Epidemiological Methods

Duanping Liao, MD, Ph.D

Professor and Vice Chair for Research
Chief, Division of Epidemiology
Department of Public Health Sciences
Penn State University College of Medicine

Email: Dliao@phs.psu.edu

Phone: 531-4149

Cross-Sectional Study

Prevalence of a Disease is defined as the total number of cases of the disease in the study population in a specific period of time. It is used as an estimate of how common a disease is within a population over a certain period of time. Thus, the burden of a disease in the population.

Measures of association from a cross-sectional study:

Prevalence of a disease in population that are “exposed” to a risk factor (X%). vs.
Prevalence of a disease in population that are “not-exposed” to a risk factor (Y%).
X^2 test, with p value

Prevalence Ratio

Odds ratio, with 95% Confidence interval and p-value

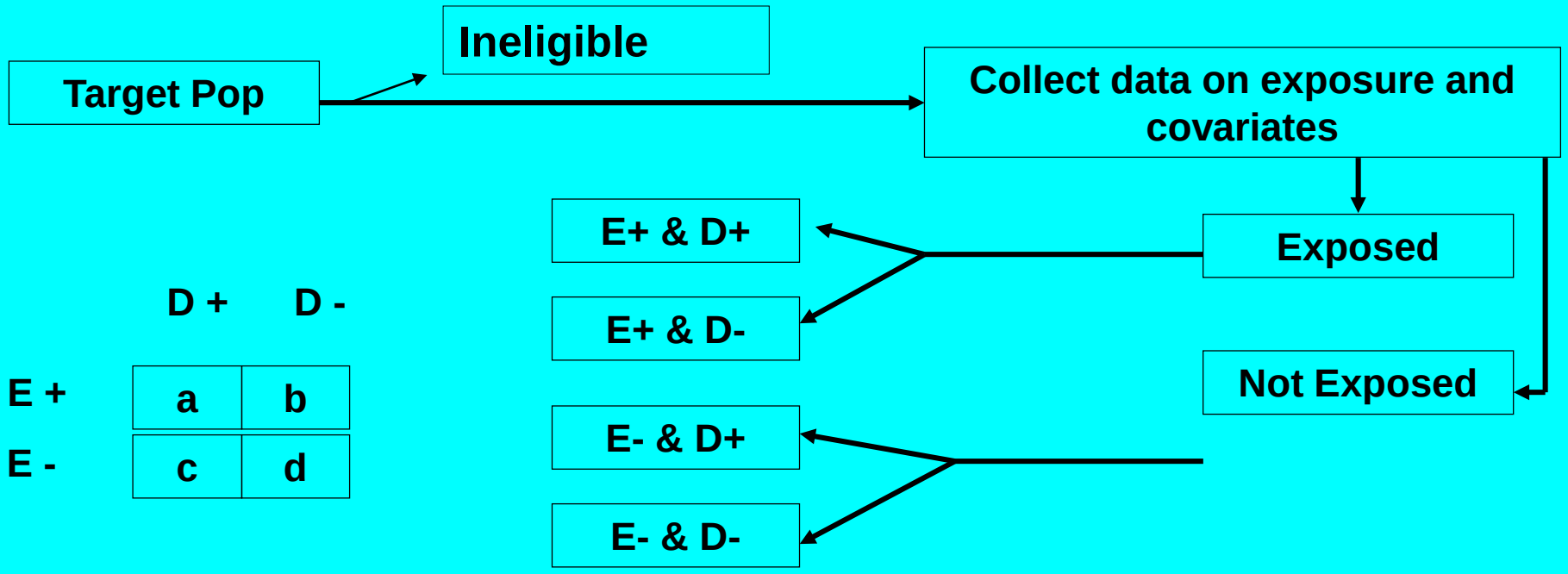
Good for investigating the burden of a disease in the population

Note: The association tested DID NOT imply any temporal relationship.

Cohort Study

(Prospective study, Follow-up Study)

An epidemiological study in which a group or groups of individuals **free** of the disease of interest are identified on the basis of presence or absence of exposure to a suspected risk factor for the disease of interest. At the initiation (baseline), often defined as the time in which the exposure status are assessed, data on exposure, cofactors, and disease/co-morbidity are collected, the eligible subjects are followed over a period of time, and the occurrence of the outcome is assessed overtime.



