

# Epidemiology Concepts and Study Designs

Martha S. Linet, M.D.

Radiation Epidemiology Branch  
Division of Cancer Epidemiology  
and Genetics

May 17, 2011

# Outline

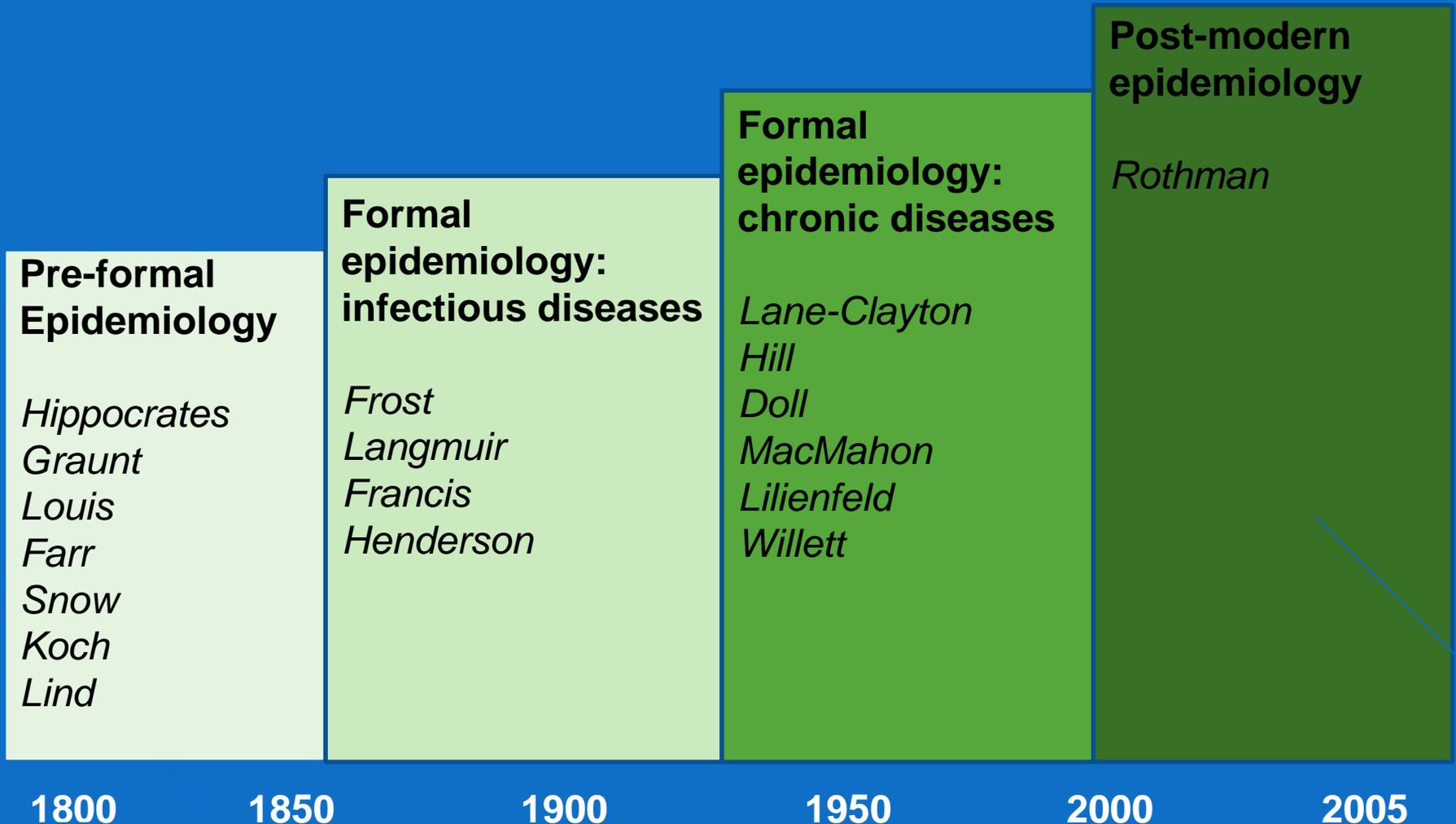
---

- **Brief history and definitions**
- **Descriptive patterns**
- **Disease models: infectious, chronic, genetic**
- **Disease causation: causal criteria and risk factors**
- **Sources of outcome information**
- **Study designs**
  - **Cohort**
  - **Case-control**
  - **Cross-sectional**
- **Exposure assessment**

# History and Definitions

# History: Epidemiology from 1850- present

---



Based on A. McMichael

# Key definitions in Epidemiology - 1

---

- **Epidemiology:** the study of the distribution of a disease or condition in human populations and the factors that influence the distribution
- **Disease:** any departure from perfect health
- **Endemic:** habitual presence of a disease within a given geographic area (also, the usual prevalence of a given disease within such an area)
- **Epidemic:** occurrence in a community or region of a group of illnesses of a similar nature in excess of normal expectancy

# Key definitions in Epidemiology - 2

---

- **Chronic diseases:** characterized by a long natural history or frequent recurrence; likely multi-factorial in etiology
- **Rate:** a measure of change in a quantity per unit time
- **Risk:** the probability of disease developing in an individual in a specified time interval
- **Incidence:** the total number of new-onset disease events divided by the total person-time at risk during a given time period
- **Mortality:** the total number of deaths from a disease divided by the total person-time at risk during a given time period

# Key definitions in Epidemiology - 3

---

- **Correlation:** the degree to which variables change together (no direction assumed)
- **Association:** the observed frequency of a disease varies by the level of an exposure; the disease occurs more (or less) frequently in the presence of an exposure than in its absence
- **Causation:** in an individual, an exposure caused a given disease; within an exposed population, at least some cases of the disease would not have occurred in the absence of the exposure
- **Effect:** the result of a cause

# Key definitions in Epidemiology - 4

---

- **Relative risk:** the incidence of disease in an exposed group divided by the incidence of disease in a non-exposed group
- **Attributable risk:** the maximum proportion of a disease attributable to a given exposure
- **Absolute risk:** the observed or calculated probability of occurrence of an event in a population related to a specific exposure
- **Exposure:** an agent or substance presumed to be causal of a disease or event (exposure surrogate is a factor indicating exposure potential, e.g., job title)

