CERIAS

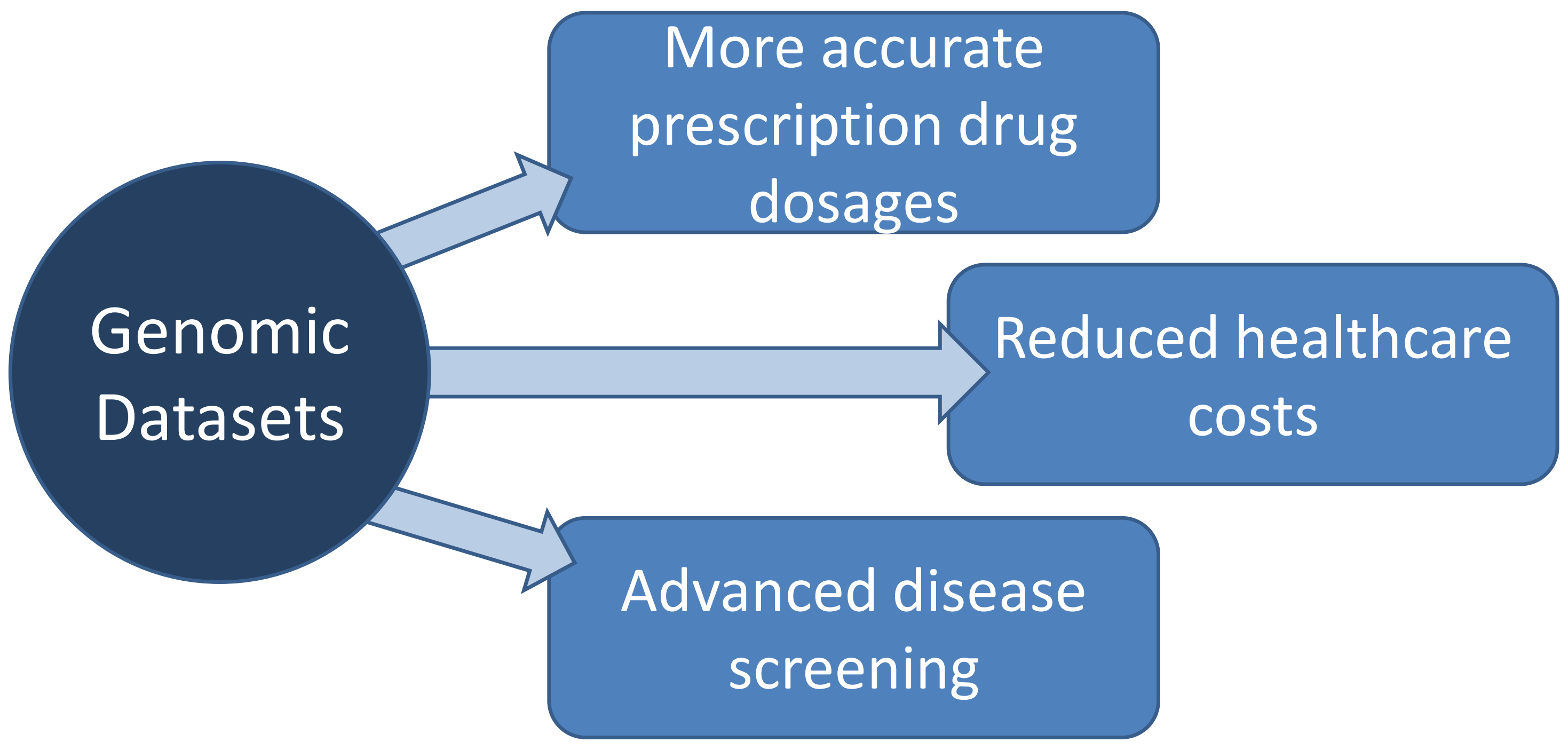
The Center for Education and Research in Information Assurance and Security

Anonymity and Security in Genomic Datasets

Kristine M. Arthur, *Purdue University*; Advisor: John A. Springer, *Purdue University;* Principal Investigator: Melissa J. Dark, *Purdue University*

Introduction

The purpose of this research is to define best practices for preserving anonymity and security in genomic datasets from currently existing privacy methods.



