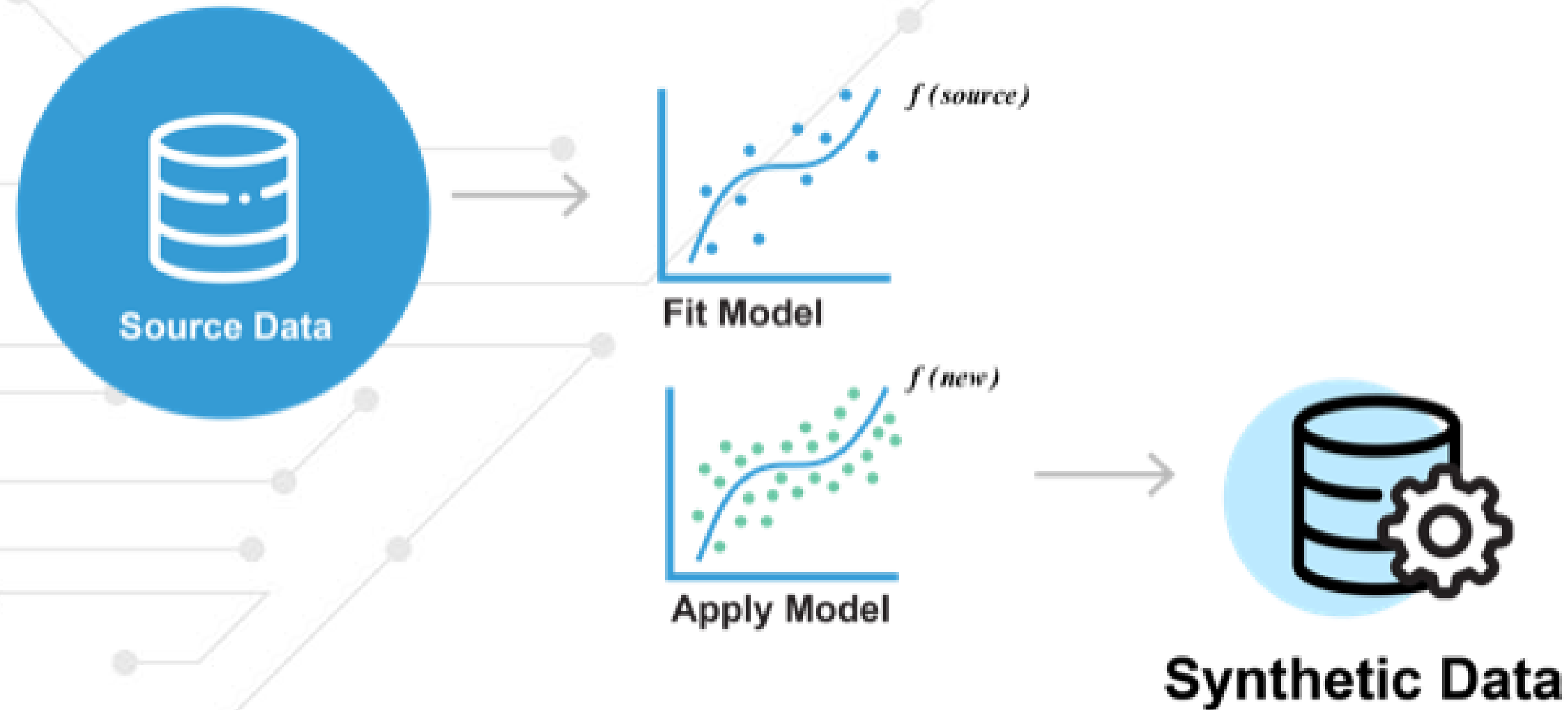


Enabling COVID-19 Data Access Using Data Synthesis

Khaled El Emam

27th October 2020

The Synthesis Process

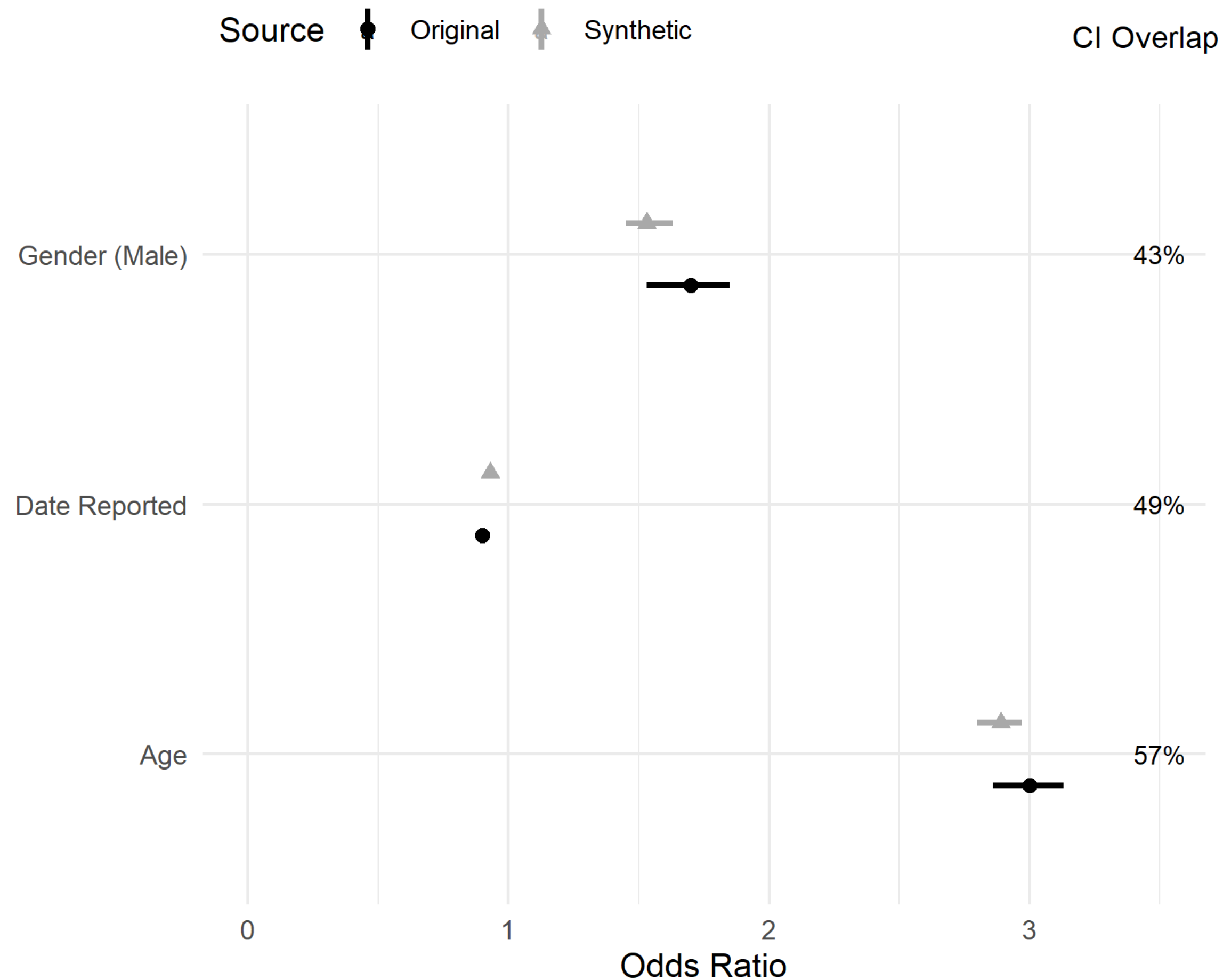


Privacy Assurance

- Matching a synthetic record to a real person
- Given the match, would an adversary learn something new

Real Data	Synthetic Data
0.034	0.0056

Logistic Regression Model



Variable	Real Data	Synthetic Data
Age	3*	2.89*
Gender (Male)	1.7*	1.53*
Date Reported	0.9*	0.93*