Study Design 101

The Dos and Don'ts of Ascertainment in Genomic Studies

Farren Briggs, PhD ScM
Department of Epidemiology and Biostatistics
School of Medicine
Case Western Reserve University

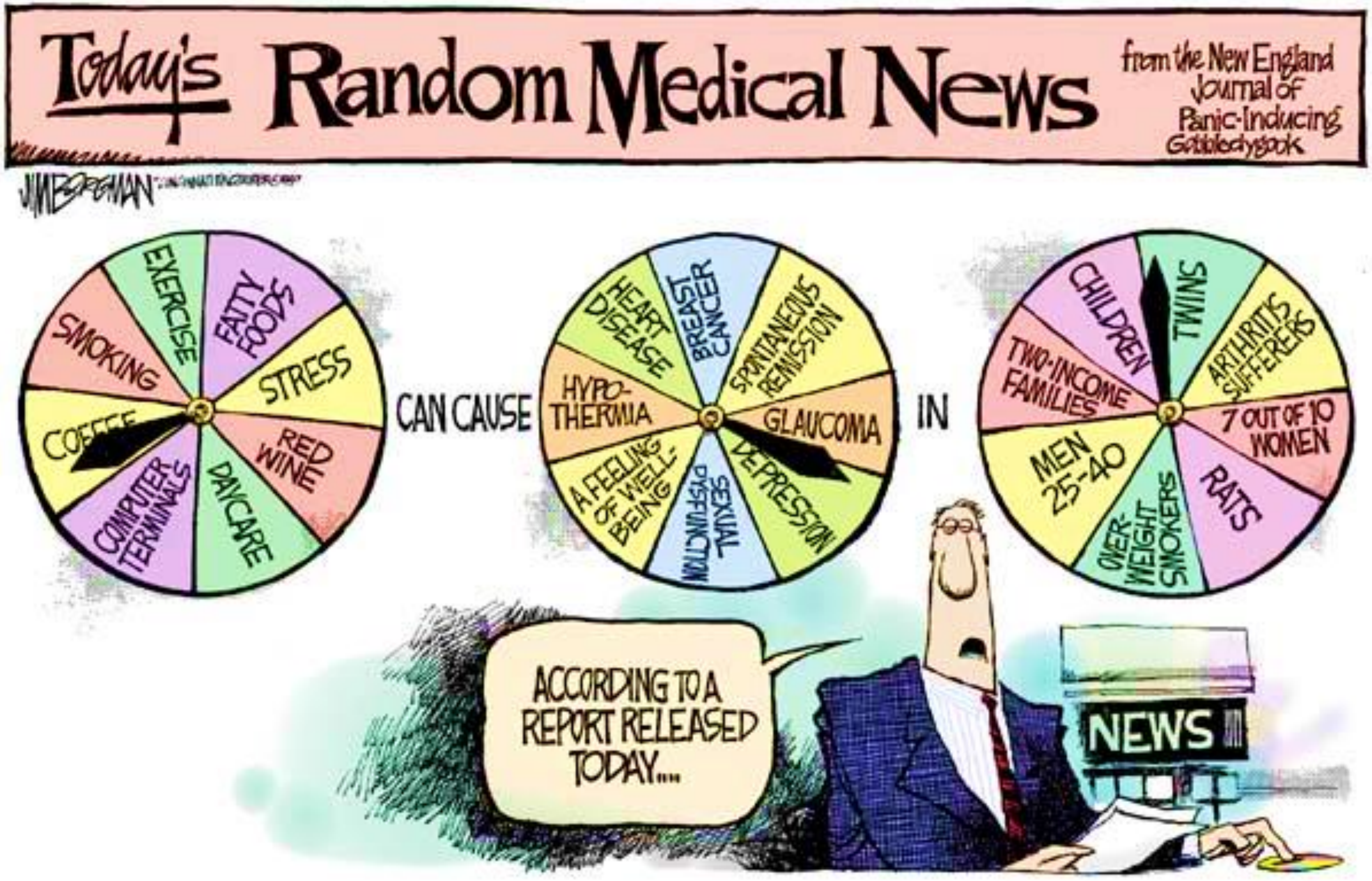
October 14th, 2016

My History


School of
Public Health
UNIVERSITY OF CALIFORNIA, BERKELEY
MS
National
Multiple Sclerosis
Society
nationalMSsociety.org
JOIN THE MOVEMENT
CASE
WESTERN
RESERVE
EST. 1826 UNIVERSITY
think beyond the possible
ST. JOHN'S COLLEGE
BELIZE CITY
Our Stolen
Future
ARE WE THREATENING OUR FERTILITY,
INTELLIGENCE, AND SURVIVAL?
A SCIENTIFIC DETECTIVE STORY

My History

Study Design 101



What is Epidemiology?

Study Design 101: Epidemiology

- "the branch of medical science that treats epidemics"

- the study of how often diseases occur in different groups of people and why

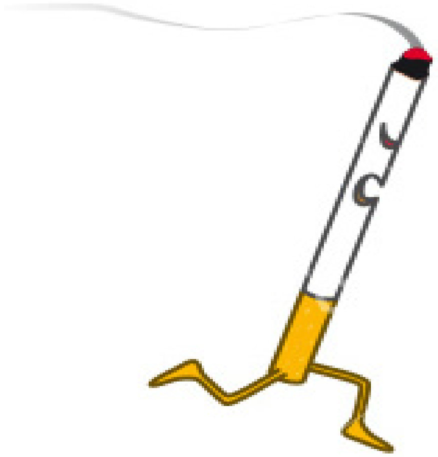
- Patterns of disease occurrence in human populations according to person, place, and time

- the study of how disease is **distributed in populations** and the **factors that influence or determine this distribution**

Study Design 101: Epidemiology

Exposure → Health Outcome ?

