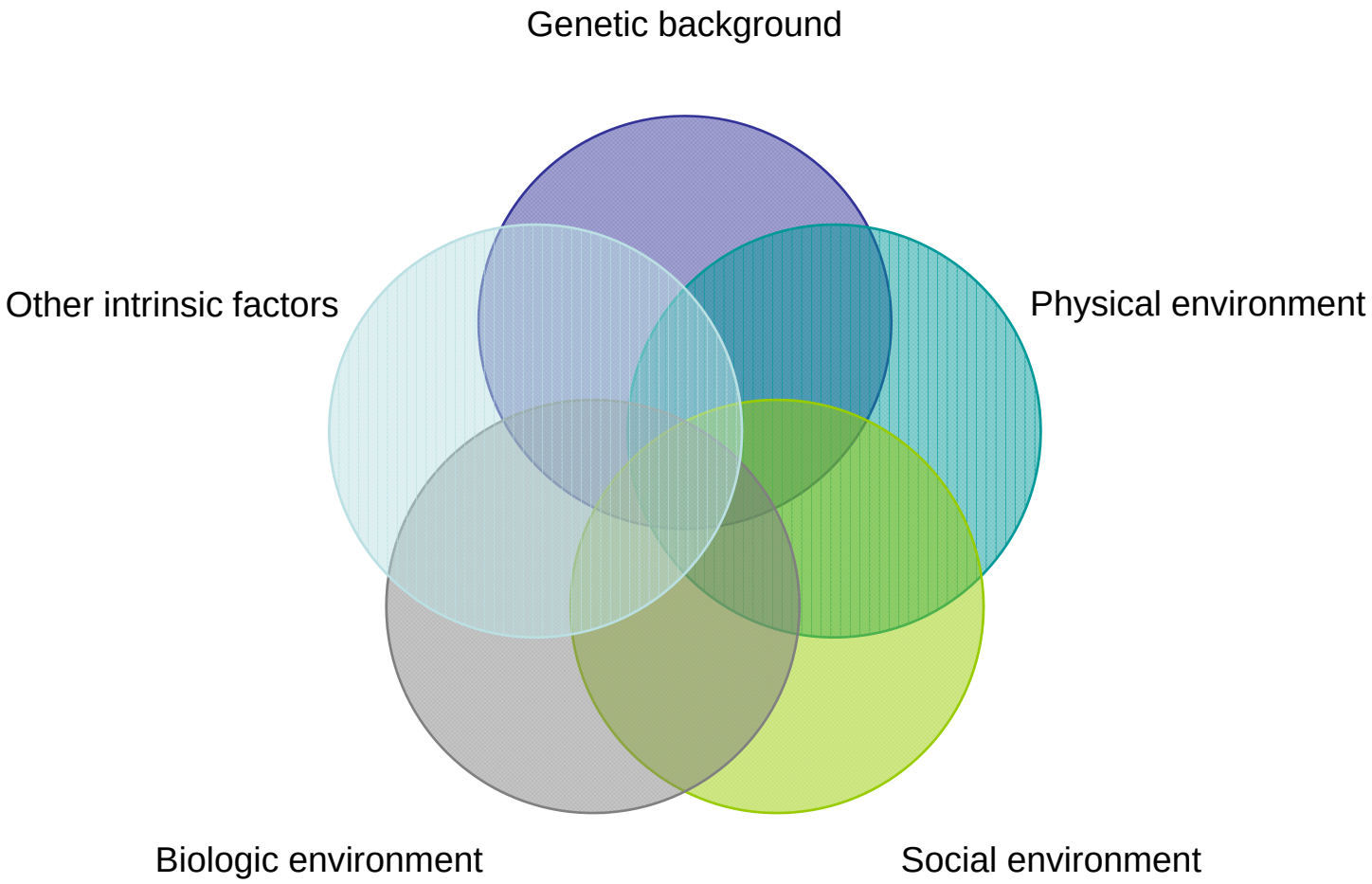
Linking Exposures and Endpoints: Measures of Association and Risk

Emily L. Harris, PhD, MPH
Epidemiologist
Office of Population Genomics
NHGRI

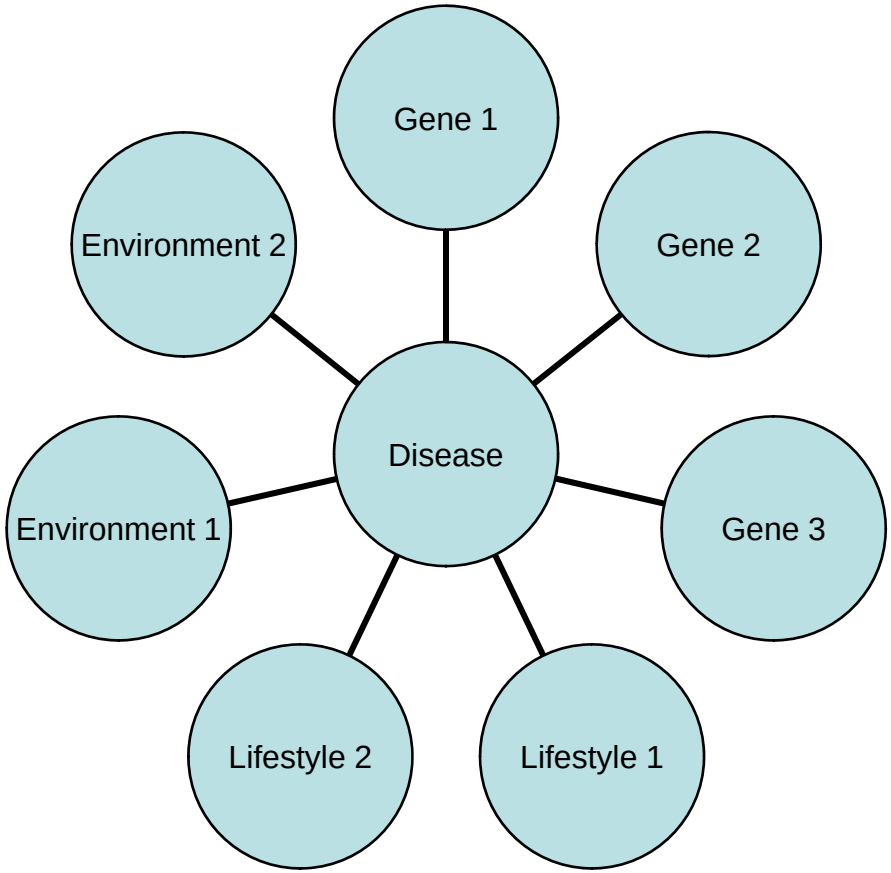
Learning Objectives

- Understand common measures of disease risk, and of association in epidemiologic designs.
- Understand the concept of interaction (gene-gene or gene-environment) in epidemiologic designs.
- Understand the concept of hypothesis testing, including p-values and statistical power.