Cumulative Incidence of a Disease is defined as the total number of NEWLY diagnosed cases of the disease in the study population who were disease-free at the baseline in a specific period of time. It is used as an estimator of how likely a disease is to occur in the population over a certain period of time. Thus, the "risk" of an individual suffering from a disease. n / N (%).

Incident Density: Number of new cases **/** Total person-time at risk. (n/10,000 person-year). Thus, the burden of disease in the general population.

Measures of association from a cohort study:

Cumulative Incidence in population that are "exposed" to a risk factor (X%). vs.
Cumulative Incidence in population that are "not-exposed" to a risk factor (Y%).
X^2 test, with p value

Cumulative Incidence Ratio: with 95% Confidence interval and p-value

Incidence Density Ratio: with 95% Confidence interval and p-value

Odds ratio, with 95% Confidence interval and p-value

Note: The associations tested ARE temporal relationship.

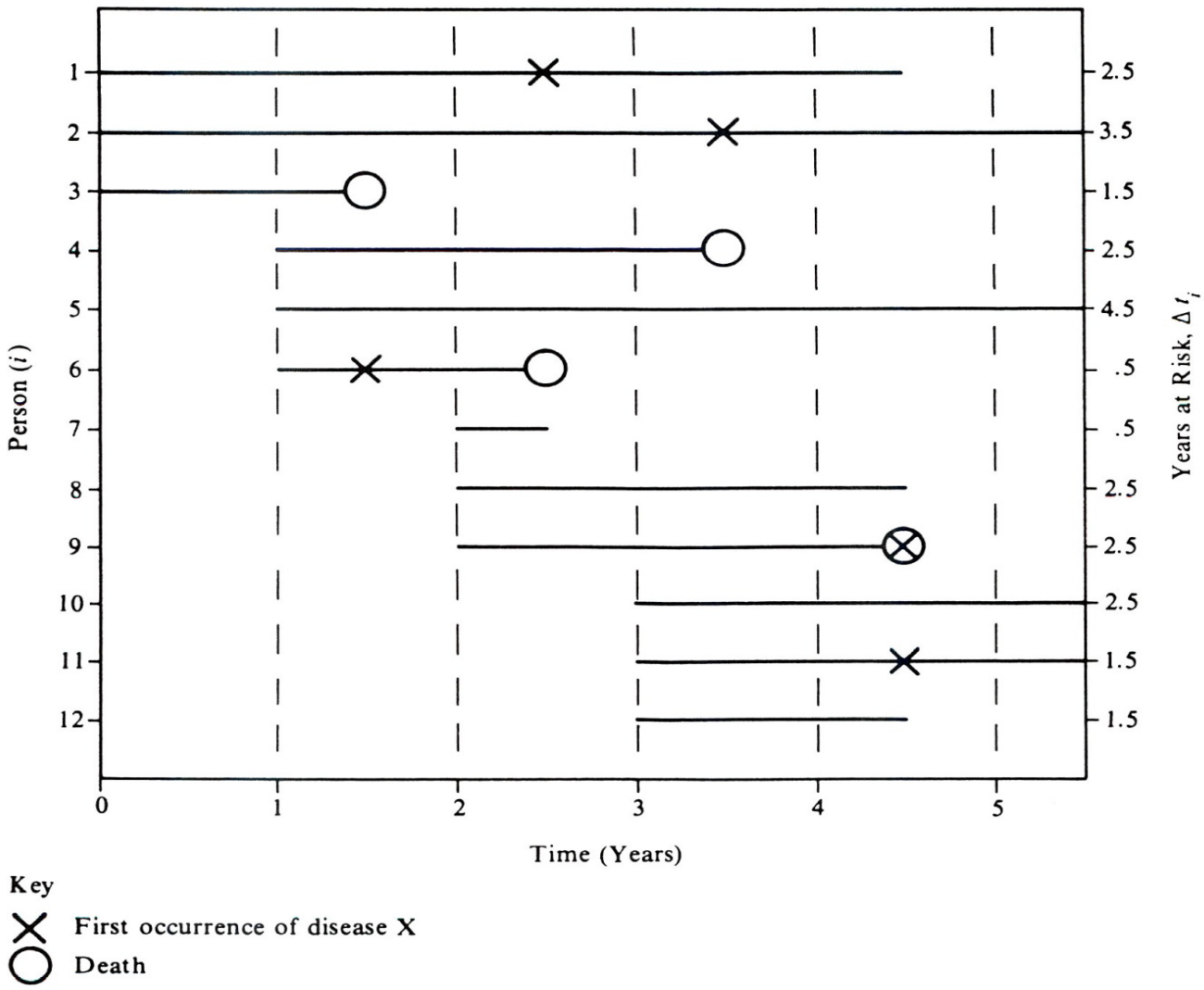


Figure 6.1 Diagrammatic Representation of the 5.5-Year Follow-up of a Hypothetical Cohort of 12 Subjects Initially Free of Disease X