Study Design 101: Testing association



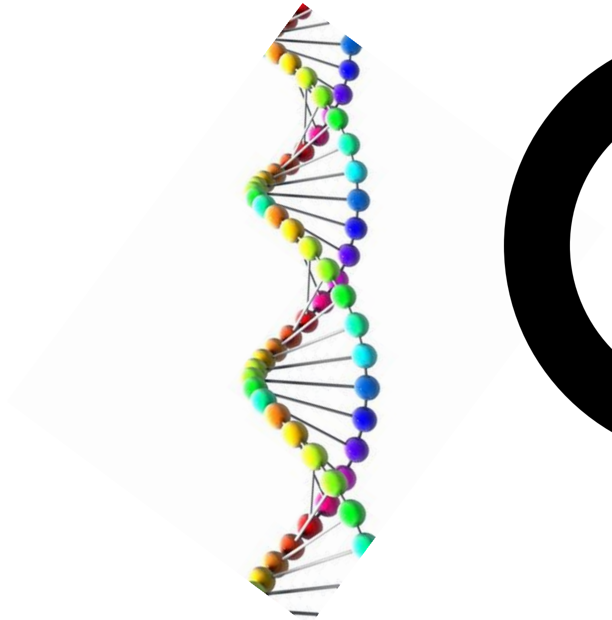
Exposure	Heart disease	No heart disease
Smokers	700	500
Non-smokers	300	500

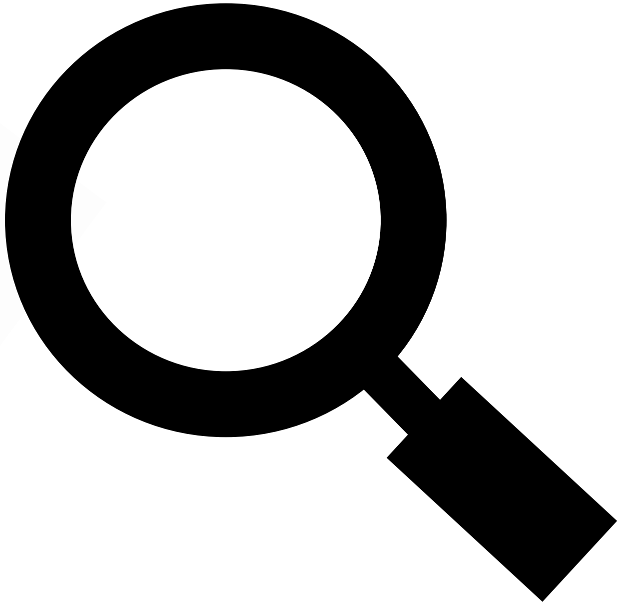
Study Design 101: Genetic Epidemiology

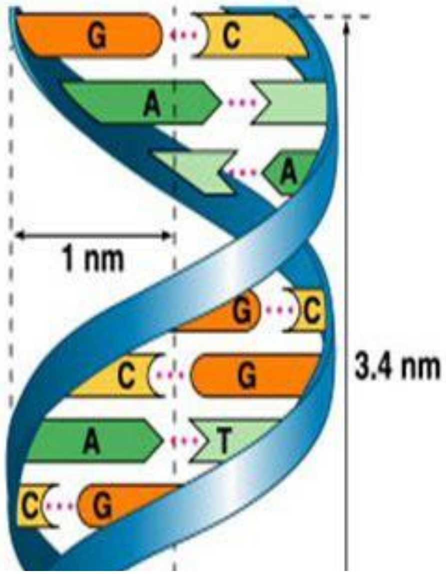
Genetic Variation → Health Outcome

?