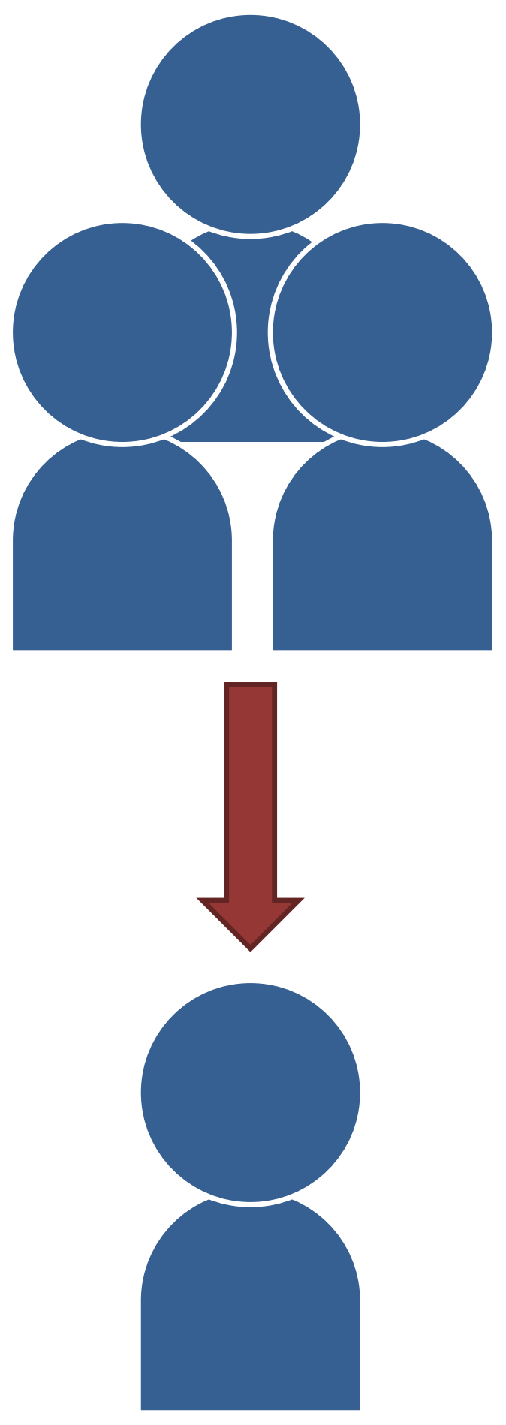
Genomic Datasets

More accurate prescription drug dosages

Reduced healthcare costs

Advanced disease screening

Threat

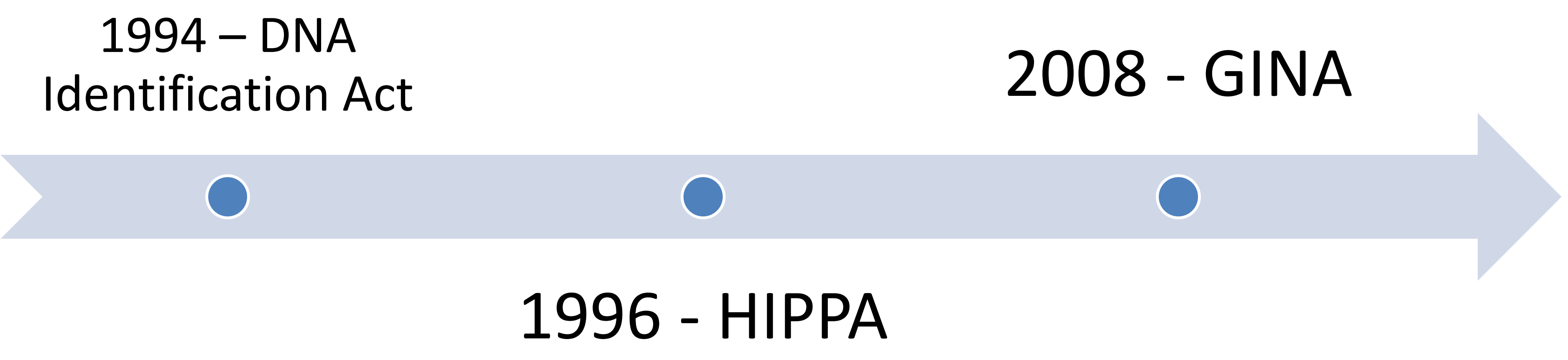


Denial of health coverage

Negatively affect consideration for employment

Increased healthcare costs

Policy Response



1994 – DNA Identification Act

1996 - HIPPA

2008 - GINA

De-Identification Methods

k-Anonymity
Records are arranged into equivalence classes with at least k-1 other records

k-Anonymity alone does not sufficiently protect against attribute disclosure

Gender	Age	DNA
Female	<25	CTGA
Female	<25	CTGA
Male	<25	ACTG
Male	<25	TCGA

l-Diversity
Each equivalence class represents as least l distinct sensitive attributes