Consider a hypothetical cohort of 12 subjects **(N)** followed for a total duration (Δt) of 5.5 years. Three subjects enter the study at the beginning of each of the first 4 years and that all subsequent events (X) of interest occur at interval midpoints. There are 7 withdrawals among the noncases, including 3 who are lost to follow-up (persons 7, 8, and 12), 2 who die (persons 3 and 4), and 2 who are terminated by the end of the study (persons 5 and 10). The individual follow-up periods (Δt_i), until either disease (X) occurrence or withdrawal, are given on the right.

Person-time at risk =
(2.5 + 3.5 + …+1.5) =26 person-yr.

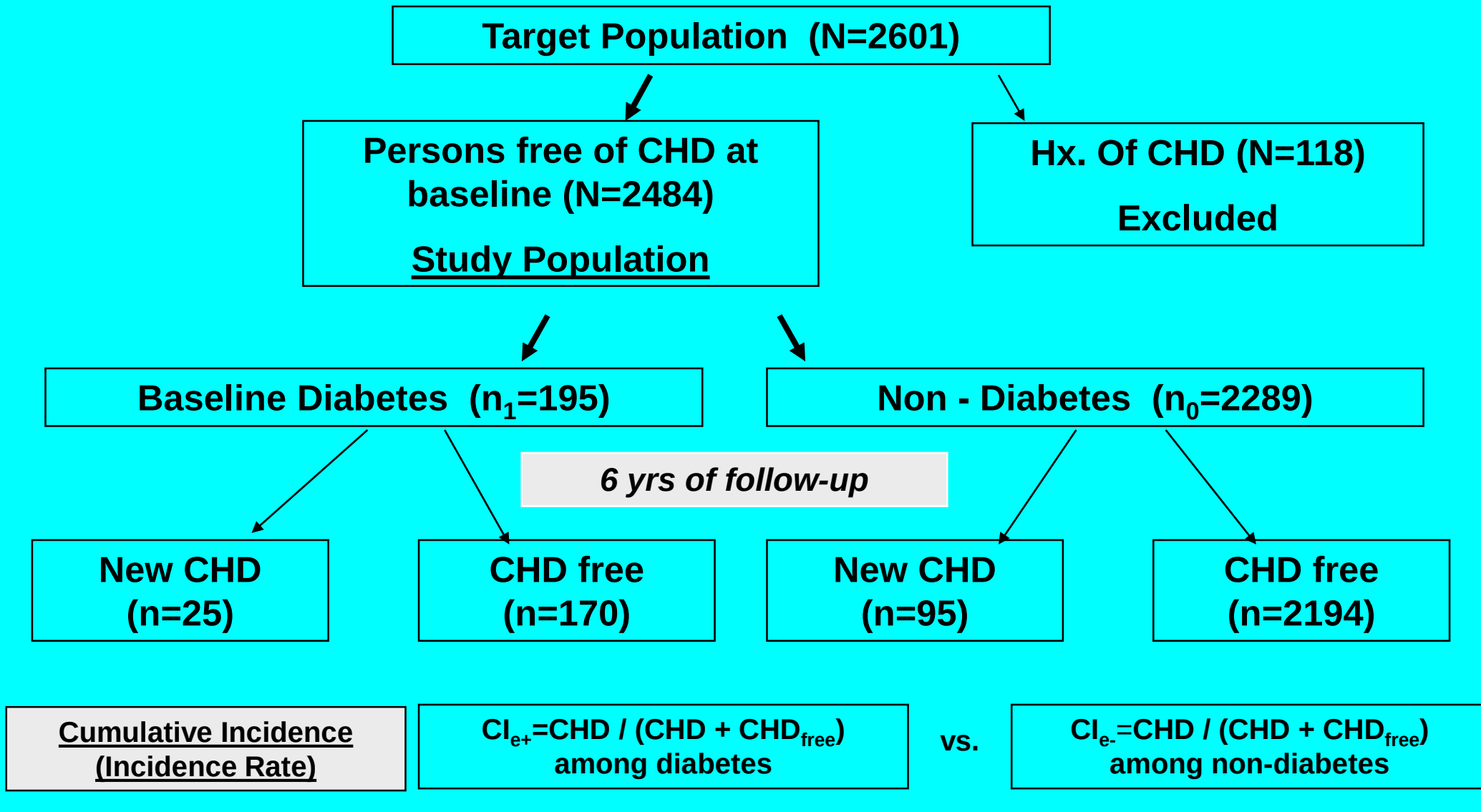
n of disease X = 5

ID = 5 /26 = 0.192/person-year.

