Measurement in Epidemiology

Course: Study designs and measurements

Introduction Frequency Association Impact Conclusion

Measurement in Epidemiology

- Measures of disease frequency
- Measures of association
- Measures of potential impact

RACE 612

Measures of disease frequency

- Measures of disease frequency in mathematical quantity
 - Count
 - Fraction
 - Rate
 - Ratio
 - Proportion (percentage)
- Measures of disease frequency in epidemiology
 - Prevalence
 - Incidence

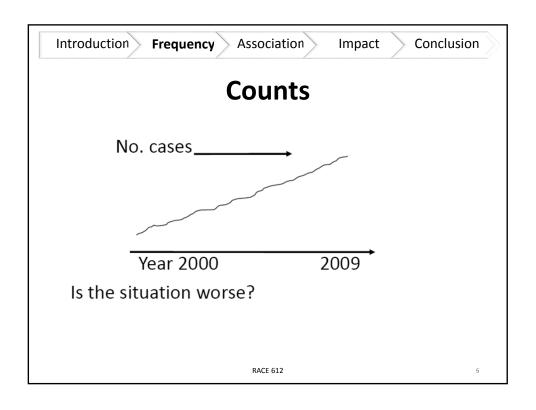
RACE 612

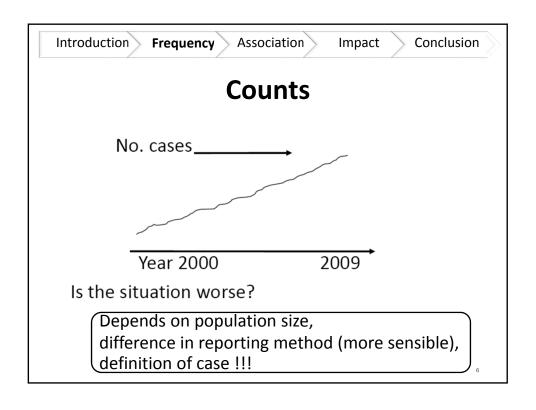
Introduction Frequency Association Impact Conclusion

Counts

- Simplest & most basic measure absolute number of persons who have disease or characteristic of interest.
- Useful for health planners & administrators: for allocation of resources (e.g. quantity of ORS needed by diarrheal cases)
- Count of No. cases of a disease, is used for surveillance of infectious disease for early detection of outbreaks.

RACE 612 4





Limited values of counts

- Number of persons with characteristic, e.g., cases of malaria depends on the size of the population at risk of the disease in an area.
 - The bigger this group, the higher is the expected number of cases.
- The duration of observation also affects the frequency of cases; the longer the observation period, the more cases can occur.

Count does not contain these elements!

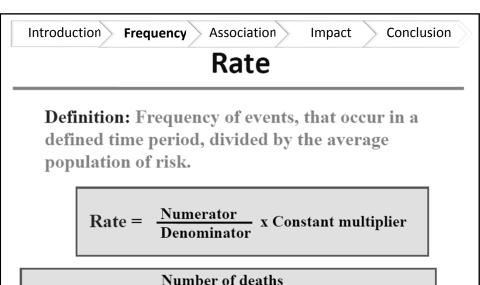
RACE 612

Introduction Frequency Association Impact Conclusion

Measurement fractions

- Rate
 - Measures the frequency of an event in a population.
 - Time and multiplier
 - Incidence
- Ratio
 - A value obtained by dividing one number by another (either related or unrelated)
 - Fraction that numerator is not a part of denominator
- Proportion
 - Numerator and denominator have the same units (dimensionless).
 - Prevalence

RACE 612



Crude death rate = $\frac{\text{(defined place and time period)}}{\text{Mid-period population}} \times 1000$ (same place and population)

Conclusion Introduction Frequency Association **Impact Crude Mortality Rates** Advantages Number of deaths -Actual Summary rates in a specified year -Easy calculation for international comparisons Number of individuals in the population **Disadvantages** in the specified year Since population vary in composition (e.g., age)differences in crude rates (Mid-period population) difficult to interpret