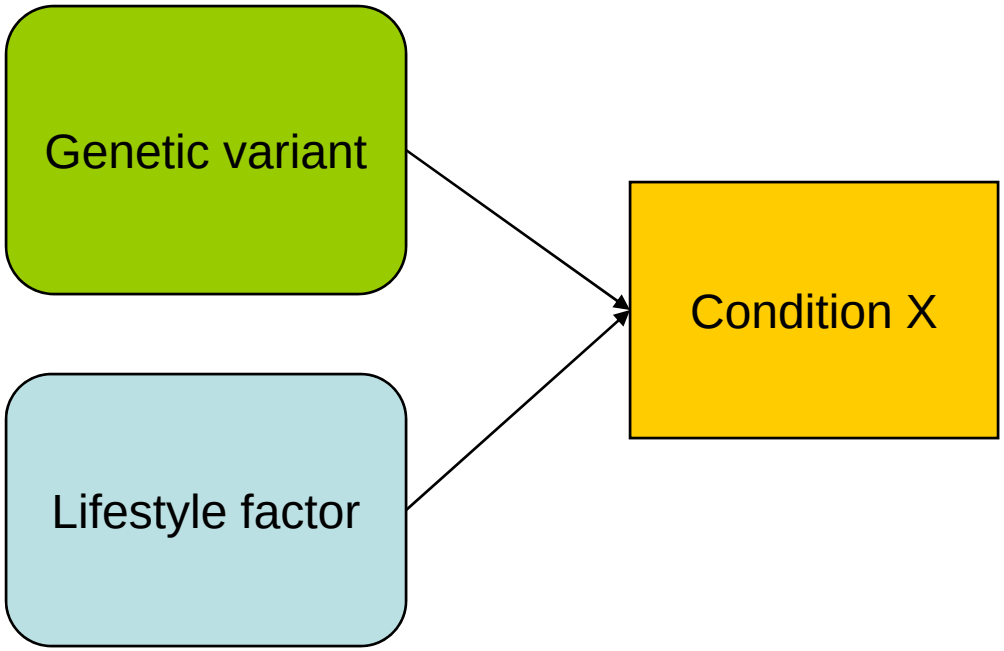
Etiologic Models



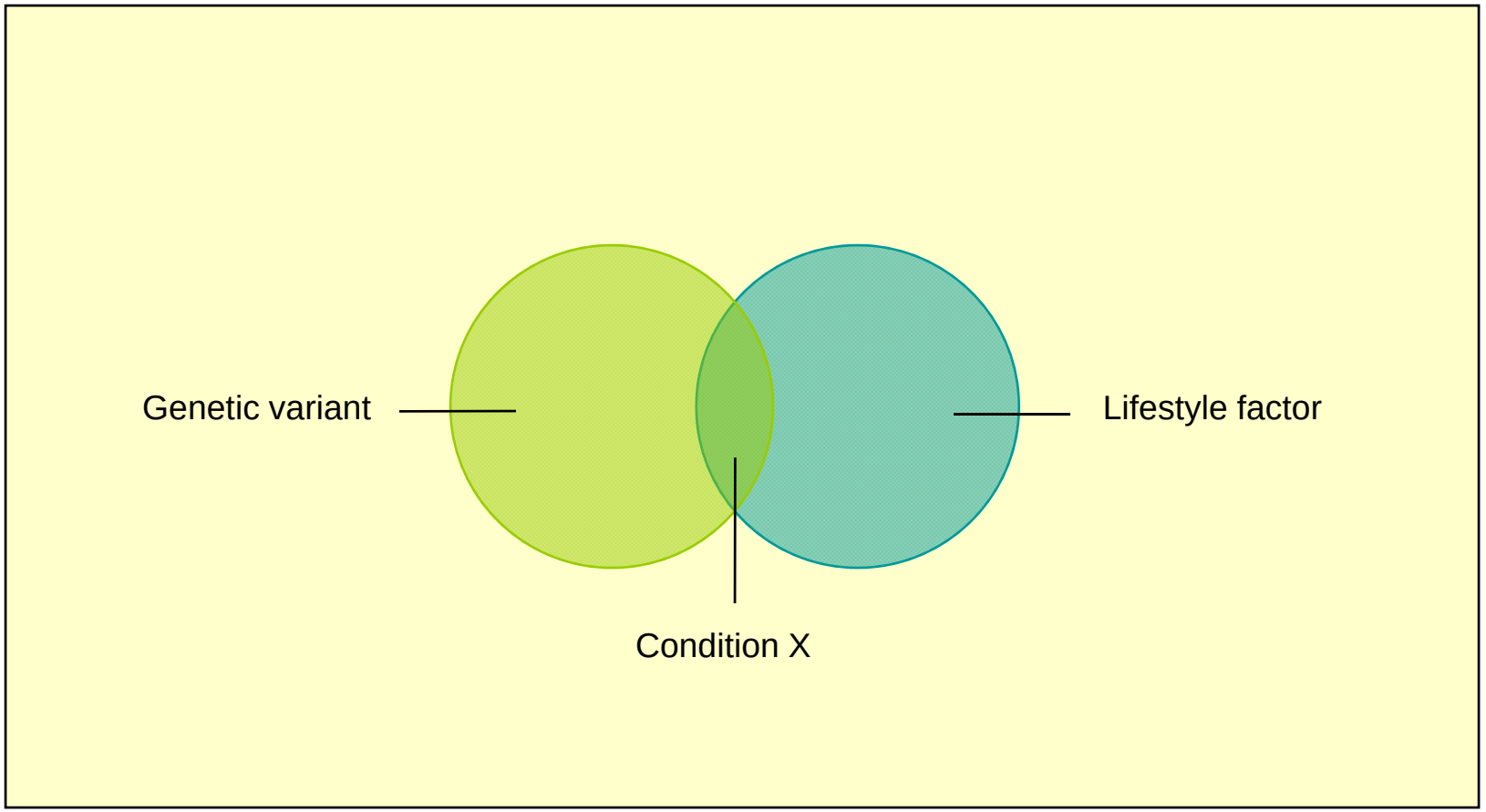
Web of Disease Causation



Etiologic Model

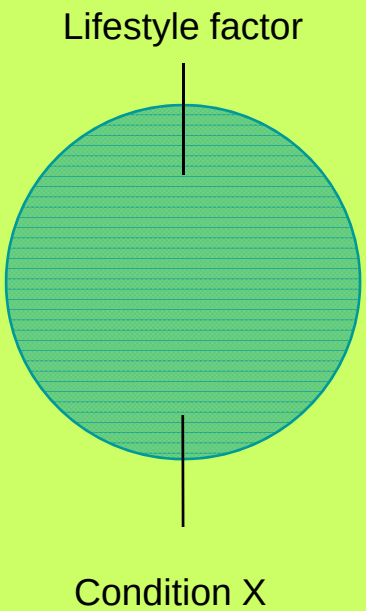


Detecting Associations



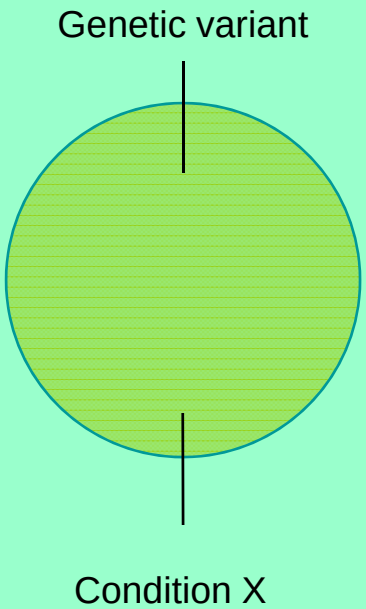
Detecting Associations: What can we detect?

Genetic variant
(universal)