For CI, we have 5/12, or
5/(12-3 lost to follow up) or
5/(12-5 lost to follow up or died), depending on your treatment of withdraws.

The follow-up time for each person is from entering the study to the "censored time":
•Become a "diseased/case";
•The end of study period;
•The last contact with the study – lost-to follow-up or other causes of mortality.
•Interval censoring – ½ of the known interval.

Example of recreating a cohort study from the ARIC subsample data.
Study question: Is type 2 diabetes (exposure) associated with the development of coronary heart disease (outcome)?



Summary of data -- 2 x 2 Tables

	CHD +	CHD -	Incidence rate of D	Total Person-time
E +	a	b	a / (a + b)	PY_{e+}
E -	c	d	c / (c + d)	PY_{e-}

or

	CHD +	CHD -	CHD Incidence(%)	Total Person-years
DIAB +	25	170	12.8	1087
DIAB -	95	2194	4.2	14048

Cumulative incidences – count data : 12.8% vs. 4.2%
Incidence densities – person-time data : 23 vs. 6.8 (/1000 person-yr)
Odds of exposure - in incident CHD group: a/c = 25/95 vs. in CHD free group: b/d =170/2194

CIR: incidence in E+ / incidence in E- –- count data

{a / (a + b)} / {c / (c + d)} : 12.8 / 4.2 = 3.05

IDR: Incident density in E / Incident density in E -

$(a / PY_{e+}) / (c / PY_{e-})$: (25 / 1087) / (95 / 14048) = 3.40

OR: Odds ratio of disease given exposure: (a/c) / (b/d) =

(a x d) / (b x c) = (25 x 2194) / (170 x 95) = 3.40

Note: If follow-up is short, and outcome is rare, relative risk estimations from CIR, IDR, and OR are very close. Thus, logistic regression, Poisson regression, and Proportional Hazard model will produce similar estimates.

We will work from one real-life project to study the application of a cohort study, with focus on relative risk calculation.

Metabolic syndrome and incident CHD (Journal of Internal Medicine 2007, 262; 113–123).

Introduction

- Metabolic Syndrome (MetS) is the clustering of several interrelated conditions:
  - Elevated blood pressure (BP)
  - Elevated fasting glucose
  - Elevated serum triglycerides (TG)
  - Decreased HDL-cholesterol (HDL)
  - Increased waist circumference (WC)
- Insulin resistance and obesity are believed to be responsible for Mets
- Each of the five Mets components is a risk factor for cardiovascular disease
- Cardiovascular disease remains the No. 1 killer in the US – about 1 million Americans died from CVD each year
- >50 million Americans have the MetS, and the burden of MetS in the US is likely to continue to rise, largely due to changes in lifestyle, including decreased levels of physical activity and increased prevalence of overweight and obesity

