Public Health Sciences 310
Epidemiologic Methods

Lectures 10 & 11
Cohort Studies

February 6 & 8, 2024

Brian C.-H. Chiu

Types of Cohorts

- General population or convenience
  - Geographic area (Framingham, Multiethnic Cohort)
  - Place of medical care (HMOs)
  - Profession (Nurses' Health Study, California Teachers Study)
- Double cohort
  - Occupational exposure (US Firefighters, Shipbuilders)
  - Specific exposure (Atomic Bomb Survivors Study, Childhood Cancer Survivor Study)
- Prognostic (disease natural history or clinical course)

Choice of Cohort

Selected based on certain characteristics

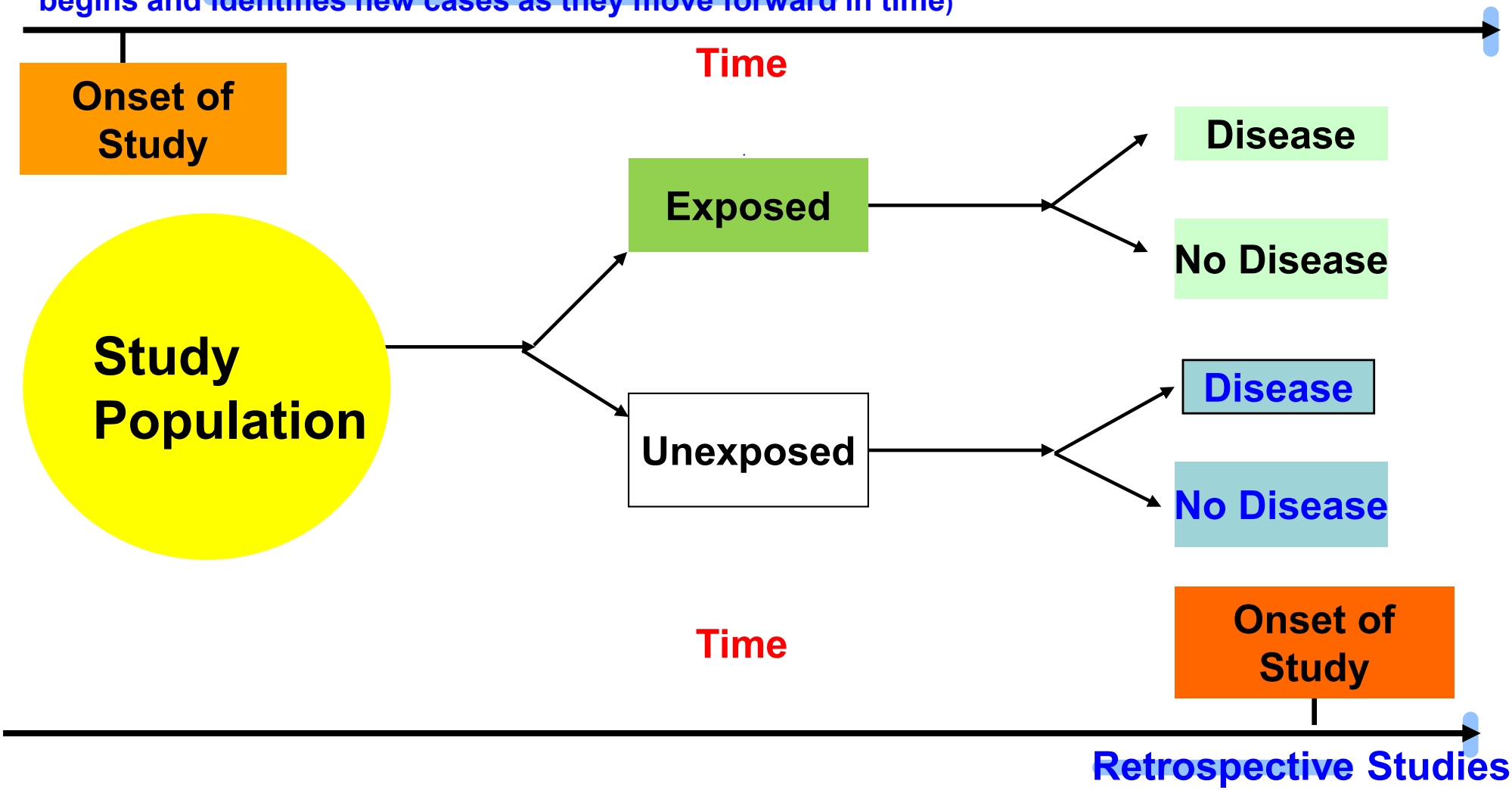
- Cohort studies are often conducted among groups specifically chosen not for their exposure status, but for their ability to facilitate the collection of relevant information (e.g., nurses, union members, veterans, students or alumni of particular colleges, residents of well-defined communities, etc.)
- When cohorts not strictly representative of the general population are selected, caution should be exercised in applying the results to the general population

