Cross-Sectional Studies

Madhukar Pai, McGill University, Montreal

Acknowledgement: Dr Kris Filion, McGill University

COVID-19 Antibody Seroprevalence in Santa Clara County, California

Eran Bendavid[1], Bianca Mulaney[2], Neeraj Sood[3], Soleil Shah[2], Emilia Ling[2], Rebecca Bromley-Dulfano[2], Cara Lai[2], Zoe Weissberg[2], Rodrigo Saavedra-Walker[4], Jim Tedrow[5], Dona Tversky[6], Andrew Bogan[7], Thomas Kupiec[8], Daniel Eichner[9], Ribhav Gupta[10], John P.A. Ioannidis[1,10], Jay Bhattacharya[1]

Methods

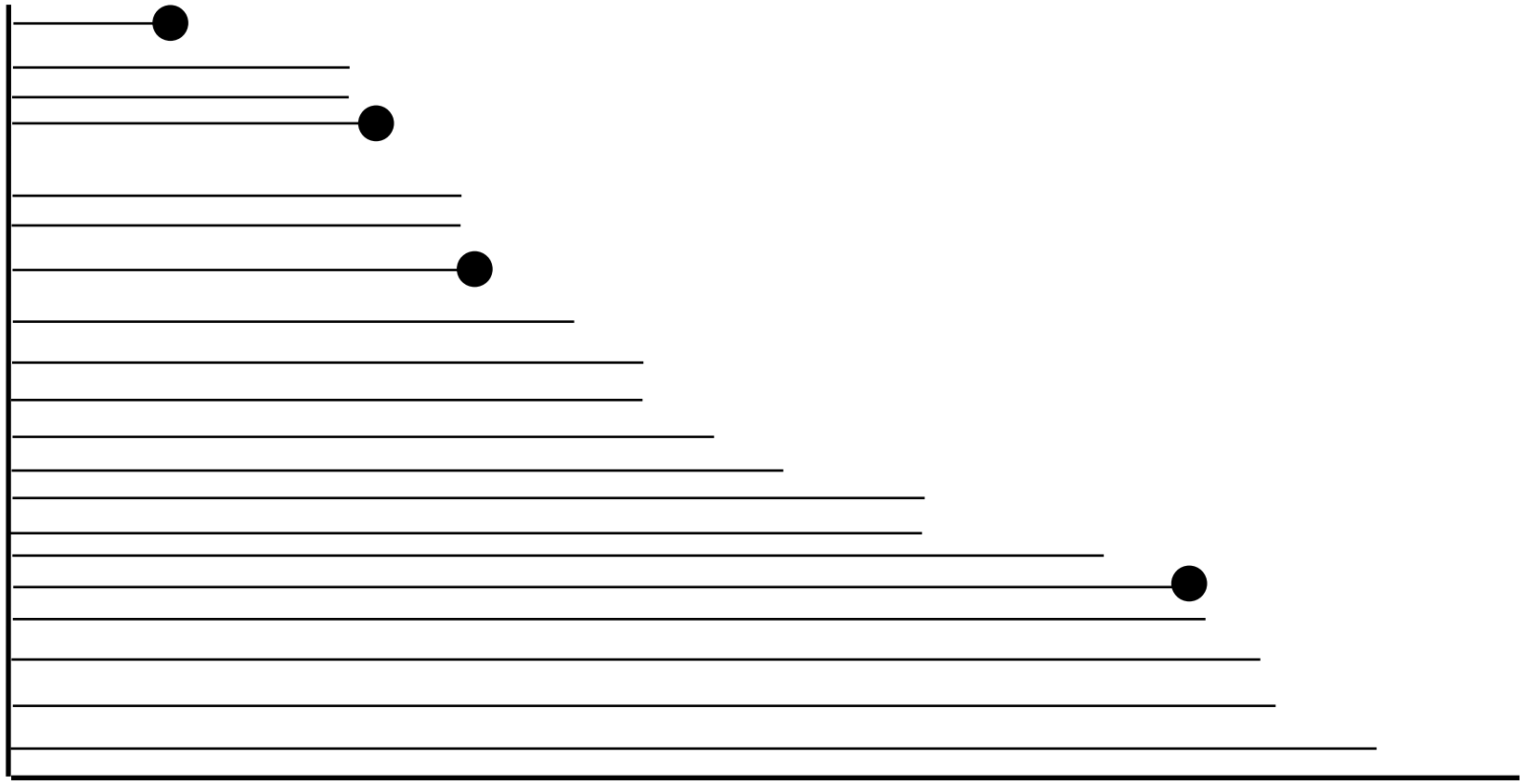
On April 3-4, 2020, we tested county residents for antibodies to SARS-CoV-2 using a lateral flow immunoassay. Participants were recruited using Facebook ads targeting a sample of individuals living within the county by demographic and geographic characteristics. We estimate weights to adjust our sample to match the zip code, sex, and race/ethnicity distribution within the county. We report both the weighted and unweighted prevalence of antibodies to SARS-CoV-2. We also adjust for test performance characteristics by combining data from 16 independent samples obtained from manufacturer's data, regulatory submissions, and independent evaluations: 13 samples for specificity (3,324 specimens) and 3 samples for sensitivity (157 specimens).

Results

The raw prevalence of antibodies to SARS-CoV-2 in our sample was 1.5% (exact binomial 95CI 1.1-2.0%). Test performance specificity in our data was 99.5% (95CI 99.2-99.7%) and sensitivity was 82.8% (95CI 76.0-88.4%). The unweighted prevalence adjusted for test performance characteristics was 1.2% (95CI 0.7-1.8%). After weighting for population demographics of Santa Clara County, the prevalence was 2.8% (95CI 1.3-4.7%), using bootstrap to estimate confidence bounds. These prevalence point estimates imply that 54,000 (95CI 25,000 to 91,000 using weighted prevalence; 23,000 with 95CI 14,000-35,000 using unweighted prevalence) people were infected in Santa Clara County by early April, many more than the approximately 1,000 confirmed cases at the time of the survey.