Study Questions

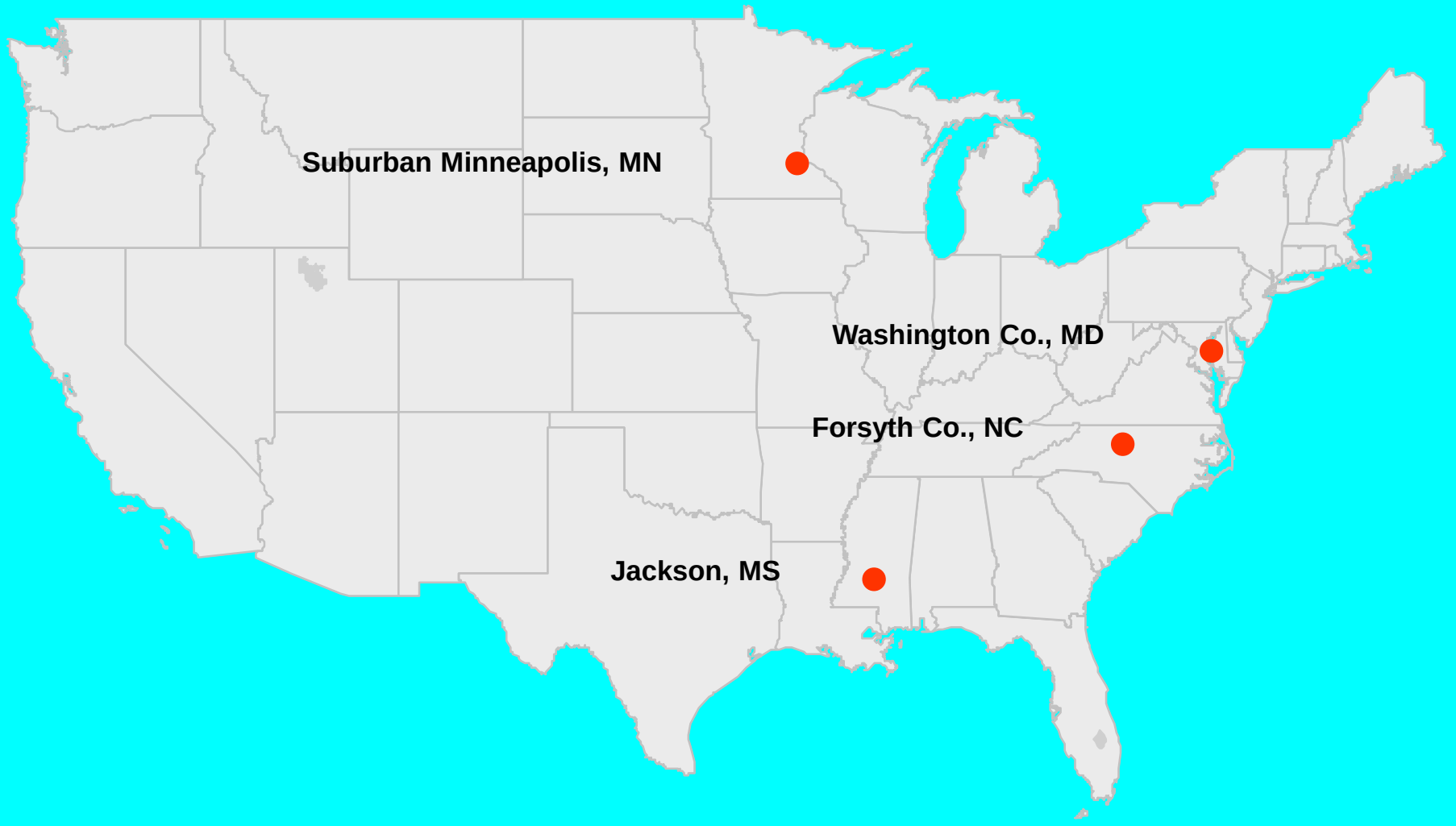
- **Whether Mets is related to the development of CHD?**
- Is there a does-response relationship between the number of Mets components and the risk of CHD?
- **What are the effect modifiers for the Mets and CHD relationship?**
- What are the Mets clusters which poise the greatest risks for the development of CHD?

Study Population

- **ARIC: Atherosclerosis Risk in Communities study is an NHLBI-supported study of atherosclerosis and its manifestations in communities.**

- **A cohort of 15,800 African American and white men and women aged 45-64, sampled from 4 communities. Baseline exam in 1987-1989 was followed by triennial re-exams. The 4^{th} exam was conducted in 1996-1998.**

- **Public use data from the 1^{st} clinical examination and follow-up data on CVD events were used for this study.**


Suburban Minneapolis, MN
Washington Co., MD
Forsyth Co., NC
Jackson, MS

Data Collection 1

- Data were collected via standardized questionnaires, clinical examination procedures, and special measurement procedures
- All data collection instruments were validated and standardized by the study coordinating center and were subject to regular, ongoing quality assurance checks
- Technicians and nurses responsible for data collection and processing were centrally trained and certified. Quality control checks were standardized and implemented by the coordinating center. Central retraining and re-certifications of study staff took place every year

Data Collection 2

- Cohort members who were free of CHD or stroke at baseline were identified for the prospective analysis, with the follow-up for incident events up to 1997
- The average follow-up time was 9 years
- Incident MI, CHD, fatal CHD, and death were identified via a well-established community surveillance system
- Final diagnoses were validated/adjudicated by a panel of physician reviewers according to standardized criteria

Statistical Methods

- Poisson regression: to estimate the **incidence density** of CVD events associated with Mets and to adjust for other established CVD risk factors
- Cox regression models: to estimate the hazards ratios for each outcome associated with Mets and the number of Mets disorders
- Multiplicative interactions: assessed by Cox's models, using $p < 0.10$ to identify significant interactions, based on a relative risk scale.