Biased towards certain population

Prospective vs. Retrospective Cohort Studies

Prospective Studies

(Investigators collects data on exposures, confounders, and/or biological samples at time study begins and identifies new cases as they move forward in time)



Retrospective Studies

The exposure, follow-up, and the outcome have already occurred (hence retrospective). Investigator has access to existing records (e.g., hospital, employment, etc.)

Origin Time

Prospective Cohort Studies – Zero Time

Enrolled to the study, data collection begins.

- Time of baseline examination
- Sometimes time of first exposure begins at baseline
  When participants are fist considered at-risk for outcome of interest.
- More often, exposure has been occurring some time
  - Cannot assign person-time back to start of exposure; must always begin counting at baseline
- Should initially include only individuals free of disease(s), but at zero time, the disease process in some individuals will have already begun (subclinical disease)

Setting proper Time Zero to avoid time-related biases.

Prospective Cohort Studies – Exposure Information

- Pre-existing records (e.g., medical records, employer records)
  - Advantages:
    - relatively inexpensive
    - unbiased classification of exposure
  - Disadvantages:
    - exposure info may be insufficient to address specific research questions
    - frequently do not contain data on potential confounding variables
- Interviews and questionnaires completed by cohort subjects
  - Advantages:
    - allows for data collection on exposure that are not routinely recorded
  - Disadvantages:
    - Information bias
    - Time-consuming
    - Expensive

Prospective Cohort Studies – Exposure Information

- Direct measurement (e.g., lab determinations, physical exam)
  - Advantages:
    - More accurate classification
  - Disadvantages:
    - Basically pertains to the point in time at which the measurement was obtained
    - Time-consuming
    - Expensive
- Periodic reexamination or resurvey to Updated
  - Allow for revision of exposure categories
  - Allow data to be collected on exposures or risk factors that were not recognized to be of interest at the initiation of the study ex) long-term follow-up

Re-assessing the Exposure Status - Example

Example 8.10. *In the British doctors study described in Example 8.3, the first questionnaire was sent to all registered British doctors in 1951. Further inquiries about changes in smoking habits were made in 1957, 1966, 1972, 1978 and 1990–91. The two last questionnaires also included additional questions on alcohol consumption and some other personal characteristics. To assess the extent of changes in smoking habits during the 40-year follow-up period, the smoking habits of the men who replied to the 1990–91 questionnaire were compared with the habits they reported in the initial questionnaire in 1951 (Table 8.2) (Doll* et al.*, 1994a).*

	Smoking habits, 1990-91						
Smoking habits, 1951	Non-smoker	Former smoker	Current smoker			No. (%) in 1951	
			Cigarette only	Cigarette and other	Pipe or cigar		
Non-smoker	2361	198	17	4	86	2666	(25)
Former smoker	0	1374	10	3	66	1453	(13)
Current smoker							
Cigarettes only	0	3355	535	47	446	4383	(41)
Cigarettes and other	0	897	74	31	308	1310	(12)
Pipe or cigar	0	695	16	2	287	1000	(9)
No. (%) in 1990–91	2361 (22)	6519 (60)	652 (6)	87 (1)	1193 (11)	10 812	(100)

[a] Data from Doll *et al.* (1994a).