# **Descriptive Patterns & Trends and Disease Classification**

# Descriptive Epidemiology

---

## Why study disease patterns and trends?

- Explain local disease occurrence
- Describe natural history of disease
- Provide guidance in administration of health services
- Suggest hypotheses to elucidate causal mechanisms

## What is the purpose of disease classification?

- Group ill persons into categories that permit distinguishing persons in one category from another
- Arrange disease entities into groups with common characteristics

# International Classification of Childhood Cancer

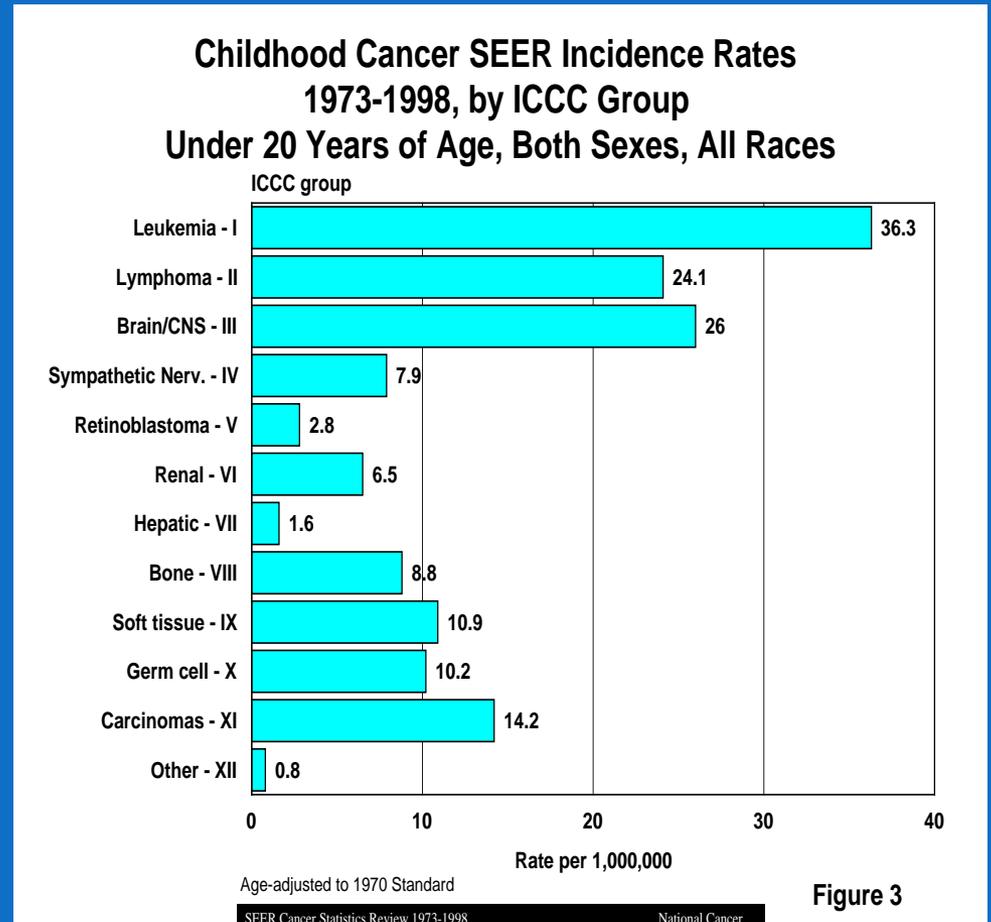
---

- I. Leukemia
- II. Lymphomas and reticuloendothelial neoplasms
- III. CNS and other intracranial and intraspinal neoplasms
- IV. Sympathetic nervous system tumors
- V. Retinoblastoma
- VI. Renal tumors
- VII. Hepatic tumors
- VIII. Malignant bone tumors
- IX. Soft tissue sarcomas
- X. Germ cell, trophoblastic, & other gonadal neoplasms
- XI. Carcinomas & other malignant epithelial neoplasms
- XII. Other and unspecified malignant neoplasms

# Childhood Cancer Statistics

Total childhood cancer estimated for 2007:

- 10,400 incident cases
- 1,545 deaths
- 5-yr survival 80%



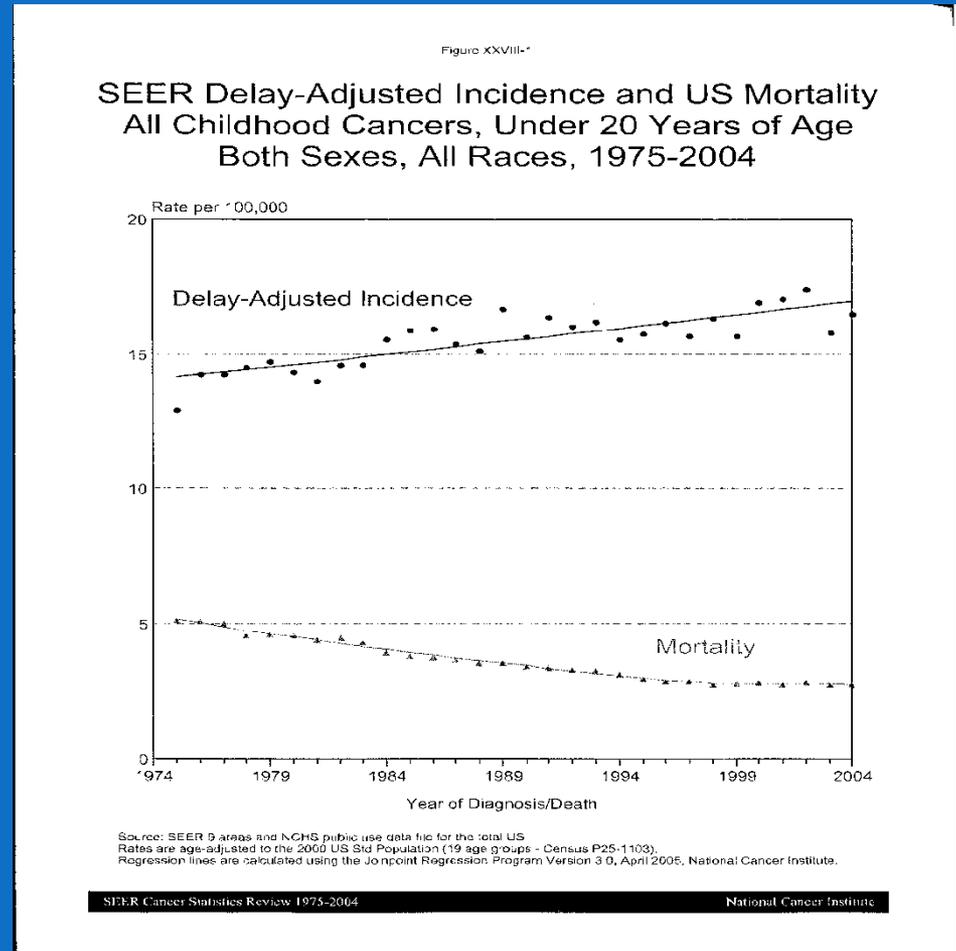
# Pediatric Cancer Types Vary in Age, Gender, and Race Patterns