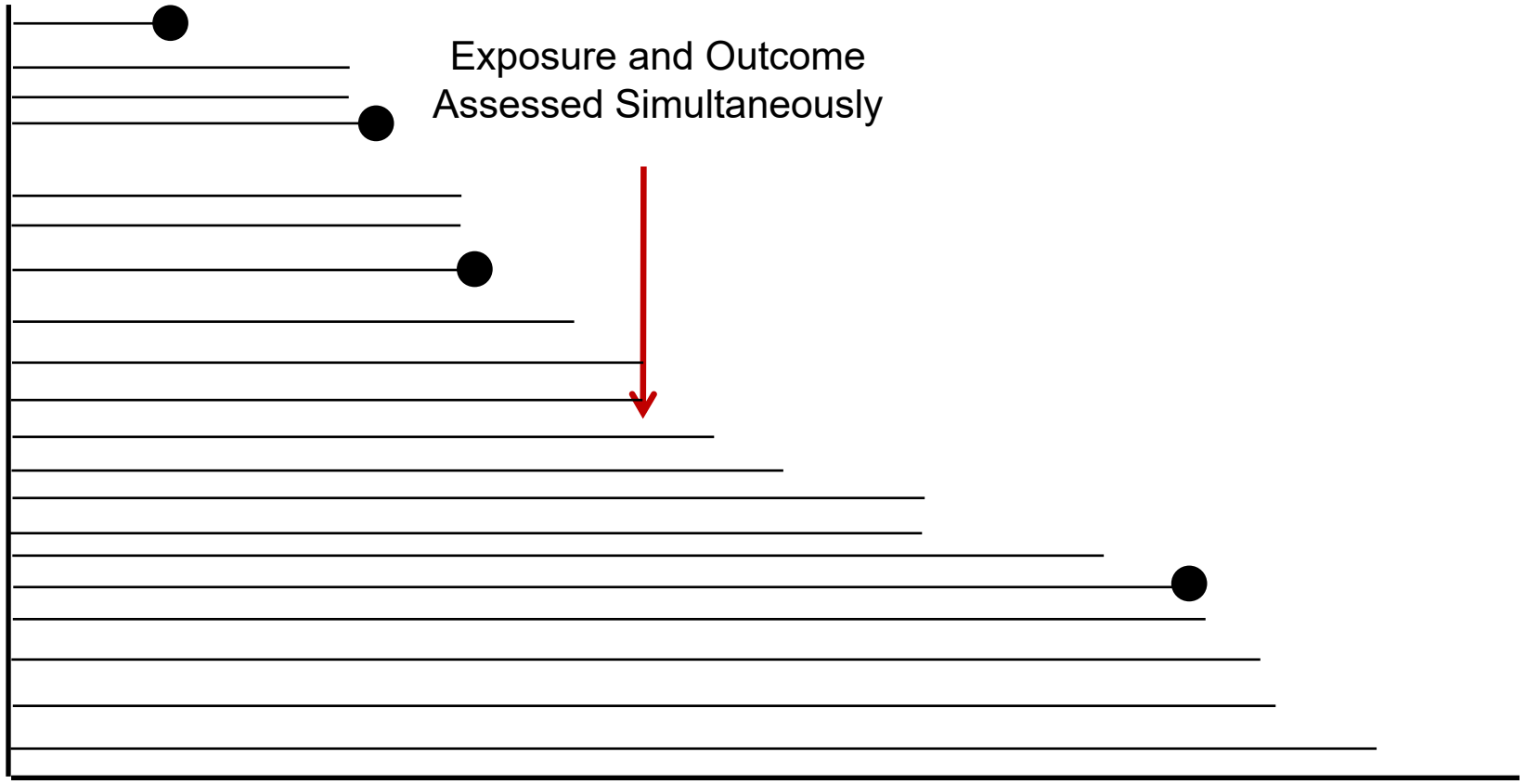
https://www.medrxiv.org/content/10.1101/2020.04.14.20062463v2.full.pdf

Cross-Sectional Study



Time

Cross-Sectional Study



Cross-Sectional Study

- Also known as prevalence study or survey
- A snapshot or picture of health experience at one time
- Sampling independent of both exposure and outcome
- Commonly used in public policy and public health
- Used to provide measure of disease burden

Recall: Prevalence

- Two types of Prevalence
  - Point prevalence
  - Period prevalence

Point Prevalence

$$P = \frac{C}{N}$$

C = # of observed cases at time t

N = Population size at time t

Measures the frequency of disease at a given point in time

Point Prevalence Example

COVID-19 Antibody Seroprevalence in Santa Clara County, California

Eran Bendavid[1], Bianca Mulaney[2], Neeraj Sood[3], Soleil Shah[2], Emilia Ling[2], Rebecca Bromley-Dulfano[2], Cara Lai[2], Zoe Weissberg[2], Rodrigo Saavedra-Walker[4], Jim Tedrow[5], Dona Tversky[6], Andrew Bogan[7], Thomas Kupiec[8], Daniel Eichner[9], Ribhav Gupta[10], John P.A. Ioannidis[1,10], Jay Bhattacharya[1]

Version 2, April 27, 2020
(revised in response to comments received. This remains a preliminary report of the work.)