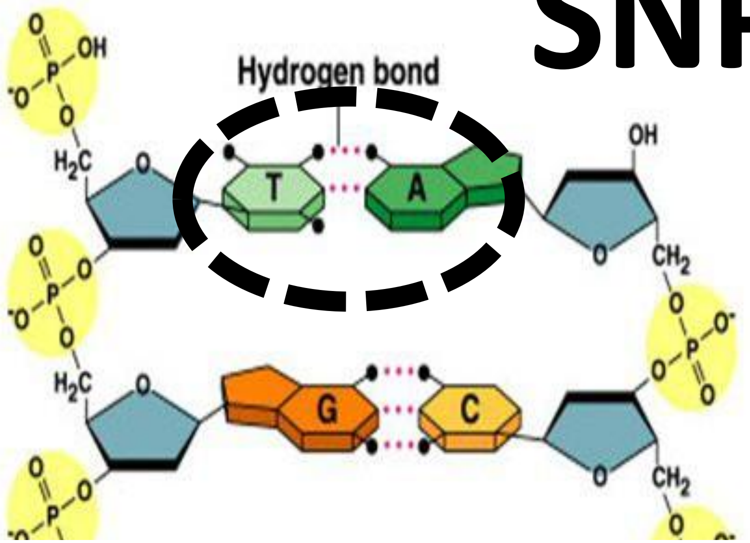
Study Design 101: Genetic Variation







SNP



SNP	Disease	Non-diseased
A allele	800	500
T allele	200	500

Study Design 101: Outcome

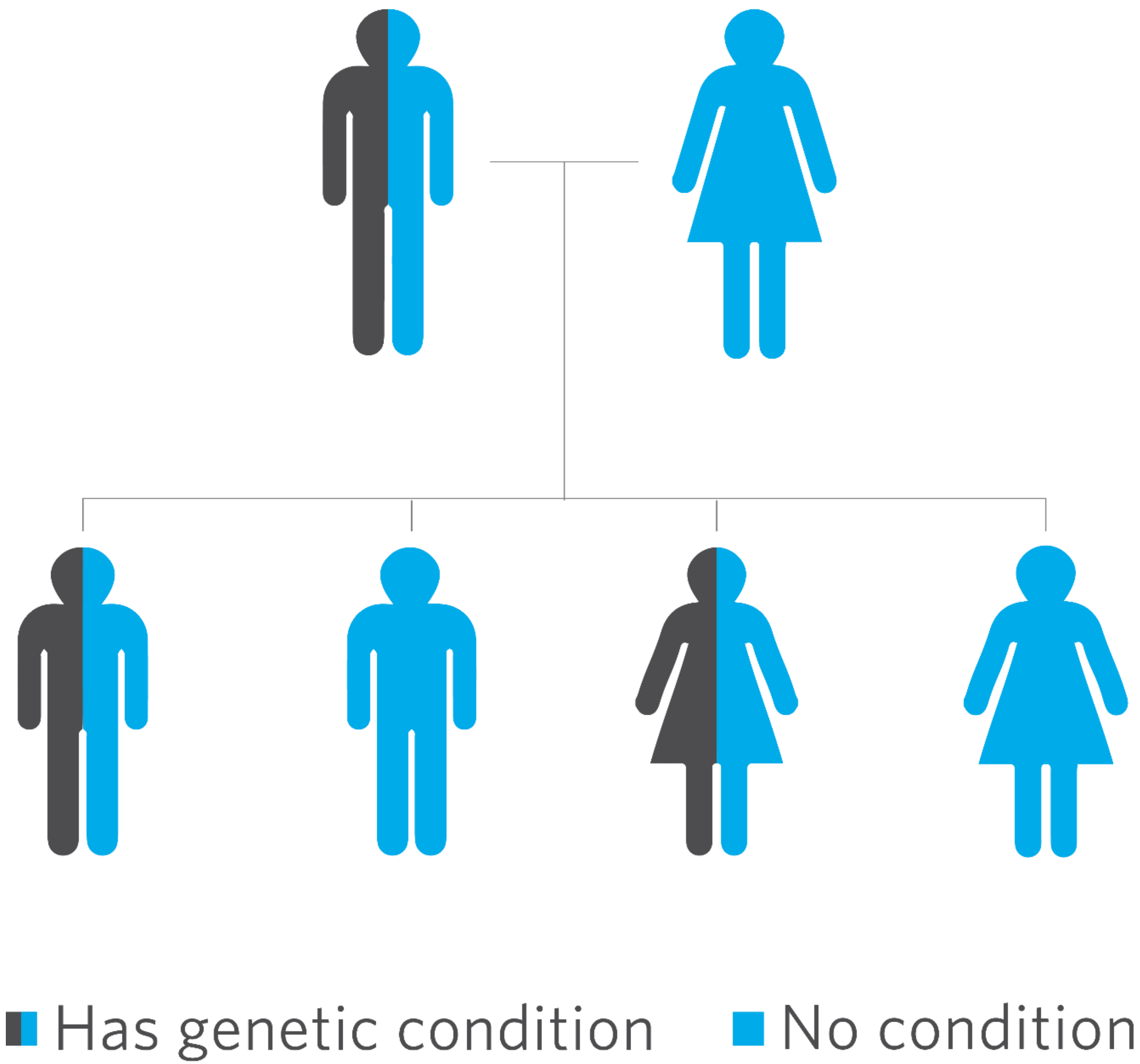
What **is the outcome of interest?**

Use *objective* and *clear* definitions

- Ask the experts
- Use standardized medical codes (ICD-10)

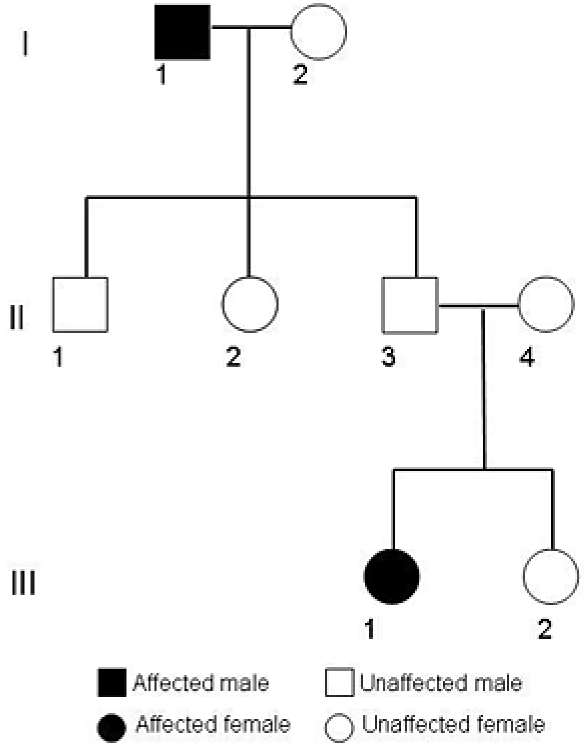
If a disease has a genetic component, where do we start?