---

<u>Characteristic</u>	<u>Subgroup</u>	<u>↑ Risk by Cancer Types</u>
- Age	infancy	neuroblastoma, CNS, leukemia, retinoblastoma
	adolescence	Hodgkin lymphoma, germ cell cancers, CNS, leukemia
- Gender	male	lymphoma
- Race	Caucasian	Ewing's sarcoma, acute lymphoblastic leukemia
	African-American	Wilms' tumor, retinoblastoma
	African	endemic Burkitt's lymphoma

# Trends in Total U.S. Childhood Cancer Incidence Children $\leq$ 20 Years Old, 1975-2004

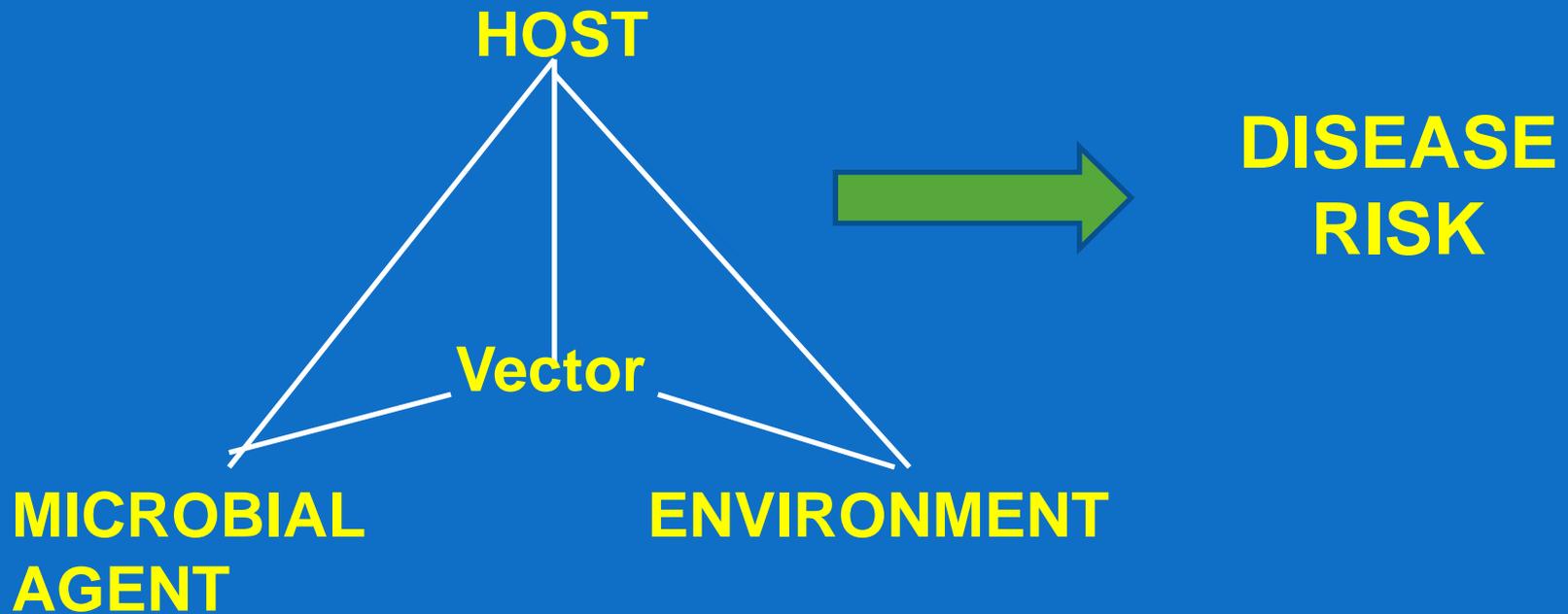
- Incidence rose about 1% per year for all childhood cancers, 1974-2004
- Rate of increase was lower (e.g., 0.2% per year) during 1990-2004
- Mortality steadily declined since chemotherapy in 1960s, but decrease has leveled off