Age-Specific Mortality Rate

- Provide a broader view of mortality for sub-groups stratified by age
- Numerator and denominator are limited to a specific age group
- Comparable across populations

Introduction F	requency Association Impact Conclusion				
Age	Age-Specific Mortality Rate				
	Number of deaths among persons aged 0-14 in a given year				
Aged 0 –14	= × 100000				
years	Total number of persons aged 0-14 in the same year				



Age-specific Rates

Advantages

- •Homogenous subgroups
- Detailed rates useful for public health and
- •Epidemiological aims

Disadvantages

•Cumbersome to compare subgroups of two or more populations

Introduction Frequency Association Impact Conclusion

Standardization of Rates (Adjusting rates)

- Used to reduce distortion in comparisons between crude areas
- Allows comparisons of rates between populations that differ by variables that can influence the rate (e.g., age)
- Two types: Direct and Indirect

Direct Adjustments of Rates

- •Requires a **standard population**, to which the estimated age-specific rates can be applied
- •Multiply standard population by age-specific rates for populations A and B to determine the

standardized rates

•Compare standardized rates

Introduction Association Conclusion Frequency **Impact** Population, Deaths, and Death Rate by **Community and by Age Community A** Community B Age Population **Deaths** Death Rate Population **Deaths Death Rate** (year) (per 1000) (per 1000) Under 1 1,000 15 15.0 5,000 100 20.0 1 - 143,000 3 20,000 35 1.0 1.0 15 - 346,000 6 1.0 35,000 35 1.0 35 - 5413,000 52 4.0 17,000 85 5.0 55 - 647,000 105 15.0 8,000 160 20.0 1,600 15,000 1,350 90.0 20,000 80.0 Over 64 50,000 1,781 35.6 100,000 1,740 17.4 All ages

tandard	Population	n by Ag	e and Ag	ge-Specific	Death Rate
Age (years)	Standard population				
	Combined				
Under 1	6,000				
1 – 14	23,000				
15 – 34	41,000				
35 – 54	30,000				
55 – 64	15,000				
Over 64	35,000				
Total	150,000				
Age – adjusted death rate					

andard	Populatio	n by Age a	nd Age-	Specific D	eath Rate
Age	Standard	Death rate		Death rate	Expected
(years)	population	in A	deaths at	in B	deaths at
	Combined	(per 1,000)	A's rate	(per 1,000)	B's rate
Under 1	6,000	15.0	90	20.0	120
1 – 14	23,000	1.0	23	1.0	23
15 – 34	41,000	1.0	41	1.0	41
35 – 54	30,000	4.0	120	5.0	150
55 – 64	15,000	15.0	225	20.0	300
Over 64	35,000	80.0	2,800	90.0	3,150
Total	150,000	35,6	3,299	17.4	3,784
Age – adjusted			22.0		25.0

	Population	n by Age a	and Age-	Specific D	eath Rate
Age (years)	Standard population	Death rate in A (per 1,000)	Expected deaths at A's rate	Death rate in B (per 1,000)	Expected deaths at B's rate
Under 1	6,000	15.0	90	20.0	120
1 – 14	23,000	1.0	23	1.0	23
15 – 34	41,000	1.0	41	1.0	41
35 – 54	30,000	4.0	120	5.0	150
55 – 64	15,000	15.0	225	20.0	300
Over 64	35,000	80.0	2,800	90.0	3,150
Total	150,000	35,6	3,299	17.4	3,784
Age – adjusted death rate (per 1000)			22.0		25.0

Indirect Adjustments of Rates

- Indirect adjustment is used less frequently than direct adjustment
- Use when age-specific numbers of deaths in the study population are either unavailable or small in number

RACE 612



Indirect Adjustment of Rates

Based on applying the **age-specific rates of the standard population** to the population of interest to determine the number of **"expected"** deaths