ATP III Definition of Mets

- Elevated BP as systolic BP $\geq$ 130, or diastolic BP $\geq$ 85 mmHg
- Elevated fasting glucose as fasting glucose $\geq$ 110 mg/dl
- Elevated serum triglycerides(TG) as TG $\geq$ 150 mg/dl
- Decreased HDL-cholesterol as HDL <40 mg/dl for men and < 50 mg/dl for women
- Increased waist circumference (WC) as WC >102 cm for men and >88 cm for women

Mets = having 3 or more of the above abnormalities (ATP III definition)

Definition of Incident CHD

Incident CHD included:

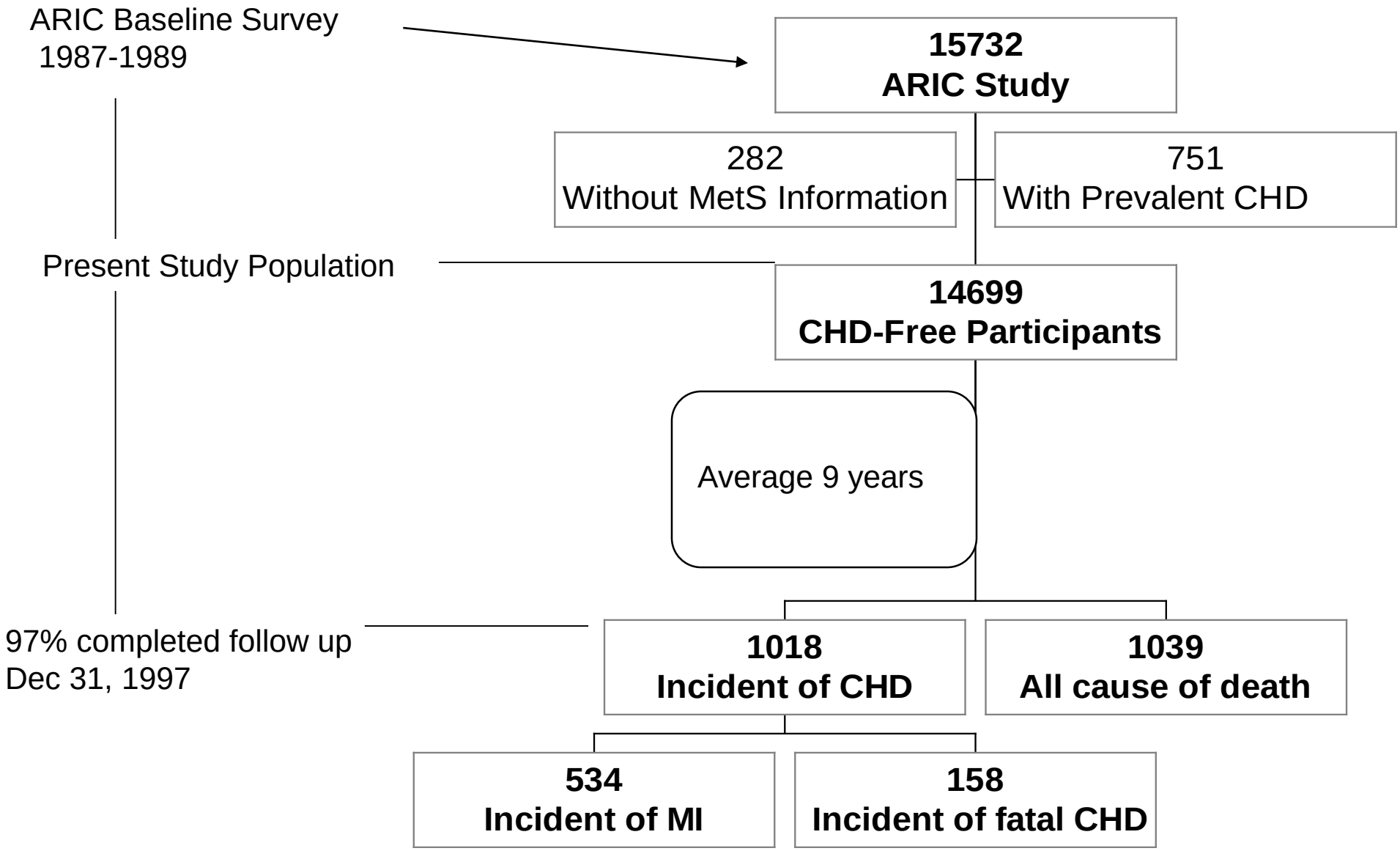
- Hospitalized MI
- Fatal CHD
- Revascularization Procedures
  - CABG
  - PTCA
- ECG detected silent MI
  - Significant Q wave
  - Borderline Q wave with significant ST segment or T wave abnormalities, in the absence of ventricular conduction defects that interfere with Q wave coding