Study Design 101: Family-based studies



Study Design 101: Family-based studies

Do…



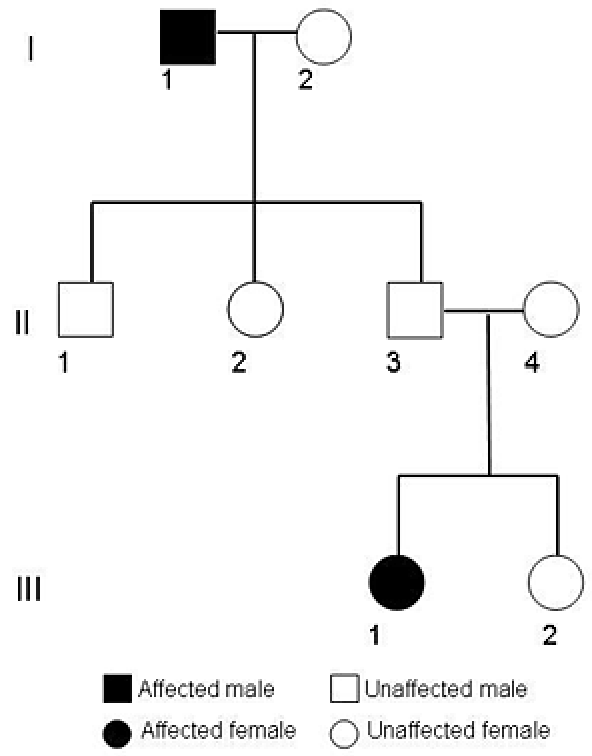
1. Recruit multiple family members

2. Recruit across generations

3. Compare genomes of affected vs unaffected family members

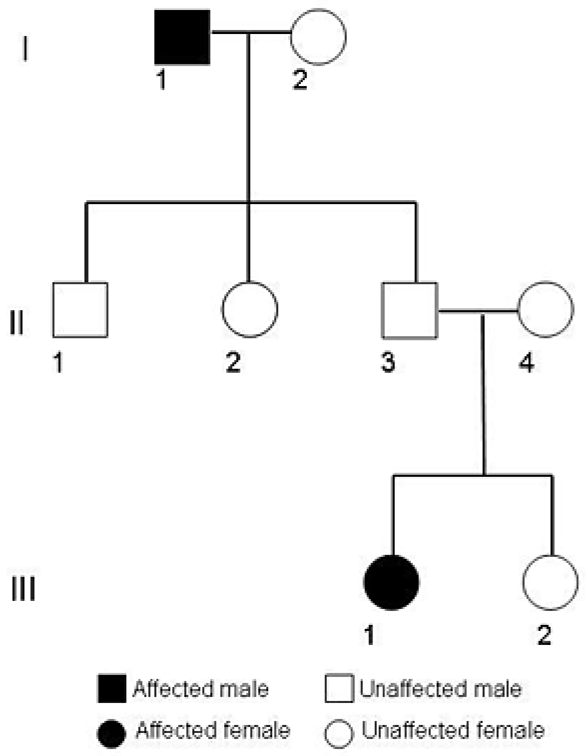
Advantages

1. Enriched for the ‘causal’ variant
2. Useful to detect rare variants



Study Design 101: Family-based studies

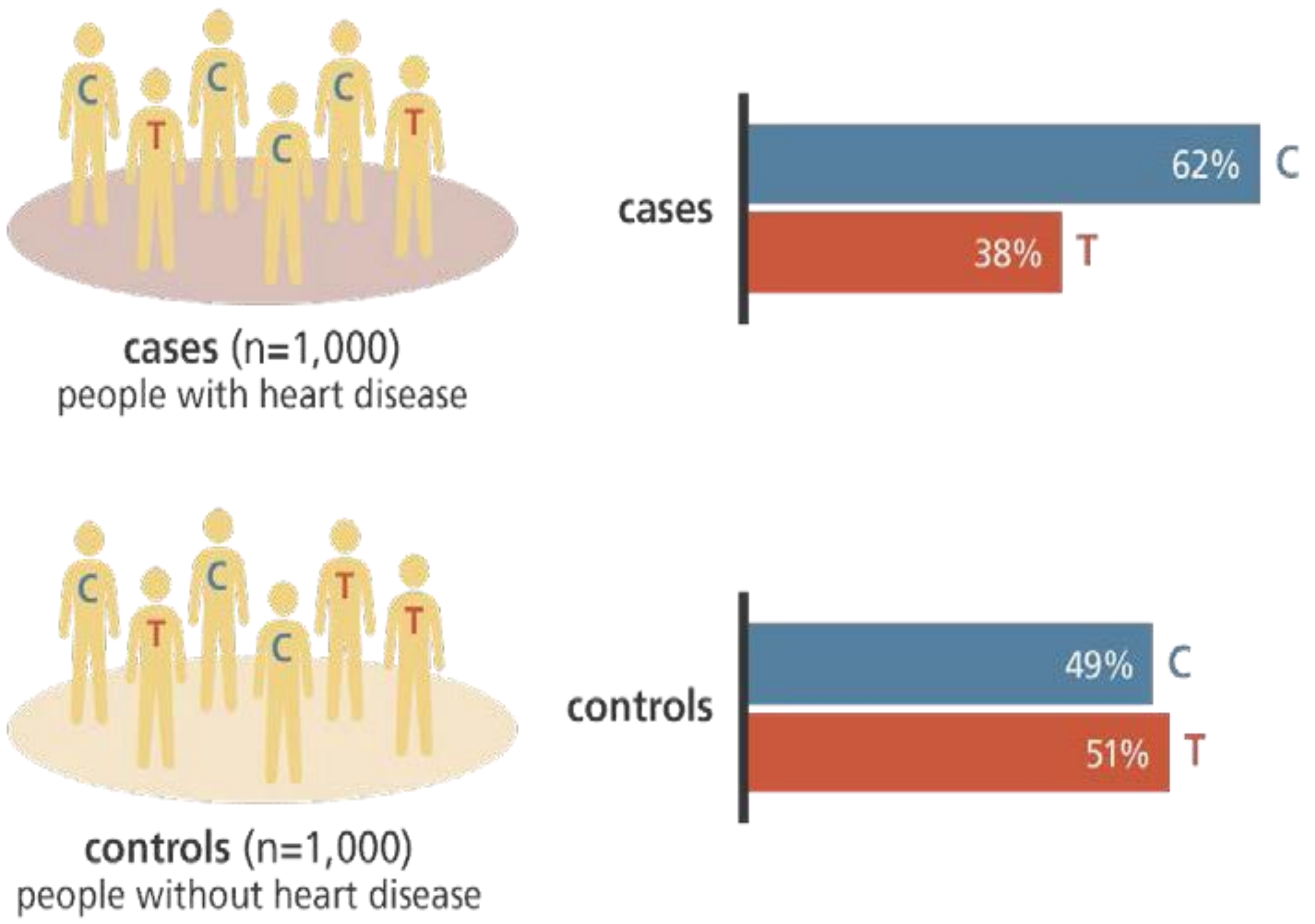
Challenges



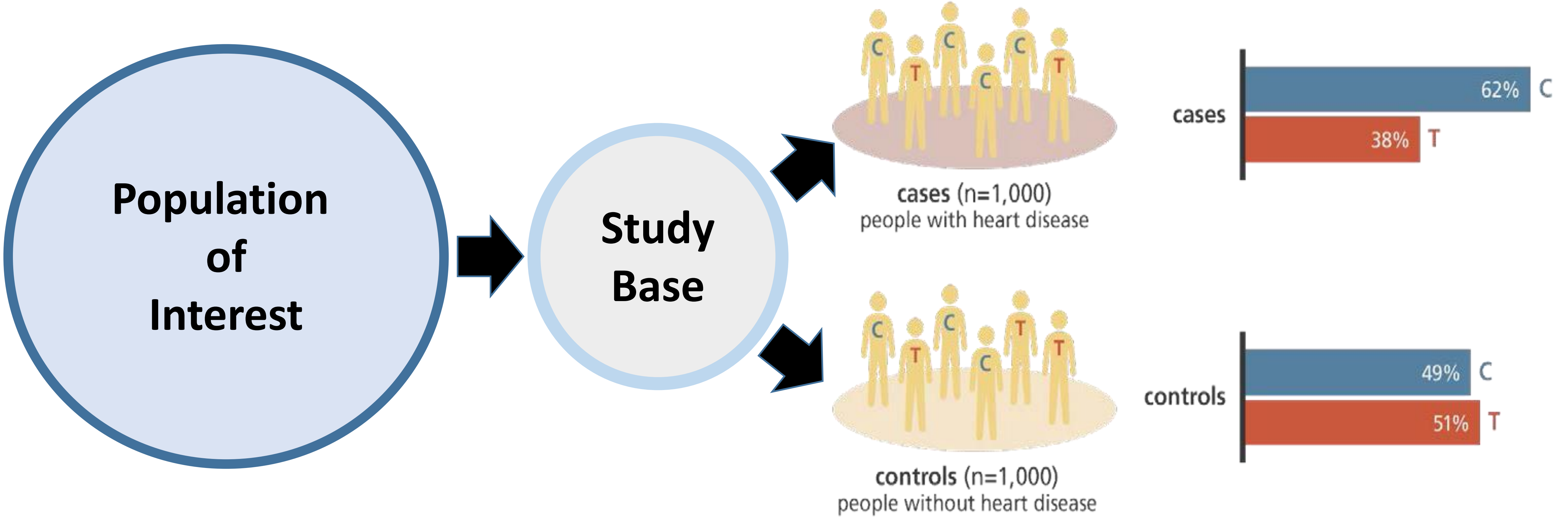
1. Computational complex
2. Difficult to find large pedigrees
3. Hard to study late-onset/age-related diseases.