Standardized Mortality Ratio **SMR**

Standardized Mortality Ratio

Total observed deaths in a population

Total expected deaths in a population

F	opula	ation of	Communi	ity A by A	\ge
_	aı	nd Stan	dard Deat		
	Age (years)	Populati on in A	Standard death rate		
	()		(per 1,000)		
ī	Under 1	1,000	20.0		
·	1 – 14	3,000	1.0		
Ī.	15 – 34	6,000	1.0		
;	35 – 54	13,000	5.0		
!	55 – 64	7,000	20.0		
Ī	Over 64	20,000	90.0		
-	Total	50,000	17.4		

Introd	luction Fre	equency	Association	Impact	Conclusion			
	Population and Expected Deaths of Community A by Age							
	Age (years) Populati on in A (per 1,000) Expected deaths in A at standard rates							
	Under 1	1,000	20.0	20.0				
	1 – 14	3,000	1.0	3.0				
	15 – 34	6,000	1.0	6.0				
	35 – 54	13,000	5.0	65.0				
	55 – 64	7,000	20.0	140.0				
	Over 64	20,000	90.0	1,800.0				
	Total	50,000	17.4	2,034.0				
	SMR _A = 1781 / 2034 = 0.876 SMR _B = 1.0							

Standardized Mortality Ratio

If the SMR is greater than 1, more deaths have occurred than anticipated

If the SMR is less than 1, fewer deaths have occurred than anticipated

Introduction Frequency Association Impact Conclusion

Ratio

 Ratio: A fraction in which the numerator is not part of the denominator.

- a and b are two mutually exclusive frequency
- Example:
- Number of hospital beds per 100,000
- Male and female dengue infection ratio = 70/35 or
 2 males to one female (2:1)

Ratio

- A ratio is the relative magnitude of two quantities or a comparison of any two values.
- It is calculated by **dividing** one variable by the other.
- The numerator and denominator need not be related. Therefore, one could compare apples with oranges or apples with number of physician visits.

RACE 612

Introduction Frequency Association Impact Conclusion

Example

${\bf EXAMPLE: Calculating \ a \ Ratio-Different \ Categories \ of \ Same \ Variable}$

Between 1971 and 1975, as part of the National Health and Nutrition Examination Survey (NHANES), 7,381 persons ages 40–77 years were enrolled in a follow-up study. At the time of enrollment, each study participant was classified as having or not having diabetes. During 1982–1984, enrollees were documented either to have died or were still alive. The results are summarized as follows.

Original Enrollment (1971–1975)	Dead at Follow-U (1982–1984)
189	100
3,151	811
218	72
3,823	511
	(1971–1975) 189 3,151 218

Of the men enrolled in the NHANES follow-up study, 3,151 were nondiabetic and 189 were diabetic. Calculate the ratio of non-diabetic to diabetic men.

Ratio = 3,151 / 189 x 1 = 16.7:1

05/08/59 RACE 612 28

Proportion (percentage, frequency)

- Proportion: + a included in the denominator
- a a + b
- + No measurement unit; > 0 to ≤ 1
- + Often expressed as %

Example: From 7,999 females aged 16 – 45 y, 2,496 use modern contraceptive methods.

The proportion of those who use modern contraceptive methods = $2,496 / 7,999 \times 100 = 31.2\%$

Introduction Frequency Association Impact Conclusion

Example

EXAMPLE: Calculating a Proportion

Example A: Calculate the proportion of men in the NHANES follow-up study who were diabetics.

Numerator = 189 diabetic men

Denominator = Total number of men = 189 + 3,151 = 3,340

Proportion = (189 / 3,340) x 100 = 5.66%

Example B: Calculate the proportion of deaths among men.

Numerator = deaths in men

= 100 deaths in diabetic men + 811 deaths in nondiabetic men

= 911 deaths in men

Notice that the numerator (911 deaths in men) is a subset of the denominator.