- A single classification at baseline is appropriate for factors that do not change with time
- Frequently, changes in exposure levels occur during the course of long-term follow-up

Prospective Cohort Studies – Follow-Up

- Role of follow-up
  - To determine cohort member under observation (at-risk): denominator information # of Total Participants
  - To determine disease outcomes (events): numerator info # of Outcomes.
- Complete follow-up important
  - Loss to follow-up could lead to a serious underestimation of the true incidence of a disease
- Type of information
  - Date begin observation
  - Date end observation (death, event, move/drop-out, medical procedure, or close of study)
  - Used to calculate person-time at risk
- Update risk factor and susceptibility information
  - At-risk pool or exposure may change (e.g., drug use, diet change, etc.)

May facilitate/inhibit cause of outcome (modify)

Prospective Cohort Studies – Approaches to Minimize Losses

- Exclude subjects that will be hard to follow or who are unwilling to return
- Collect tracking information

  Contact Route
  - Name and address, birthdate, contact persons (friends, relatives, physicians)
- Period contact

  Monitor presence/absence
  - Follow-up surveys
  - Phone
  - Email, newsletter

Avoid loss to follow up

Prospective Cohort Studies – Assessment of Outcome

- Blind assessment Data analyzer unaware of key information: ↓ objective result
  - Standard (e.g., registry) and explicit criteria
  - Review panel (blind to subject characteristics)
- Death certificates/Cancer registration
- Self-report (when no form of disease surveillance system exists) Last choice
  - Confirm with medical records to avoid recall bias
- Key: method chosen to ascertain the outcome(s) of interest should be used identically for subjects exposed and those not exposed

Prospective Cohort Studies – Strengths

- Only way to directly establish incidence and natural history of a disease
- Can assess the relationship between a given exposure and a variety of diseases/outcomes
- Recall bias can be limited Able to observe results
  - Differential Misclassification
    - no opportunity since outcome has not yet occurred
    Unable to know when certain groups have certain outcome: Prospective (Need to know result inorder to analyze)
  - Nondifferential Misclassification
    - Possible (many exposures just cannot accurately be recalled over time)
    - Can still have misclassification (many variables, e.g. diet, cannot be measured accurately)

    systematic error for all participants that doesn't favor over a certain group: Common trait / characteristic.

Prospective Cohort Studies – Strengths (cont)

- Selection bias can be limited
  - As long as follow-up mechanisms do not favor particular exposure groups
- Correct time sequence Allows follow-up and observe of cases for temporality.
  - Exposure occurred before clinical disease
  - Often the only way to establish temporal relationships for physiologic variables which are altered by disease

Gathered at baseline

- Pre-disease information on biological parameters available
  - Disease often modifies physiologic or biochemical measures, case-control methods unable to address

For example, if researchers are investigating the risk factors for cardiovascular disease in a cohort study, they may collect baseline data on biological parameters such as blood pressure, cholesterol levels, glucose levels, body mass index (BMI), smoking status, genetic markers, and other relevant factors from participants who are initially free of cardiovascular disease.

Determine other potential risk factors

Prospective Cohort Studies – Weaknesses

- Extremely expensive and time-consuming
- Inefficient for studying rare diseases Case Control Study
- Results not know for a long time
- Still susceptible to confounding
  - Known confounders must be measured
  - Unknown confounders could still be present
- Diseases with (potentially) long preclinical phases Length Time Bias Lead Time Bias
  - Unknown or undetected preclinical disease might affect baseline measurements

Prospective Cohort Studies – Example

Exposure Outcome

Allergic Diseases and Risk of Hematopoietic Malignancies in a Cohort of Postmenopausal Women: A Report from the Iowa Women's Health Study

Amy M. Linabery[1,2], Anna E. Prizment[2,3], Kristin E. Anderson[2,3], James R. Cerhan[4], Jenny N. Poynter[1,2], and Julie A. Ross[1,2]

Abstract

Background: Allergic diseases signify immune dysregulation attributable to underlying genetics and environmental exposures. Associations between various allergies and hematopoietic cancers have been observed, albeit inconsistently; however, few prospective studies have examined the risk, and even fewer among older adults.

Outcome Exposure

Methods: We examined risk of incident hematopoietic cancers in those reporting allergic diseases in a population-based cohort of 22,601 older women (Iowa Women's Health Study). Self-reported allergic status, including asthma, hay fever, eczema, and/or other allergies, was determined via questionnaire in 1997 (mean age, 72 years; range, 63–81 years). Incident cancers were ascertained by linkage with the Iowa Cancer Registry from 1997 to 2011. Cox proportional hazards regression was performed to estimate multivariate-adjusted HR and 95% confidence intervals (CI) for myeloid ($N = 177$) and lymphoid ($N = 437$) malignancies, respectively.