# Disease Models

# Infectious Diseases

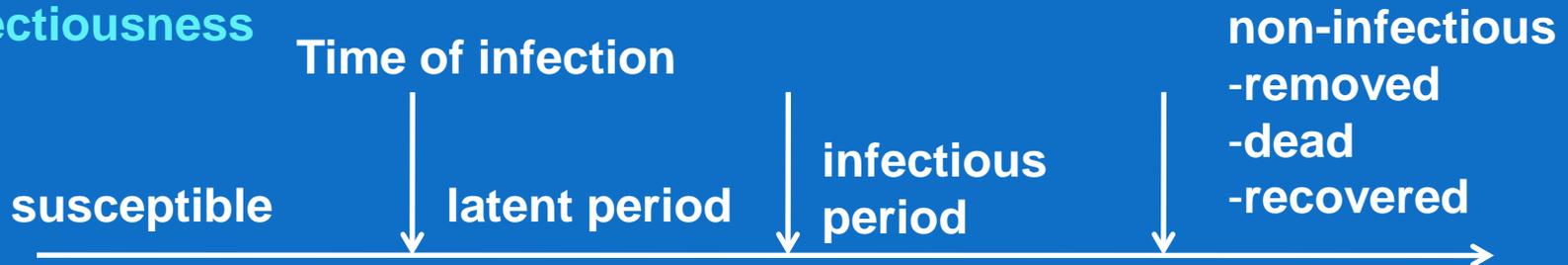
---



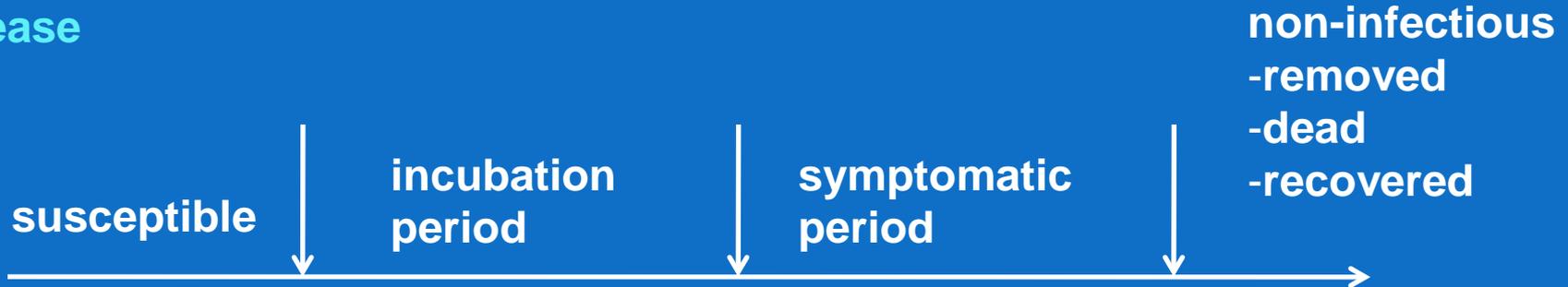
# Dynamics of Infection and Disease

---

## Dynamics of Infectiousness

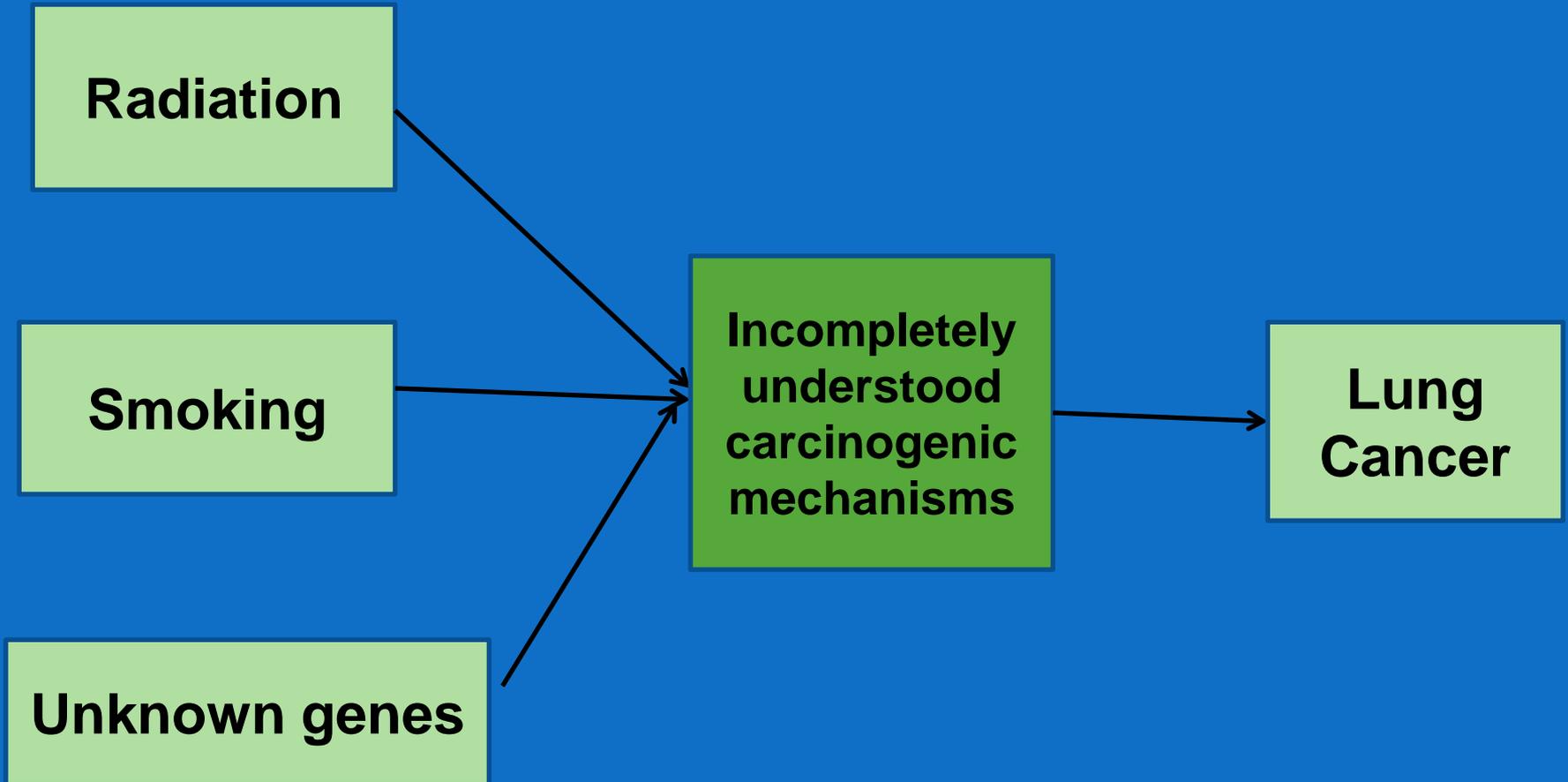


## Dynamics of Disease

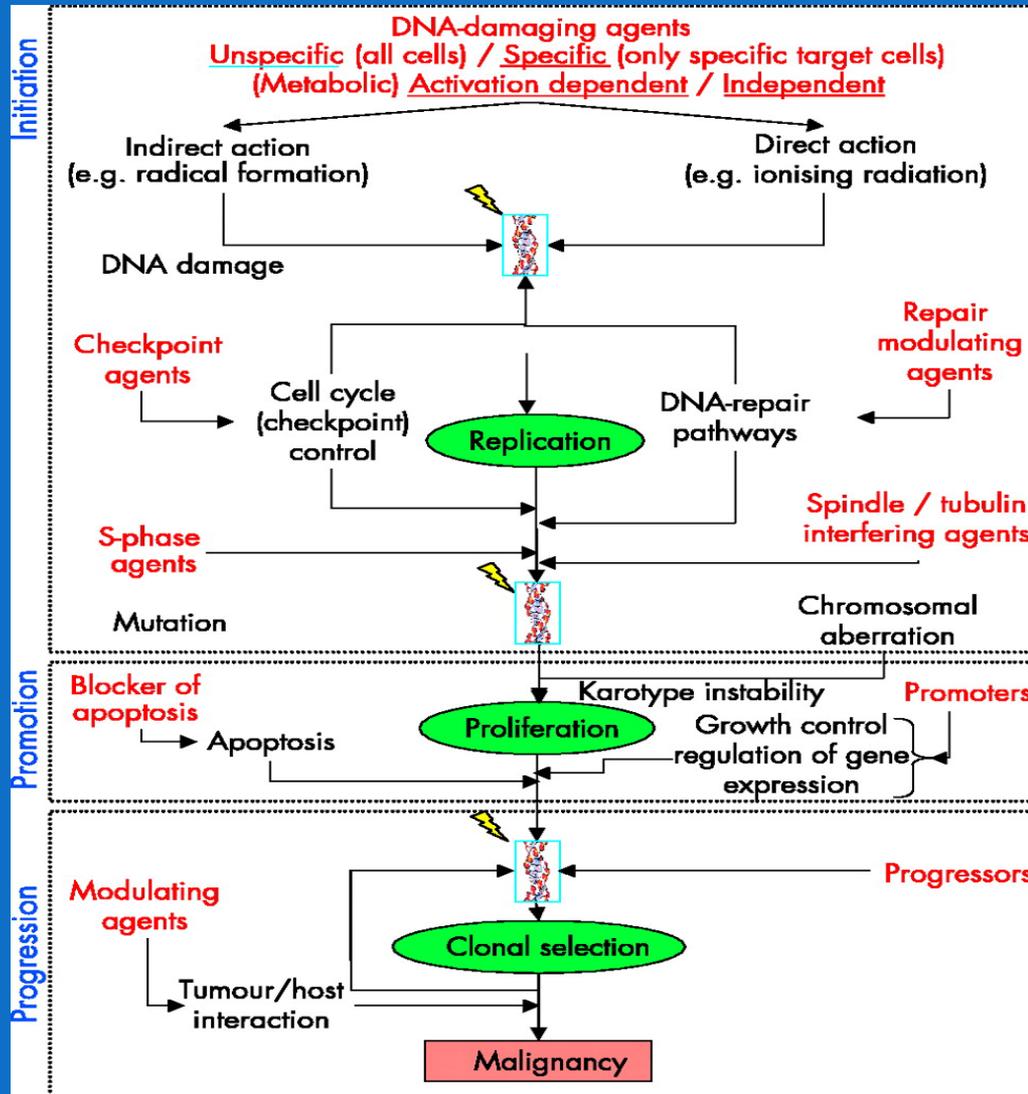


# Chronic Diseases

---

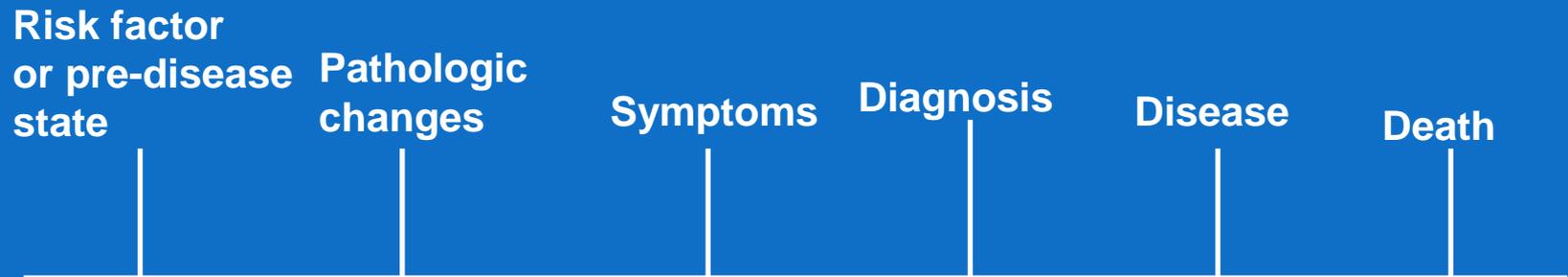


# Steps in Malignant Transformation



# Natural History of Chronic Disease

---



# Diseases with Genetic Component - 1

---

## Evidence of familial occurrence

### **Familial occurrence of certain diseases**

- > Rare diseases that are common within affected families
- > Small but significantly increased risk in families

**Onset of some cases at much younger ages**

# Diseases with Genetic Component - 2

---

## Approaches to identifying genes as causes

**Familial aggregation/segregation analysis**

**Population-based association studies**

**Genetic pathways**

**Genome-wide association studies**

# Multi-Level Disease Causation

---

## Societal Factors

- Economic
- Social
- Political
- Neighborhood
- Family/peer  
group norms

## Individual – Level Factors

- Behavioral
- Environmental
- Occupational
- Medical
- Physiologic
- Genetic predisposition

# **Disease Causation: Causal Criteria and Risk Factors**

# Statistical Association

---

**Definition of association:** Statistical dependence between two or more events, characteristics or other variables. An association is present if the probability of occurrence of an event or characteristic, or the quantity of a variable, depends upon the occurrence of one or more other events, the presence of one or more other characteristics, or the quantity of one or more other variables.

# Criteria for Causation: Bradford Hill - 1

---

- **Strength of the association**
  - Level of risk
- **Consistency of the association**
  - Repeatedly observed in different populations
- **Specificity of the association**
  - “If...limited to specific workers and to specific types of disease...then clearly that is a strong argument in favor of causation”
- **Plausibility**
  - “What is biologically plausible depends on the biological knowledge of the day”

Hill AB. The Environment and Diseases. Association or Causation? Proc R Soc Med 1965;58:295-300.

