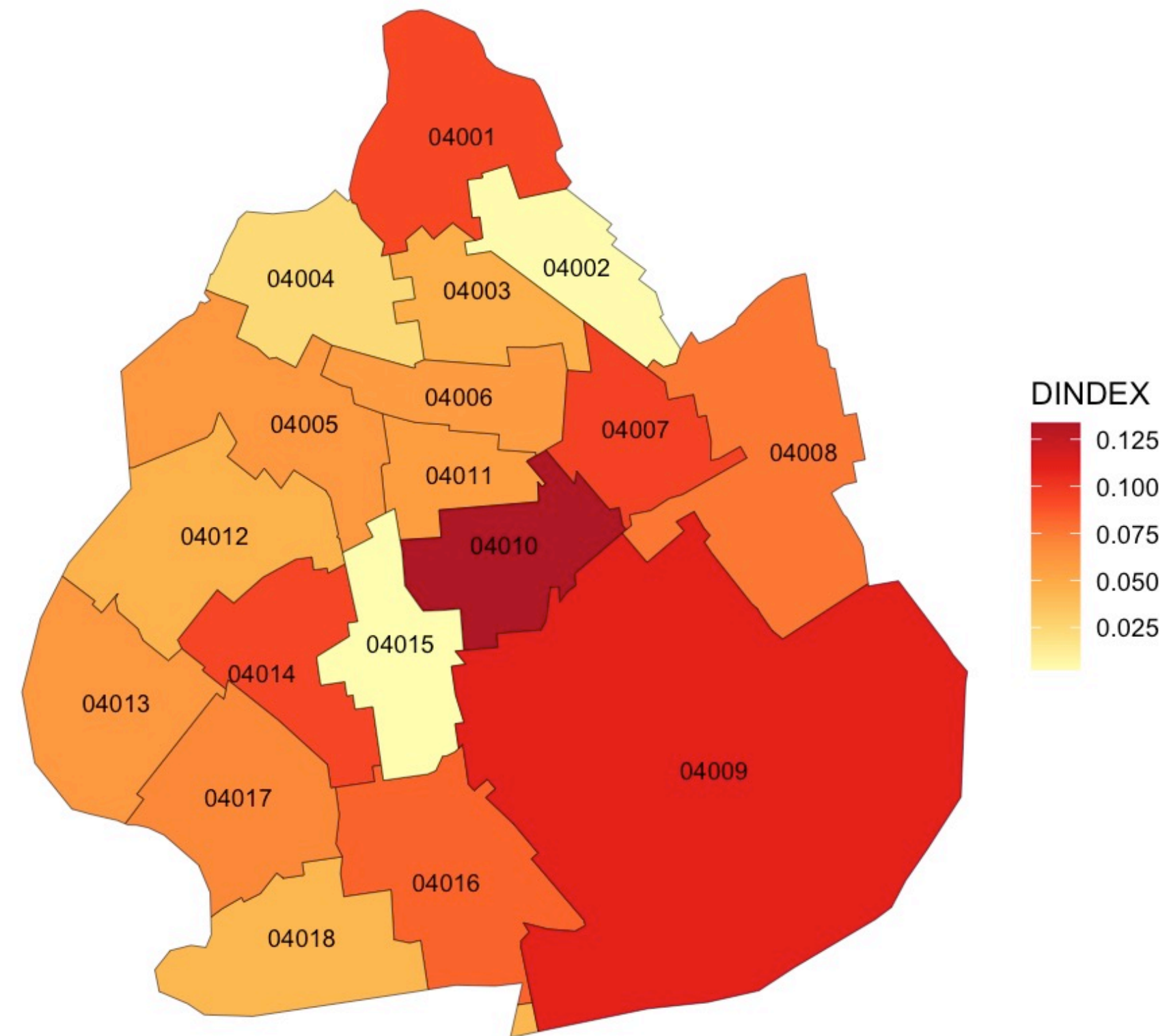
158	40	256
98	104	144
256	144	

The diagram illustrates a data swap operation. An arrow labeled "Data Swap" points from the initial table to the resulting table. In the initial table, the top-left cell (161) and bottom-right cell (107) are swapped. In the resulting table, the top-left cell (158) and bottom-right cell (104) are swapped. The row and column totals remain unchanged.