	N	Person-years	Age-adjusted HR[a] (95% CI)	Multivariable-adjusted HR[b] (95% CI)
MM				
No allergy[c]	53	183,443	Ref.	Ref.
Allergy[d]	26	81,029	1.17 (0.70–1.79)	1.15 (0.72–1.85)
No allergy[c]	53	183,443	Ref.	Ref.
Asthma[e]	5	18,518	0.95 (0.38–2.37)	0.94 (0.38–2.36)
Hay fever[e]	3	21,086	0.50 (0.16–1.60)	0.51 (0.16–1.62)
Skin allergy[e]	10	31,905	1.09 (0.56–2.15)	1.13 (0.57–2.23)
Other allergy[e]	14	45,334	1.08 (0.60–1.95)	1.11 (0.61–2.01)

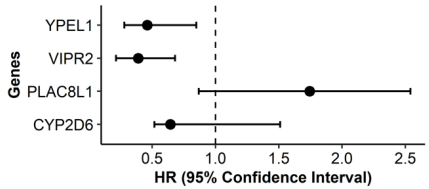
Linabery AM et al., CEBP 2014

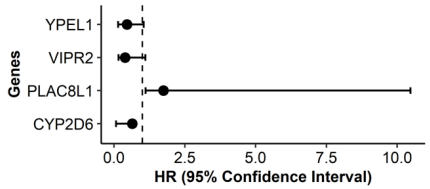
Prospective Cohort Studies – Example

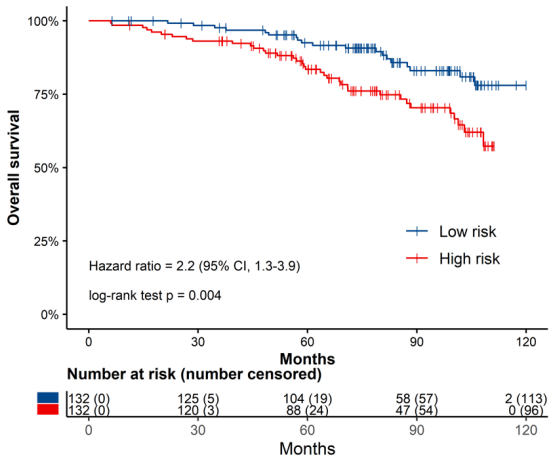
5-Hydroxymethylcytosines (Exposure) in circulating cell-free DNA and overall survival in patients with multiple myeloma (Outcome)

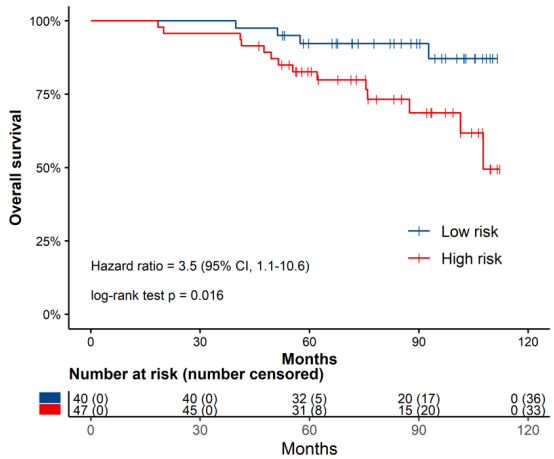
Study Participants: 586 newly dx MM pts prospectively enrolled at UCMC between 2010-19 with blood samples collected at dx. Plasma for 354 subjects were included in this analysis. Follow-up ended 12/31/2020. Vital status ascertained in 351 patients (deaths=73) using the National Death Index.

5hmC in cfDNA and MM prognosis: Genome-wide 5hmC in cfDNA was profiled using 5hmC-Seal and NGS. Machine learning and Cox with elastic net regularization in training set to identify differentially modified genomic features. Weighted prognostic scores computed from a four-gene marker panel, validated in the testing set.