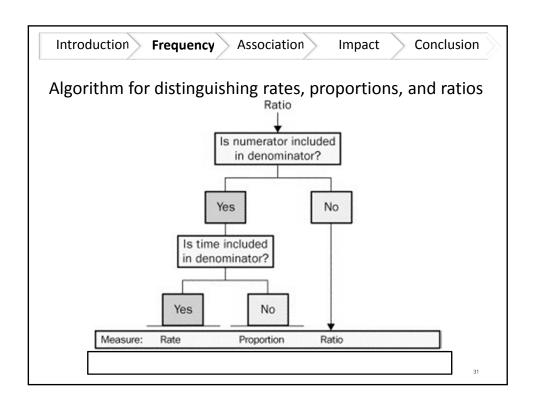
Denominator = all deaths

= 911 deaths in men + 72 deaths in diabetic women + 511 deaths in nondiabetic women

= 1,494 deaths

Proportion = 911 / 1,494 = 60.98% = 61%

RACE 612



Measures of disease
Frequency in epidemiology

• Incidence (I): Measures new cases of a disease that develop over a period of time.

1.Cumulative incidence (incidence)

2.Incidence rate = incidence density

• Prevalence (P): Measures existing cases of a disease at a particular point in time or over a period of time.

1. Point prevalence

2. Period prevalence

Cumulative Incidence (CI) = Incidence

No. of individuals who get the disease during a certain period

CI =

No. of individuals in the population at the beginning of the period

- A proportion
- Has no dimension
- Varies between 0 and 1

Introduction Frequency Association Impact Conclusion

Example of Cumulative Incidence

- The population statistic of Lab Lair District in 2001 revealed that there were 5,572 women aged 20-39 years who were sex workers. Based on the record of CHAS, among those women, 45 were HIV + ve during 2002-2005.
- What is the cumulative incidence of HIV + ve among those women during a period of 4 years?
- Cumulative incidence = 45 / 5,572 = 0.008 or 0.8%

05/08/59 RACE 612

Incidence Rate

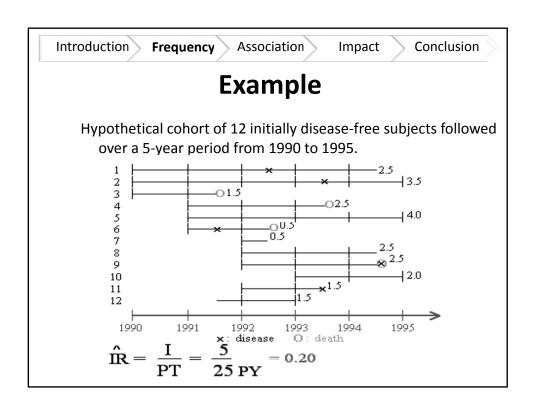
$$IR = \frac{I}{PT}$$

I = # of new cases during follow-up

PT = total time that disease-free individuals in the cohort are observed over the study period (time at risk).

Synonyms: hazard rate, incidence density rate.

Measures the **rapidity** with which new cases are occurring in a population



Incidence Rate

$$|\hat{IR}| = \frac{I}{PT} = \frac{5}{25PY} = 0.20$$

= 20 new cases per 100 person - years

Study questions:

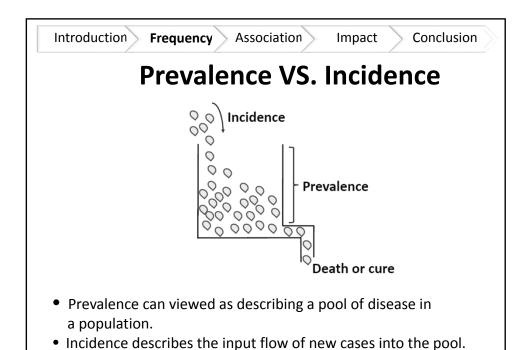
- 1) Is the value of 0.20 a proportion?
- 2) Does the value of 0.20 person-year represent the risk of developing disease?

Introduction Frequency Association Impact Conclusion

Prevalence

 $\frac{\text{Number of exiting cases}}{\text{Number of population}}$

Number of existing cases = New + preexisting cases/ Number of population during the same time period



• Fatality and recovery reflects the output flow from the pool.