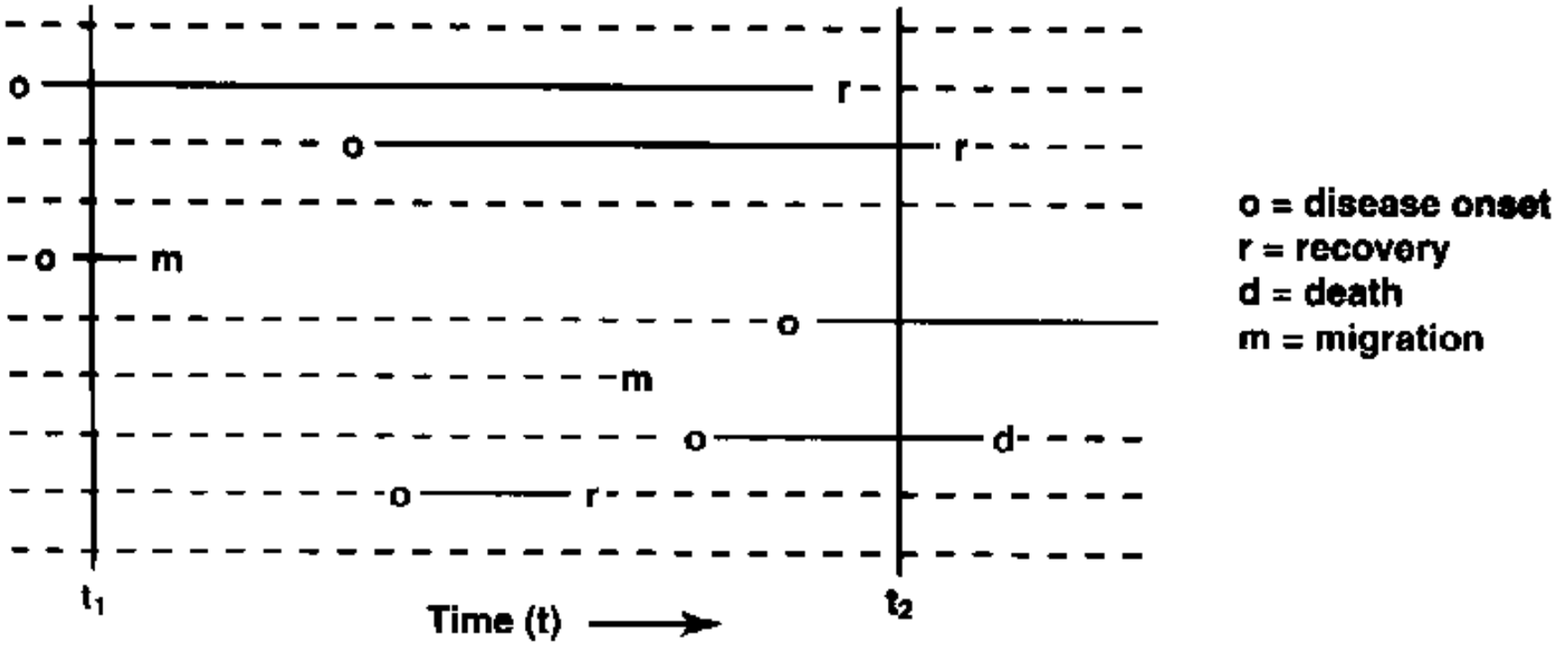
- April 3rd and 4th, 2020, researchers did serologic testing for SARS-CoV-2 antibodies in 3,330 adults and children in Santa Clara County
- Total number of positive cases by either IgG or IgM = 50
- Crude point prevalence = 50/3330 = 1.5% (95 CI 1.1-2.0%)

Period Prevalence

$$PP = \frac{C + I}{N}$$

- C = the # of prevalent cases at the beginning of the time period.
- I = the # of incident cases that develop during the period.
- N = size of the population for this same time period.

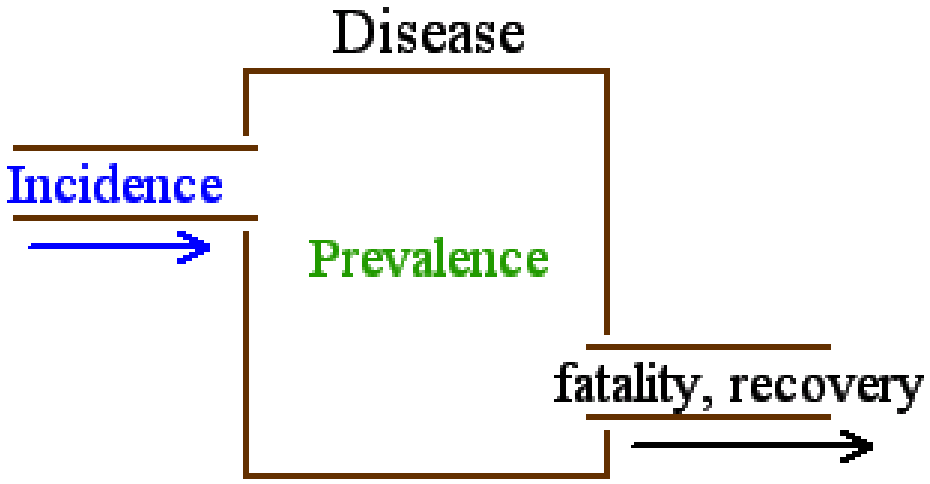
Prevalence: example



Point prevalence at time t_1 = 2/10 = 20%
Point prevalence at time t_2 = 3/8 = 38%
Period prevalence between t_1 and t_2: 6/10 = 60%

Source: Silva 1999

Recall: Prevalence vs. Incidence

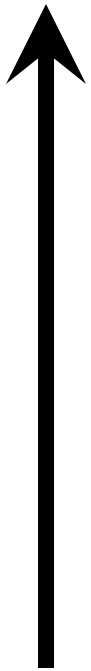


- Prevalence can be viewed as describing a pool of disease in a population.
- Incidence describes the input flow of new cases into the pool.
- Fatality and recovery reflects the output flow from the pool.

What factors can increase prevalence?

Prevalence

Longer duration
Prolongation of life without cure
Increased incidence
In-migration of cases
Out-migration of healthy people
In-migration of susceptible people
Better diagnosis/reporting



Source: Beaglehole, 1993

What factors can decrease prevalence?