Applications of Data Swapping

American Community Survey

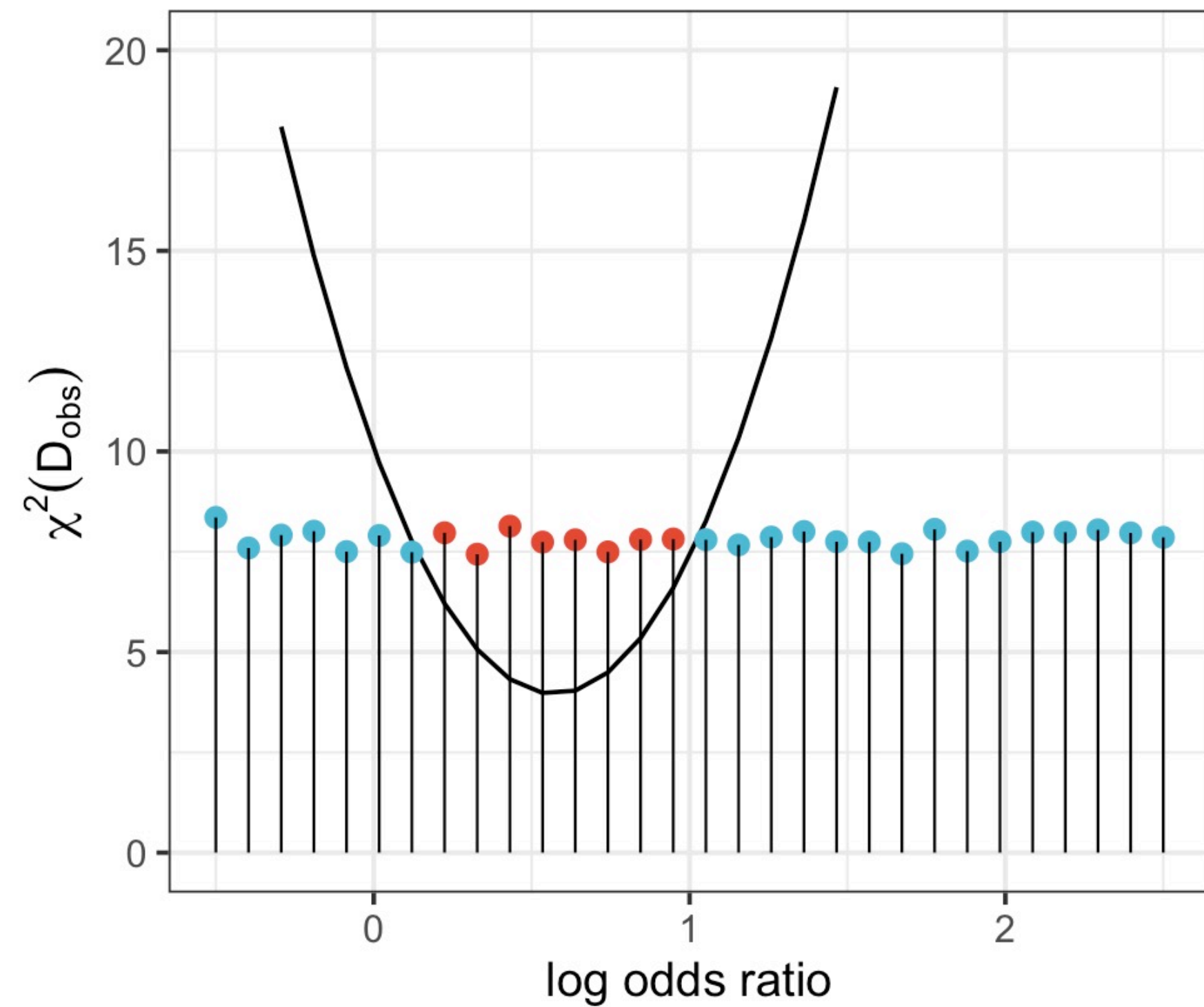
PUMAs in Kings County, NY
For 2014-2019 ACS Data



Simulation Based Inference (SBI)