Chiu BC et al., J Clin Oncol 2021 (suppl 15; abstr 8032)

Prediagnostic serum organochlorine insecticide concentrations and primary liver cancer: A case–control study nested within two prospective cohorts

Lawrence S. Engel[1], Emily C. Zabor[2], Jaya Satagopan[2], Anders Widell[3], Nathaniel Rothman[4], Thomas R. O'Brien[4], Mingdong Zhang[4], Stephen K. Van Den Eeden[5] and Tom K. Grimsrud[6]

Study populations

Kaiser Permanente Multiphasic Health Checkup cohort. The MHC cohort comprises Kaiser Permanente Northern California (KPNC) health plan members who underwent an MHC at facilities in Oakland, CA or San Francisco, CA between 1964 and 1992. Kaiser Permanente members presenting for a routine health examination, for any reason, were invited to complete a detailed questionnaire and standardized physical examination.[20,21] Fasting serum samples were collected from almost 129,000 of these participants during 1964–1969 and from 57,963 individuals during the 1980s; the earlier sera were initially stored at −23 °C, but all sera have been stored at −40 °C since 1980. The Kaiser Permanente population generally reflects the region's underlying census demographics.[22,23]

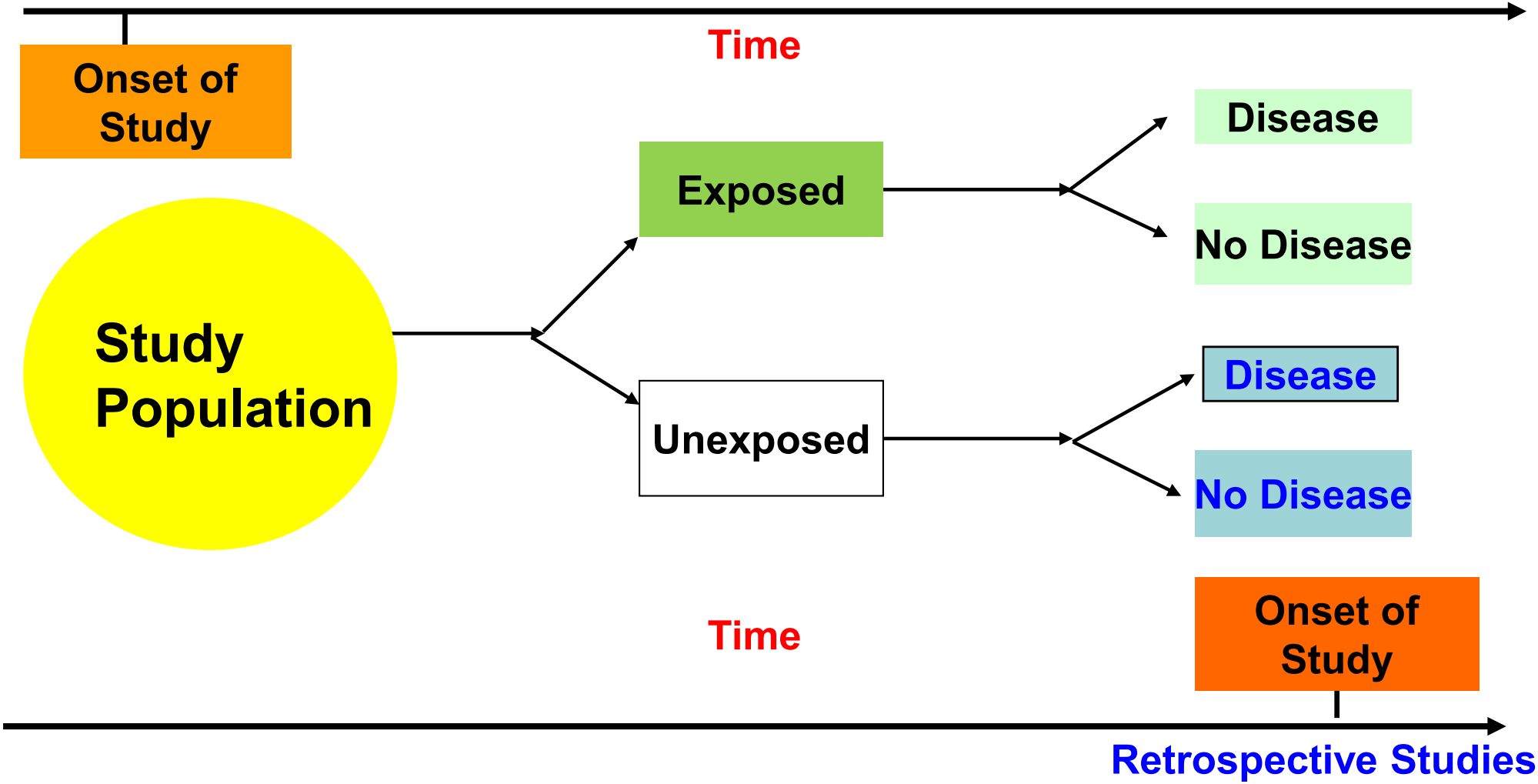
Janus Serum Bank cohort. The Janus Serum Bank Cohort is a population-based biobank with specimens and data from approximately 319,000 Norwegian men and women. Participants were recruited primarily from county cardiovascular health examinations conducted in the 1970s and 1980s.[24] The cohort includes almost 90% of persons aged 35–49 years and a 10% random sample of persons aged 20–34 years at the time of their examination who were residing in the sampled areas. Approximately 10% of cohort members were recruited from Red Cross blood donors aged 18–65 years and living in Oslo, Norway during 1972–2004. Serum, from nonfasting blood samples, was separated and stored at −25 °C.