Prevalence

Useful for:

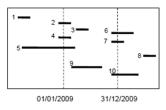
- Assessing the health status of a population.
- Planning health services.

Not Useful for:

Identifying risk factors



Prevalence divided into two types:



Point prevalence

01/01/2009: case No. 2, 4, 5

31/12/2009: case No. 6, 7, 10

Period prevalence between 01/01-31/12/2009:

Case No. 2, 3, 4, 5, 6, 7, 9, 10

Introduction Frequency Association Impact Conclusion

Example

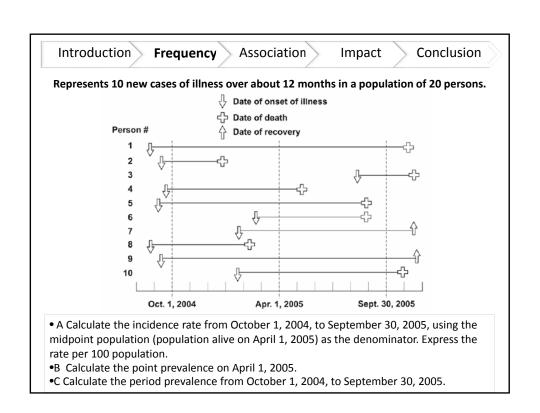
Suppose we followed a population of 150 persons for one year, and 25 had a disease of interest at the start of follow-up and another 15 new cases developed during the year.

- 1) What is the period prevalence for the year?
- 2) What is the point prevalence at the start of the period?
- 3) What is the cumulative incidence for the one year period?

Example (Answer)

Suppose we followed a population of 150 persons for one year, and 25 had a disease of interest at the start of follow-up and another 15 new cases developed during the year.

- 1) What is the period prevalence for the year?
 - pp = (25+15)/150 = 0.27 or 27%
- 2) What is the point prevalence at the start of the period?
 - p = 25/150 = 0.17 = 17%
- 3) What is the cumulative incidence for the one year period?
 - CI = 15/125 = 0.12 = 12%



Example A: Calculate the incidence rate from October 1, 2004, to September 30, 2005, using the
midpoint

population (population alive on April 1, 2005) as the denominator. Express the rate per 100 population.

Incidence rate numerator = number of new cases between October 1 and September 30

= 4 (the other 6 all had onsets before October 1, and are not included)

Incidence rate denominator = April 1 population

= 18 (persons 2 and 8 died before April 1)

Incidence rate = $(4/18) \times 100 = 22$ new cases per 100 population

Example B: Calculate the point prevalence on April 1, 2005.

Point prevalence is the number of persons ill on the date divided by the population on that date.

On April 1, seven persons (persons 1, 4, 5, 7, 9, and 10) were ill.

Point prevalence = $(7 / 18) \times 100$

= 38.89%

• Example C: Calculate the period prevalence from October 1, 2004, to September 30, 2005.

The numerator of period prevalence includes anyone who was ill any time during the period. In Figure 3.1, the first 10 persons were all ill at some time during the period.

Period prevalence = (10 / 20) x 100 = 50%

= 50

05/08/59 RACE 612

Introduction Frequency Association Impact Conclusion

Measures of association

Absolute

Risk difference, excess risk Attributable risk

Exposed - unexposed

Relative

Risk ratio, rate ratio Odds ratio

Exposed /unexposed

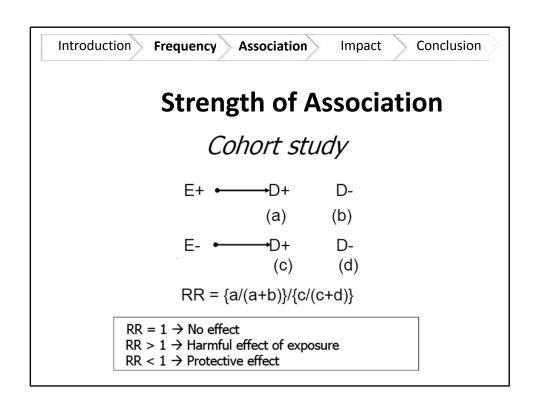
Measures of association

• Strength (magnitude) of association

- Cohort ---> Relative risk (RR)