Prevalence

Shorter duration

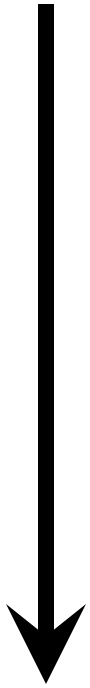
High case fatality

Decreased incidence

In-migration of healthy people

Out-migration of cases

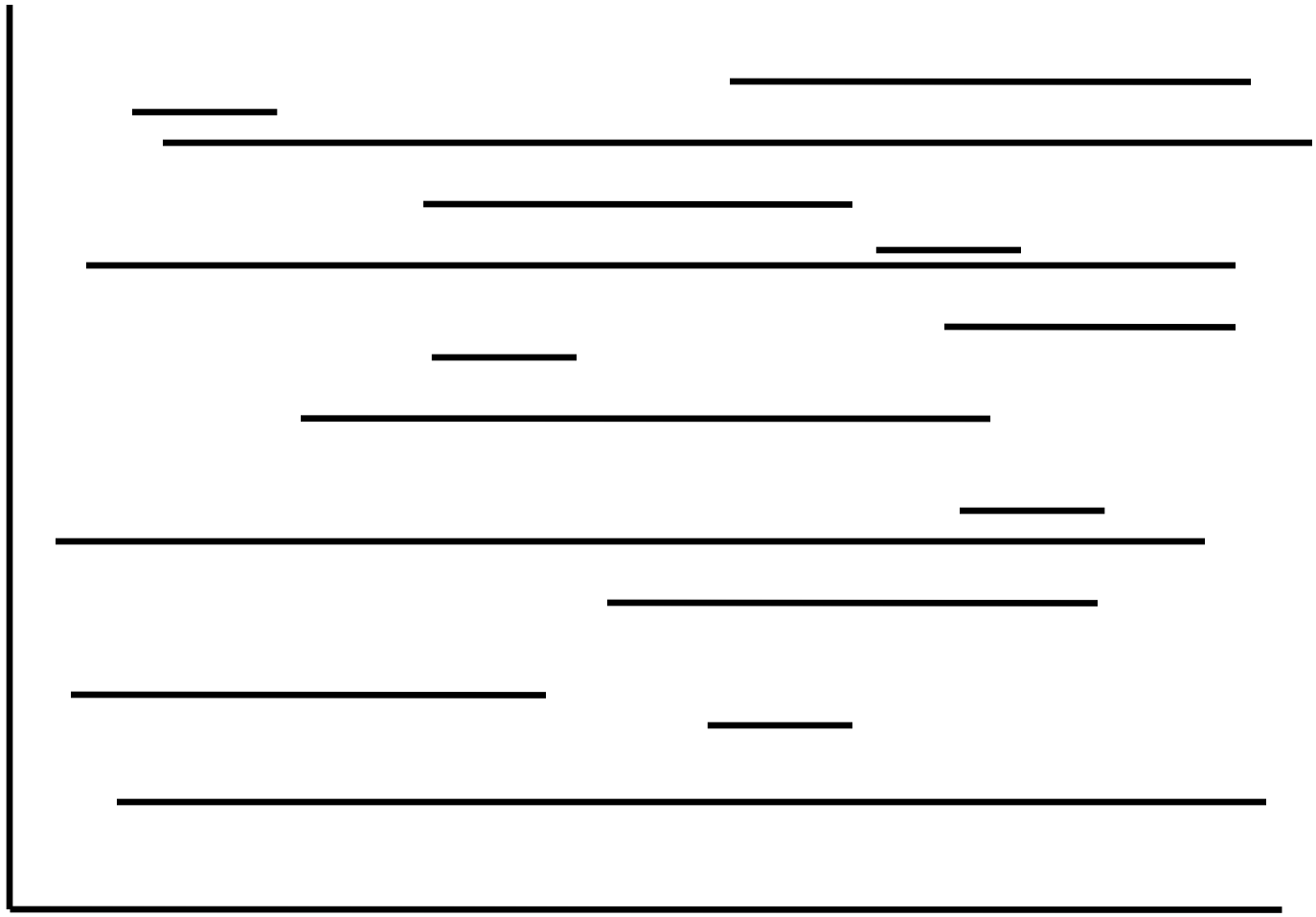
Improved cure rates

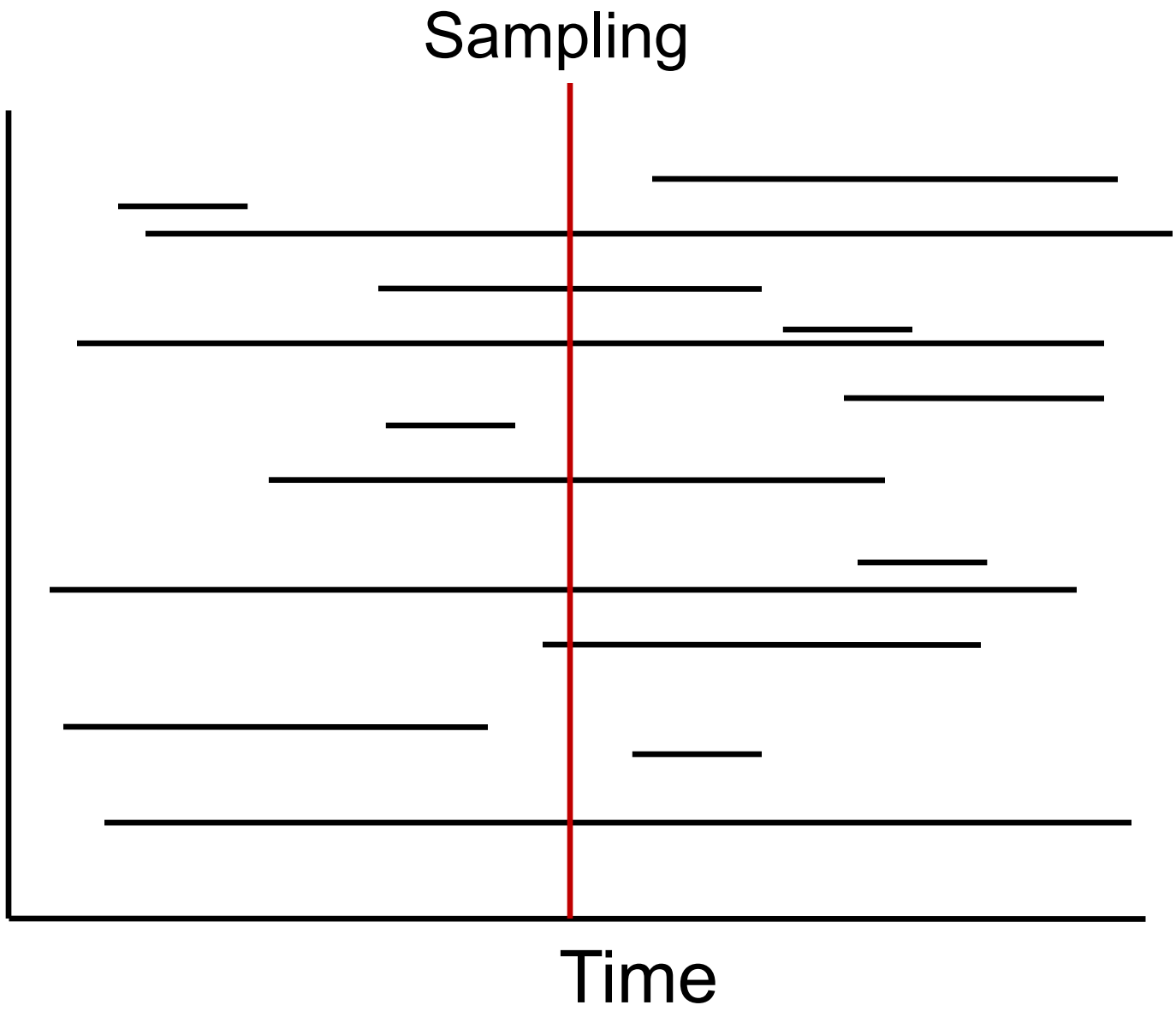