Incident MI included only Hospitalized MI (hard endpoint)

Incident Fatal CHD included only Fatal CHD

Other Covariates

- Age
- Ethnicity
- Sex
- Education level
- Cigarette smoking status
- LDL-cholesterol level


ARIC Baseline Survey
1987-1989
15732
ARIC Study
282
Without MetS Information
751
With Prevalent CHD
Present Study Population
14699
CHD-Free Participants
Average 9 years
97% completed follow up
Dec 31, 1997
1018
Incident of CHD
1039
All cause of death
534
Incident of MI
158
Incident of fatal CHD

Demographic Characteristics

Age (years)	54.0 (5.7)
Gender (% Men)	44
Race (% Black)	26
Current Smoker (%)	26
Less than high school (%)	23
Mets(%)	31

The Frequency (%) of MetS Components in the ARIC Baseline CHD-free Members (1987-1989)

	Mets	Increased WC	Elevated TG	Low HDL	Elevated BP	Elevated Glucose
Total	31	50	27	38	46	22
All Whites	30	47	31	40	39	19
Women	29	57	27	36	39	15
Men	30	35	35	46	39	24
All Blacks	35	58	18	32	66	31
Women	40	76	16	35	67	31
Men	26	28	21	27	65	31

Incidence Density (10,000 person-year), Hazard Ratio (95% CI) of MI Associated with MetS

	% (No. of Events)	Incidence Density	Hazard Ratio (95% CI)
Entire Cohort			
No-MetS	2.65 (269)	23.3 (20.4, 26.7)	1.00
MetS	5.83 (265)	48.3 (42.0, 55.6)	**2.07 (1.74, 2.47)**
Women			
No-MetS	1.42 (80)	18.0 (14.4, 22.5)	1.00
MetS	4.46 (120)	43.2 (35.3, 52.7)	**2.41 (1.80, 3.22)**
Men			
No-MetS	4.17 (189)	48.5 (41.7, 56.4)	1.00
MetS	7.81 (145)	92.1 (77.2, 109.9)	**1.91 (1.53, 2.38)**

Sex-MetS Interaction was statistically significant, $p<0.01$ for Sex-Mets interaction term
All results were adjusted for age, race, education, smoking status, LDL cholesterol, and sex when appropriate

Multivariable Adjusted Hazard Ratios and 95% CI of Incident MI Associated with the Number of Metabolic Syndrome Components

Number of MetS Components	Number of Participants	% (No. of Events)	Incidence Density	Hazard Ratio (95% CI)
0	2804	1.4 (39)	13.7 (10.0, 18.8)	1.00
1	3894	2.6 (101)	22.1 (18.0, 27.2)	1.66 (1.14, 2.40)
2	3455	3.7 (129)	31.6 (26.2, 38.0)	2.39 (1.67, 3.43)
3	2554	4.7 (120)	39.2 (32.4, 47.5)	2.98 (2.07, 4.30)
4	1422	6.7 (95)	55.2 (44.4, 68.6)	4.25 (2.91, 6.20)
5	570	8.8 (50)	74.4 (54.3, 102.0)	5.84 (3.75, 9.10)

The p-value for linear trend < 0.001
All models adjusted for age, sex, ethnicity, education, smoking status, and LDL cholesterol