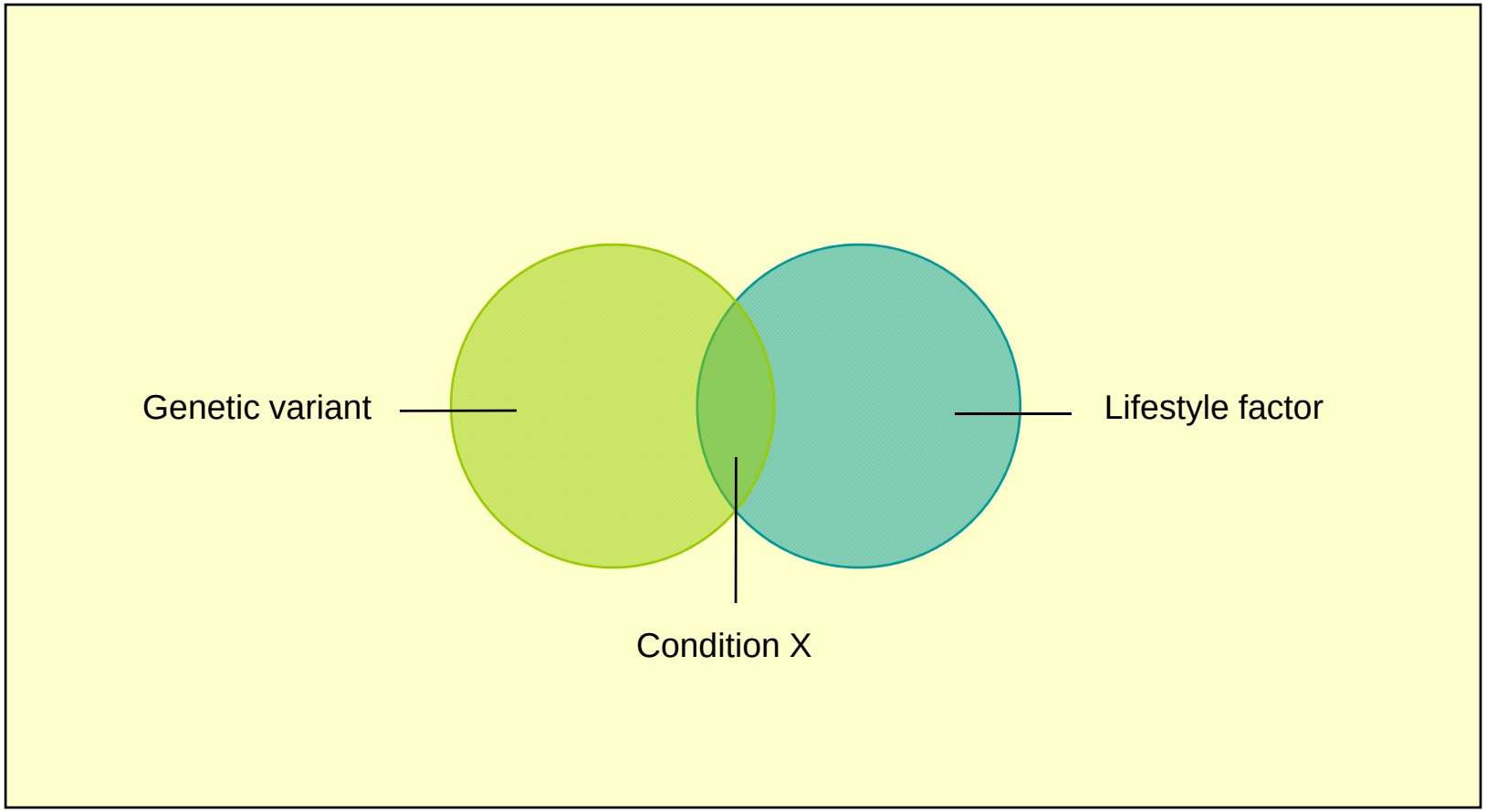


Detecting Associations: What can we detect?

Lifestyle factor
(universal)



Detecting Associations: What can we detect?

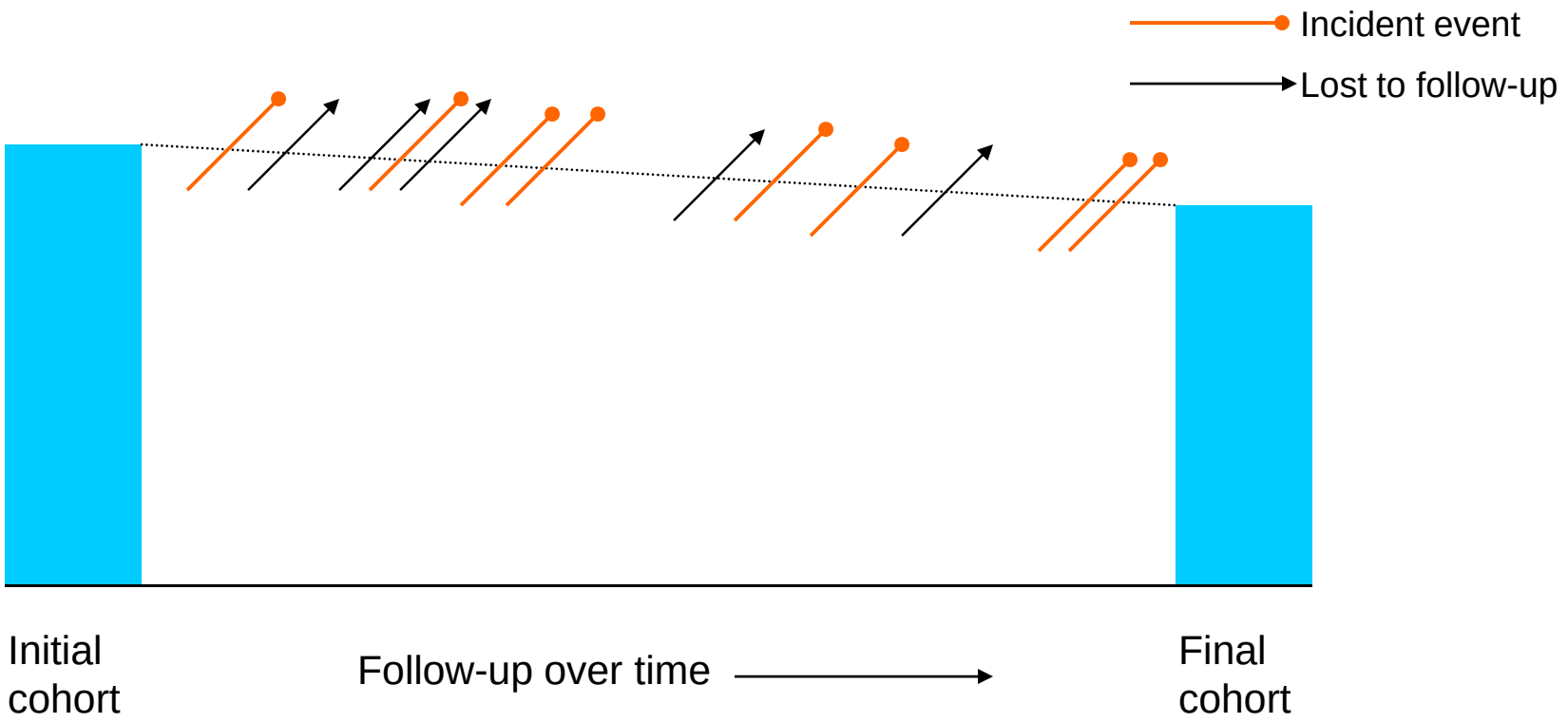


Measures of Risk

- Incidence
  - Newly-diagnosed cases
- Prevalence
  - Existing cases at a point in time
- Cumulative risk
  - Probability of developing disease over time