Source: Beaglehole, 1993

Advantages

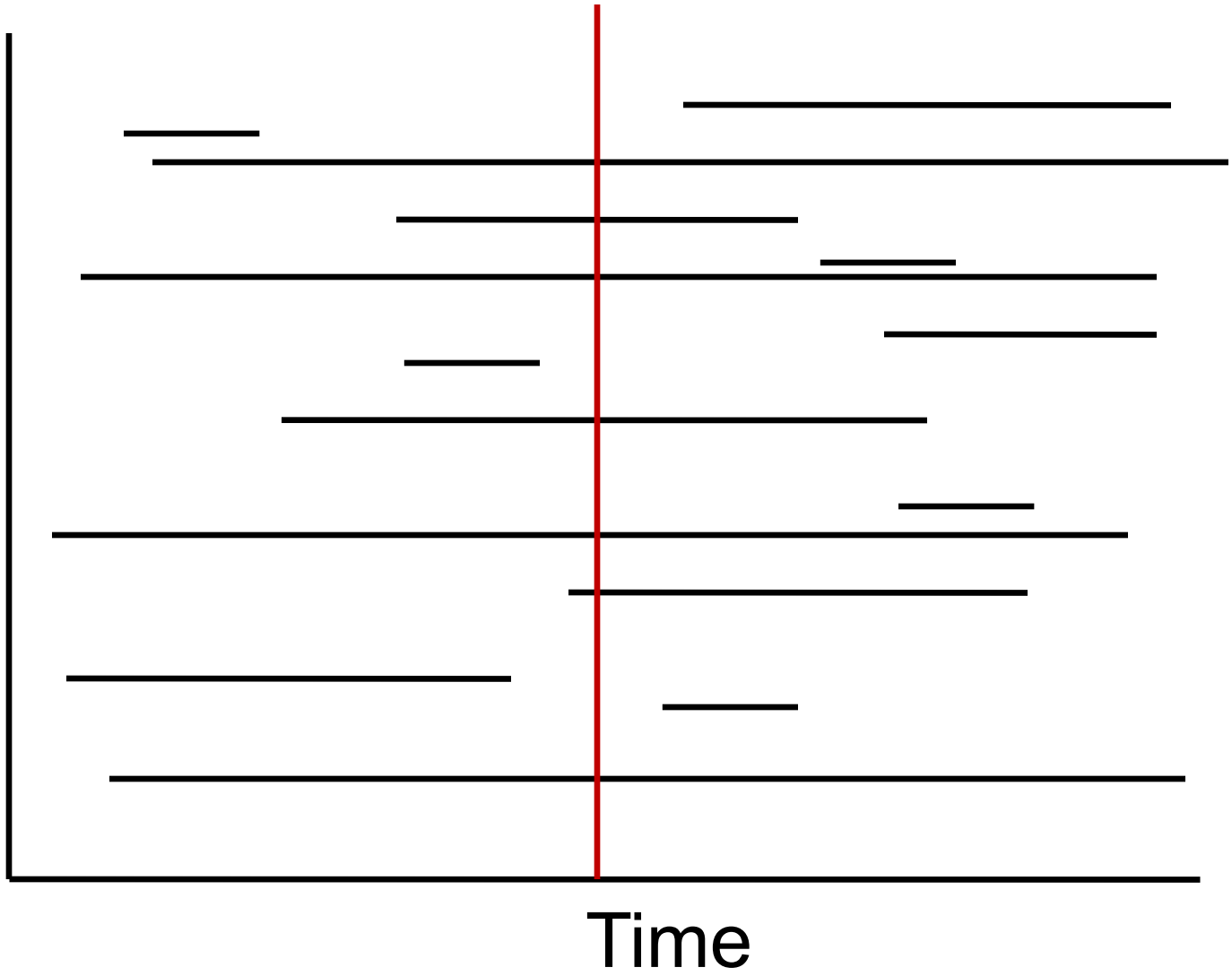
- Representative of general population
- Convenient, inexpensive, and fast
- Can assess several exposures and several outcomes
- Common diseases with long duration: generate hypotheses for future studies