# Criteria for Causation: Bradford Hill - 2

---

- **Coherence**
  - “...the cause and effect interpretation... should not...conflict with the...known... natural history and biology of the disease”
- **Experiment**
  - “Occasionally is it possible to appeal to experimental or semi-experimental evidence?”
- **Analogy**
  - “With the effects of thalidomide and rubella before us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy”

Hill AB. The Environment and Diseases. Association or Causation? Proc R Soc Med 1965;58:295-300.

# Types of Causal Associations

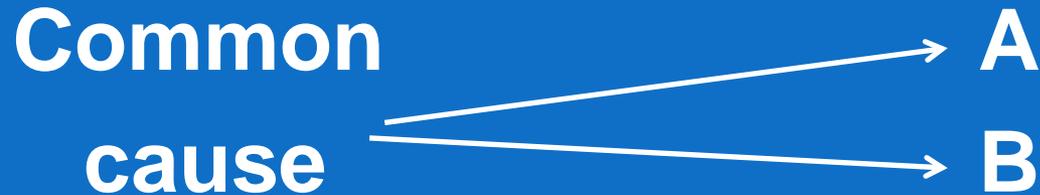
---



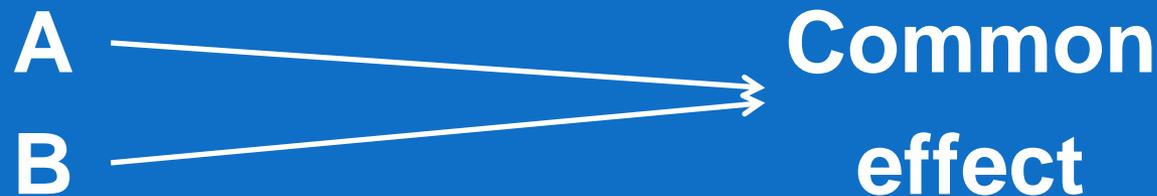
“A causes B”



“B causes A”



“A and B have a common cause”



“A and B have a common effect”

# Causal Model

---

- **Necessary vs sufficient**

- Necessary: must be present to cause disease
- Sufficient: can independently cause disease
- Example: smoking is neither a necessary or sufficient cause of lung cancer

	Sufficient (S+)	Not sufficient (S-)
Necessary (N+)	N+S+ (necessary & sufficient)	N+S- (necessary but not sufficient)
Not necessary (N-)	N-S+ (sufficient but not necessary)	N-S- (neither necessary nor sufficient)

# Non-Causal Associations and Related

---

- **Types of non-causal associations**
  - Chance association
  - Bias
    - > Selection bias (differential selection or participation of exposed vs. unexposed or controls vs. cases)
    - > Recall bias (differential recall by exposed vs. unexposed or controls vs. cases)
    - > Confounding (association of disease and an exposure with a third variable may introduce spurious associations)
- **Misclassification**
  - > Disease outcomes
  - > Exposure assessment

# Sources of Outcome Information

# Sources of Outcome Information

- Vital records
  - > death certificates
  - > birth certificates
- Morbidity surveys
  - > Health Interview Survey
  - > Health Examination Survey
- Disease notification and registration
  - > cancer registries
  - > infection notification
- Interviews
  - > validate self-reported diagnoses

COPY

STATE OF OHIO  
DEPARTMENT OF HEALTH  
DIVISION OF VITAL STATISTICS  
CERTIFICATE OF DEATH

1 PLACE OF DEATH  
County Wayne Registration District No. 1328 File No. 665  
Township Lebanon Primary Registration District No. 3407 Registered No. 65  
or Village Lebanon No. \_\_\_\_\_ St. \_\_\_\_\_ Ward \_\_\_\_\_  
(If death occurred in a hospital or institution, give its NAME instead of street and number)  
or City of \_\_\_\_\_

2 FULL NAME Hannah Waggoner Did Deceased Serve in U. S. Navy or Army No  
(a) Residence No. \_\_\_\_\_ St. \_\_\_\_\_ Ward \_\_\_\_\_  
(Usual place of abode) (If nonresident give city or town and State)  
Length of residence in city or town where death occurred yrs. mos. ds. How long in U. S., if of foreign birth? yrs. mos. ds.

PERSONAL AND STATISTICAL PARTICULARS					MEDICAL CERTIFICATE OF DEATH	
3 SEX <u>Female</u>	4 COLOR OR RACE <u>White</u>	5 Single, Married, Widowed or Divorced (write the word) <u>Widowed</u>	6 DATE OF DEATH (month, day and year) <u>July 4, 1929</u>		16 DATE OF DEATH (month, day and year) <u>July 4, 1929</u>	
7a If married, widowed or divorced HUSBAND or (or) WIFE of <u>Josee B. Waggoner</u>			7b I HEREBY CERTIFY, That I attended deceased from <u>June 15, 1929</u> to <u>July 4, 1929</u> that I last saw him alive on _____, 19____ and that death occurred, on the date stated above, at _____m.		The CAUSE OF DEATH* was as follows: <u>Diabetic Coma</u>	
8 DATE OF BIRTH (month, day, and year) <u>March 26, 1857</u>			7 AGE Years <u>72</u> Months <u>2</u> Days <u>8</u> If LESS than 1 day _____ hrs. or _____ min.		17 (duration) _____ yrs. _____ mos. _____ ds.	
8 OCCUPATION OF DECEASED (a) Trade, profession, or particular kind of work <u>Home Duties</u> (b) General nature of Industry, business, or establishment in which employed (or employer): (c) Name of employer			18 Where was disease contracted If not at place of death? Did an operation precede death? <u>No</u> Date of _____ Was there an autopsy? <u>No</u> What test confirmed diagnosis? <u>Chemical Exam</u>		CONTRIBUTORY (SECONDARY) <u>Diabetes</u> (duration) _____ yrs. _____ mos. _____ ds.	
9 BIRTHPLACE (city or town) <u>Near Loveland</u> (State or country) <u>Ohio</u>			10 NAME OF FATHER <u>Wm. H. Parker</u>		18 (Signed) <u>Edward Blais, M. D.</u> <u>July 5, 1929</u> (Address) <u>Lebanon, O.</u>	
11 BIRTHPLACE OF FATHER (city or town) <u>Virginia</u> (State or country)			12 MAIDEN NAME OF MOTHER <u>Mary Tignor</u>		*State the DISEASE CAUSING DEATH, or in deaths from Violent Causes, state (1) MEANS AND NATURE OF INJURY, and (2) whether ACCIDENTAL, SUICIDAL or HOMICIDAL. (See reverse side for additional space).	
13 BIRTHPLACE OF MOTHER (city or town) <u>New Jersey</u> (State or country)			14 Informant <u>Chas. J. Waggoner</u> (Address) <u>Lebanon, O.</u>		19 PLACE of Burial, Cremation, or Removal <u>Western O. Cemetery</u> DATE OF BURIAL <u>July 8, 1929</u>	
15 Filled <u>7/6, 1929</u> <u>M. E. Gustin</u> REGISTRAR			20 UNDERTAKER <u>W. A. Donald</u> ADDRESS <u>Lebanon, O.</u>		20a WAS THE BODY EMBALMED? <u>Yes</u> EMBALMER'S LICENSE NO. <u>12919</u>	