Engel LS et al., Int J Cancer 2019

Retrospective Cohort Studies

Prospective Studies

(Investigators collects data on exposures, confounders, and/or biological samples at time study begins and identifies new cases as they move forward in time)



The exposure, follow-up, and the outcome have already occurred (hence retrospective). Investigator has access to existing records (e.g., hospital, employment, etc.)

Retrospective Cohort Studies

- Strengths
  - Can establish time sequence (predictor variable preceded outcome) All events have already occured.
  - No bias in baseline measurement by later outcome status
  - Less costly and time-consuming than prospective cohort studies

- Weaknesses
  - Exposure info may not be available (major limitation)
  - If available, no control over the nature and quality of the measurements made
    - Assessment of predictor variables outdated
    - Minimal or incomplete information available on variables of interest
    - Lack of relevant confounders and effect modifiers

Retrospective Cohort Studies - Example

Original Article

Myelodysplastic Syndrome and Acute Myeloid Leukemia After Receipt of Granulocyte Colony-Stimulating Factors in Older Patients With Non-Hodgkin Lymphoma

Gregory S. Calip, PharmD, MPH, PhD[1,2]; Kellyn M. Moran, PharmD[1]; Karen I. Sweiss, PharmD[3]; Pritesh R. Patel, MD[4]; Zhaoju Wu, MD, PhD[1]; Sruthi Adimadhyam, MS[1]; Todd A. Lee, PharmD, PhD[1]; Naomi Y. Ko, MD, MPH, AM[5]; John G. Quigley, MD[4]; and Brian C.-H. Chiu, PhD[6]

BACKGROUND: Granulocyte colony-stimulating factors (G-CSFs), which are used for the prevention of complications from chemotherapy-related neutropenia, are linked to the risk of developing second primary myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML). The objective of this study was to examine the correlation between using a specific G-CSF agent and the risk of MDS/AML among older patients with non-Hodgkin lymphoma (NHL). **METHODS:** This was a retrospective cohort study of adults aged >65 years who were diagnosed with first primary NHL between 2001 and 2011. With data from the Surveillance, Epidemiology, and End Results-Medicare–linked database, adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) were estimated for the risk of MDS/AML associated with the receipt of G-CSF (filgrastim and pegfilgrastim) in Cox proportional-hazards models, which were stratified according to treatment accounting for confounding by indication. **RESULTS:** Among 18,245 patients with NHL patients who had a median follow-up of 3.5 years, 56% received chemotherapy and/or immunotherapy, and G-CSF was most commonly used in those who received rituximab plus multiple chemotherapy regimens (77%). Subsequent MDS/AML diagnoses were identified in 666 patients (3.7%). A modest increased risk of MDS/AML was observed with the receipt of G-CSF (HR, 1.28; 95% CI, 1.01-1.62) and a trend was observed with increasing doses (P_{trend} < .01). When specific agents were analyzed, an increased risk of MDS/AML was consistently observed with filgrastim (≥10 doses: HR, 1.67; 95% CI, 1.25-2.23), but not with pegfilgrastim (≥10 + doses: HR, 1.11; 95% CI, 0.84-1.45). **CONCLUSIONS:** A higher of MDS/AML was observed in patients with NHL risk among those who received G-CSF that was specific to the use of filgrastim (≥10 doses), but not pegfilgrastim. Neutropenia prophylaxis is an essential component of highly effective NHL treatment regimens. The differential risk related to the types of G-CSF agents used warrants further study given their increasing use and newly available, US Food and Drug Administration-approved, biosimilar products. ***Cancer*** **2019;125:1143-1154.**