Study Design 101: Population-based studies

Study Design 101: Population-based studies

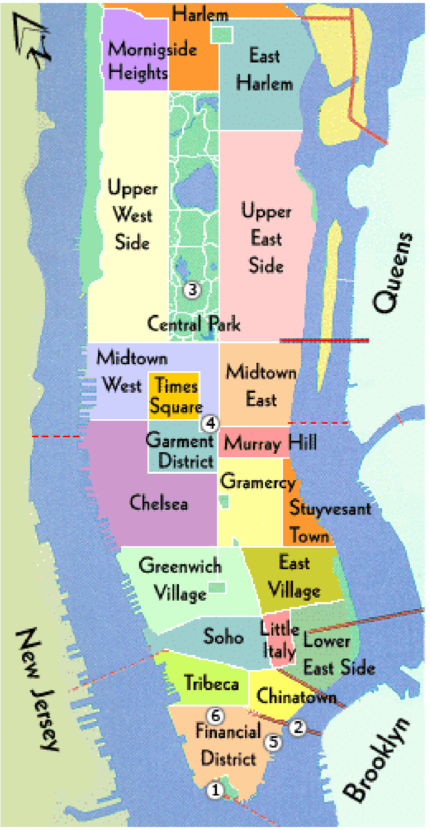


Study Design 101: Population-based studies



Study Design 101: Defining the study base

Primary study base



Population of interest is defined ***FIRST***

Eligible cases are from among all cases

Controls would be a random sample

Study Design 101: Defining the study base

Secondary study base

Cases are identified ***FIRST***

(e.g. cancer patients at a hospital)

What population do cases represent?

(e.g. a single city? a county? a region?)

Controls are to reflect this “*population*”

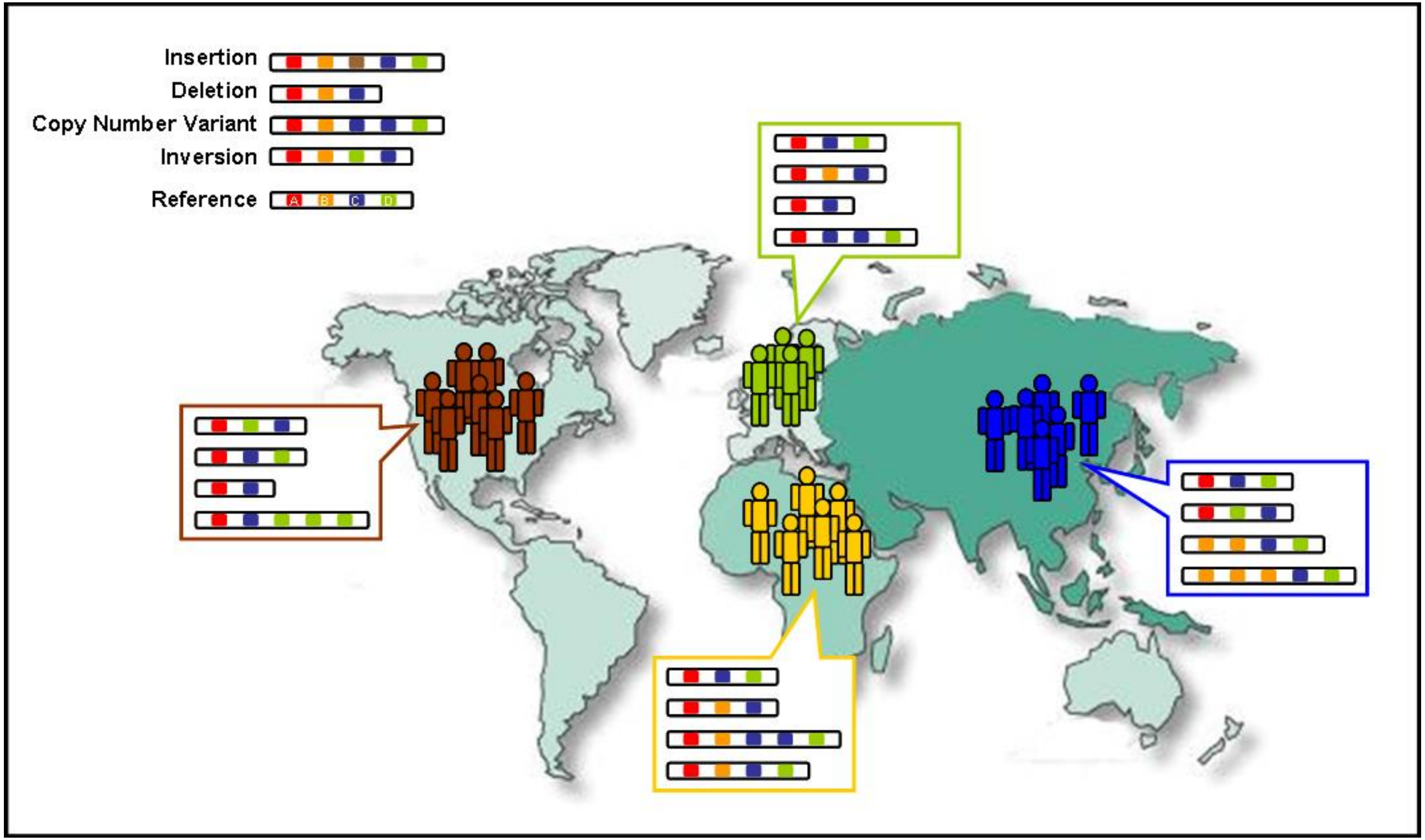
(e.g. would been a case if had disease)