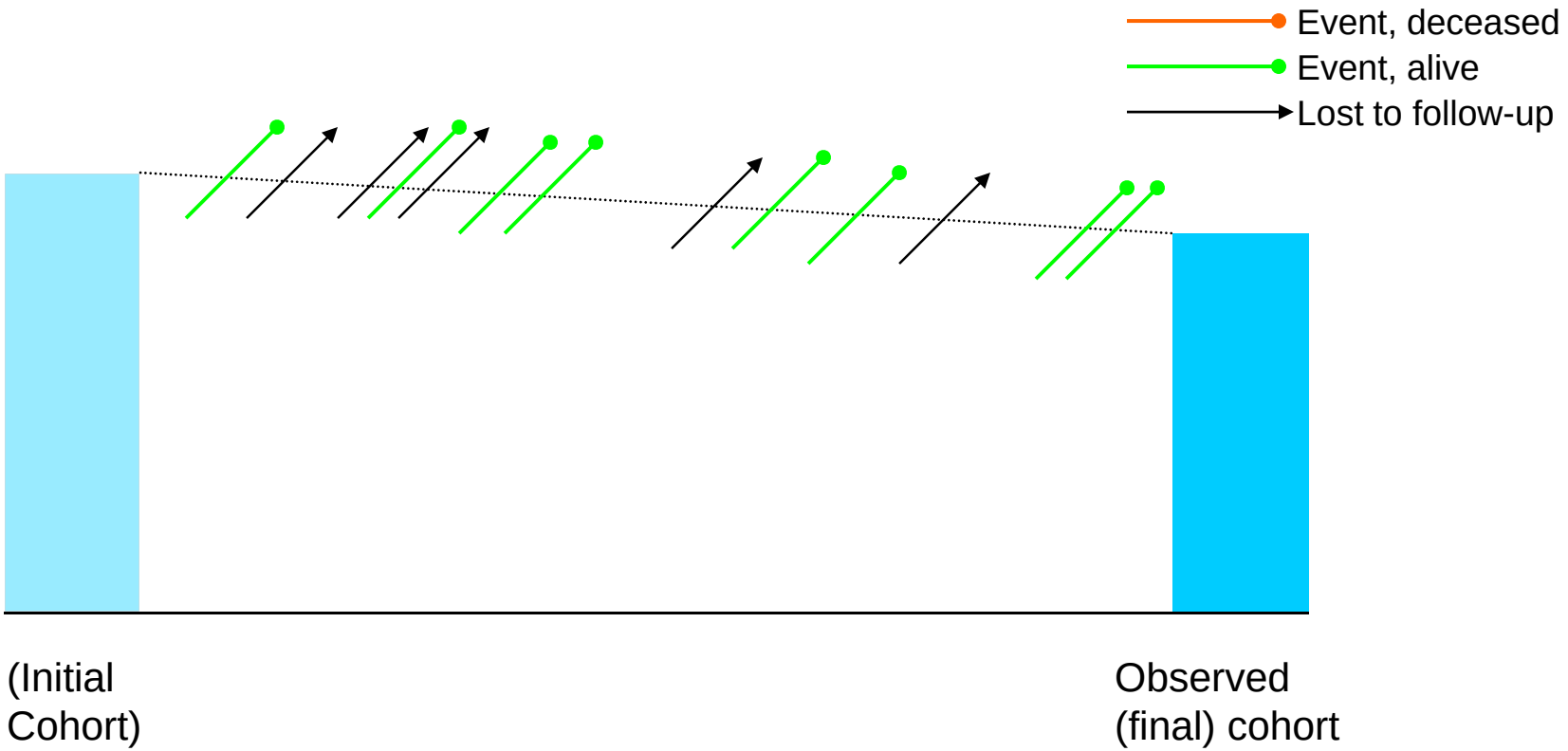
Incidence

New cases of a condition over a specified period of time (e.g., 5/100,000 per year)



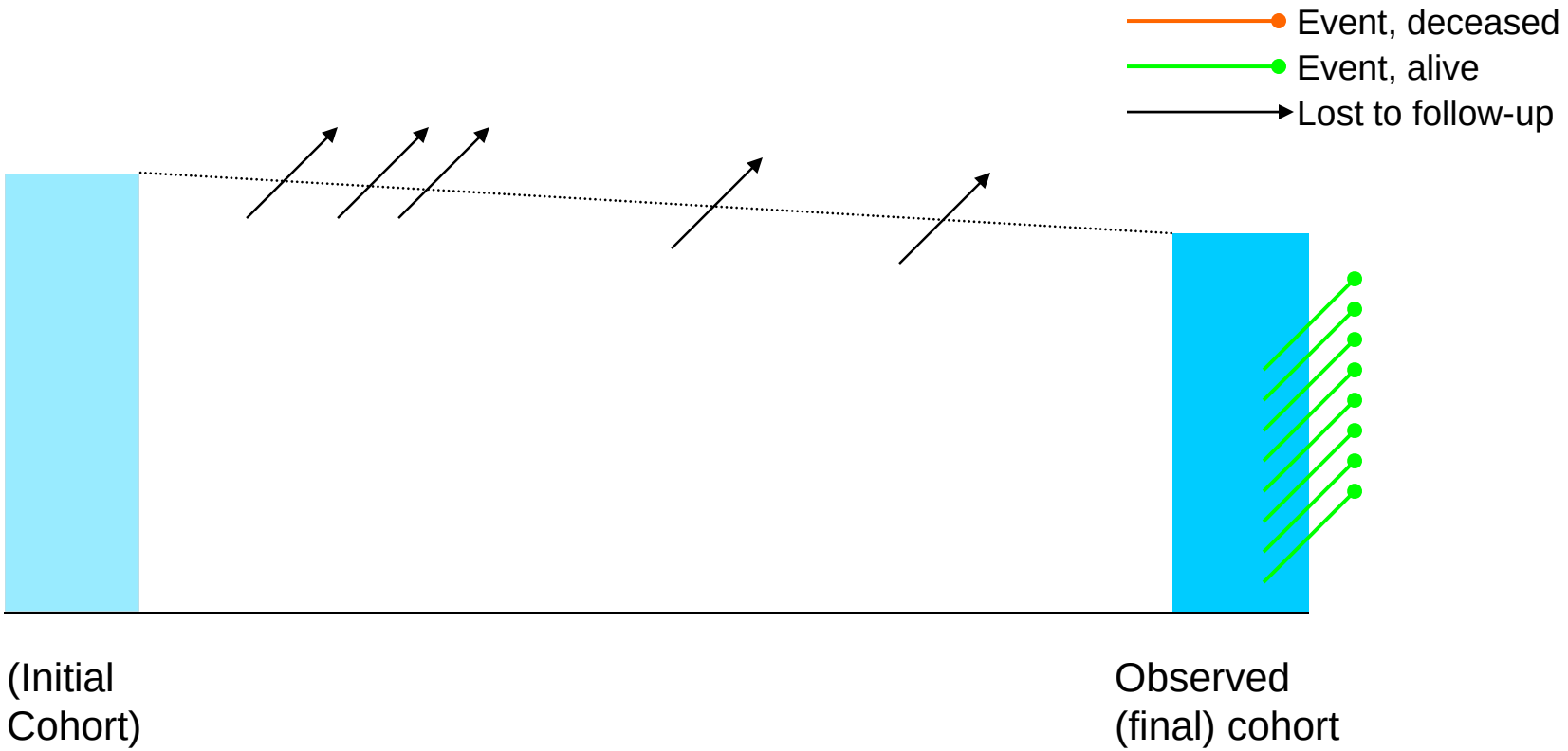
Prevalence

Cases of a condition at a point in time (e.g., 8/1,000)



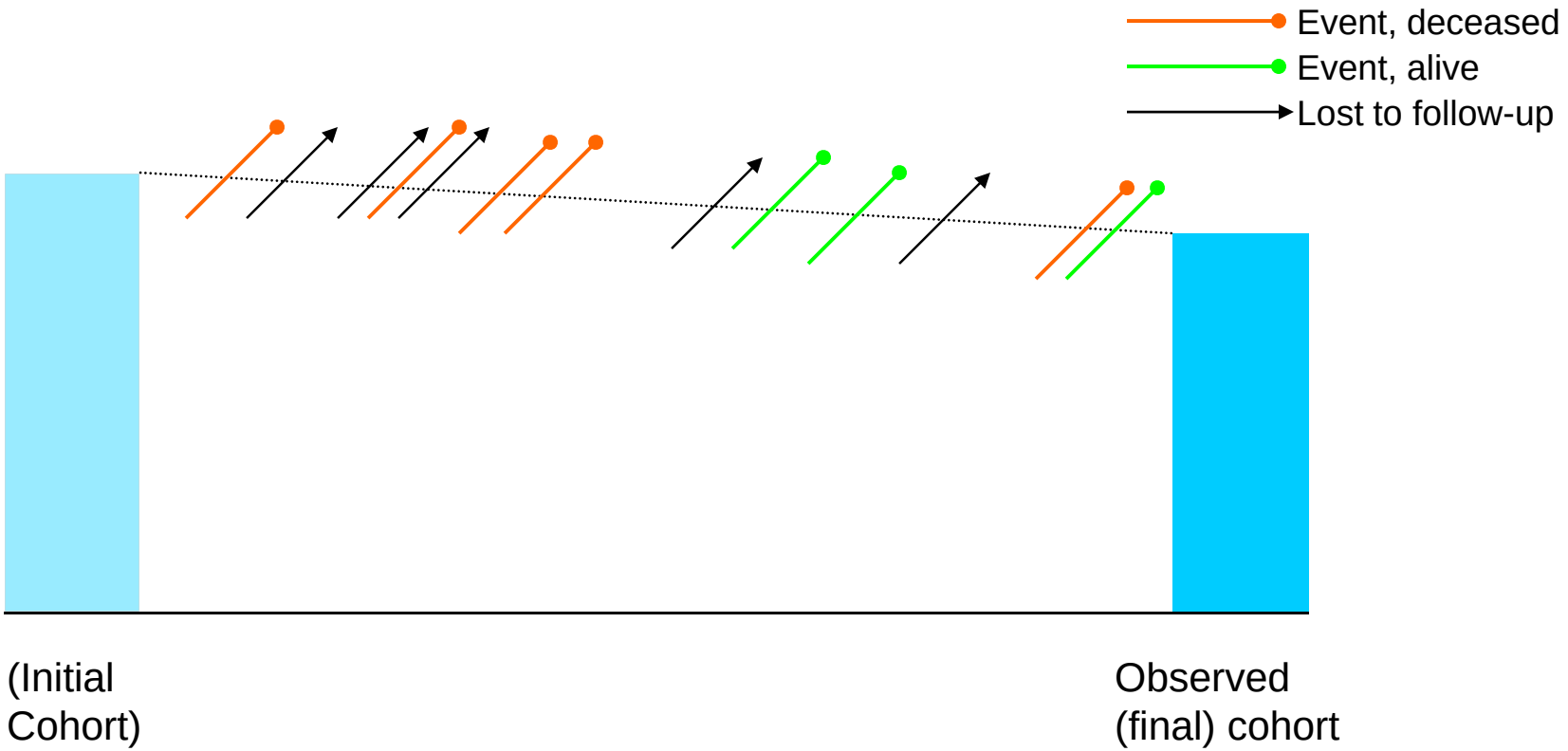
Prevalence

Cases of a condition at a point in time (e.g., 8/1,000)