Double Cohort Studies (Use an External Control Group)

- Structure Exposure
  - Define two distinct (separate) cohorts of subjects
    - One cohort exposed to a potential risk factor
    - One (or more) cohort(s) with low or no exposure to the risk factor (i.e., external control group(s))
    - E.g., similar workers in another occupation, general population of the geographical area in which the exposed individuals resides
  - How is this different from a case-control study? Outcome

    Case control study — One cohort diseased
    — One cohort controlled
- Measures of association:
  - Standardized mortality ratio (SMR)
  - Standardized incidence ratio (SIR)

Double Cohort Studies

- Strengths
  - In general, same advantages as cohort studies
  - May be the only feasible approach to study rare exposures (e.g., benzene, atomic bomb)

- Limitations
  - In occupational studies, "Healthy Worker Effect"
    - Employed persons are on average healthier than the general population, better economic circumstances
  - Confounding
    - Cohorts may be different in other important ways besides exposure to the predictor variable

Variability in different populations, prone to ↑ influence from factors

Bias in Cohort Studies

- Selection bias
  - Nonparticipation bias
  - Migration bias
- Information bias
  - Observer bias
  - Report bias (differential, non-differential)
- Measurement bias

Nonparticipation Bias

Will not be included in the study, but have characteristics that differ from participants. (same group)

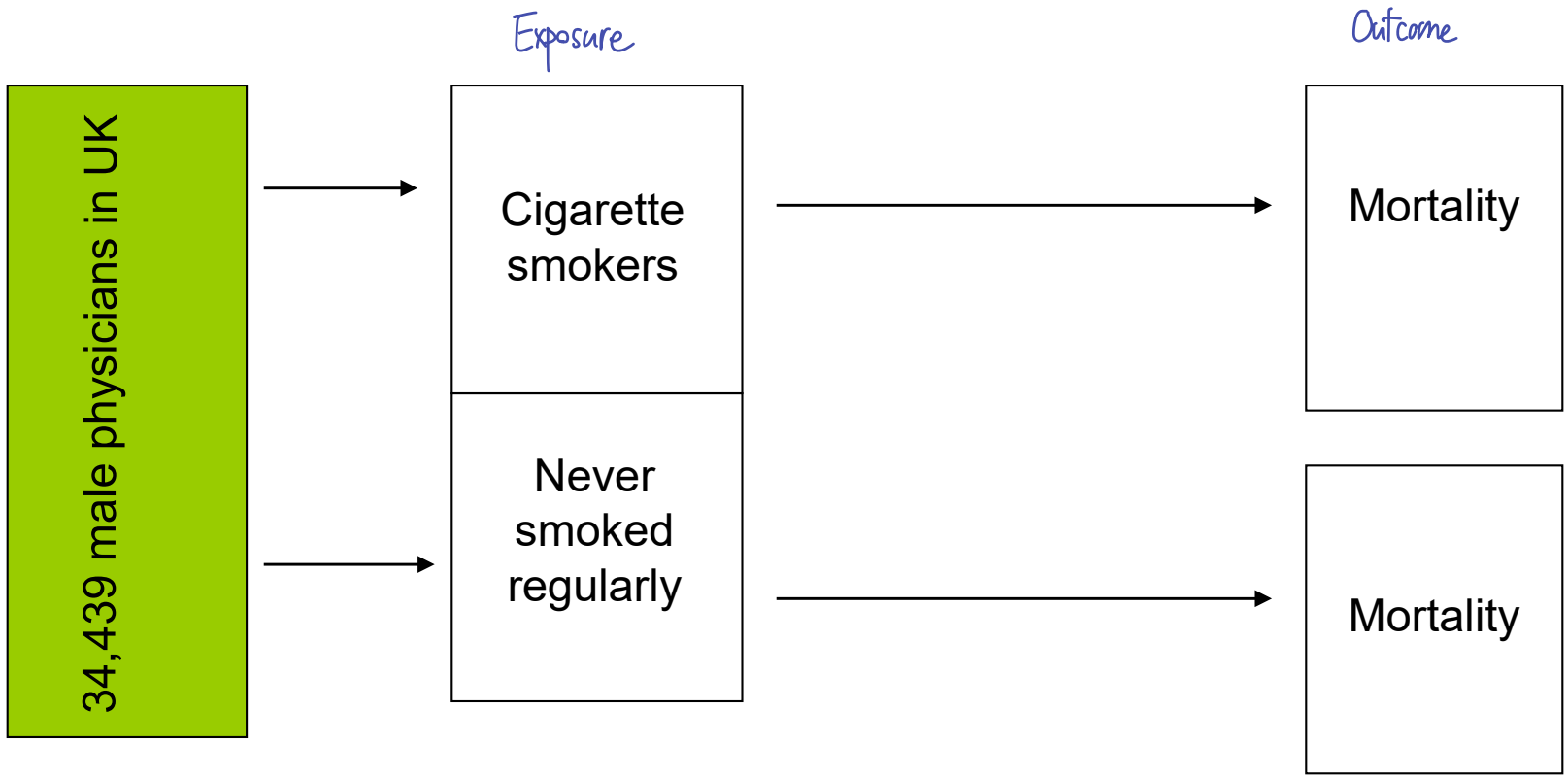
- Only a portion of eligible will participate
- Participants are likely differ from nonparticipants
  - Motivation
  - Attitudes toward health
  - Risk factor profile
- Generally does not affect study validity
  - Bias only if nonresponse related to both the exposure, and independent of the exposure, to the outcome under study
- Does affect generalizability