- Case control ---> Odds ratio (OR)

- Cross sectional ---> Odds ratio (OR)



Introduction Conclusion Frequency Association **Impact** Measurement of association **Expression** Question Definition Absolute risk What is the incidence of I = #new case disease in a group initially #People in group free of the condition? What is the incidence of $AR = I_{E^+} - I_{E^-}$ Attributable risk disease attributable to Risk difference exposure? Relative risk How many times more likely $RR = I_{E^+}$ are exposed persons to IE-Risk ratio become disease, relative to nonexposed persons?

RACE 612

05/08/59

Relative risk or risk ratio

 Compares the risk of a heath event among one group with the risk among another group.

Risk of disease (Cumulative incidence) in exposure group
Risk of disease (Cumulative incidence) in non-exposure group

RACE 612 50

Risk	smoking	а	b	a+b		
INION	non-smoking	С	d	c+d		
Risks of CA in smoking (I_{E+}) = a/a+b Risks of CA in non-smoking (I_{E-}) = c/c+d Relative risk (risk ratio) $(I_{E+})/(I_{E-})$ = a/a+b $\overline{c/c+d}$ Absolute risk reduction (ARR) = $(I_{E+}) - (I_{E-})$ Number needed to treat (NNT) = $\overline{1}$ \overline{ARR}						
05/08/59		RACE 612			51	

Risk	smoking	20	180	200	
1/15/	non-smoking	5	495	500	
Risks of CA in smoking (I _{E+}) = Risks of CA in non-smoking (I _{E-}) = Relative risk (risk ratio) (I _{E+})/(I _{E-}) =					
Absolute risk reduction (ARR) = Number needed to treat/harm = (NNT/NNH)					
05/08/59		RACE 612			52

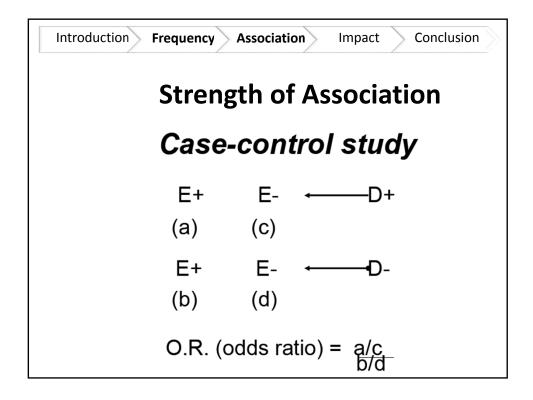
Rate ratio

 Compares the incidence rates (person-time rates) of two groups

Mortality rate for exposure group

Mortality rate for non-exposure group

512



Probability vs. Odds

- Probability (P)
 - The **proportion** of people in whom a particular characteristic, such as a positive test, is present.
- Odds
 - The **ratio** of two probabilities of an event to that of 1the probability of the event
- Odds = \underline{P} or $P = \underline{Odds}$ 1-P 1+Odds

RACE 612

Introduction Frequency Association Impact Conclusion

Example

- Probability of win = 0.8
- Odds of win = 0.8/(1-0.8)= 0.8/0.2= 4



RACE 612

56

Example

Among 100 people at baseline, 20 develop influenza over a year.

- The risk is 1 in 5 (i.e. 20 among 100)
- The odds is 1 to 4 (i.e. 20 compared to 80)

Thus a risk is a *proportion*, But an odds is a *ratio*.