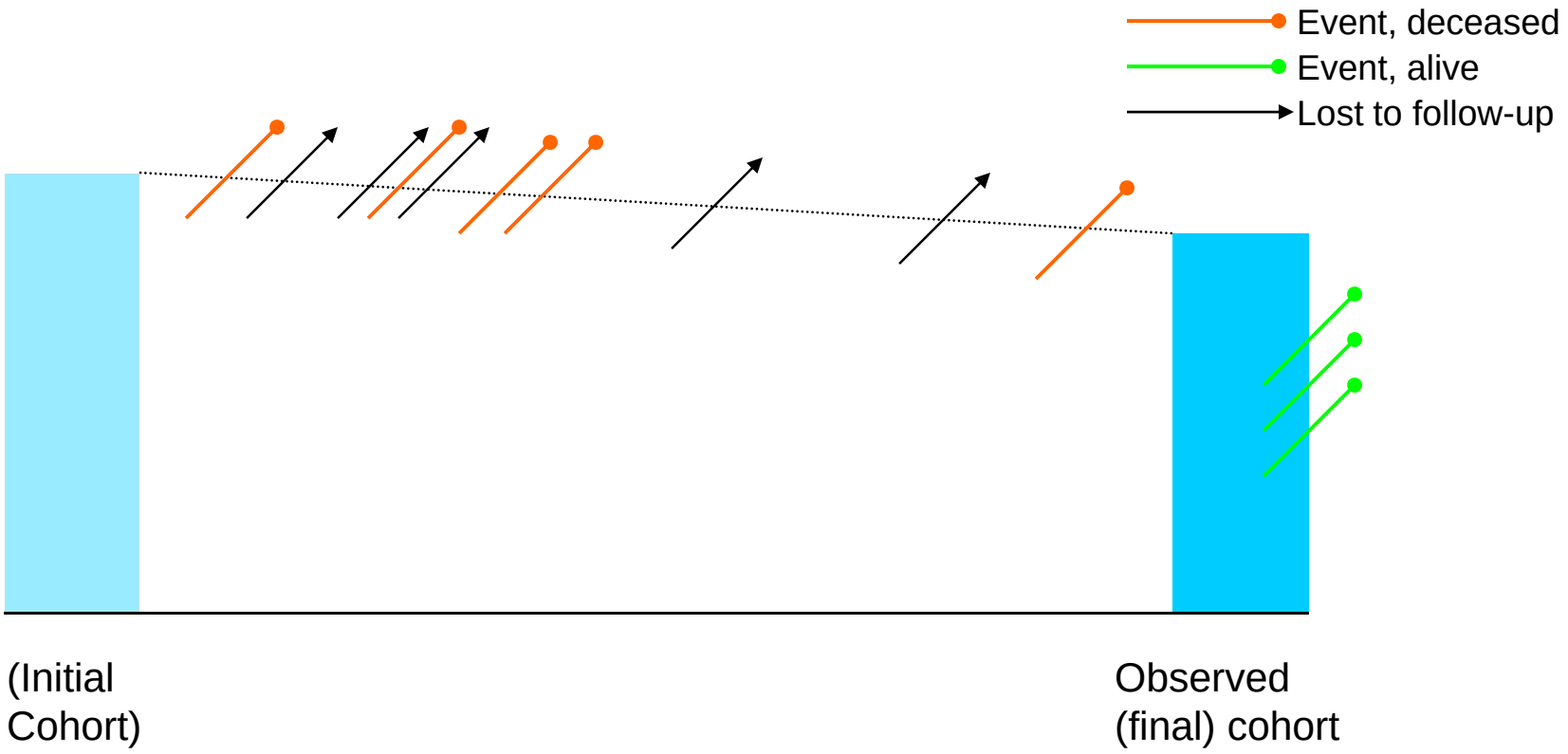
Prevalence

Cases of a condition at a point in time (e.g., 3/1,000)



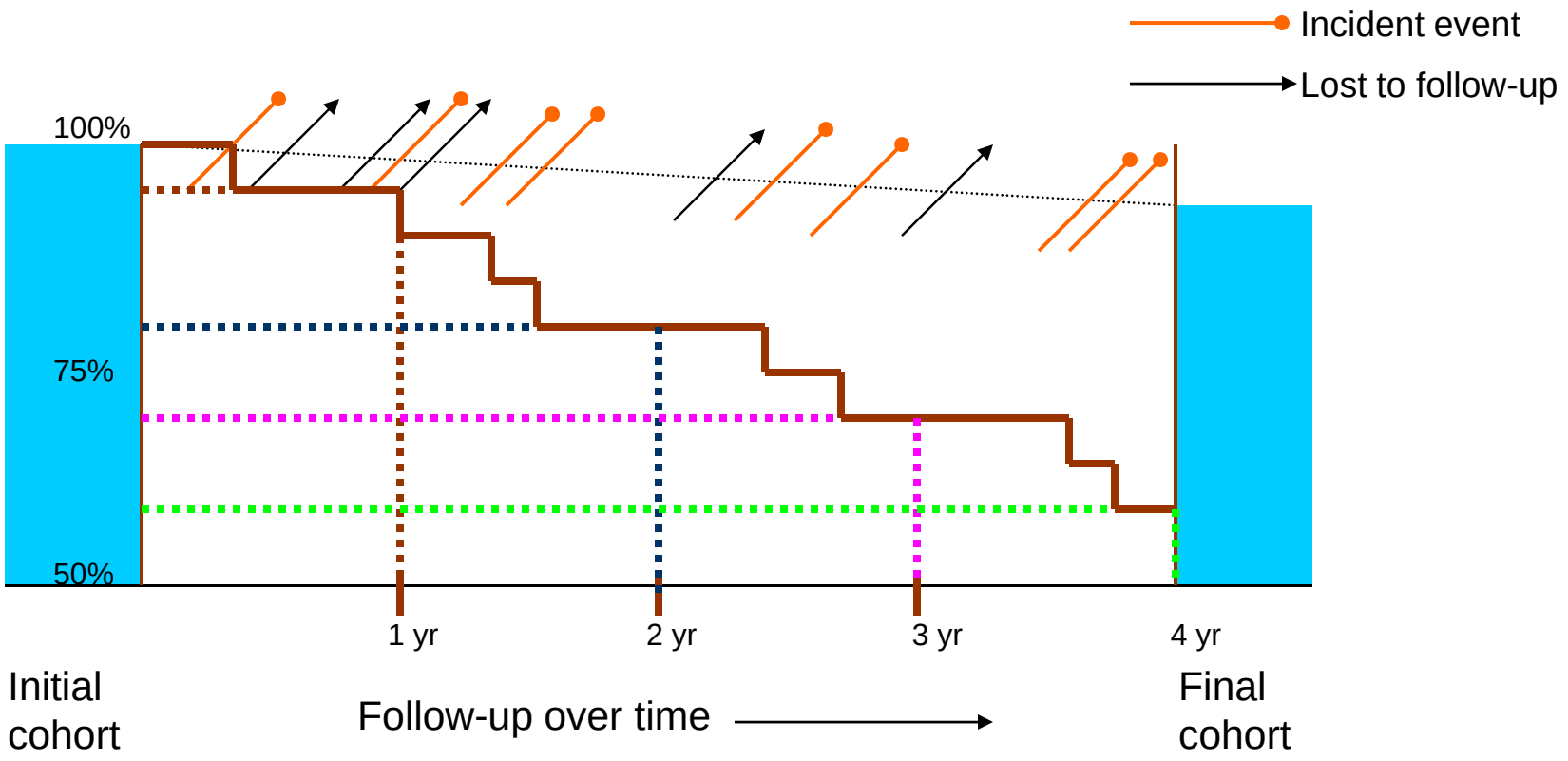
Prevalence

Cases of a condition at a point in time (e.g., 3/1,000)