V. S. 11-A-892M-4 25--Books of 100  
MARGIN RESERVED FOR BINDING  
N. B.—WRITE PLAINLY, WITH UNFADING INK—THIS IS A PERMANENT RECORD. Every item of information should be carefully supplied. AGE should be stated EXACTLY. PHYSICIANS should state the CAUSE OF DEATH in plain terms, so that it may be properly classified. Exact statement of OCCUPATION is very important. See instructions on back of certificate.

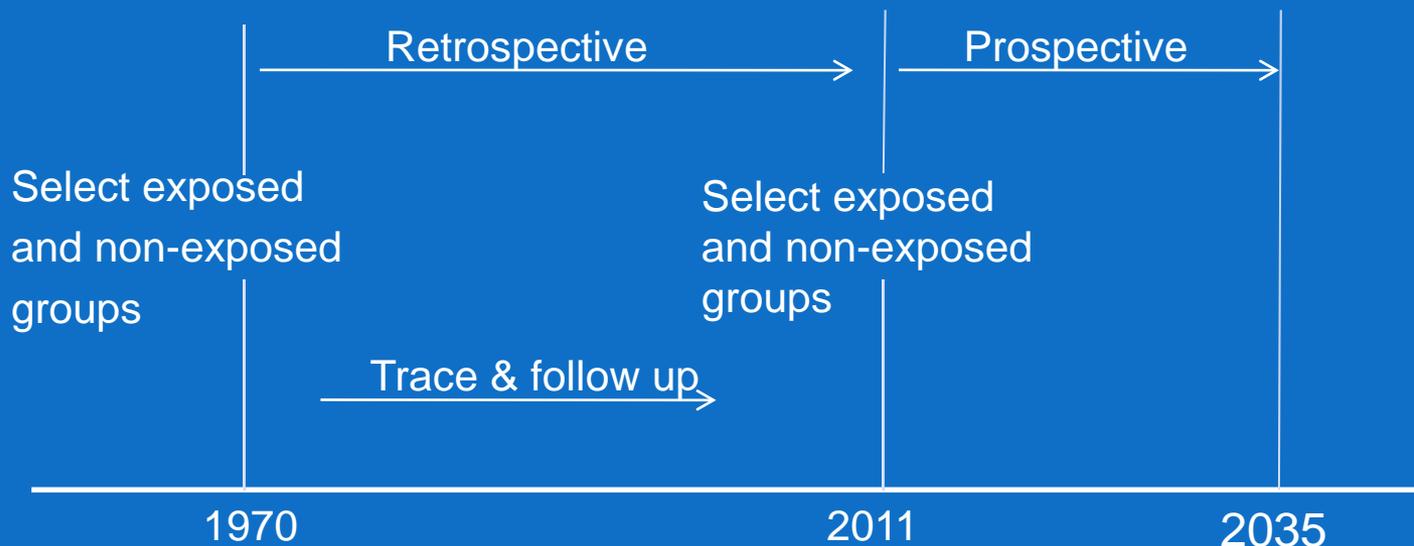
# **Epidemiologic Study Designs**

# Cohort Studies

- Distinguishing features

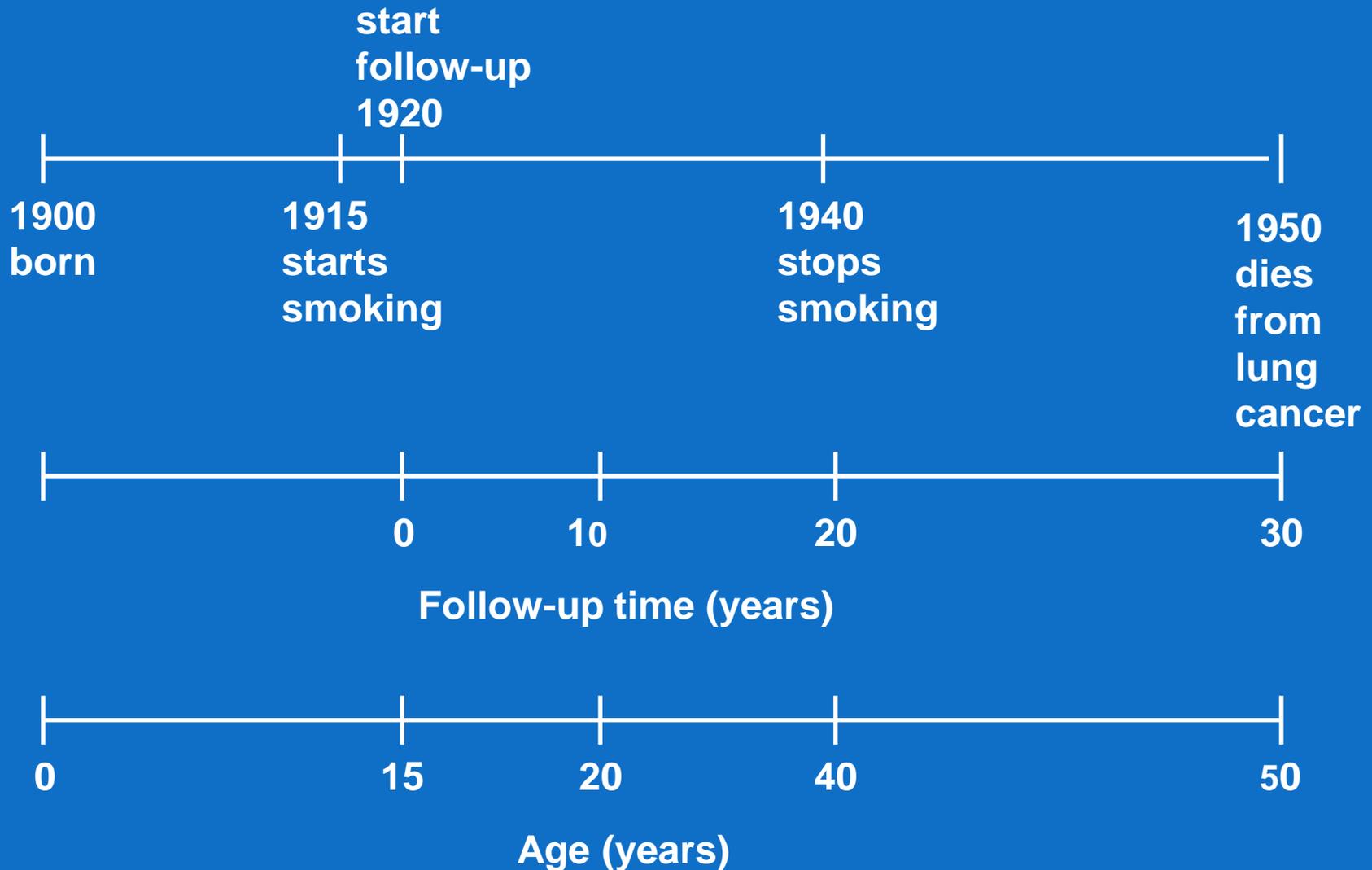
- > population defined by exposures prior to onset of disease
- > population followed over time to estimate disease/death rate
- > compare rates in exposed vs unexposed groups