Study Design 101: Defining the study base

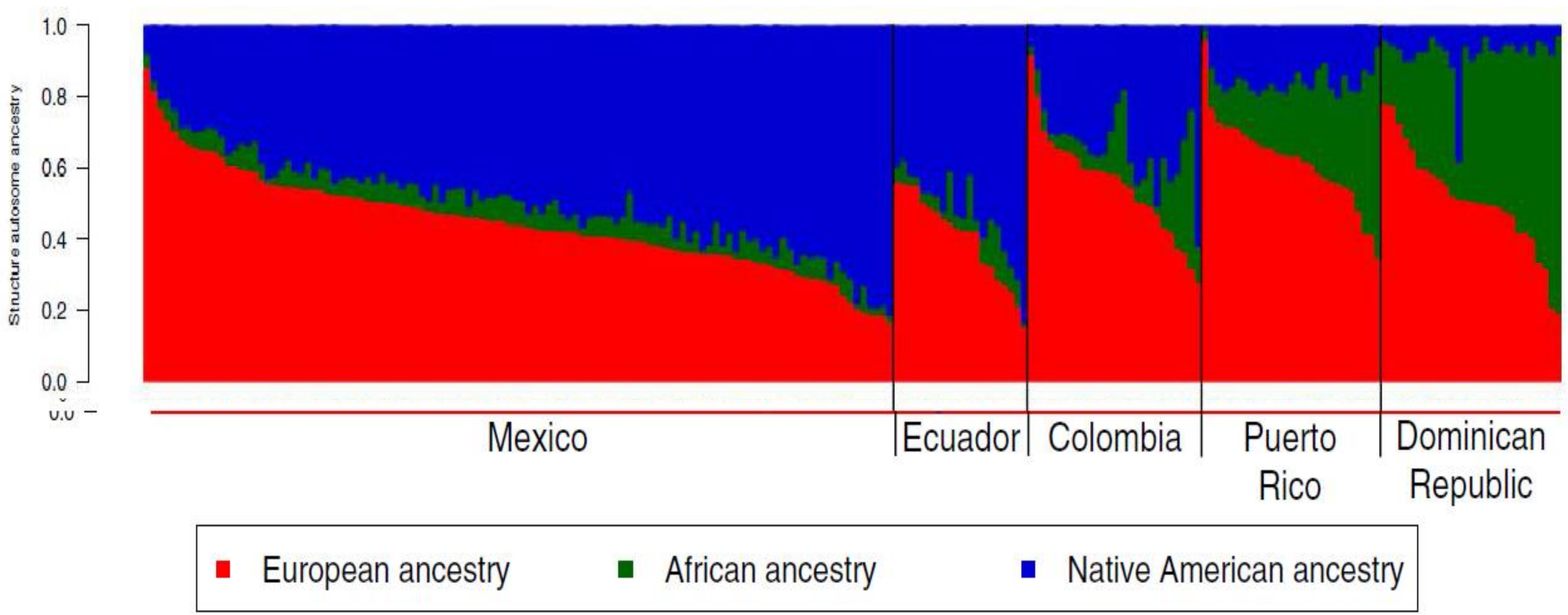
Yes

No

Study Design 101: Genetic variation VARIES



Study Design 101: Population stratification



Bryc 2010, PNAS

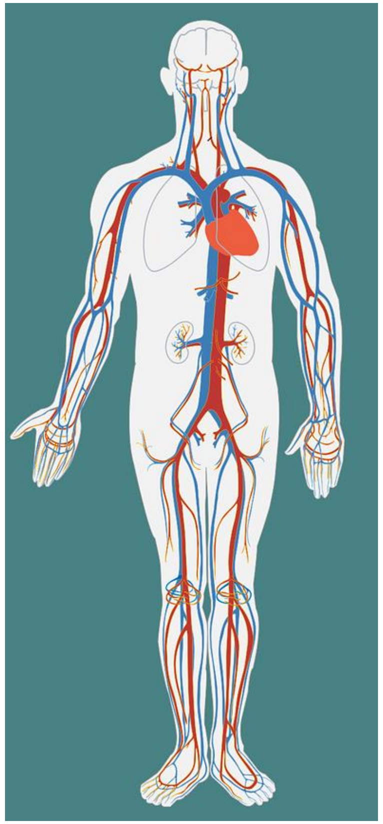
Study Design 101: Population stratification

SNP

… C **T** C G A A A…

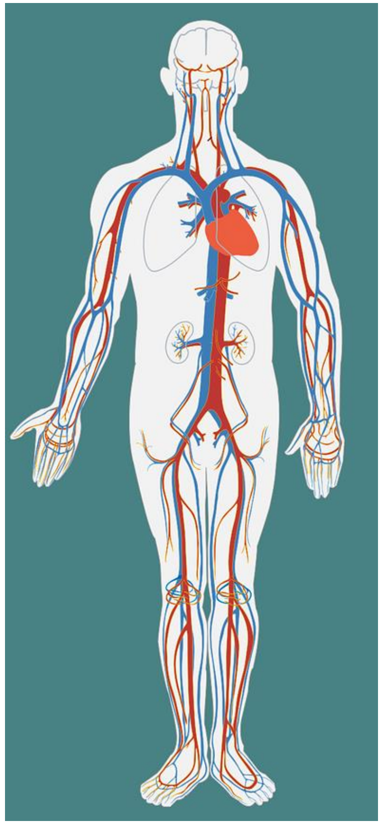
… C **A** C G A A A…

?