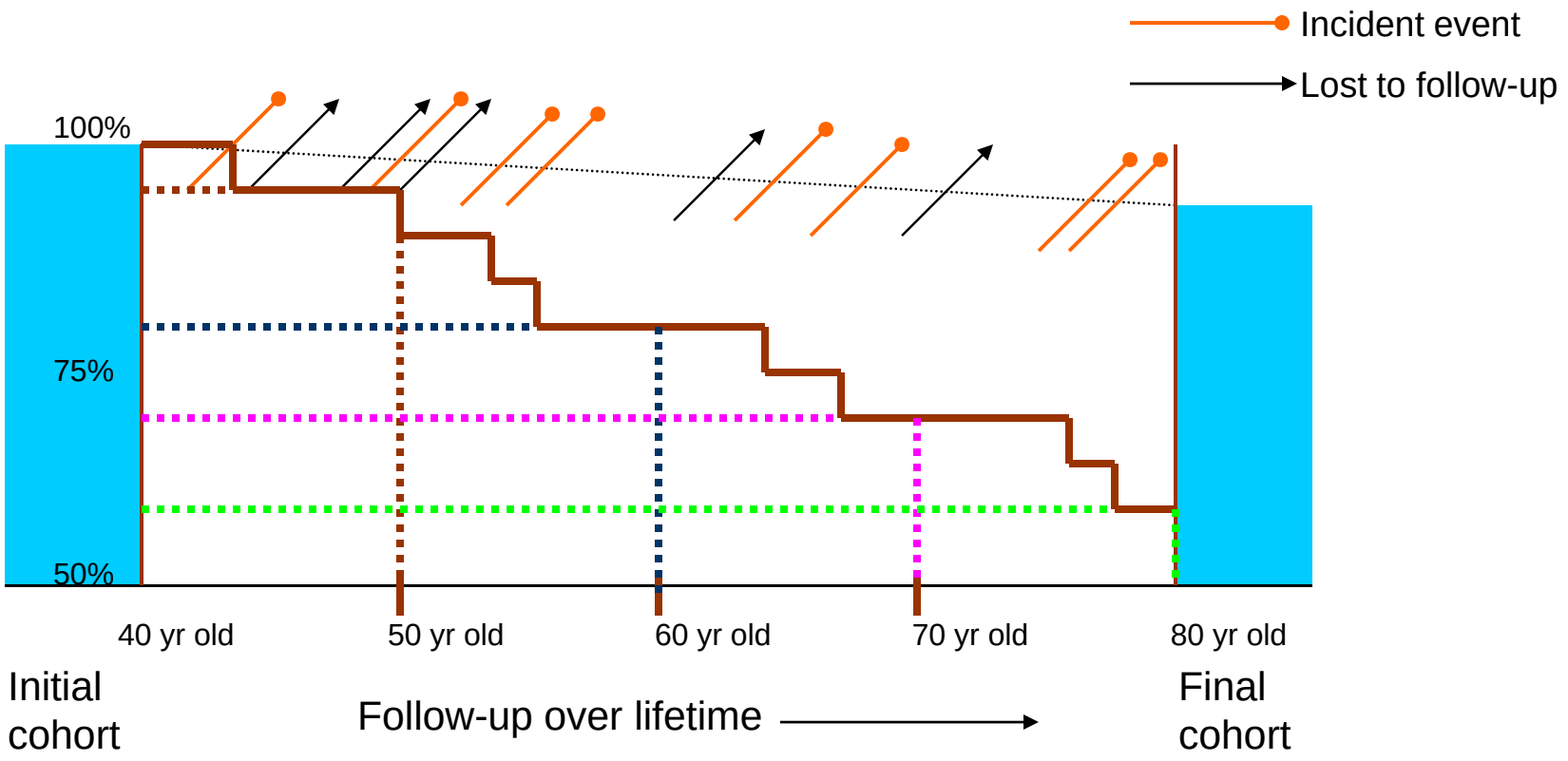
Cumulative Risk

Probability of developing a condition over a certain period of time



Cumulative Risk

Probability of developing a condition over a certain period of time



Measures of Association

- Strength of the association
  - Valuable for etiologic research & hypothesis testing
  - Measures:
    - Relative risk
    - Odds ratio
- Importance in the population
  - Applicable in clinical practice and public health
  - Measures:
    - Population attributable risk

Hypothesis Testing

- Idea that we want to evaluate:

  Is Genotype A is associated with Disease Z?

- How to test?

  Reject something: “null hypothesis”

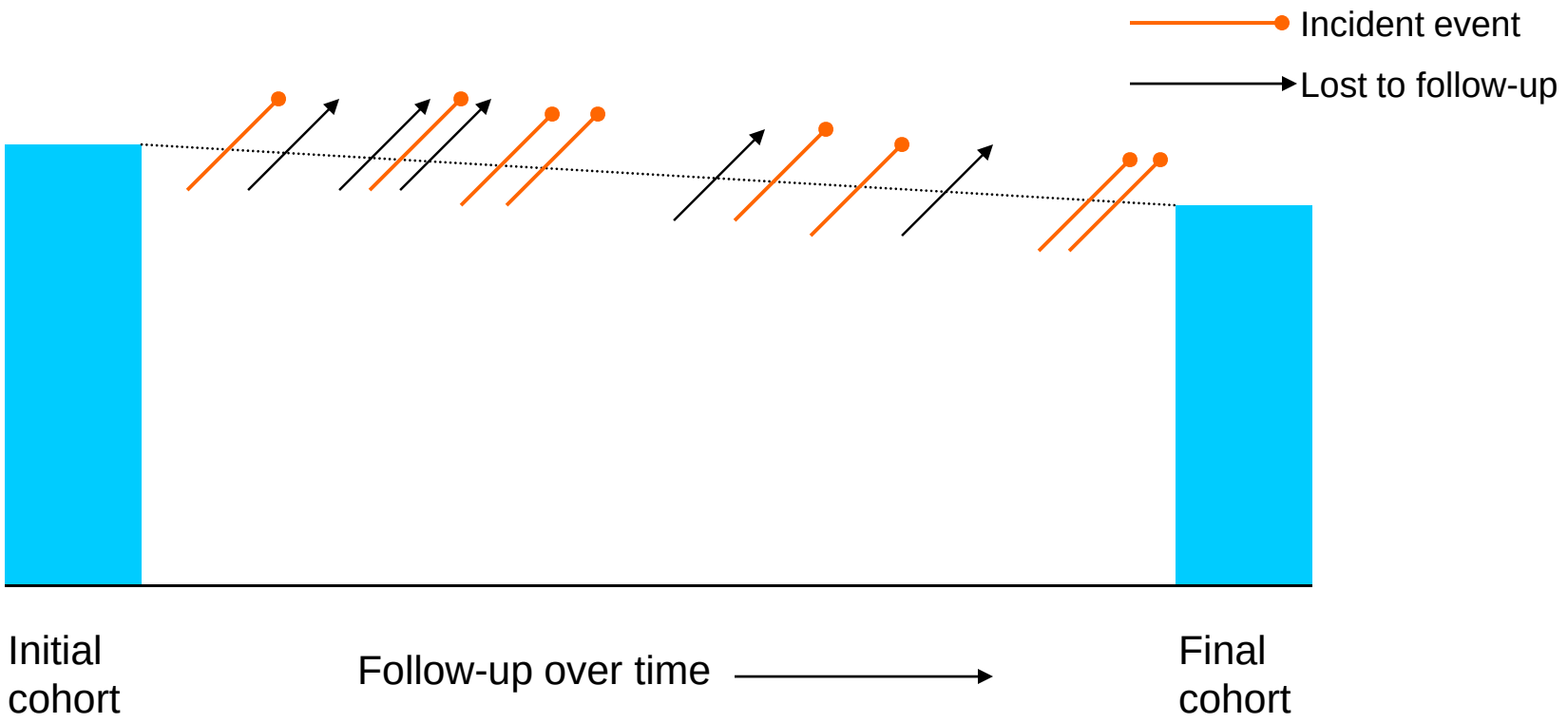
  - Null hypothesis (H_0): There is no association.
  - Alternative hypothesis (H_A): Genotype A is associated with Disease Z.