Result not based on "true" representation of population.

Migration Bias

- Occurs when patients in one exposure group leave their original group under study, either move to another exposure group or to drop out ex) Geographically, Status
  - Random movement among exposure groups does not bias study
  - If random movement is large and differential, can invalidate study
  - Reasons for dropping out (death, recovery, side effects) are often related to future disease status/prognosis, and may differentially affect exposure groups within the cohort
  - Changing exposure groups leads to misclassification, and would attenuate risk estimates

A classic cohort study: Mortality in relation to smoking



Male British Doctors Study

1951 — 1991

Doll et al., BMJ 1994

a. Pretend that physicians who smoked were more likely to move out of the country and become lost to follow-up than physicians who didn't smoke. (We are interested here in exploring possible selection bias.) In particular, how might it affect the following (state in which direction, if any, the estimate would change and why):

i. Their estimates of smoking prevalence over time?

Underestimate

More frequent
↑
Both smoker and non-smoker move out

Differential migration bias

ii. Their estimate of lung cancer mortality in the smokers?

Underestimate

iii. Their estimate of lung cancer mortality in the non-smokers?

Underestimate

iv. Their estimate of relative risk of lung cancer comparing smokers with non-smokers?

Undetermined: Unable to determine mortality and causes, or if the migration leads to extra exposure.
(Minimize migration bias from lost to follow-up.)

Measurement Bias

Way to detect the outcome

- Subjects in one of the exposure groups have a better chance of having their outcome detected
  - Unlikely to miss major outcomes: death, major cancer, etc
  - More likely for less clear-cut outcomes: subclinical disease, side effects, disability, etc.

Likely to detect ↓

- Minimize
  - Ensure that method by which observations are made are blind to patient group
  - Set careful rule for deciding whether or not an outcome event has occurred Test result, Symptoms
  - Apply efforts to discover events equally in all parts of the study

Estimates made from Cohort Studies

- The incidence rate of disease in those exposed and those not exposed (cumulative incidence, incidence density)
- The relative risk of disease
- The attributable risk of disease
- Standardized mortality or incidence ratio
- Need to account for any known potential confounding variables
- Need to describe effect modifiers
- Need to evaluate the role of chance

- Sampling variability
- Random error
- Sampling error ... etc