Introduction Frequency Association Impact Conclusion

Odds ratio

- The ratio of the odds of a condition in the exposed compared with the odds of the condition in the unexposed
- Usually applied to prevalence studies rather than incidence studies

OR = Odds of exposed among disease
Odds of exposed among non-disease

KACE 012

		Outcome CA lung Yes No				
Risk	smoking	а 🔪	, b			
INISK	non-smoking	c	d			

Odds of smoking in CA = a/c
Odds of smoking in non-CA = b/d
Odds Ratio = a/c
b/d = ad
cb

RACE 612 59

		Outc CA I		
		Yes	No	
Risk	smoking	30	5	
	non-smoking	10	35	

Odds of smoking in CA = a/c = 30/10 = 3Odds of smoking in non-CA = b/d = 5/35 = 0.14Odds Ratio = a/c = a

The odds of smoking in CA lung/the odds of smoking in control (odds of having CA lung comparing smoking and non-smoking)

RACE 612 60

Example

Assume that among the 100 people at risk, 50 are men and 50 women. If 15 men and 5 women develop influenza,

The relative risk of developing influenza in men, as compared with women, is:

Risk in men = 15/50 divided by Risk in women = 5/50

15/50:5/50=3.0

The odds ratio is a ratio of two odds
The odds in men = 15/35

the odds of women getting

divided by The odds in women = 5/35

5:5/45=3.9

We conclude that the odds of men getting influenza over the year are 3.9 times as high as

influenza.

Introduction Frequency Association Impact Conclusion

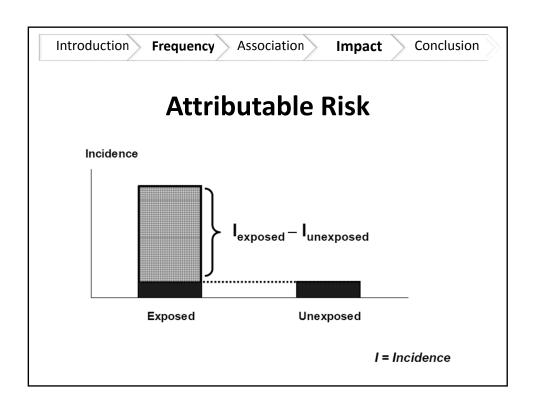
Measures of Impact

- Reflects the burden that an exposure contribute to the frequency of disease in the population
- Impact of exposure removal
- Two concepts
 - Attributable risk among exposed
 - Population attributable risk



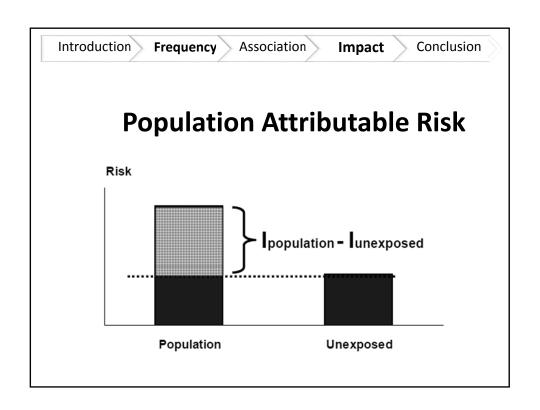
Attributable Risk (AR)

- Quantifies disease burden in exposed group attributable to exposure.
- Provides answer to
 - What is the risk which can be attributed to the exposure?
 - What is the excess risk due to the exposure?
- Calculated as risk difference (RD)



Population Attributable Risk (PAR)

- Excess risk of disease in total population attributable to exposure.
- Reduction in risk which would be achieved if population entirely unexposed.
- Helps determining which exposures relevant to public health in community.
- PAR = AR*P





Attributable Risk fraction

· Attributable risk in the exposed group

$$AR\% = \underbrace{I_{\text{exposed}} - I_{\text{unexposed}}}_{\text{exposed}}$$

· Attributable risk in the total population

$$PAR\% = \frac{I_{population} - I_{unexposed}}{I_{population}}$$

Note:

 I_p can be linked to I_e and I_u if one knows the proportions of the population who are exposed (P) and unexposed (Q), (P and Q add to 1).