Hypothesis Testing

Statistical test result	Reality	
	H_0 is true (No association)	H_0 is not true (Association)
Not statistically significant (H_0)	Conclusion true	Type II error (1-power)
Statistically significant (Reject H_0)	Type I error (p-value)	Conclusion true (power)

Incidence

New cases of a condition over a specified period of time (e.g., 5/100,000 per year)



Relative Risk

Comparing the incidence rate or risk for two or more cohorts defined by the characteristic under study

Characteristic		Total	Event	No event
Genotype AA	→	5,000	50 (a)	4,950 (b)
Genotype AB or BB	→	15,000	50 (c)	14,950 (d)

$$\text{Relative Risk} = \frac{a/(a+b)}{c/(c+d)} = \frac{50/5{,}000}{50/15{,}000} = 3.0$$

Odds Ratio

Comparing the odds of exposure for cases to the odds of exposure for controls

Characteristic
Genotype AA
Genotype AB or BB

	Event (Cases)	No event (Controls)
	50 a	50 b
	50 c	150 d
Total	100	200

$$\text{Odds Ratio} = \frac{a/b}{c/d} = \frac{50/50}{50/150} = 3.0$$

Relative Risk vs. Odds Ratio

Odds ratio ~= Relative risk

if disease is relatively rare

$$\text{Relative Risk} = \frac{a/(a+b)}{c/(c+d)} = \frac{50/5{,}000}{50/15{,}000} = 3.0$$

$$\text{Odds Ratio} = \frac{a/b}{c/d} = \frac{50/50}{50/150} = 3.0$$

Population Attributable Risk

Proportion of disease risk attributable to a certain exposure

$$\text{Population attributable Risk (PAR)} = \frac{\text{Risk factor prev popn} * (\text{Relative risk} - 1)}{\text{Risk factor prev popn} * (\text{Relative risk} - 1) + 1}$$

$$\text{PAR} = \frac{0.25*(3.0-1)}{0.25*(3.0-1) + 1} = 0.33$$

Measures of Interaction

(Gene-Gene or Gene-Environment)

Ways to think about interaction:

- Combined effect of two or more risk factors are different from what you would predict from their individual effects.
  - E.g., Odds ratios: 2 x 2 = 6
- Effect of a risk factor differs across subgroups.
  - E.g., Odds ratio for a genotype = 3.0 for those with a certain non-genetic risk factor, vs. 1.0 for those without the risk factor

Example from Genetics:

- Single locus: Dominance
- More than one locus: Epistasis

Gene-Environment Interaction

Table 15–14. Estimated Population Incidence per 10,000 Person-Years of First Venous Thrombosis in Women Aged 15 to 49 Years According to Presence of Factor V Leiden Mutation and Use of Oral Contraceptives

	Factor V Leiden Mutation	
	Absent	*Present*
Did not use oral contraceptives	0.8	5.7
Used oral contraceptives	3.0	28.5

Adapted from Vandenbroucke JP, Koster T, Briët E, et al: Increased risk of venous thrombosis in oral contraceptive users who are carriers of factor V Leiden mutation. Lancet 344:1453–1457, 1994.

Gordis. *Epidemiology*. Philadelphia: W.B. Saunders Company, 1996.

Gene-Environment Interaction

Using incidence rates from previous slide:
Relative risks for venous thrombosis

	Factor V Leiden mutation	
	Absent	Present
Did not use oral contraceptives	1 (Reference)	7.1
Used oral contraceptives	3.7	35.6

Gene-Environment Interaction

Using incidence rates from previous slide:
Relative risks for venous thrombosis

	Factor V Leiden mutation	
	Absent	Present
Did not use oral contraceptives	1 (Reference)	1 (Reference)
Used oral contraceptives	3.7	5.0

Summing Up

- Etiologic models
  - To detect an etiologic factor, we need variation
  - Differences on studies on etiologic factors among studies may reflect differences in the distribution of factors across population

Summing Up

- Measures of Risk
  - Incidence: new cases
  - Prevalence: existing cases, a combination of incidence and survival
  - Cumulative risk: probably of developing disease over risk

Summing Up

- Strength of association
  - Relative risk or odds ratio
  - Important for etiologic studies
  - Consider gene-gene and gene-environment interaction
- Importance of a risk factor in the population
  - Population attributable risk
  - Important for public health
- Hypothesis testing
  - Consider type I and type II errors in designing and interpreting epidemiologic studies

Classic Measures of Nature versus Nurture

Emily L. Harris, PhD, MPH
Epidemiologist
Office of Population Genomics
NHGRI

Learning Objectives

- Learn about common designs for detecting genetic variation in common diseases or traits
- Appreciate the strengths and limitations of these designs
- Understand the concept of heritability
- Appreciate the limitations of heritability estimates

Detecting Genetic Variation

- Ecologic studies
- Migrant studies
- Adoption studies
- Twin studies

Ecologic Studies

- Comparison based on “group data,” not individual-level information
  - Comparing the frequency of a characteristic with disease risk across populations
  - Comparing disease incidence rates or risk across population groups
    - Across countries
    - Among race/ethnic groups within countries
- Ecologic fallacy
  - Ascribing characteristics to members of a group, when we do not know whether individuals have the characteristic

Ecologic Study

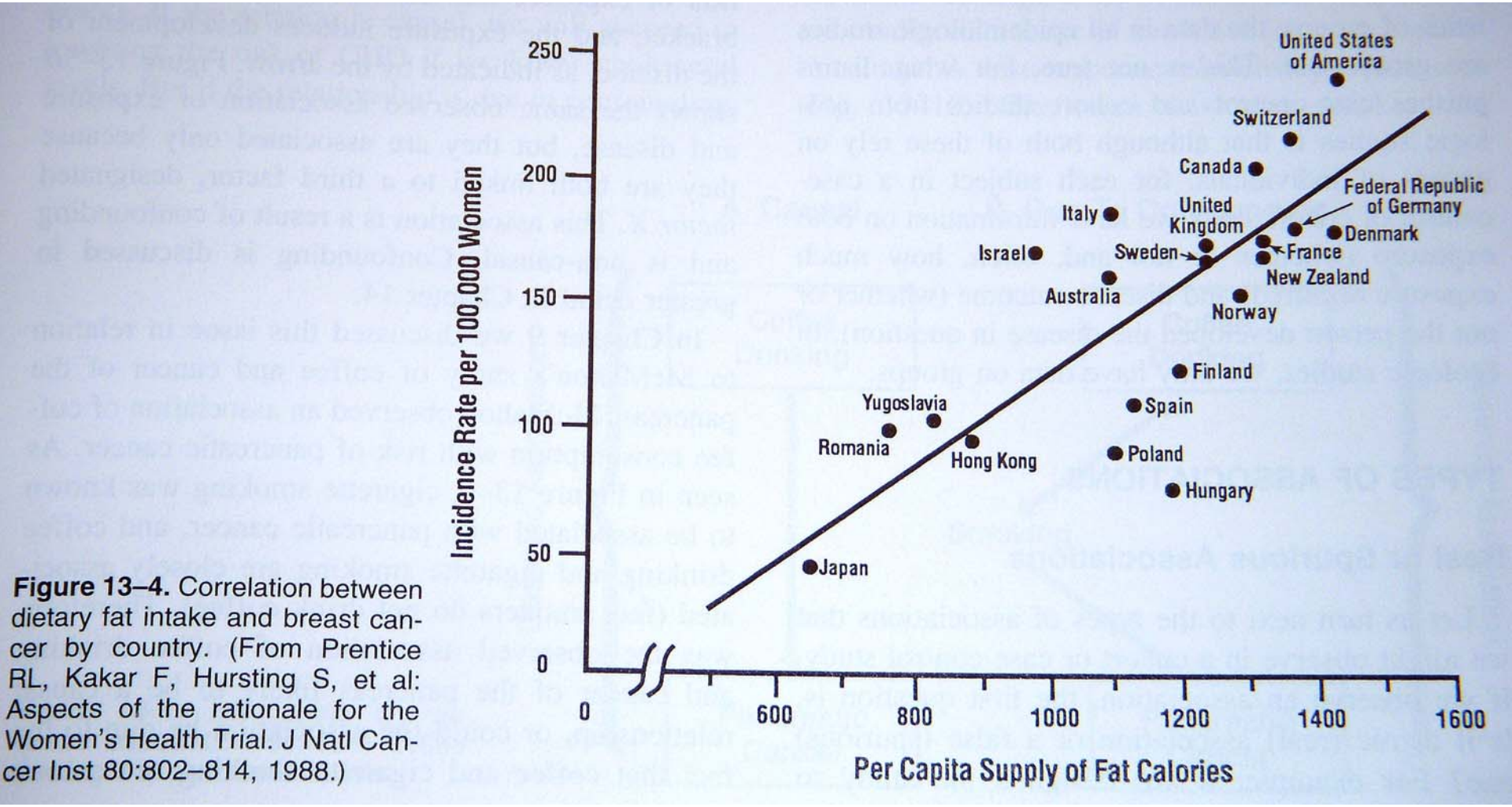


Figure 13–4. Correlation between dietary fat intake and breast cancer by country. (From Prentice RL, Kakar F, Hursting S, et al: Aspects of the rationale for the Women's Health Trial. J Natl Cancer Inst 80:802–814, 1988.)