Study Design 101: Population stratification

False association

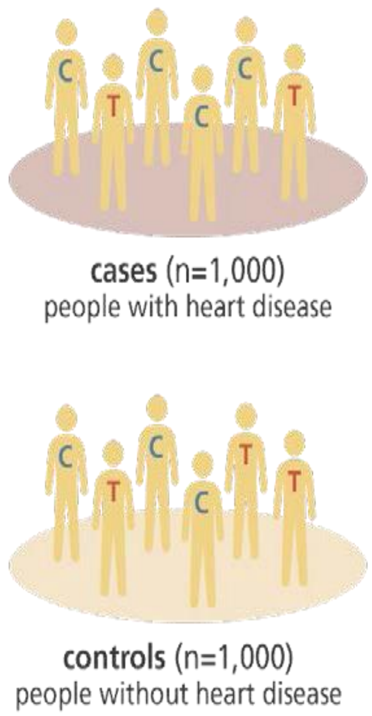


SNP

... C T C G A A A...

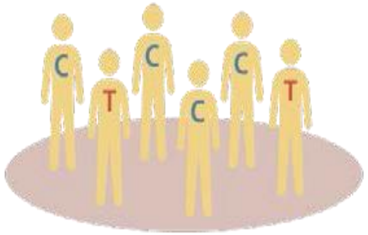
... C A C G A A A...

Advantages



1. Simpler statistics
2. Ideal for studying late-onset diseases
3. By increasing sample size, we increase power to detect an association

Challenges



cases (n=1,000)
people with heart disease

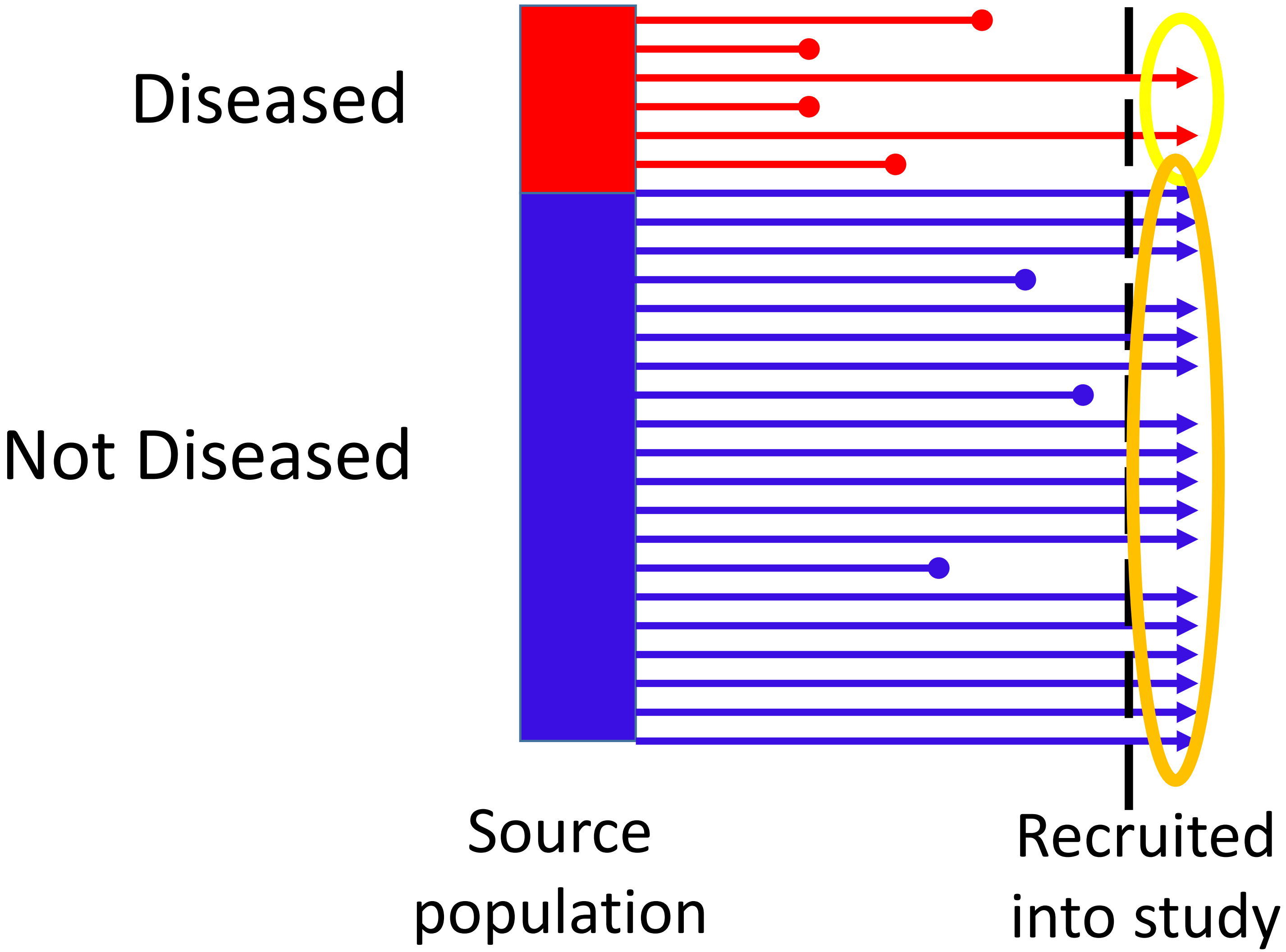
1. Defining study base might be difficult