Time


Sampling
Time

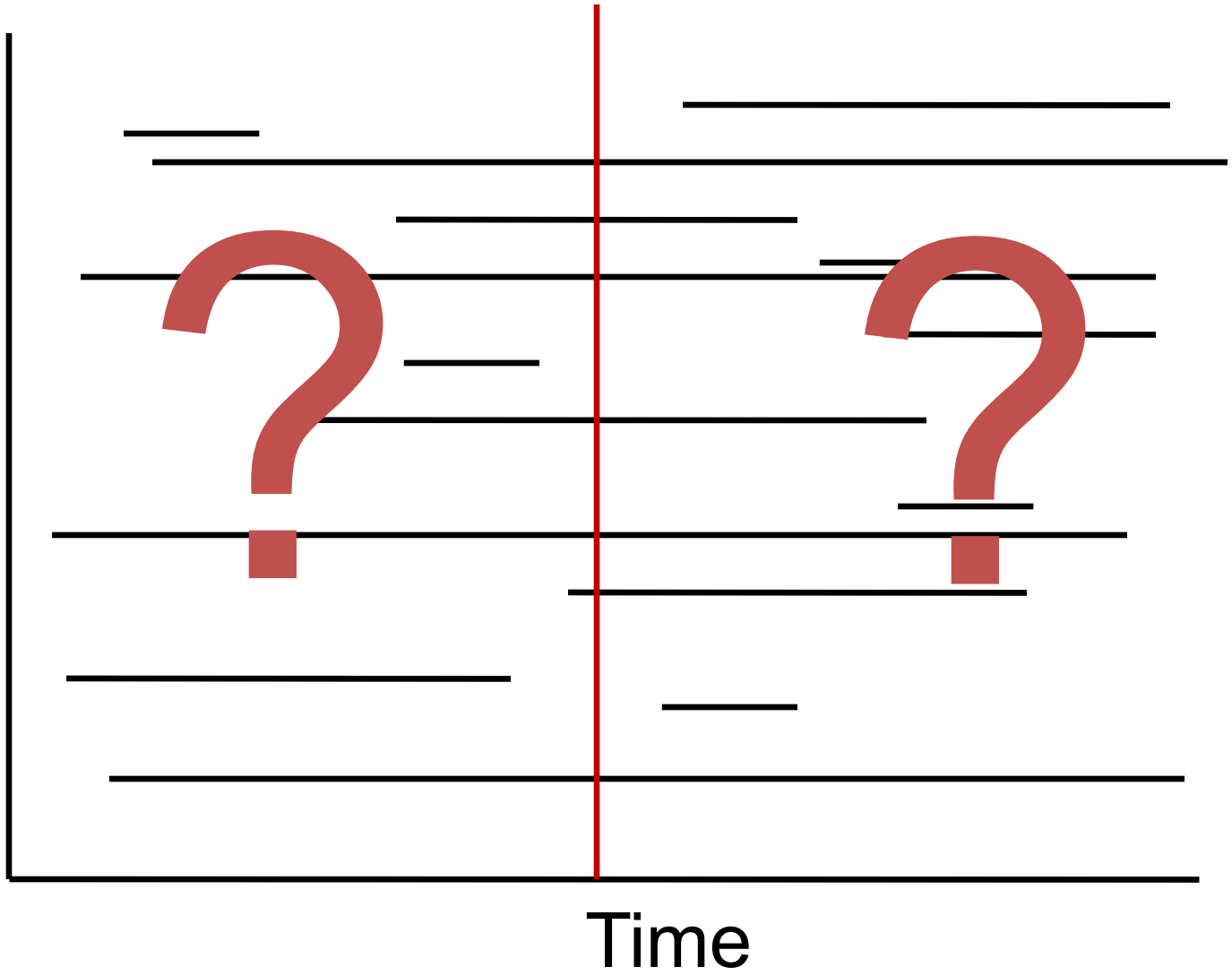
Sampling Prevalent Cases

Existing Cases at Single Point in Time Only