Gordis. *Epidemiology*. Philadelphia: W.B. Saunders Company, 1996.

Ecologic Study

National Program of Cancer Registries, 2004

(http://apps.nccd.cdc.gov/uscs)

Cancer incidence for U.S. men

	White	Black	Asian/ Pacific Islander	American Indian/ Alaska Native	Hispanic
Stomach	8.7	16.1	17.5	9.5	14.5
Lung and Bronchus	84.4	104.5	49.7	51.1	48.5
Prostate	134.5	217.5	79.8	76.6	121.9

Rates are per 100,000 persons and are age-adjusted to the 2000 U.S. standard population.

Migrant Studies

- Compare disease incidence rates or risk for populations of the same ethnic background:
  - Living in the “home” country (or country of origin)
  - Migrated to a different country

Interpreting Migrant Studies

	Home country vs. New country
Genes	Shared
Environment	
Macro	Different
Micro	May be shared to some degree

Migrant Study

Table 15–11. Standardized Mortality Ratios (SMRs) for Cancer of the Stomach in Japanese Men, Issei, Nisei, and U.S. White Males

	SMRs
Japanese men	100
Issei*	72
Nisei*	38
U.S. white males	17

*Issei and nisei are first- and second-generation Japanese migrants, respectively.

From Haenzel W, Kurihara M: Studies of Japanese migrants: I. Mortality from cancer and other disease among Japanese in the United States. J Nat Cancer Inst 40:43–68, 1968.

Gordis. *Epidemiology*. Philadelphia: W.B. Saunders Company, 1996.

Adoption Studies

- Compare biologic and adoptive relatives
  - Disease risk
  - Similarity in traits

	Adoptive relatives	**Biologic relatives**
Genes	Not shared	Shared
Environment	Shared	Shared or Not shared*

* Depends on the design

Adoption Study

Table 15–10. Correlation Coefficients for Parent-Child Aggregation of Blood Pressure

	Between Parents and	
	Biologic Child	***Adopted Child***
Systolic	.32 (P<.001)	.09 (NS)
Diastolic	.37 (P<.001)	.10 (NS)

Abbreviation: NS, not significant.

Adapted from Biron P, Mongeau JG, Bertrand D: Familial aggregation of blood pressure in 558 adopted children. CMAJ 115:773–774, 1975.

Gordis. *Epidemiology*. Philadelphia: W.B. Saunders Company, 1996.

Twin Studies

- Compare the similarity of identical twins vs. fraternal twins
  - Similarity in traits
  - Disease concordance
- Correlations within twin families
  - Identical twins and their offspring
    - “First cousins” who are genetic half-siblings
    - Maternal and paternal effects

Interpreting Twin Studies

	MZ twins	**DZ twins**
Genes	100% shared	50% shared (average)
Additive	100%	50%
Dominance	100%	25%
Environment	Shared to some degree	Shared to some degree

Interpreting Twin Studies

What makes individuals in twin-pairs similar?

MZ twins

- Genetics
  - Additive: 100%
  - Dominance: 100%
- Shared environment

DZ twins

- Genetics
  - Additive: 50%
  - Dominance: 25%
- Shared environment

The importance of shared genes and shared environment in total variation is reflected in the similarity of MZ and DZ twins.

Measuring Similarity

Intraclass correlation coefficient (r_i)

- Proportion of the total variation due to variation among twin-pairs
- Range: 0 to 1
  - $r_i = 1$: Individuals within each pair are exactly alike
  - $r_i = 0$: Individuals within each pair are no more alike than two randomly selected individuals
- Thus, there's an internal "control" for random similarity.

Twin Study

Table 3

Twin Correlations for Height by Country and Zygosity Group

	Australia	Denmark	Finland	Italy	Netherlands	Norway	Sweden	UK
MZm	0.87	0.89	0.92	0.94	0.89	0.87	0.89	n.a.
DZm	0.42	0.47	0.53	0.57	0.47	0.49	0.56	n.a.
MZf	0.84	0.89	0.87	0.94	0.90	0.89	0.89	0.88
DZf	0.49	0.55	0.53	0.49	0.49	0.49	0.49	0.56
DOS	0.46	0.50	0.49	0.30	0.43	0.44	n.a.	n.a.

Note: MZm = male monozygotic twins, DZm = male dizygotic twins, MZf = female monozygotic twins, DZf = female dizygotic twins, DOS = opposite sex twin pairs

Silventoinen et al. Twin Research 2003; 6:399-408.

Measuring Similarity

Concordance rates

- Probability of an individual having a disease if her/his co-twin has the disease
- Range: 0 to 1
  - Concordance = 1: Individual will get the disease co-twin has it
  - Concordance = Population risk: Individual not at increased risk of getting the disease because co-twin has it (no familial clustering)
- Problem: Need information about population risk to accurately interpret concordance rates
  - Can convert concordance rates to correlation in underlying risk, if population risk is known

Concordance Rates

Table 15–5. Concordance Rates of Anencephaly and Spina Bifida (ASB) in New York State, 1955–1974

Incidence of ASB	1.3/1,000	
Concordance rates		
Among co-twins	4/59	(6.8%)
Among full siblings	19/1,037	(1.8%)
Among half siblings	1/133	(0.8%)

From Janerich DT, Piper J: Shifting genetic patterns in anencephaly and spina bifida. J Med Genet 15:101–105, 1978.

Gordis. *Epidemiology*. Philadelphia: W.B. Saunders Company, 1996.

Heritability

Proportion of total phenotypic variation that is attributable to genetic variation

Common formula:

$2*(r_{iMZ} - r_{iDZ})$

Genetic variation represented:

$2*[(V_A+V_D)-(0.5\ V_A + 0.25\ V_D)]$

$= V_A + 1.5\ V_D$

Twin Study: Heritability

2*(riMZ – riDZ)

<u>Australia</u>

Males: 2*(0.87-0.42) = 0.90

Females: 2*(0.84-0.49) = 0.70

<u>Denmark</u>

Males: 2*(0.89-0.47) = 0.84

Females: 2*(0.89-0.55) = 0.68

Table 3

Twin Correlations for Height by Country and Zygosity Group

	Australia	Denmark	Finland
MZm	0.87	0.89	0.92
DZm	0.42	0.47	0.53
MZf	0.84	0.89	0.87
DZf	0.49	0.55	0.53
DOS	0.46	0.50	0.49

Note: MZm = male monozygotic twins, DZm = male dizygotic twins, MZf = fe
DZf = female dizygotic twins, DOS = opposite sex twin pairs

Silventoinen et al. *Twin Research* 2003; 6:399-408.

Heritability

Interpret with caution

- Requires assumptions:
  - All estimates require important assumptions about the type of genetic variation, shared environment, and gene-gene and gene-environment interaction.
  - Estimates may be right-on, overestimates, or underestimates.
- Limited precision:
  - Confidence intervals around estimates are typically large.

Summing Up

- Ecologic, migrant, adoption, and twin studies can provide evidence of genetic variation contributing to disease risk.
- Each design has limitations that may affect interpretation.
- Heritability estimates should be viewed with caution, because of the assumptions underlying the estimates.