   Thus identifying suitable controls may be hard

controls (n=1,000)
people without heart disease

2. Population stratification can muddle effects

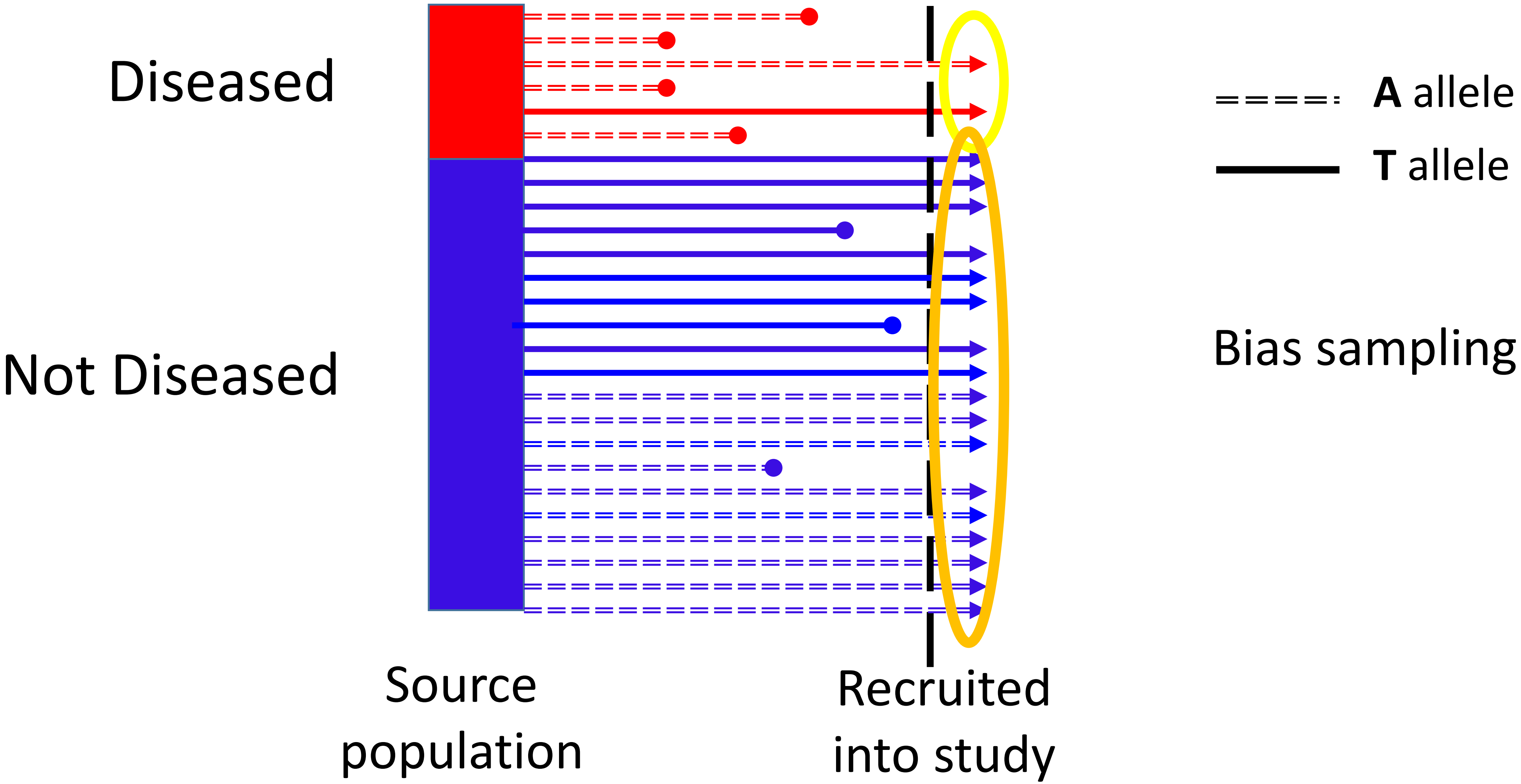
Study Design 101: Who are we *really* studying?



Survivors?

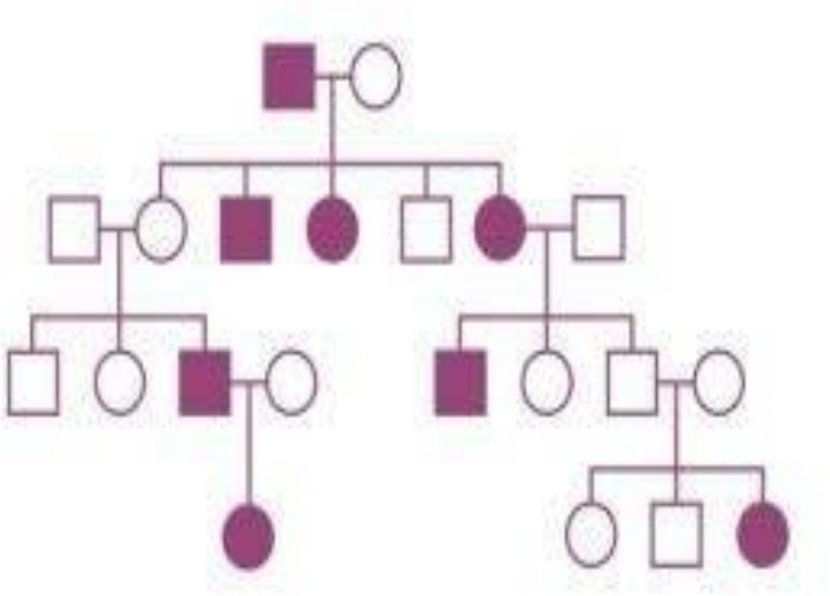
Do they represent the source population?

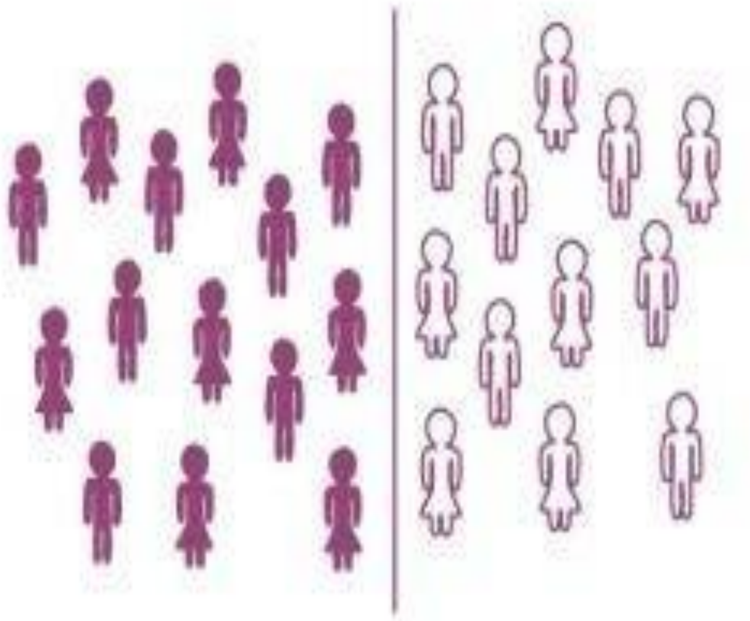
Study Design 101: Who are we *really* studying?



Bias sampling!

Study Design 101: Recap





Thank you!