- Retrospective vs prospective follow-up



# Follow-up: Multiple Axes of Time

---



# Case-Control Studies

---

- **Distinguishing features**

- > determine exposures prior to diagnosis/referent date using interviews, medical records or other records
- > compare proportion of cases with exposure to proportion of controls with exposure

- **Framework**

Characteristics	With disease	Without disease	Total
With exposure	a	b	a + b
Without exposure	c	d	c + d
Total	a + c	b + d	a + b + c + d

# Cross-Sectional Studies

---

- **Distinguishing features**
  - > compare proportion of cases with a characteristic or exposure to proportion of controls with exposure at the time of the study
- **Example: cross-sectional vs. case-control studies**
  - > cross-sectional: compare total white blood count in cases with prevalent benzene hematotoxicity vs. benzene-exposed controls
  - > case-control: compare total white blood counts 5+ years before diagnosis of benzene hematotoxicity vs. referent time in controls

# Exposure Assessment

# Multiple Exposure Data Sources Often Used

---

- **Biological measures of exposure**
- **Available measurements**
- **Exposure surrogate measures**
- **Data from records**
- **Questionnaire data**

# Epidemiologic Approaches to Exposure Assessment - 1

---

- **Relevant biologic measurements of exposure available or feasible?**
  - > **goal for assessing relevant exposures for chronic diseases is a measure reflecting cumulative and past exposures**
  - > **biologic measure of cumulative or past exposures must be validated (e.g., chromosomal translocations characteristic of radiation exposure, bone lead levels, asbestosis in lung)**

# Epidemiologic Approaches to Exposure Assessment - 2

---

- **Relevant historical measurements available?**
  - > **occupational** : workplace individual or job-specific measurements
  - > **environmental**: routine monitoring (for air or water levels); residential monitoring (radon); measurements following accidents (chemical spills, nuclear accidents)
  - > **medical**: medication dose data; radiotherapy dose planning; vaccinations; hormone replacement therapy; laboratory examinations
  - > **body measurements**: height, weight, BMI

# Epidemiologic Approaches to Exposure Assessment - 3

---

- **Exposure surrogate data from sources other than questionnaires?**
  - > **occupational** : job titles, work history, tasks, procedures, and raw materials used
  - > **environmental**: containers of pesticides used on farms or homes, paints used for indoor or outdoor painting, art supplies
  - > **medical**: radiology reports
  - > **socioeconomic**: census data from small area unit where subject resides

# Epidemiologic Approaches to Exposure Assessment - 4

---

- **Data from records?**

- > **occupational** : salary; personnel or human resources; factory records of raw materials; professional society membership lists
- > **environmental**: records of pesticide application, roadwork, major construction documentation
- > **medical**: history and physical examination; medical treatment data in cancer registries (incomplete)
- > **mobile phone subscriber lists**

# Epidemiologic Approaches to Exposure Assessment - 5

---

- **Questionnaire data?**
  - > food frequency questionnaire data
  - > self-reported physical activity
  - > self-reported time outdoors in the sun
  - > history of allergies
  - > environmental: containers of pesticides used on farms or homes, paints used for indoor or outdoor painting, art supplies
  - > medical: radiology reports

# Exposure Validation

---

- **Even if biological exposure measures available, not feasible or affordable to obtain for all subjects**
- **Similarly, validate questionnaire data with record-based information in a subset**

# Dose Metrics

---

- **Duration:** year last minus year first exposed
- **Frequency:** number of times exposed in a given period of time
- **Intensity:** exposure concentration at given time
- **Peak dose:** highest dose(s)
- **Cumulative exposure:** frequency and/or intensity times duration of exposure

# Summary

---

- Descriptive epidemiologic studies explain local disease occurrence, natural history and suggest new hypotheses
- Disease characteristics, infectious and chronic disease models, and causal criteria drive decisions about optimal epidemiologic study design, and approaches for outcome and exposure assessment

# Acknowledgements

---

Assistance from Drs. Hannah Yang, Gila Neta, and Sara Schonfeld is gratefully acknowledged.

# References

---

- MacMahon B, Pugh T. Epidemiology - Principles and Methods. Boston: Little Brown, 1970.
- Lilienfeld AM, Lilienfeld DE. Foundations of Epidemiology, 2<sup>nd</sup> Edition. New York: Oxford University Press, 1980.
- Rothman K, Greenland S. Modern Epidemiology, Second Edition. Philadelphia: Lippincott Williams & Wilkins, 1998.
- Cancer Epidemiology and Prevention, 3<sup>rd</sup> Edition. Schottenfeld D and Fraumeni JF Jr, eds. New York: Oxford University Press, 2006.