$$I_p = P(I_e) + Q(I_u)$$

Example

- Consider a cohort study of risk of ischemic stroke, taken in 1
 year, with 500 subjects with atrial fibrillation (AF) controlled
 against 500 subjects without AF.
- Given the proportion of AF in general population is 30%.
- The results are summarized as follow:

	Ischemic stroke	Ischemic stroke	Total
	present	absent	
AF	2	498	500
No AF	1	499	500
	3	997	

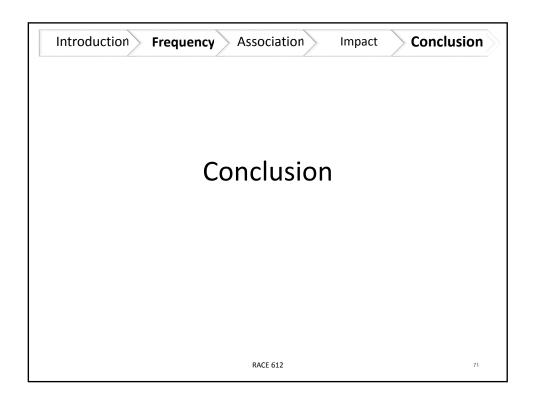
RACE 612

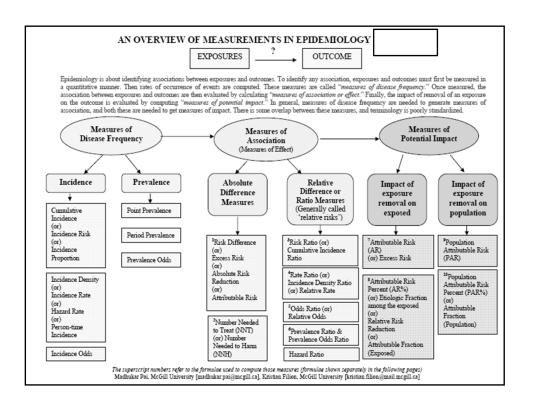
69

Answer

- Attributable risk
 - $-I_E I_{NonE} = 0.004 0.002 = 0.002$
 - The incidence of ischemic stroke that is attributable to AF is 2 in 1000
- Attributable risk fraction
 - $-I_{E}-I_{nonE}/I_{E} = 0.002/0.004 = 50\%$
 - If AF patient were controlled, we could expect 50% reduction in ischemic stroke.
- Population attributable risk
 - -AR*P = 0.002*0.3=0.0006
 - The risk of ischemic stroke in the total population that is attributable to AF is 6 in 10000.
- population attributable fraction
 - PAR /[(I_E*P) +(InonE*(1-P)]
 - = 0.0006/[(0.004*0.3) + (0.002*0.7)] = 0.23
 - According to these results, control AF in population would reduce the overall risk of ischemic stroke by 23%.

RACE 612





Frequently used measures of morbidity

Measure	Numerator (X)	Denominator (Y)
Crude death rate	Total number of deaths during a given time interval	Mid-interval population
Cause-specific death rate	Number of deaths assigned to a specific cause during a given time interval	Mid-interval population
Attack rate	No. of new cases of a specified disease reported during an epidemic period	Population at start of the epidemic period
Secondary attack rate	No. of new cases of specified disease among contacts of known cases	Size of contact population at risk

Introduction Frequency	Association	Impact	Conclusion
------------------------	-------------	--------	------------

Other measures of mortality

Measure	Numerator (X)	Denominator (Y)
Death-to-case-ratio	No. of deaths assigned to a specific cause during a given time interval	No. of new cases of that disease reported during the same time interval
Neonatal mortality rate	No. of deaths under age 28 days during a given time interval	No. of live births during the same time interval
Postneonatal mortality rate	No. of deaths from ages 28 days to, but not including, 1 year during a given time interval	No. of live births during the same time interval
Maternal mortality rate	No. of deaths assigned to pregnancy-related causes during a given time interval	No. of live births during the same time interval