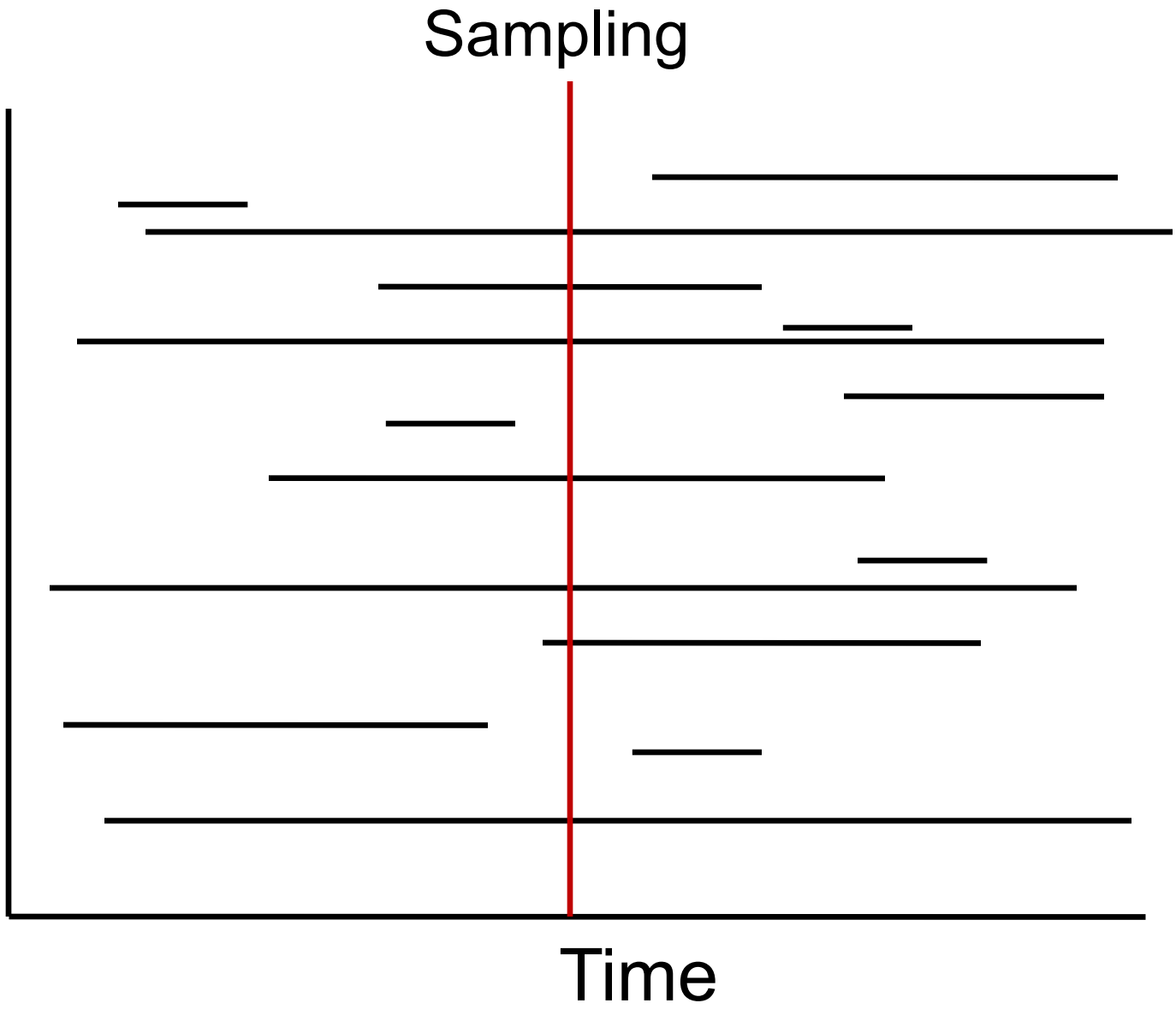


Sampling Prevalent Cases

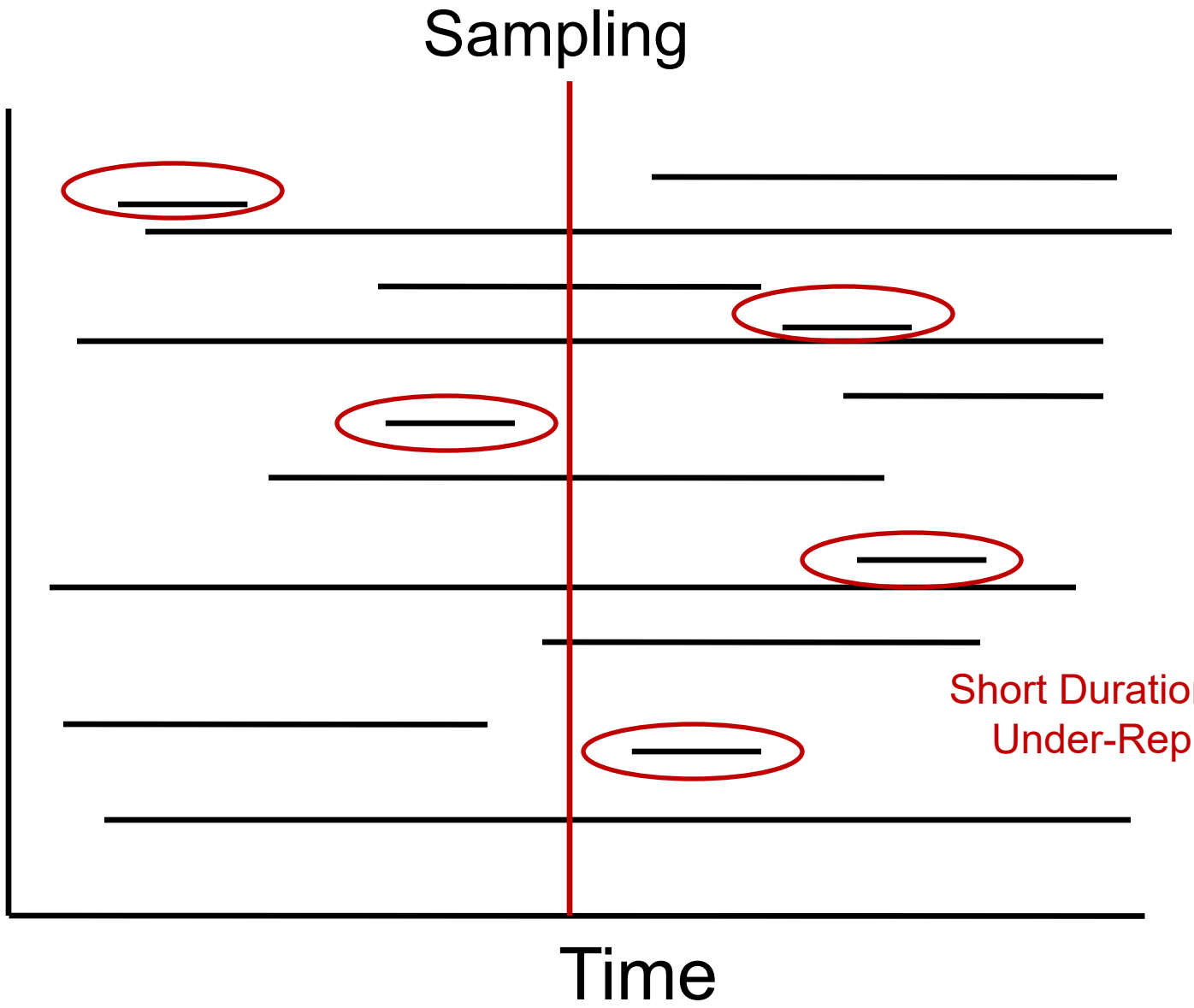
Existing Cases at Single Point in Time Only