Components Cluster	Events/Participants	HR (95%CI)
Reference Group	39/2804	1.00
HBP+HG+TG	8/90	5.45 (2.52,11.76)
HBP+HG+HDL	15/100	9.77 (5.34,17.86)
HBP+HG+WC	20/465	3.00 (1.72,5.21)
HBP+TG+HDL	22/269	4.84 (2.86,8.22)
HBP+TG+WC	9/285	2.19 (1.05,4.53)
HBP+HDL+WC	17/536	2.48 (1.39,4.41)
HG+TG+HDL	5/85	3.34 (1.31,8.50)
HG+TG+WC	1/65	1.00 (0.14,7.26)
HG+HDL+WC	3/162	1.42 (0.44,4.58)
TG+HDL+WC	20/497	3.12 (1.82,5.36)
HBP+HG+TG+HDL	16/141	7.36 (4.09,13.25)
HBP+HG+TG+WC	17/177	6.54 (3.68,11.64)
HBP+HG+HDL+WC	23/345	5.14 (3.03,8.69)
HBP+TG+HDL+WC	28/528	3.88 (2.38,6.32)
HG+TG+HDL+WC	11/231	3.46 (1.77,6.76)
HBP+HG+TG+HDL+WC	50/570	6.76 (4.42,10.33)

Adjusted Hazard Ratios (95% CI) of Incident MI Associated with MetS Clusters

Ref. Group 0 MetS component

HBP = Elevated BP

HG = Elevated fasting glucose

TG = Elevated triglycerides

HDL = Low HDL-Cholesterol level

WC = Elevated Waist circumference

All 16 possible clusters of MetS components were entered into the models and compared to individuals without any MetS component (reference group). All models were adjusted for age, race, and sex.

Summary

- Over 30% of the middle-aged population in United States have Mets.
- Mets is associated with the development of incident CHD, MI, and all cause mortality
- Women with Mets are at greater risk of developing incident CHD and MI than men with Mets, in another words, Mets is a stronger risk factor for incident CHD and MI in women than in men
- There is a clear does-response relationship between the number of Mets components and the risk of developing CHD
- Mets clusters with both elevated BP and elevated fasting glucose in the clusters poise the greatest risk of incident CHD, MI, and all cause mortality

Conclusion

This study highlighted the needs to target Mets, especially the highest risk clusters of Mets, to reduce the population burden of cardiovascular disease, especially CHD and MI

Cohort Study

Cumulative Incidence of a Disease is defined as the total number of NEWLY diagnosed cases of the disease in the study population who were disease-free at the baseline in a specific period of time. It is used as an estimator of how likely a disease is to occur in the population over a certain period of time. Thus, the "risk" of an individual suffering from a disease. n / N (%).

Incident Density: Number of new cases **/** Total person-time at risk. (n/10,000 person-year). Thus, the burden of disease in the general population.

Measures of association from a cohort study:

Cumulative Incidence in population that are "exposed" to a risk factor (X%). vs.
Cumulative Incidence in population that are "not-exposed" to a risk factor (Y%).
X^2 test, with p value

Cumulative Incidence Ratio: with 95% Confidence interval and p-value

Incidence Density Ratio: with 95% Confidence interval and p-value

Odds ratio, with 95% Confidence interval and p-value

Note: The associations tested ARE temporal relationship.

Cohort Study

Cumulative Incidence (CI) of a Disease is defined as the total number of NEWLY diagnosed cases of the disease in the study population who were disease-free at the baseline in a specific period of time. It is used as an estimate of likely a disease is to occur in the population over a certain period of time. Thus, the "risk" of suffering from a disease in the population. n / N (%).

Incident Density: Number of new cases **/** Total person-time at risk. (n / 1000 person-year).

Measures of association from a cohort study:

CI of a disease in population that are "exposed" to a risk factor (X%). vs.
Incidence of a disease in population that are "not-exposed" to a risk factor (Y%).
X^2 test, with p value

Cumulative Incidence Ratio: with 95% Confidence interval and p-value

Incidence Density Ratio: with 95% Confidence interval and p-value

Odds ratio, with 95% Confidence interval and p-value

Note: The associations tested ARE temporal relationship.

Cohort Study

Cumulative Incidence (CI) of a Disease is defined as the total number of NEWLY diagnosed cases of the disease in the study population who were disease-free at the baseline in a specific period of time. It is used as an estimate of likely a disease is to occur in the population over a certain period of time. Thus, the "risk" of suffering from a disease in the population. n / N (%).

Incident Density: Number of new cases **/** Total person-time at risk. (n / 1000 person-year).

Measures of association from a cohort study:

CI of a disease in population that are "exposed" to a risk factor (X%). vs.
Incidence of a disease in population that are "not-exposed" to a risk factor (Y%).
X^2 test, with p value

Cumulative Incidence Ratio: with 95% Confidence interval and p-value

Incidence Density Ratio: with 95% Confidence interval and p-value

Odds ratio, with 95% Confidence interval and p-value

Good for studying the determinants of a disease.

Note: The associations tested ARE temporal relationship.