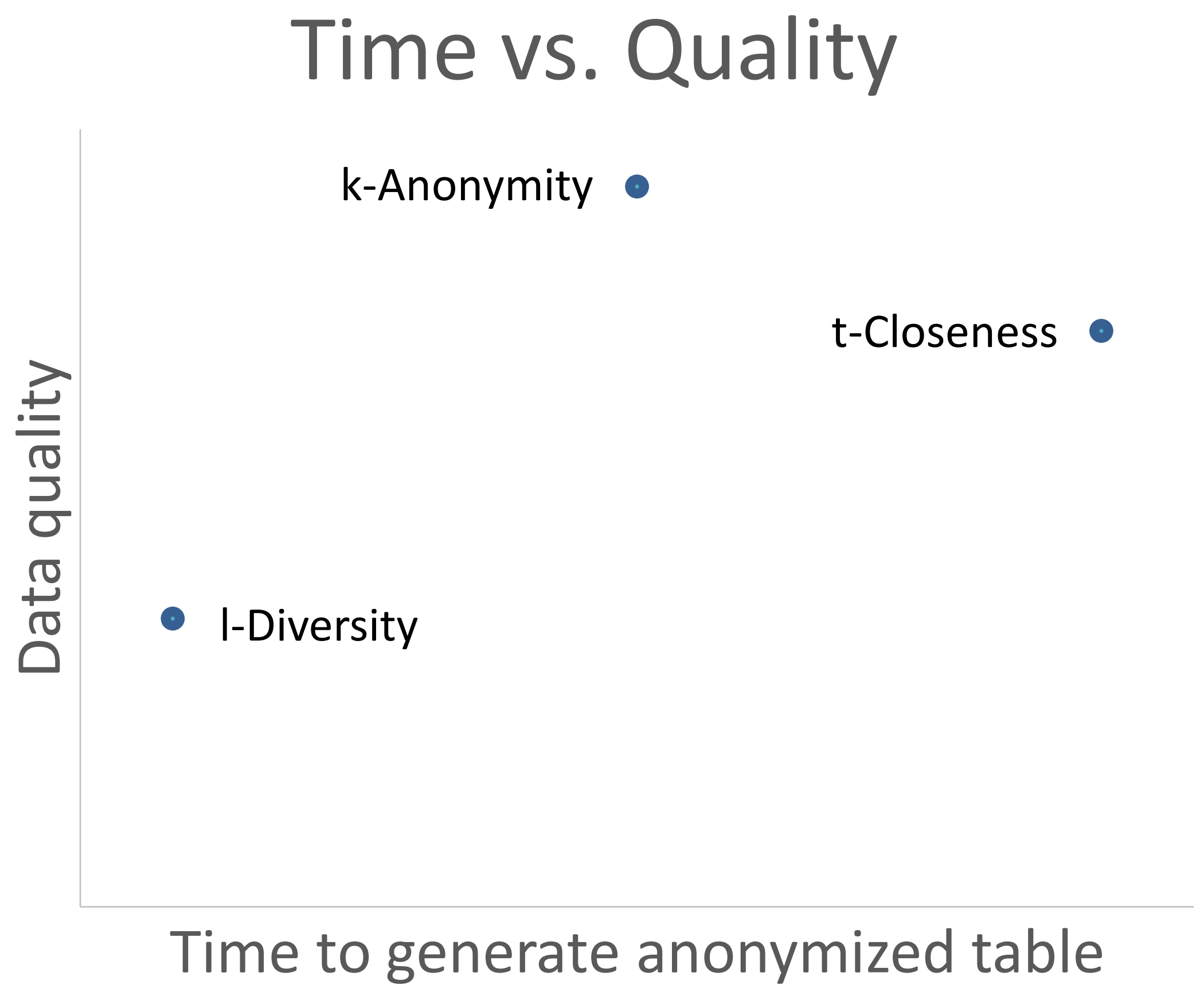
L-Diversity can misrepresent the occurrence of an attribute in the population

Gender	Age	DNA
Male	<25	CTGA
Male	<25	CTGA
Male	<25	ACTG
Male	<25	CTGA

t-Closeness
The distribution of sensitive attributes in each equivalence class is within a threshold t of the overall distribution

t-Closeness releases data which is close in distribution to the overall population

Gender	Age	DNA
Female	>25	CTGA
Female	>25	GTAC
Female	>25	ACTG
Female	>25	CGTA



Both k-Anonymity and t-Closeness require longer running times to generate anonymized tables.

However, l-Diversity produces lower data quality.

This material is based upon work supported by the National Science Foundation under grant #1062970. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation

CERIAS

PURDUE UNIVERSITY