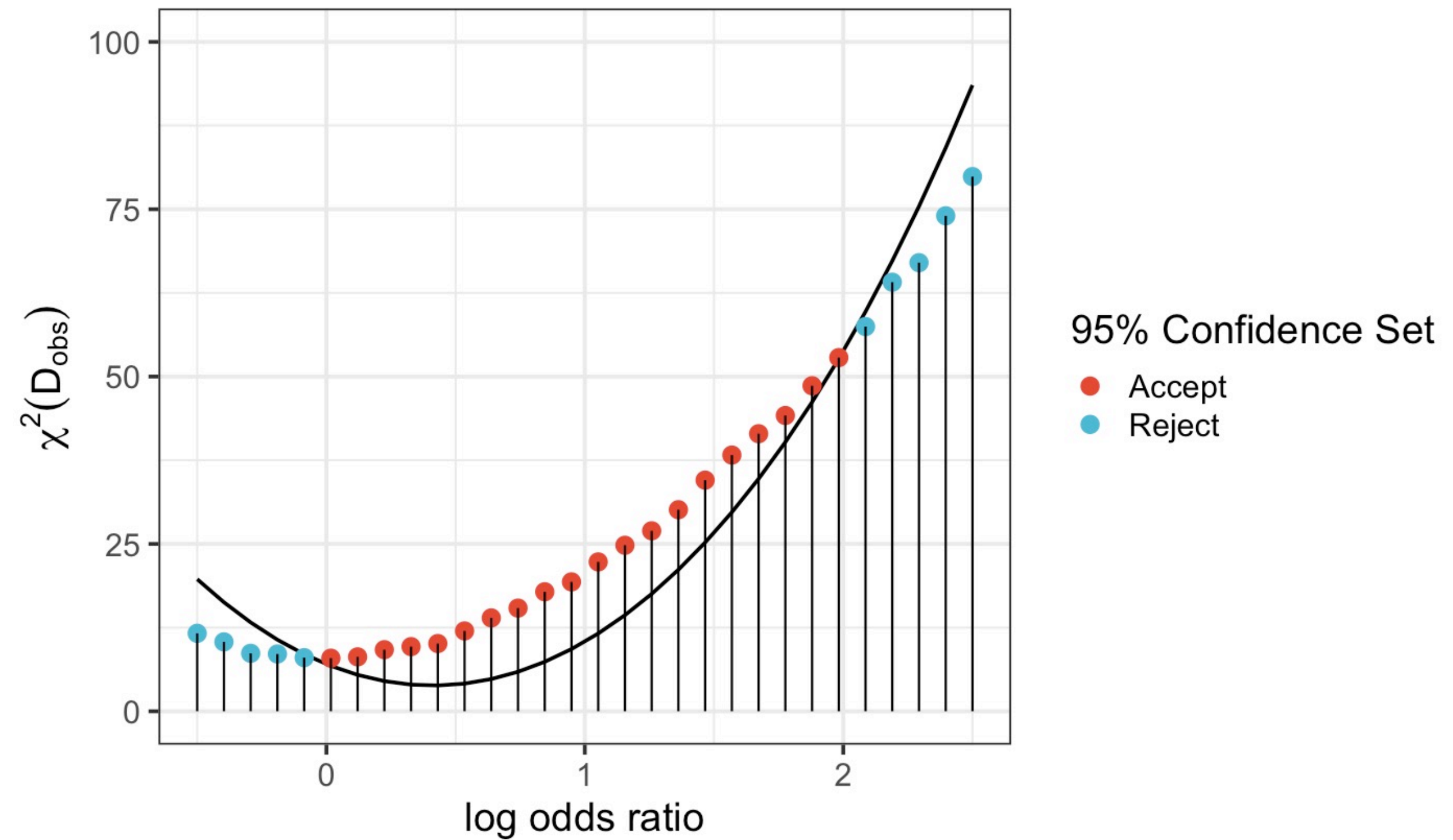
Calibrating Cut-off Values with Monte Carlo

swap rate = 0



95% CI: (1.25, 2.58)

swap rate = .60



95% CI: (1.01, 7.26)

State-of-the-art SBI

- Machine learned test statistic, and calibrated critical values.
- Valid finite sample confidence intervals.
- Computationally tractable in high-dimension spaces.