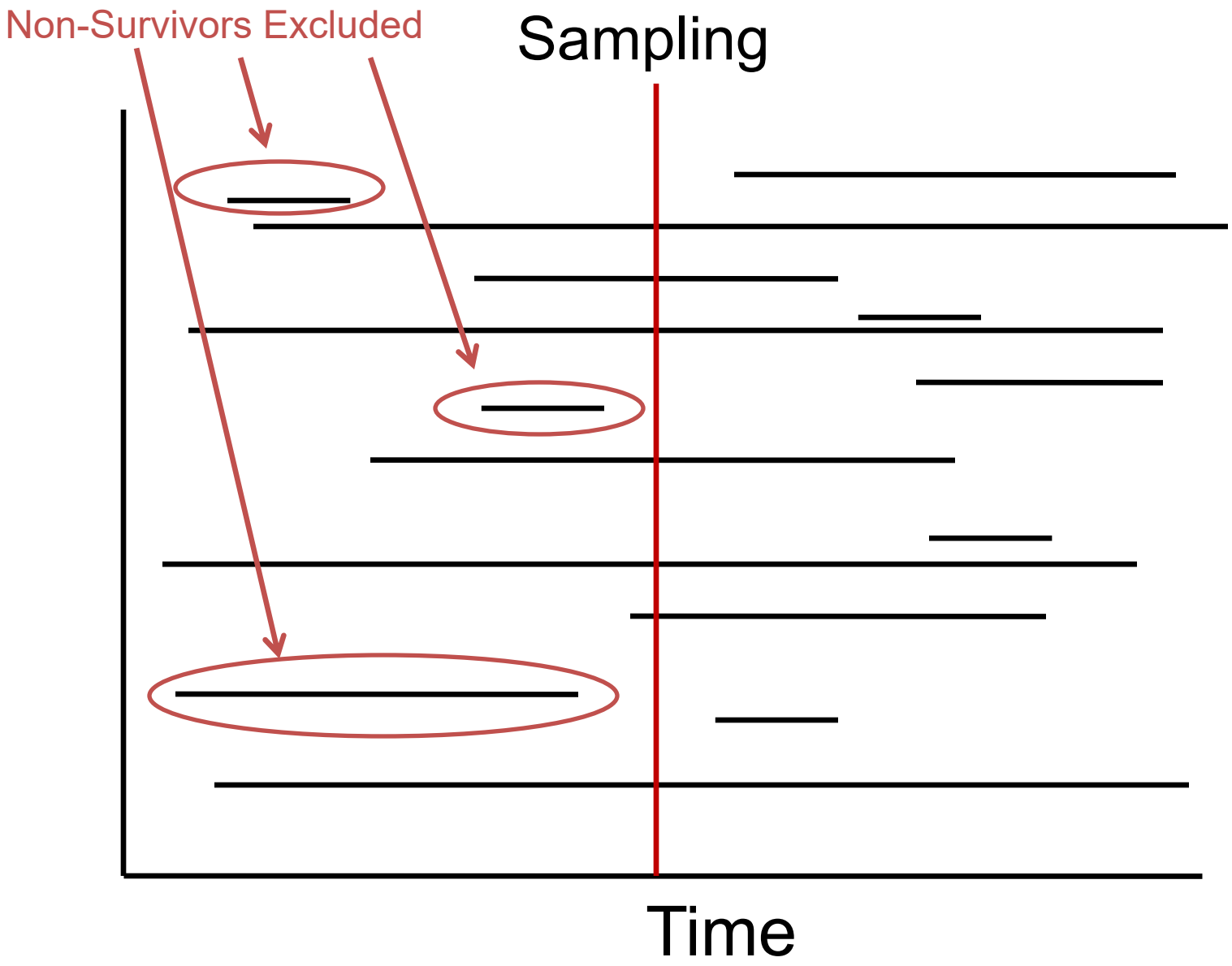
Prevalent Cases



Prevalent Cases



Prevalent Cases



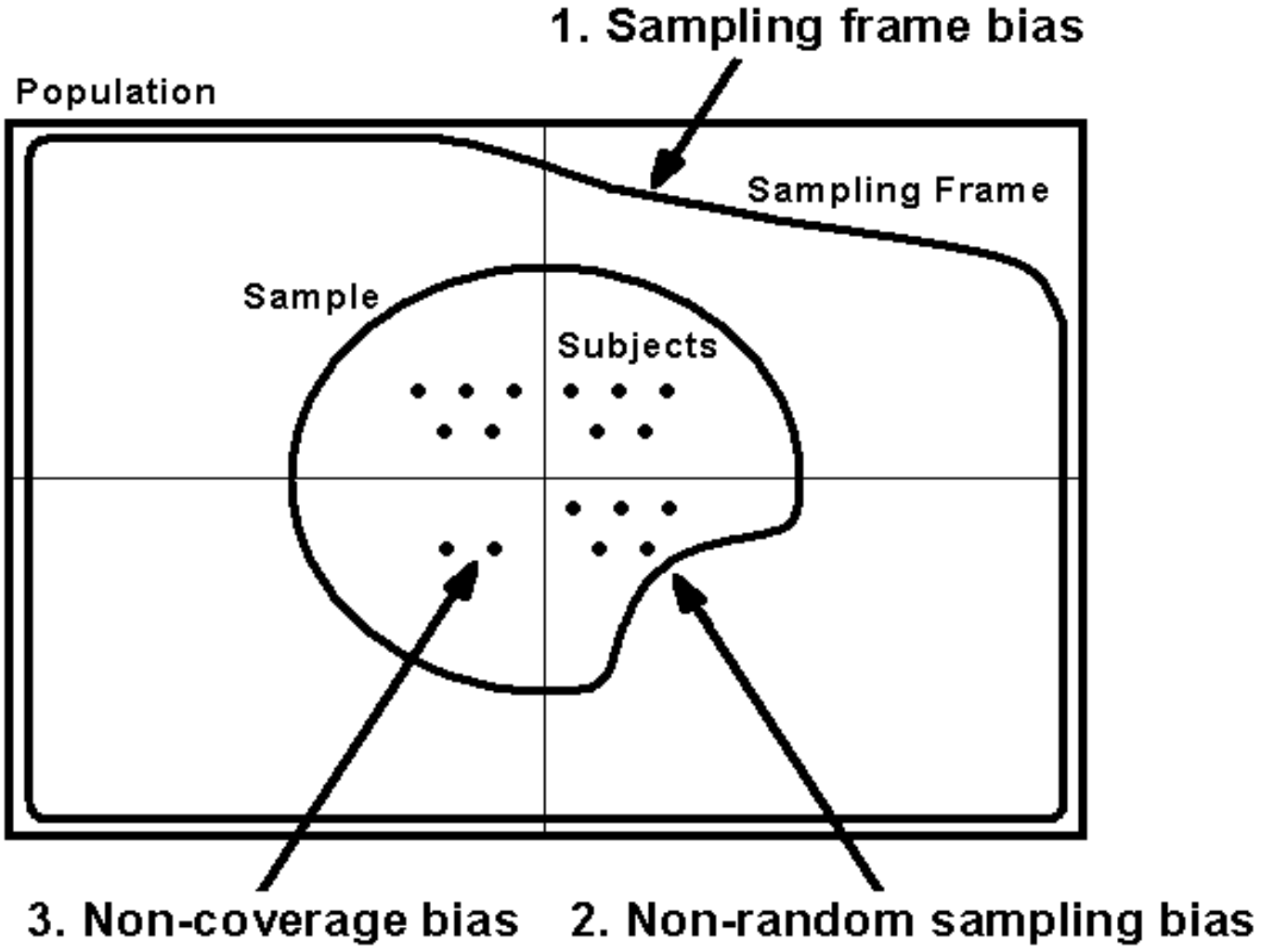
Disadvantages

- Examines existing cases at a single point in time (not new cases over follow-up period)
- Includes only cases that survive to be available for study (i.e., prevalent cases)
- Short duration diseases may be under-represented
- Cannot establish temporality and directionality
  - Does E → D? Does D → E?

Selection bias in cross-sectional studies

- Sources:
  - Bias due to sampling
    - Selection of “survivors” or “prevalent” cases
    - Non-random sampling schemes
    - Volunteer bias
    - Membership bias
  - Bias due to non-participation
    - Non-response bias

Selection bias in sample surveys



Information bias in cross-sectional studies

- Sources:
  - Quality of the tool, instrument or test used to collect data
    - Questionnaires
    - Tests, etc.
- Example: If you did a sero-prevalence survey on Covid-19, and used a test with low specificity, most positive people will be false-positives.